



CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

UM ESTUDO SOBRE A PRODUÇÃO CIENTÍFICA BRASILEIRA A PARTIR DE DADOS DA PLATAFORMA LATTES

THIAGO MAGELA RODRIGUES DIAS

Orientador: Prof. Dr. Gray Farias Moita
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientador: Prof. Dr. Alberto Henrique Frade Laender
Universidade Federal de Minas Gerais

Belo Horizonte
Setembro de 2016

THIAGO MAGELA RODRIGUES DIAS

**UM ESTUDO SOBRE A PRODUÇÃO CIENTÍFICA
BRASILEIRA A PARTIR DE DADOS DA PLATAFORMA
LATTES**

Tese apresentada ao Programa de Pós-graduação em Modelagem Matemática e Computacional do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito parcial para a obtenção do título de Doutor em Modelagem Matemática e Computacional.

Área de concentração: Modelagem Matemática e Computacional

Linha de pesquisa: Sistemas Inteligentes

Orientador: Prof. Dr. Gray Farias Moita
Centro Federal de Educação Tecnológica de Minas Gerais

Coorientador: Prof. Dr. Alberto Henrique Frade Laender
Universidade Federal de Minas Gerais

Belo Horizonte
Setembro de 2016



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL

“UM ESTUDO SOBRE A PRODUÇÃO CIENTÍFICA BRASILEIRA A PARTIR DE DADOS DA PLATAFORMA LATTES”

Tese de Doutorado apresentada por **Thiago Magela Rodrigues Dias**, em 18 de outubro de 2016, ao Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG, e aprovada pela banca examinadora constituída pelos professores:

Prof. Dr. Gray Farias Moita (Orientador)
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Alberto Henrique Frade Laender (Coorientador)
Universidade Federal de Minas Gerais

Prof. Dr. Nelson Francisco Favilla Ebecken
Universidade Federal do Rio de Janeiro

Prof. Dr. Wagner Meira Júnior
Universidade Federal de Minas Gerais

Prof. Dr. Rodrigo Tomás Nogueira Cardoso
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Paulo Eduardo Maciel de Almeida
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida a impressão,

Profª. Drª. Elizabeth Fialho Wanner
Coordenadora do Programa de Pós-Graduação em
Modelagem Matemática e Computacional

Dedico esta tese à minha esposa Patrícia,
e aos meus filhos Guilherme e Eduarda.

AGRADECIMENTOS

Agradeço a todos que me incentivaram, me motivaram e contribuíram de alguma forma no desenvolvimento desta tese. Em especial, agradeço aos meus orientadores que estão nessa jornada comigo há muito tempo:

- Professor Gray Farias Moita, que me acolheu e que tem me orientado com muita competência desde o mestrado, sempre foi exemplo de comprometimento, conhecimento e dedicação. Agradeço pela orientação, pela paciência e, principalmente, pela amizade. Ao longo de dez anos de orientação, tenho o privilégio de ser a aresta mais densa em sua rede de colaboração, resultado de vários trabalhos que me possibilitaram crescer muito como orientado e acadêmico. Hoje, um amigo.
- Professor Alberto Henrique Frade Laender, que me recebeu e abriu as portas do LBD, Laboratório de Bancos de Dados do Departamento de Ciência da Computação da UFMG para que eu pudesse aprender grande parte do que foi utilizado no desenvolvimento desta tese. Seu comprometimento e sua dedicação em suas pesquisas foi motivação na busca de resultados cada vez melhores. Sua capacidade em realizar pesquisas de excelência e sua paciência em minha orientação proporcionaram obter resultados significativos em nossa pesquisa. Suas considerações e seu empenho nas revisões foram fundamentais para que pudesse alcançar os resultados apresentados.

Agradeço, ainda, ao CEFET-MG, instituição na qual tenho o privilégio de ser docente e discente, e que sempre proporcionou um ambiente favorável para que eu pudesse exercer as minhas atividades como aluno e professor. Em especial, agradeço a todos os professores do PPGMMC (Programa de Pós-Graduação em Modelagem Matemática e Computacional) que contribuíram significativamente em minha formação. Agradeço, também, ao DCC (Departamento de Ciência da Computação) da UFMG. As diversas disciplinas que lá cursei tiveram contribuição significativa no desenvolvimento desta tese.

Também importante registrar meu agradecimento aos Professores Paulo Eduardo Maciel de Almeida e Wagner Meira Júnior, que participaram da minha banca de defesa de qualificação, pelas importantes considerações, análises e sugestões que incentivaram a continuidade e o aprimoramento dos estudos descritos nesta tese.

Por fim, agradeço imensamente aos familiares e amigos que, mesmo alguns não compreendendo de forma clara o que é um doutorado, sabiam que era necessário incentivar, e isso foi importante na caminhada.

“Se enxerguei mais longe, foi por estar
sobre ombros de gigantes ”.
(Isaac Newton)

RESUMO

Os estudos sobre dados de produções científicas têm recebido atenção de pesquisadores, de diversas áreas, que visam obter conhecimento sobre a evolução das pesquisas em geral. Tais estudos possibilitam a análise da produção científica para diversos propósitos e um dos desafios neste tipo de análise está na diversidade de repositórios contendo dados em formatos e estruturas distintas. Além disso, problemas como redundância dos dados, ambiguidade e dados incorretos tornam as análises ainda mais complexas. Logo, a maioria dos trabalhos atuais visa realizar análises sobre conjuntos restritos de dados científicos que, frequentemente, são originados de repositórios específicos e com uma quantidade limitada de registros. Apesar de tais análises apresentarem resultados significantes, estes não conseguem expressar, de forma geral, grandes conjuntos com extensa quantidade de pesquisadores. Assim, esta tese apresenta um estudo sobre a produção científica brasileira, tendo como fonte de dados os currículos cadastrados na Plataforma Lattes. Os currículos da Plataforma Lattes se caracterizam, atualmente, como importante ferramenta para que pesquisadores, acadêmicos e estudantes registrem seus dados, sendo amplamente utilizados, caracterizando-se como um dos maiores repositórios de dados sobre produção científica, técnica, artística e profissional, contendo milhões de indivíduos cadastrados. Para a coleta, o tratamento e a análise dos dados, foi desenvolvido um arcabouço denominado LattesDataXplorer, responsável pela extração de todo o conjunto de dados dos currículos compondo um grande repositório de dados científicos. No arcabouço, também são implementadas técnicas para análises bibliométricas dos dados e métricas baseadas em análises de redes sociais para estudo das redes de colaboração científica. Como resultado, inicialmente, é apresentada uma caracterização de todo o repositório de currículos cadastrados na Plataforma Lattes e, ainda, um retrato da produção científica brasileira. Esses estudos objetivam apresentar uma visão detalhada sobre a produção científica brasileira e como as pesquisas estão sendo realizadas nas diversas áreas do conhecimento. Posteriormente, análises bibliométricas e baseadas em redes sociais são realizadas com o conjunto de publicações dos indivíduos com doutorado concluído e, finalmente, a produção científica dos docentes que têm atuado em programas de pós-graduação, bem como, dos pesquisadores de excelência em suas pesquisas é analisada. Os resultados a serem apresentados possuem caráter inédito tendo em vista a abrangência das análises realizadas e pela grande quantidade de indivíduos considerados.

PALAVRAS-CHAVE: Plataforma Lattes, Bibliometria, Produção Científica, Redes de Colaboração Científica.

ABSTRACT

The investigations on scientific productions data have received the attention of researchers from various fields that aims to acquiring knowledge about the evolution of the research in general terms. Such studies allow for the analysis of the scientific production for various purposes and one of the challenges in this type of analysis is in the diversity of the repositories containing data in distinct formats and structures. In addition, problems such as data redundancy, ambiguity and incorrect data make the analyses even more complex. Hence, most of the current works intend to carry out analyses on restricted sets of scientific data that are often originated from specific repositories and with a limited amount of records. Even though such analyses give significant results, they fail to encompass, in general, large datasets with large amount of researchers. Therefore, this thesis presents a study of the Brazilian scientific production having the curricula registered in the Lattes Platform as data source. The Lattes Platform curricula are currently characterised as an important tool in which researchers, scholars and students record their academic and technical data. It is widely used and is characterised as one of the largest data repositories of scientific, technical, artistic and professional production, containing millions of registered people. For the collection, processing and analysis of data, a framework called LattesDataXplorer was developed; it is responsible for the extraction of the entire dataset of curricula yielding a large repository of scientific data. In the framework, techniques for bibliometric analyses of the data and metrics based on social networks analyses to the study of scientific collaboration networks are also implemented. As a result, a characterization of the entire repository of curricula registered in the Lattes Platform is initially presented, in addition to a snapshot of the Brazilian scientific production. These studies aim to present a detailed view of the Brazilian scientific production and how the researches are being carried out in the different areas of knowledge. Furthermore, bibliometric analyses (based on social networks) are performed with the set of publications of doctors and finally the scientific production of teachers that are working in postgraduate programs as well as researchers of excellence in their research work are analyzed. The presented results hold unprecedented nature in the view of the scope of analyses performed and from the amount of individuals considered.

KEYWORDS: Lattes Platform, Bibliometrics, Scientific Production, Scientific Collaboration Networks.

LISTA DE FIGURAS

FIGURA 2.1: EXEMPLO DE UMA REDE DE COAUTORIA COM CONCEITOS BÁSICOS DE REDES SOCIAIS	31
FIGURA 2.2: EXEMPLO DE UMA REDE COM CÁLCULO DAS MÉTRICAS E SEUS RESPECTIVOS VALORES	34
FIGURA 3.1: REDE DE PALAVRAS-CHAVE PARA A IDENTIFICAÇÃO DE NÓS COM MAIOR GRAU (ZHU ET AL., 2013)	40
FIGURA 4.1: CRESCIMENTO DA PLATAFORMA LATTES ENTRE MAIO DE 2014 E JUNHO DE 2016	49
FIGURA 4.2: VISÃO GERAL DO LATTESDATAEXPLORER	50
FIGURA 4.3: PROCESSO DE EXTRAÇÃO E SELEÇÃO DE DADOS	50
FIGURA 4.4: INTERFACE DE BUSCA DA PLATAFORMA LATTES	51
FIGURA 4.5: CABEÇALHO DE UM CURRÍCULO LATTES	52
FIGURA 4.6: PROCESSO DE TRATAMENTO DOS DADOS	55
FIGURA 4.7: EXEMPLO DE LISTA DE COLABORADORES VINCULADOS A SEUS IDENTIFICADORES	58
FIGURA 4.8: COMPONENTE PARA CÁLCULO DE ESTATÍSTICAS E MÉTRICAS	61
FIGURA 5.1: DISTRIBUIÇÃO DOS CURRÍCULOS PELA DATA DA ÚLTIMA ATUALIZAÇÃO	64
FIGURA 5.2: NÍVEL MAIS ALTO DE FORMAÇÃO CONCLUÍDA DOS INDIVÍDUOS	64
FIGURA 5.3: TOTAL DE PUBLICAÇÕES POR ANO	66
FIGURA 5.4: ARTIGOS EM PERIÓDICO DOS 20 PAÍSES MAIS PRODUTIVOS. (SCIMAGO (2016))	67
FIGURA 5.5: MÉDIA DE AUTORES POR TRABALHO	69
FIGURA 5.6: PROFICIÊNCIA DOS INDIVÍDUOS EM IDIOMAS	70
FIGURA 5.7: NÍVEL DE PROFICIÊNCIA DOS INDIVÍDUOS	71
FIGURA 5.8: DISTRIBUIÇÃO DOS INDIVÍDUOS POR GRANDE ÁREA DE ATUAÇÃO	72
FIGURA 5.9: DISTRIBUIÇÃO DOS INDIVÍDUOS POR ÁREA DE ATUAÇÃO	73
FIGURA 5.10: DISTRIBUIÇÃO DO NÚMERO DE PUBLICAÇÕES EM ANAIS DE CONGRESSO	77
FIGURA 5.11: DISTRIBUIÇÃO DO NÚMERO DE PUBLICAÇÕES EM PERIÓDICO	77
FIGURA 5.12: DISTRIBUIÇÃO DA QUANTIDADE DE CURRÍCULOS PELO TOTAL DE ARTIGOS EM ANAIS DE CONGRESSO	78
FIGURA 5.13: DISTRIBUIÇÃO DA QUANTIDADE DE CURRÍCULOS PELO TOTAL DE PUBLICAÇÕES EM PERIÓDICO	78
FIGURA 5.14: DISTRIBUIÇÃO DOS ARTIGOS EM ANAIS DE CONGRESSO NOS CURRÍCULOS COM PRODUÇÃO	80
FIGURA 5.15: DISTRIBUIÇÃO DOS ARTIGOS EM PERIÓDICO NOS CURRÍCULOS COM PRODUÇÃO	81
FIGURA 6.1: REPRESENTATIVIDADE DOS DOUTORES NA PLATAFORMA LATTES	85
FIGURA 6.2: DATA DA ÚLTIMA ATUALIZAÇÃO DOS CURRÍCULOS DOS DOUTORES	86
FIGURA 6.3: TAXA DE PUBLICAÇÃO DOS DOUTORES POR TIPO DE PRODUÇÃO	86
FIGURA 6.4: DISTRIBUIÇÃO DA QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO POR DOUTOR	87
FIGURA 6.5: DISTRIBUIÇÃO DA QUANTIDADE DE ARTIGOS EM PERIÓDICO POR DOUTOR	87
FIGURA 6.6: QUANTIDADE DE DOUTORES FORMADOS POR ANO	88
FIGURA 6.7: PRODUÇÃO CIENTÍFICA DOS DOUTORES POR ANO	89
FIGURA 6.8: PRODUÇÃO DE ARTIGOS EM ANAIS DE CONGRESSO DOS GRUPOS CONSIDERADOS	91
FIGURA 6.9: PRODUÇÃO PER CAPITA DOS DOUTORES POR ANO	91
FIGURA 6.10: QUANTIDADE DE PUBLICAÇÕES E QUANTIDADE DE AUTORES POR ANO	92
FIGURA 6.11: PRODUÇÃO DOS DOUTORES FORMADOS NO BRASIL E EM INSTITUIÇÕES ESTRANGEIRAS	93
FIGURA 6.12: DISTRIBUIÇÃO DA QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO PÓS-DOUTORADO	94
FIGURA 6.13: DISTRIBUIÇÃO DA QUANTIDADE DE ARTIGOS EM PERIÓDICO PÓS-DOUTORADO	94
FIGURA 6.14: QUANTIDADE DE COAUTORES EM ARTIGOS PUBLICADOS EM ANAIS DE CONGRESSO E EM PERIÓDICO DO CONJUNTO DE DOUTORES	95
FIGURA 6.15: QUANTIDADE DE COAUTORES EM ARTIGOS DO CONJUNTO DE DOUTORES NAS GRANDES ÁREAS	96
FIGURA 6.16: MÉDIA DE AUTORES POR TRABALHO COM PELO MENOS UM DOUTOR	97
FIGURA 6.17: REDES GERAIS DE COLABORAÇÃO CIENTÍFICA CONTENDO TODOS OS NÓS	99
FIGURA 6.18: COMPONENTES GIGANTES	100
FIGURA 6.19: EVOLUÇÃO TEMPORAL DO COMPONENTE GIGANTE	101
FIGURA 6.20: EVOLUÇÃO DO DIÂMETRO DA REDE	101
FIGURA 6.21: EVOLUÇÃO DO CAMINHO MÍNIMO MÉDIO	102
FIGURA 6.22: DISTRIBUIÇÃO DOS DOUTORES POR GRANDE ÁREA DE ATUAÇÃO	106
FIGURA 6.23: DISTRIBUIÇÃO DOS DOUTORES POR ÁREA DE ATUAÇÃO	107
FIGURA 6.24: TAXA DE PUBLICAÇÃO DE ARTIGOS EM ANAIS DE CONGRESSO POR ÁREA	109

FIGURA 6.25: TAXA DE PUBLICAÇÃO DE ARTIGOS EM PERIÓDICO POR ÁREA	109
FIGURA 6.26: DISTRIBUIÇÃO DOS TIPOS DE PUBLICAÇÃO POR ÁREA	111
FIGURA 6.27: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO POR GRANDE ÁREA	111
FIGURA 6.28: QUANTIDADE DE ARTIGOS EM PERIÓDICO POR GRANDE ÁREA	112
FIGURA 6.29: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO POR ÁREAS	113
FIGURA 6.30: QUANTIDADE DE ARTIGOS EM PERIÓDICO POR ÁREAS	113
FIGURA 7.1: DISTRIBUIÇÃO DOS DOCENTES DE PÓS-GRADUAÇÃO PELAS SUAS GRANDES ÁREAS DE ATUAÇÃO	121
FIGURA 7.2: DISTRIBUIÇÃO DAS ORIENTAÇÕES CONCLUÍDAS CONSIDERANDO A GRANDE ÁREA DO ORIENTADOR	122
FIGURA 7.3: CORRELAÇÃO ENTRE NÚMERO DE ORIENTADORES E ORIENTAÇÕES CONCLUÍDAS NAS GRANDES ÁREAS	123
FIGURA 7.4: DISTRIBUIÇÃO DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA POR MODALIDADE DE BOLSA	126
FIGURA 7.5: TEMPO DE DOUTORADO EM ANOS DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA	127
FIGURA 7.6: DISTRIBUIÇÃO DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA POR GRANDES ÁREAS	128
FIGURA 7.7: PRODUÇÃO CIENTÍFICA DOS DOCENTES DE PÓS-GRADUAÇÃO	130
FIGURA 7.8: PRODUÇÃO CIENTÍFICA DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA	131
FIGURA 7.9: PRODUÇÃO CIENTÍFICA DOS DOCENTES DE PÓS-GRADUAÇÃO E BOLSISTAS DE PRODUTIVIDADE EM PESQUISA	133
FIGURA 7.10: TEMPO DE CARREIRA EM ANOS APÓS A CONCLUSÃO DO DOUTORADO	133
FIGURA 7.11: CORRELAÇÃO ENTRE NÚMERO DE ARTIGOS EM ANAIS DE CONGRESSO E ORIENTAÇÕES CONCLUÍDAS	135
FIGURA 7.12: CORRELAÇÃO ENTRE NÚMERO DE ARTIGOS EM PERIÓDICO E ORIENTAÇÕES CONCLUÍDAS	135
FIGURA 7.13: DISTRIBUIÇÃO DOS ARTIGOS EM ANAIS DE CONGRESSO POR MODALIDADE DE BOLSA DE PESQUISA	135
FIGURA 7.14: DISTRIBUIÇÃO DOS ARTIGOS EM PERIÓDICO POR MODALIDADE DE BOLSA DE PESQUISA	136
FIGURA 7.15: DISTRIBUIÇÃO DA MEDIANA DO FATOR DE IMPACTO DAS PUBLICAÇÕES DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA	138
FIGURA 7.16: DISTRIBUIÇÃO DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA PELOS RESPECTIVOS VALORES DO ÍNDICE-H	143
FIGURA 7.17: DISTRIBUIÇÃO DOS VALORES DOS ÍNDICES-H PELAS MODALIDADES DAS BOLSAS	143
FIGURA 7.18: ANÁLISE DOS VALORES DOS ÍNDICES-H DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA DE ACORDO COM AS RESPECTIVAS GRANDES ÁREAS	145
FIGURA A.1: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO – CIÊNCIAS AGRÁRIAS	165
FIGURA A.2: QUANTIDADE DE ARTIGOS EM PERIÓDICO – CIÊNCIAS AGRÁRIAS	165
FIGURA A.3: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO – CIÊNCIAS BIOLÓGICAS	166
FIGURA A.4: QUANTIDADE DE ARTIGOS EM PERIÓDICO – CIÊNCIAS BIOLÓGICAS	166
FIGURA A.5: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO – CIÊNCIAS DA SAÚDE	167
FIGURA A.6: QUANTIDADE DE ARTIGOS EM PERIÓDICO – CIÊNCIAS DA SAÚDE	167
FIGURA A.7: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO – CIÊNCIAS EXATAS E DA TERRA	168
FIGURA A.8: QUANTIDADE DE ARTIGOS EM PERIÓDICO – CIÊNCIAS EXATAS E DA TERRA	168
FIGURA A.9: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO – CIÊNCIAS HUMANAS	169
FIGURA A.10: QUANTIDADE DE ARTIGOS EM PERIÓDICO – CIÊNCIAS HUMANAS	169
FIGURA A.11: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO – CIÊNCIAS SOCIAIS APLICADAS	170
FIGURA A.12: QUANTIDADE DE ARTIGOS EM PERIÓDICO – CIÊNCIAS SOCIAIS APLICADAS	170
FIGURA A.13: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO – ENGENHARIAS	171
FIGURA A.14: QUANTIDADE DE ARTIGOS EM PERIÓDICO – ENGENHARIAS	171
FIGURA A.15: QUANTIDADE DE ARTIGOS EM ANAIS DE CONGRESSO – LINGUÍSTICA, LETRAS E ARTES	172
FIGURA A.16: QUANTIDADE DE ARTIGOS EM PERIÓDICO – LINGUÍSTICA, LETRAS E ARTES	172
FIGURA A.17: MÉDIA DE CITAÇÕES EM PERIÓDICO – CIÊNCIAS AGRÁRIAS	177
FIGURA A.18: MÉDIA DE CITAÇÕES EM PERIÓDICO – CIÊNCIAS BIOLÓGICAS	177
FIGURA A.19: MÉDIA DE CITAÇÕES EM PERIÓDICO – CIÊNCIAS DA SAÚDE	178
FIGURA A.20: MÉDIA DE CITAÇÕES EM PERIÓDICO – CIÊNCIAS EXATAS E DA TERRA	178
FIGURA A.21: MÉDIA DE CITAÇÕES EM PERIÓDICO – CIÊNCIAS HUMANAS	179
FIGURA A.22: MÉDIA DE CITAÇÕES EM PERIÓDICO – CIÊNCIAS SOCIAIS APLICADAS	179
FIGURA A.23: MÉDIA DE CITAÇÕES EM PERIÓDICO – ENGENHARIAS	180
FIGURA A.24: MÉDIA DE CITAÇÕES EM PERIÓDICO – LINGUÍSTICA, LETRAS E ARTES	180
FIGURA A.25: MÉDIA DE CITAÇÕES – BOLSISTAS	181

LISTA DE TABELAS

TABELA 4.1: EXEMPLO DE TRANSFORMAÇÃO DOS TÍTULOS EM CHAVES PARA O DICIONÁRIO DE IDENTIFICAÇÃO	57
TABELA 4.2: EXEMPLO DE DICIONÁRIO DE IDENTIFICAÇÃO	57
TABELA 5.1: QUANTITATIVO DOS DADOS DA PLATAFORMA LATTES EM ABRIL DE 2015	63
TABELA 5.2: DISTRIBUIÇÃO GEOGRÁFICA DOS INDIVÍDUOS	65
TABELA 5.3: TOTAL DE PUBLICAÇÕES E TAXA DE CRESCIMENTO POR ANO	67
TABELA 5.4: ÁREAS DE PESQUISA COM A MAIOR QUANTIDADE DE CURRÍCULOS	74
TABELA 5.5: DISTRIBUIÇÃO DOS CURRÍCULOS POR GRANDE ÁREA E MAIOR NÍVEL DE FORMAÇÃO CONCLUÍDA	75
TABELA 5.6: PUBLICAÇÕES POR GRANDE ÁREA DE PESQUISA E MÉDIA POR INDIVÍDUO	76
TABELA 5.7: QUANTIDADE E PERCENTUAL DE CURRÍCULOS SEM PRODUÇÃO REGISTRADA	76
TABELA 5.8: CÁLCULO DA MÉDIA, DESVIO PADRÃO E MEDIANA	79
TABELA 5.9: CURRÍCULOS SEM PRODUÇÃO REGISTRADA	80
TABELA 5.10: INDIVÍDUOS MAIS PRODUTIVOS EM CADA GRANDE ÁREA	82
TABELA 5.11: NÚMERO DE PUBLICAÇÕES POR MAIOR NÍVEL DE FORMAÇÃO	83
TABELA 6.1: PRODUÇÃO CIENTÍFICA DOS DOUTORES E TAXA DE CRESCIMENTO POR ANO	89
TABELA 6.2: TAXA DE CRESCIMENTO DOS TRABALHOS PUBLICADOS E DOS AUTORES	93
TABELA 6.3: RESULTADO DAS MÉTRICAS ADOTADAS	102
TABELA 6.4: ÁREAS COM A MAIOR QUANTIDADE DE DOUTORES	108
TABELA 7.1: DISTRIBUIÇÃO DOS DOCENTES DE PÓS-GRADUAÇÃO POR ESTADOS E NAS PRINCIPAIS INSTITUIÇÕES	119
TABELA 7.2: QUANTITATIVO DE ORIENTAÇÕES CONCLUÍDAS IDENTIFICADAS EM TODOS OS CURRÍCULOS DA PLATAFORMA LATTES E AS ORIENTADAS PELO CONJUNTO DE DOCENTES	120
TABELA 7.3: QUANTITATIVO DE ORIENTAÇÕES EM ANDAMENTO IDENTIFICADAS EM TODOS OS CURRÍCULOS DA PLATAFORMA LATTES E AS ORIENTADAS PELO CONJUNTO DE DOCENTES	120
TABELA 7.4: ÁREAS COM A MAIOR QUANTIDADE DE DOCENTES	121
TABELA 7.5: DOCENTES COM O MAIOR NÚMERO DE ORIENTAÇÕES CONCLUÍDAS	123
TABELA 7.6: INSTITUIÇÕES COM O MAIOR NÚMERO DE BOLSISTAS DE PRODUTIVIDADE EM PESQUISA	125
TABELA 7.7: ÁREAS DE ATUAÇÃO COM A MAIOR QUANTIDADE DE BOLSISTAS DE PRODUTIVIDADE EM PESQUISA	128
TABELA 7.8: QUANTIDADE DE BOLSISTAS DE PRODUTIVIDADE EM PESQUISA POR GRANDE ÁREA EM CADA MODALIDADE	129
TABELA 7.9: DISTRIBUIÇÃO DA PRODUÇÃO CIENTÍFICA DOS INDIVÍDUOS DOS PRINCIPAIS CONJUNTOS ANALISADOS	132
TABELA 7.10: DOCENTES DE PÓS-GRADUAÇÃO COM O MAIOR NÚMERO DE ORIENTAÇÕES CONCLUÍDAS	134
TABELA 7.11: MÉDIA, MEDIANA E DESVIO PADRÃO DAS CITAÇÕES POR ESTRATO DO FATOR DE IMPACTO DOS PERIÓDICOS DE CADA GRANDE ÁREA	139
TABELA 7.12: PERCENTUAL DE ARTIGOS QUE NÃO POSSUEM CITAÇÕES EM CADA GRANDE ÁREA	141
TABELA 7.13: ARTIGOS COM MAIOR QUANTIDADE DE CITAÇÕES	141
TABELA 7.14: BOLSISTAS DE PRODUTIVIDADE EM PESQUISA COM OS MAIORES VALORES DE ÍNDICE-H EM CADA MODALIDADE DE BOLSA	144
TABELA A.1: PERCENTUAL DE CRESCIMENTO NA PRODUÇÃO DE ARTIGOS EM ANAIS DE CONGRESSO	173
TABELA A.2: PERCENTUAL DE CRESCIMENTO NA PRODUÇÃO DE ARTIGOS EM PERIÓDICO	175

LISTA DE ALGORITMOS

ALGORITMO 4.1: ETAPAS DO MÉTODO PARA CARACTERIZAÇÃO DE REDES DE COLABORAÇÃO	56
---	----

LISTA DE ABREVIATURAS E SIGLAS

CA	Comitê de Assessoramento
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
C&T	Ciência e Tecnologia
DBLP	<i>Digital Bibliography & Library Project</i>
DOI	<i>Digital Object Identifier</i>
FI	Fator de Impacto
Fiocruz	Fundação Oswaldo Cruz
INCT	Institutos Nacionais de Ciência e Tecnologia
ISI	<i>Institute for Scientific Information</i>
ISSN	<i>International Standard Serial Number</i>
PUC-RJ	Pontifícia Universidade Católica do Rio de Janeiro
UEL	Universidade Estadual de Londrina
UEM	Universidade Estadual de Maringá
UERJ	Universidade Estadual do Rio de Janeiro
UFBA	Universidade Federal da Bahia
UFC	Universidade Federal do Ceará
UFES	Universidade Federal do Espírito Santo
UFF	Universidade Federal Fluminense
UFG	Universidade Federal de Goiás
UFMG	Universidade Federal de Minas Gerais
UFPA	Universidade Federal do Pará
UFPB	Universidade Federal da Paraíba
UFPE	Universidade Federal de Pernambuco
UFPR	Universidade Federal do Paraná
UFRGS	Universidade Federal do Rio Grande do Sul
UFRJ	Universidade Federal do Rio de Janeiro
UFRN	Universidade Federal do Rio Grande do Norte
UFSC	Universidade Federal de Santa Catarina
UFSCAR	Universidade Federal de São Carlos
UFSM	Universidade Federal de Santa Maria
UFU	Universidade Federal de Uberlândia
UFV	Universidade Federal de Viçosa
UNB	Universidade de Brasília
UNESP	Universidade Estadual Paulista Júlio de Mesquita Filho
UNICAMP	Universidade Estadual de Campinas
UNIFESP	Universidade Federal de São Paulo
USP	Universidade de São Paulo
URL	<i>Uniform Resource Locator</i>

XML	<i>eXtensible Markup Language</i>
XPath	<i>XML Path Language</i>
Web	<i>World Wide Web</i>
WOS	<i>Web of Science</i>

SUMÁRIO

1. INTRODUÇÃO	18
1.1 DESCRIÇÃO DO PROBLEMA.....	21
1.2 OBJETIVOS E CONTRIBUIÇÕES	22
1.3 ESTRUTURA DA TESE	23
2. FUNDAMENTAÇÃO CONCEITUAL.....	25
2.1 BIBLIOMETRIA	25
2.2 ANÁLISE DE REDES SOCIAIS.....	28
2.2.1 CONCEITOS BÁSICOS.....	30
2.2.2 MÉTRICAS PARA ANÁLISE DE REDES SOCIAIS	32
3. TRABALHOS RELACIONADOS	36
3.1 ANÁLISES BIBLIOMÉTRICAS E BASEADAS EM REDES.....	36
3.2 TRABALHOS ENVOLVENDO DADOS DA PLATAFORMA LATTES	41
4. COLETA E TRATAMENTO DOS DADOS.....	48
4.1 MOTIVAÇÕES	48
4.2 LATTESDATAEXPLORER	49
4.2.1 SELEÇÃO DE GRUPOS	54
4.2.2 IDENTIFICAÇÃO E CARACTERIZAÇÃO DE REDES DE COLABORAÇÃO CIENTÍFICA	55
4.2.3 CÁLCULO DE ESTATÍSTICAS E MÉTRICAS.....	60
5. CARACTERIZAÇÃO DOS DADOS DA PLATAFORMA LATTES.....	62
5.1 ESTATÍSTICAS GERAIS.....	62
5.2 ESTATÍSTICAS POR ÁREAS E GRANDES ÁREAS.....	71
6. ANÁLISE DA PRODUÇÃO CIENTÍFICA DOS DOUTORES	84
6.1 CARACTERIZAÇÃO GERAL.....	84
6.2 ANÁLISE DA PRODUÇÃO CIENTÍFICA DOS DOUTORES	89
6.3 ANÁLISE DAS REDES DE COLABORAÇÃO CIENTÍFICA	95
6.4 ANÁLISE POR ÁREAS DE ATUAÇÃO	106
6.5 CONSIDERAÇÕES FINAIS.....	115
7. ANÁLISE DA PRODUÇÃO CIENTÍFICA DOS DOCENTES DE PÓS-GRADUAÇÃO E DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA DO CNPQ.....	117
7.1 CARACTERIZAÇÃO GERAL DOS CONJUNTOS	117
7.1.1 DOCENTES DE PÓS-GRADUAÇÃO	118
7.1.2 BOLSISTAS DE PRODUTIVIDADE EM PESQUISA DO CNPQ	124
7.2 ANÁLISE QUANTITATIVA DA PRODUÇÃO CIENTÍFICA DOS DOCENTES DE PÓS-GRADUAÇÃO E DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA.....	130
7.3 ANÁLISE QUALITATIVA DA PRODUÇÃO CIENTÍFICA DOS BOLSISTAS DE PRODUTIVIDADE EM PESQUISA	137
7.4 CONSIDERAÇÕES FINAIS.....	146
8. CONCLUSÕES.....	148

8.1 CONTRIBUIÇÕES	148
8.2 PUBLICAÇÕES	150
8.3 EXTENSÕES PARA POSSÍVEIS TRABALHOS FUTUROS	153
REFERÊNCIAS.....	155
APÊNDICES	165

Capítulo 1

Introdução

A produção e a publicação de trabalhos científicos têm aumentado significativamente nos últimos anos, sendo a *World Wide Web* (Web) o principal facilitador para acesso e distribuição das publicações científicas. Serviços como bibliotecas digitais, redes sociais acadêmicas e sítios para registro individual de produção científica são alguns exemplos de ferramentas tecnológicas que têm contribuído para aumentar a quantidade de trabalhos publicados, permitindo que usuários não apenas acessem o conteúdo disponível, mas também possam registrar a sua própria produção técnica e científica interagindo diretamente com essas ferramentas. Com isso, trabalhos publicados e disponibilizados eletronicamente podem ser acessados de forma instantânea, contribuindo de forma significativa para a disseminação do conhecimento científico.

Dentre as principais ferramentas atualmente disponíveis, podemos citar a DBLP¹ (*Digital Bibliography & Library Project*), AMiner², Google Scholar³, Semantic Scholar⁴, Microsoft Academic Search⁵, PubMed⁶ e Scopus⁷ que mantêm não só registros de citação bibliográfica (*bibliographic citation records*), mas também incluem um apontador para a cópia eletrônica da respectiva publicação ou até mesmo armazenam essa cópia localmente. Mais recentemente, as redes sociais acadêmicas, como ResearchGate⁸ e Academia⁹, também têm contribuído para a disseminação da produção científica, já que possibilitam não só o compartilhamento de dados, mas também a interação entre os seus usuários.

De acordo com Ferreira (2010), a ciência produz conhecimento havendo a necessidade e o compromisso de tornar seus resultados públicos, a partir da divulgação junto à

¹ DBLP: <http://dblp.uni-trier.de/db/>

² AMiner: <http://aminer.org/>

³ Google Scholar: <http://scholar.google.com>

⁴ Semantic Scholar: <https://www.semanticscholar.org/>

⁵ Microsoft Academic Search: <http://academic.research.microsoft.com/>

⁶ PubMed: <http://www.ncbi.nlm.nih.gov/pubmed>

⁷ Scopus: <http://www.elsevier.com/online-tools/scopus>

⁸ ResearchGate: <http://www.researchgate.net/>

⁹ Academia: <http://www.academia.edu>

comunidade científica das atividades realizadas durante a pesquisa. Candotti (2002), por sua vez, destaca que disponibilizar novas descobertas e inovações originadas de pesquisas científicas de forma simples é um dos principais objetivos da ciência, independentemente de qual seja a área.

Nesse cenário, a Bibliometria (ARAÚJO, 2006), que possui grande relação com a Cientometria (SILVA; HAYASHI; HAYASHI, 2011), se destaca como uma das principais ciências métricas de análise de conteúdo, podendo ser aplicada a fontes de dados científicos com o intuito de se obter informações quantitativas sobre publicações (PRITCHARD, 1969). Diante disso, é possível analisar dados de publicações científicas com o objetivo de identificar as tendências e o crescimento do conhecimento em diversas áreas, prever padrões de pesquisa, observar a dispersão do conhecimento científico, auxiliar políticas de auxílio à pesquisa e entender como ocorre a evolução científica de uma determinada área do conhecimento ou de grupos de pesquisadores. A produção de indicadores bibliométricos considerados mais representativos tornou-se realidade concreta nas últimas décadas do século XX. Os principais motivos para a proposta de tais indicadores devem-se à construção, manutenção e informatização de repositórios de dados científicos (MUGNAINI; JANNUZZI; QUONIAM, 2004). Por sua vez, um dos fatores que têm motivado os estudos de avaliação da produção científica, principalmente com a adoção de métricas bibliométricas, tem sido a busca pela excelência em áreas de pesquisa e, também, a competição pelos recursos financeiros das agências de fomento (LOPES, 2012).

Nos últimos anos, além da análise bibliométrica de publicações científicas, diversos outros estudos têm procurado compreender como a ciência tem evoluído e como a colaboração científica ocorre. Diante disso, técnicas baseadas em análises de redes surgem como uma alternativa para verificar esse fenômeno. De modo geral, uma rede pode ser caracterizada como um grafo, que consiste de um conjunto de nós (vértices) e ligações (arestas) entre os nós (SZWARCFITER, 1986). Essas ligações podem ser direcionadas ou não e, opcionalmente, podem ainda ter um peso associado. Diversos fenômenos naturais podem ser descritos em termos de uma rede. O cérebro pode ser caracterizado como uma rede de neurônios ligados por sinapses. A Internet também é um exemplo de uma rede importante e amplamente estudada nos dias atuais. Em redes rodoviárias, as cidades e as rodovias que as interligam podem ser consideradas como um grafo, em que as cidades se caracterizam como nós e as rodovias como arestas. Já nas redes de comunicação, computadores interligados por um meio de comunicação constituem um grafo no qual nós representam computadores e arestas representam ligações físicas entre eles (LIBEN-NOWELL; KLEINBERG, 2007).

Entre os vários tipos de rede, existem as redes sociais. Uma rede social é um conjunto de pessoas ou grupos que têm algum tipo de relação entre eles (NEWMAN, 2001a). As redes sociais têm sido objeto de estudos há muitos anos. Barabási e Albert (1999) apresentam diversos estudos que motivaram e agregaram valor à teoria dos grafos, já que faziam uso do conceito de redes sociais. Em uma rede social de amigos, a relação entre duas pessoas pode representar uma amizade entre elas. Em uma rede de relações de parentesco entre pessoas, a relação pode indicar que duas pessoas pertencem à mesma família.

No domínio científico, um exemplo de uma rede social é a de colaboração científica, que pode ser observada como uma rede na qual os nós correspondem aos autores de publicações científicas e as arestas correspondem à relação de coautoria. Nesse tipo de rede, as arestas podem ou não ser ponderadas. A adição de um peso representa o número de trabalhos em que os autores relacionados pela aresta considerada participaram conjuntamente. Dessa forma, a intensidade dos relacionamentos presentes em uma rede de colaboração científica é medida pelo número de colaborações entre um par de autores. A presença do peso é útil para representar, por exemplo, a afinidade e os interesses comuns entre dois autores da rede (SONNENWALD, 2007).

Newman (2001a) apresenta uma avaliação extensa sobre características sociais das redes de colaboração científica nas áreas de computação, biologia, física e medicina no período de 1995 a 1999, encontrando diversas características sociais que são comuns a essas redes. Já em outro trabalho (NEWMAN, 2004), o autor analisa redes de coautoria nas áreas de biologia, física e matemática visando identificar propriedades estatísticas, como número médio de publicações e quantidade média de autores, de modo a determinar padrões comuns entre os elementos da rede analisada.

Apesar de existirem vários esforços no passado para análise de redes, foi com o avanço dos computadores que essa área de pesquisa ganhou grande interesse, já que se tornou possível o estudo de redes de grande porte. Tais redes podem englobar milhares de nós com milhões de arestas, o que era impossível de se modelar com as ferramentas disponíveis alguns anos atrás, já que os computadores possuíam recursos limitados se comparados aos atuais.

Com a modelagem e caracterização das redes de colaboração científica, é possível aplicar diversas técnicas que permitem entender como essas redes são estruturadas, fornecendo assim subsídios para diversos estudos como predição de vínculos entre pesquisadores, recomendação de especialistas e identificação de grupos de pesquisa. Porém, a identificação dessas colaborações em grandes repositórios de dados não é uma tarefa trivial devido a uma série de fatores, como ambiguidade entre nomes de autores, relação de autores incompleta, erros ortográficos nas citações ou nos títulos das publicações, dentre outros.

Diante disso, esta tese apresenta um estudo sobre a produção científica brasileira, tendo como base análises bibliométricas que visam quantificar a produção científica brasileira e análises de redes de colaboração que possibilitam compreender como ocorre a colaboração científica e como estão relacionados os grupos de pesquisa nas diversas áreas do conhecimento. Para isso, são analisados dados da Plataforma Lattes¹⁰ do CNPq¹¹ (Conselho Nacional de Desenvolvimento Científico e Tecnológico) responsável por armazenar informações curriculares da comunidade científica nacional. Atualmente a Plataforma Lattes é considerada uma importante fonte de dados sobre a produção científica brasileira (LANE, 2010).

¹⁰ Plataforma Lattes: <http://lattes.cnpq.br/>

¹¹ CNPq: <http://www.cnpq.br/>

Como grande parte dos programas de financiamento a pesquisa conduzidos no país pelas agências governamentais, como o próprio CNPq, utiliza os currículos dos pesquisadores cadastrados na Plataforma Lattes como um dos quesitos para avaliação das propostas de projeto submetidas aos seus editais, isso tem feito com que esses pesquisadores mantenham seus currículos atualizados, tornando a Plataforma Lattes uma fonte adequada para análise da produção científica brasileira. Por essa razão, vários trabalhos têm utilizado os currículos cadastrados na Plataforma Lattes como fonte de dados para análise e avaliação da produção científica brasileira nas diversas áreas do conhecimento (MENA-CHALCO; CESAR-JUNIOR, 2009; ALVES; YANASSE; SOMA, 2011; DIGIAMPIETRI et al., 2012; DIGIAMPIETRI et al., 2014b; BRITO; QUONIAM; MENA-CHALCO, 2016; SILVA et al., 2016).

1.1 Descrição do Problema

Uma nova geração de serviços disponíveis principalmente na Web está mudando a forma de divulgar e disponibilizar a produção científica e tecnológica. Existe, atualmente, uma tendência que reforça a troca de informações e a colaboração entre as pessoas. A forte relação entre os domínios científico e socioeconômico tem gerado um interesse crescente pela compreensão dos mecanismos que norteiam as atividades científicas, sendo possível apontar diversos trabalhos que analisam aspectos específicos como as características da linguagem e dos discursos empregados na comunicação científica (HOFFNAGEL, 2009) ou, ainda, a relação de colaboração entre pesquisadores e grupos de pesquisa (DING, 2011; REVOREDO et al., 2012; STROELE; ZIMBRÃO; SOUZA, 2012).

Para Mugnaini et al. (2014), o levantamento da produção científica de um país permite estudar diversos aspectos que podem ser qualificados como resultados mensuráveis de seu respectivo sistema de ciência, tecnologia e inovação. Acompanhar o fluxo de comunicação científica das diversas áreas facilita o processo de avaliação dos resultados de pesquisa, cujas características são tão diversificadas quanto o é a própria ciência. Para os autores, a análise quantitativa da ciência, que se dá a partir de indicadores, é tida como complementar à análise qualitativa, realizada pelos pares em diversas etapas, desde a formação, com a entrada e progressão na carreira de cientista, até a avaliação das pesquisas empreendidas, consubstanciada em manuscritos e títulos.

Com a competição cada vez mais acirrada entre as instituições de pesquisa, torna-se importante para seus integrantes determinar a abrangência de suas publicações e a qualidade de seus trabalhos quando comparados aos de outros grupos, bem como identificar potenciais colaboradores com o intuito de realizar trabalhos em colaboração visando impulsionar a sua produção científica e obter melhores resultados em suas pesquisas. É possível apontar trabalhos recentes nos quais é mostrado que grupos de pesquisa com uma rede social científica bem conectada tendem, geralmente, a ser mais produtivos (LOPES et al., 2011; BRANDÃO et al., 2013). Para Wanderley et al. (2014), evidências apontam que a forma como os pesquisadores colaboram tem forte impacto sobre a sua produtividade.

No entanto, o grande volume de dados sobre produção científica disponível em diferentes formatos e em diferentes repositórios dificulta a realização de estudos e consultas por parte de usuários que necessitam de uma visão unificada desses dados para, por exemplo, possibilitar a identificação de grupos de indivíduos que estejam trabalhando com determinado tema em diferentes instituições ou regiões. Além disso, o crescimento e a evolução da Web deram origem a uma grande quantidade de dados textuais pouco estruturados, heterogêneos e armazenados, sem nenhuma preocupação com padronização em diferentes repositórios. Tais características requerem muitas vezes a utilização de técnicas específicas para integração e conciliação de dados provenientes de diferentes fontes (LIU, 2011; CHRISTEN, 2012).

Estudos bibliométricos, principalmente em grandes repositórios bibliográficos, não são tarefas triviais tendo em vista a quantidade de dados a serem analisados e as características dos repositórios, que, em sua maioria, não possuem um padrão definido. Atualmente, grande parte desses estudos tem utilizado como principais fontes de dados resultados de consultas a repositórios internacionais que apresentam dados sobre trabalhos científicos, geralmente publicados em periódicos indexados. Entretanto, muitos desses repositórios negligenciam trabalhos publicados em periódicos nacionais que muitas vezes não são indexados e grande parte dos artigos publicados em anais de congressos, que constituem importante meio de publicação de algumas áreas do conhecimento como, por exemplo, a Ciência da Computação (LAENDER et al., 2008).

Logo, fica evidente a dificuldade que existe para se realizar estudos abrangentes que possam apresentar de forma ampla análises sobre toda a produção científica de um grande conjunto de indivíduos que estejam vinculados a diferentes instituições ou que atuam em áreas distintas, como, por exemplo, o conjunto de todos os pesquisadores de um determinado país. Diante disso, esta tese apresenta um estudo sobre a produção científica registrada nos currículos cadastrados na Plataforma Lattes, possibilitando, desta forma, apresentar um retrato dessa produção até então inédito, tendo em vista a quantidade de indivíduos e dados analisados.

1.2 Objetivos e Contribuições

Buscando um maior entendimento sobre como tem-se desenvolvido a pesquisa científica brasileira nas diversas áreas do conhecimento, esta tese tem como objetivo geral realizar um estudo da produção científica brasileira, investigando como as pesquisas têm sido realizadas e evoluído ao longo dos anos a partir de análises bibliométricas e de redes sociais realizadas sobre dados curriculares disponíveis na Plataforma Lattes.

Especificamente, para alcançar tal objetivo, são abordados os seguintes desafios:

- Construção de redes de colaboração científica a partir de dados extraídos da Plataforma Lattes;
- Adoção de métricas de análise de redes sociais para extração de conhecimento sobre as redes de colaboração científica construídas;

- Análise bibliométrica para a determinação de indicadores sobre a produção científica brasileira.

Dentro desse contexto, o propósito desta tese é explorar dados dos currículos cadastrados na Plataforma Lattes de modo a analisar como ocorre a produção e a colaboração científica no Brasil nas diversas áreas do conhecimento. Para isso, diferentes levantamentos e análises de dados foram realizados com o intuito de avaliar diversos indicadores considerados. Assim sendo, destacam-se como principais contribuições desta tese:

- A construção de um repositório de dados criado a partir da coleta de todo o conjunto de currículos da Plataforma Lattes.
- Uma caracterização geral dos dados que compõem o repositório construído.
- Uma análise aprofundada do perfil dos principais pesquisadores brasileiros considerando somente os indivíduos com doutorado concluído.
- Uma avaliação da qualidade da produção científica brasileira a partir dos dados dos atuais docentes dos programas de pós-graduação reconhecidos pela CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e dos pesquisadores contemplados com bolsa de produtividade em pesquisa do CNPq.

Nesse cenário, abordagens considerando a adoção de métricas de análises de redes sociais e de indicadores bibliométricos podem ser de grande valia para a compreensão acerca da produção científica brasileira nas diversas áreas do conhecimento. Dessa forma, esta tese explorará todo o conjunto de currículos cadastrados na Plataforma Lattes, apresentando um retrato da produção científica brasileira de forma ampla.

1.3 Estrutura da Tese

O restante desta tese está organizado da seguinte forma:

- No Capítulo 2, é apresentada uma revisão bibliográfica que aborda os assuntos permeados nesta tese. O capítulo aborda ainda conceitos relacionados a análises bibliométricas e de redes sociais.
- No Capítulo 3, são apresentados, discutidos e comparados alguns trabalhos relacionados aos assuntos abordados na tese, incluindo trabalhos que analisam dados da Plataforma Lattes.
- No Capítulo 4, é apresentada uma visão geral do LattesDataXplorer, um arcabouço desenvolvido para coleta, tratamento e análise de dados da Plataforma Lattes.

- No Capítulo 5, é apresentada uma caracterização dos dados curriculares que compõem a Plataforma Lattes, a partir da extração realizada em abril de 2015. Essa análise inclui uma visão geral do grupo de indivíduos com currículos cadastrados na Plataforma Lattes, bem como detalhes que revelam o perfil bibliométrico da pesquisa científica brasileira. Posteriormente, são apresentadas estatísticas da distribuição dos indivíduos de acordo com as áreas do conhecimento definidas pelo CNPq, identificando, ainda, os grupos mais representativos quando considerada a produção científica de seus indivíduos.
- No Capítulo 6, é apresentada uma análise geral da produção científica do conjunto de doutores com currículos cadastrados na Plataforma Lattes. Este conjunto, constituído pelos indivíduos que possuem o nível de formação mais elevado, é responsável por um percentual substancial de toda a produção científica brasileira. Inicialmente, é apresentada uma visão geral do conjunto analisado, sendo a seguir realizado um estudo sobre a evolução da colaboração científica desses indivíduos ao longo dos anos, tendo como objeto de análise os trabalhos por eles publicados com seus colaboradores. Finalmente, são realizadas análises bibliométricas considerando as grandes áreas e as áreas específicas de atuação dos doutores.
- No Capítulo 7, são apresentados estudos que consideram o conjunto de docentes que atuam em programas de pós-graduação *strictu sensu* brasileiros e o conjunto de pesquisadores que recebem bolsas de produtividade em pesquisa do CNPq. Tais estudos visam analisar o conjunto de docentes responsável pela formação nos níveis mais altos de capacitação no Brasil e dos pesquisadores com reconhecida capacidade pela excelência em suas pesquisas. É apresentada uma caracterização geral dos dois conjuntos e, ainda, um estudo sobre a qualidade das publicações dos bolsistas.
- Por fim, no Capítulo 8, são apresentadas as considerações finais da tese, destacando-se as suas principais contribuições e os resultados obtidos. Esse capítulo apresenta ainda a lista de publicações resultantes dos trabalhos realizados no decorrer da elaboração desta tese e comenta sobre extensões para possíveis trabalhos futuros.

Capítulo 2

Fundamentação Conceitual

Este capítulo apresenta um estudo bibliográfico relativo aos assuntos permeados por esta tese. Nas próximas seções, são apresentadas definições e conceitos sobre bibliometria, bem como sobre as principais métricas adotadas para análise bibliométrica (Seção 2.1). Também são apresentados conceitos e as principais métricas envolvidos nas áreas de análise de redes sociais (Seção 2.2).

2.1 Bibliometria

Tendo em vista o crescimento da produção científica mundial, cada dia mais expressiva e acessível de forma instantânea, bem como os problemas decorrentes desse crescimento, são necessários mecanismos que possibilitem o controle e monitoramento de todo este conhecimento produzido. A bibliometria, por meio de suas técnicas, procura estudar os aspectos quantitativos da ciência e da produção científica, como uma atividade que envolve aspectos sociais, econômicos e políticos. Ela fornece um instrumental para estudos que visam mapear o conhecimento científico e extrair informações para a compreensão de como a produção científica tem sido realizada (HAYASHI, 2012).

Araújo (2006) define o termo bibliometria como sendo uma ferramenta quantitativa para a análise e medição de índices de produção científica. Já Pritchard (1969) define a bibliometria como sendo estudos que visam quantificar os processos de comunicação científica. Logo, pode-se considerar a bibliometria como a área responsável por avaliar a ciência, baseando-se em dados da produção científica dos pesquisadores. Além disto, a análise bibliométrica pode ser aplicada a diferentes enfoques, podendo ser utilizada para mensurar a produção científica de indivíduos em particular, grupos de pesquisa, instituições, regiões geográficas, organizações ou eventos.

Para Vanti (2002), existem inúmeras vantagens em se utilizar análises bibliométricas para medir o registro da informação e a difusão do conhecimento. Dentre algumas possibilidades, estão:

- Identificar as tendências e o crescimento do conhecimento em uma área;
- Identificar os periódicos mais relevantes de uma determinada área;
- Mensurar a cobertura dos periódicos secundários;
- Identificar os usuários de uma disciplina;
- Prever as tendências de publicações;
- Estudar a dispersão e a obsolescência da literatura científica;
- Prever a produtividade de autores individuais, organizações e países;
- Medir o grau e os padrões de colaboração entre autores;
- Analisar os processos de citação e co-citação;
- Determinar o desempenho dos sistemas de recuperação da informação;
- Avaliar os aspectos estatísticos da linguagem, das palavras e das frases;
- Avaliar a circulação e uso de documentos em um centro de documentação;
- Medir o crescimento de determinadas áreas e o surgimento de novos tópicos.

Guedes e Borschiver (2005) apresentam as principais leis e princípios bibliométricos. As principais leis bibliométricas são: Lei de Bradford (produtividade de periódicos), Lei de Lotka (produtividade científica de autores) e Lei de Zipf (frequência de palavras).

A Lei de Bradford incide sobre conjuntos de periódicos. Com o objetivo de descobrir a extensão na qual artigos de um assunto científico específico apareciam em periódicos destinados a outros assuntos, estudando a distribuição dos artigos em termos de variáveis de proximidade ou de afastamento, Bradford realizou uma série de estudos que culminaram, em 1934, com a formulação da lei da dispersão. Verifica-se, na Lei de Bradford, que permite estimar o grau de relevância de periódicos em uma dada área do conhecimento, que os periódicos que produzem o maior número de artigos sobre um dado assunto formam um núcleo, supostamente de maior qualidade ou relevância para aquela área (ARAÚJO, 2006).

A Lei de Lotka, conhecida como Lei do Quadrado Inverso por causa de sua premissa, estabelece que o número de autores que tenham publicado (n) trabalhos é inversamente proporcional a (n^2) (FERREIRA, 2010). Para Maltrás (2003), a Lei de Lotka pode ser vista como uma função de probabilidade da produtividade. Assim, quanto mais se publica, mais fácil parece se tornar o processo de publicação de um novo trabalho, tornando os pesquisadores com publicações mais interessantes os melhores candidatos à aquisição de reconhecimento e acesso a recursos para melhorar ainda mais as suas pesquisas.

A Lei de Zipf descreve a relação entre palavras em um determinado texto e a ordem de série dessas palavras. Zipf, em um de seus estudos, encontrou uma correlação entre o número de palavras distintas e a frequência de sua utilização, concluindo que existe uma regularidade no processo de seleção e uso das palavras, em que, um pequeno número de palavras é usado muito mais frequentemente. Diante disto, foi possível formular o princípio do menor esforço, em que as palavras mais utilizadas indicariam o assunto do documento (ARAÚJO, 2006).

Além destas leis clássicas da bibliometria, existem várias outras teorias bibliométricas, como a proposta por Goffman e Newill (1964; 1967), denominada teoria epidêmica de transmissão de ideias. Os autores explicam que a propagação de ideias dentro de uma determinada comunidade é similar à transmissão de doenças infecciosas, e as ideias de um trabalho são o material infectante, e os leitores, as pessoas suscetíveis à doença.

Uma abordagem bastante utilizada é baseada no número de publicações de um autor em particular ou de um grupo de autores. Tal abordagem, cuja estratégia é baseada na contabilização de publicações, visa obter medidas como total e média de publicações para que, dessa forma, possam ser avaliados seus autores e, conseqüentemente, ranqueados, classificados, dentre outros. Digiampietri (2015) destaca que, ao se contabilizar o total de publicações de um grupo, é necessário identificar as publicações em coautoria existentes dentro do grupo que está sendo analisado, para que, dessa forma, um artigo não seja contabilizado várias vezes, inflando as medidas calculadas.

Outra abordagem comumente investigada é a análise baseada em citações. Neste tipo de análise, as citações de um conjunto de trabalhos publicados são verificadas, com o intuito de qualificar um trabalho em particular, identificar relações entre conjuntos de trabalhos ou analisar o perfil de citação em cada área de pesquisa. Araújo (2006) destaca que os dados de citações possibilitam descobrir informações valiosas para o entendimento da procedência geográfica ou institucional dos autores mais influentes em um determinado campo de pesquisa, dos pesquisadores com maior número de publicações, dentre outros. A análise de citações possibilita, ainda, estimar o fator de imediatismo de uma determinada publicação, avaliando a concentração de citações recebidas pela publicação recentemente. A hipótese é de que, em determinada área científica, artigos de periódico citados mais frequentemente são mais relevantes do que artigos menos citados (GUEDES; BORSCHIVER, 2005).

A colaboração entre autores também é uma forma muito utilizada para se compreender como trabalhos são realizados de forma colaborativa. Nestas análises, são identificados os trabalhos realizados em conjunto, quantidade média de autores, instituições que os autores de um mesmo artigo estão vinculados, e tipo de relacionamento dos autores de um mesmo trabalho. Com isso, é possível obter uma visão sobre como a colaboração científica vem sendo realizada. Além disso, medidas que avaliam a qualidade dos meios de publicações têm sido consideradas para verificar a qualidade de grupos de autores. Tais medidas visam obter o nível de abrangência de um artigo publicado em um determinado meio de publicação e seu impacto na comunidade científica.

No Brasil, os estudos bibliométricos proliferaram na década de 1970, principalmente com os estudos realizados no Instituto Brasileiro de Bibliografia e Documentação – IBBD, hoje Instituto Brasileiro de Informação Científica e Tecnológica, IBICT. Os estudos bibliométricos realizados naquele período incidiram sobre a literatura científica de vários campos científicos (ARAÚJO, 2006). Para o autor, as preocupações dos pesquisadores em bibliometria caminham para direções distintas, em que alguns estão preocupados com o aperfeiçoamento das fórmulas que expressam as leis bibliométricas e, outros, questionando a validade de estudos exclusivamente quantitativos, colocando em questão a existência da bibliometria como disciplina

científica e propondo que ela deveria passar a ser uma técnica de suporte a ser somada a outras na realização de estudos concretos.

Para Mugnaini (2013), após um período de, aproximadamente, 40 anos de bibliometria no Brasil, ela é tida como uma especialidade de análise, que tem na Ciência da Informação um lugar importante. Por outro lado, sua utilidade atende a comunidade científica como um todo, dada a necessidade de compreensão e crítica dos critérios de avaliação propostos pela política científica nacional.

Diante do exposto, é possível observar que as análises bibliométricas possibilitam a realização de estudos quantitativos e estatísticos, envolvendo técnicas diversas para a realização de análises complexas, objetivando descobrir conhecimento sobre a produção científica. As métricas bibliométricas podem ser aplicadas em áreas distintas de pesquisa, possibilitando, dessa forma, uma análise quantitativa para a medição da produção científica e dos mecanismos de disseminação do conhecimento.

2.2 Análise de Redes Sociais

Atualmente, a análise de redes sociais tem sido considerada tão importante, ou mais importante, do que a simples análise de indicadores de um profissional ou grupo de profissionais, pois a maioria das tarefas complexas não é realizada por uma única pessoa e sim por grupos (DIGIAMPIETRI et al., 2014a). Para Stroele, Zimbrão e Souza (2012), uma rede social é um conjunto de objetos, em que cada um deles está conectado a outro objeto. As redes sociais refletem uma estrutura social que pode ser representada por indivíduos ou organizações e suas relações. Uma rede social pode ser representada por uma rede na qual os nós estão relacionados ou não por arestas.

O tipo de uma rede social é definido pelo contexto em que esta está inserida. Geralmente, o tipo de uma rede está diretamente ligado ao tipo de relacionamento envolvido nessa rede. Assim, diferentes tipos de relacionamento produzem diferentes tipos de redes. Por exemplo, uma rede na qual os nós são indivíduos, as arestas podem ser caracterizadas por vínculo familiar, coautoria ou relacionamento afetivo, alterando apenas o tipo do relacionamento da rede (MENEZES, 2012).

As redes surgem e evoluem em determinados contextos específicos nos quais seus elementos se relacionam de alguma forma, como, por exemplo, professores de uma mesma instituição, alunos de um mesmo curso, pesquisadores de um mesmo projeto, cientistas de uma mesma área ou autores de um mesmo artigo, dentre outros. Quando se observa que as conexões em uma rede permitem a interação e a troca de informações entre as pessoas, é possível perceber a aplicabilidade desses conceitos na área científica, em que a interação entre os pesquisadores é um fator de suma importância para o avanço científico.

Redes sociais têm sido objeto de estudo em diversas áreas (MISLOVE et al., 2007; KUMAR; NOVAK; TOMKINS, 2010), e seus conceitos vêm sendo empregados para entendimento de redes reais, como a Internet (BARABÁSI; ALBERT; JEONG, 2000), redes de coautoria (HUANG et al., 2008) e redes biológicas (BARABÁSI; OLTVAI, 2004). O conceito de mundo pequeno (*small world*) tem tido crescente interesse

recentemente, principalmente para análise de redes sociais. As redes com pequenos caminhos entre pares de atores e com alto coeficiente de agrupamento são, geralmente, classificadas como redes de mundo pequeno. Tal fenômeno, típico das redes sociais, foi identificado pela primeira vez no estudo de Milgram (1967).

Barabási e Albert (1999) propuseram um modelo no qual os novos nós da rede se vinculam aos nós já presentes nessa rede por regras de preferência, em que nós com um maior número de vínculos possuem maior probabilidade de se vincular a esses novos nós. No modelo proposto, a distribuição da quantidade de vínculos dos nós que compõem a rede tende a seguir uma lei de potência, com poucos nós concentrando muitos vínculos e muitos nós contendo um número baixo de vínculos. Pelo fato dos nós que compõem a rede possuírem uma distribuição de ligações que segue uma lei de potência, este modelo é conhecido como um modelo *power-law*. Essas redes também são comumente descritas como redes livres de escala.

Redes como a Web e as de colaborações científica tendem a apresentar características de redes com pequeno diâmetro e alto coeficiente de agrupamento, podendo ser caracterizadas como redes de mundo pequeno, livres de escala com o grau de seus nós seguindo uma lei de potência, ou seja, poucos nós com muitos vínculos e muitos nós com poucos ou sem nenhum vínculo (ADAMIC, 1999; FALOUTSOS; FALOUTSOS, 1999; FALOUTSOS, 1999; NEWMAN, 2001a).

Para Lopes (2012), a análise mais importante é a análise que considera todos os elementos da rede social. Todas as informações de todos os relacionamentos de todos os nós são utilizadas a fim de explicar a estrutura relacional da rede social como um todo. Assim, nessa análise, todas as informações da rede são usadas, tornando-a muito mais completa e complexa.

Vários trabalhos investigam o estudo de redes sociais, uma grande quantidade dos quais realiza análises estatísticas, enquanto outros tratam de métricas de análise de redes sociais. Uma dificuldade inerente à análise de redes está na dificuldade de coletar grandes quantidades de dados que possibilitem realizar um estudo não apenas estatístico, mas também sobre as estruturas topológicas das redes. O modo como essas redes podem ser representadas influencia significativamente a sua análise, sendo que as análises que podem ser realizadas sobre as redes estão diretamente vinculadas à forma como elas são caracterizadas.

No âmbito desta tese, considera-se as redes de colaboração científica. A análise de redes de colaboração científica tem despertado o interesse da comunidade científica devido à sua capacidade de representação de problemas complexos de forma objetiva. Esse tipo de análise proporciona subsídios para o estudo das propriedades e comportamentos dos elementos que compõem as redes analisadas.

Nesse contexto, diversas métricas bibliométricas e de redes sociais podem ser aplicadas em redes de coautoria, como, por exemplo, o grau de centralidade, intermediação e proximidade para poder indicar a importância de um determinado autor (OTTE; ROSSEAU, 2002). Outros trabalhos, com ênfase no estudo das características individuais e estruturais de redes de coautoria em diferentes domínios e repositórios, foram apresentados por Newman (2001a), Newman (2004) e Huang et al.

(2008). Newman (2003) apresenta métricas relacionadas ao tamanho das redes, aos níveis de conectividade e transitividade, ao grau de miscigenação e, ainda, sobre a estrutura comunitária das redes.

A análise de redes científicas permite identificar informações importantes como os relacionamentos entre pesquisadores e instituições, a idade e frequência desses relacionamentos, por exemplo, como um determinado departamento está conectado a outros, e ainda, características como grau e intensidade de colaboração, além do tipo de relacionamento.

Considerando todos estes fatores, novas métricas surgem em consequência de novos problemas a serem solucionados. Logo, um conjunto de métricas formalmente estabelecidas, conforme descrito a seguir, são comumente utilizadas no estudo e resolução de grande parte dos problemas envolvendo análise de redes sociais.

2.2.1 Conceitos Básicos

São descritos a seguir conceitos frequentemente utilizados na análise de redes e que são fundamentais para melhor entendimento do restante desta tese. Essas definições são baseadas em Wasserman e Faust (2009).

- **Ator:** entidade social que participa de uma rede, e que é capaz de agir e formar vínculos com outros atores presentes na rede. Pode representar um indivíduo, uma organização ou um coletivo social, como uma classe em uma escola. Quando todos os atores da rede são do mesmo tipo, esta rede é denominada como rede monomodal. No entanto, existem redes que possuem atores que podem ser de tipos diferentes.
- **Vínculos:** um vínculo pode ser definido como uma conexão entre dois atores. Estas conexões podem ser de tipos diversos, como envio de e-mails em uma rede de amizade, estradas em uma rede de cidades ou rotas aéreas em uma rede de aeroportos. Estes vínculos podem, opcionalmente, possuir valor e serem direcionados.
- **Díade:** uma díade consiste em um par de atores e a possível ligação entre eles. Esta ligação é inerente ao par de atores, não pertencendo a um ator individualmente. Para o entendimento sobre vínculos entre atores, a díade é uma unidade de análise frequentemente explorada. Geralmente, a díade é a unidade de estudo em análises estatística em redes sociais.
- **Tríade:** uma tríade é um subconjunto formado por três atores e as possíveis ligações entre si. A tríade fornece conceitos importantes como transitividade e balanceamento que serão discutidos adiante. Potencialmente, em uma tríade, há duas díades. Em redes sociais, isso pode ser um fator importante, pois, se o Ator 1 mantém relações com o Ator 2, e este com o Ator 3, há um caminho possível, passando pelo Ator 2, para o Ator 1 realizar contato com o Ator 3.
- **Grupo:** um grupo pode ser definido com o conjunto de todos os atores e suas ligações, definindo um limite para o grupo. Consiste em um conjunto finito de atores que podem estabelecer vínculos. Para efeito das análises de rede sociais, a definição dos limites define o grupo.

- **Subgrupo:** um subgrupo é um subconjunto de atores e todos os vínculos entre eles. Analisar subgrupos pode ser muito importante no estudo de redes sociais complexas e extensas, com a análise de subgrupos específicos, delimitados dentro do grupo.
- **Relação:** uma relação é um conjunto de vínculos de um determinado tipo. Enquanto um vínculo se dá somente entre dois atores, a relação é definida para todo o conjunto de vínculos. Uma relação se refere à coleção de vínculos de um dado tipo estabelecida por atores de um conjunto específico.
- **Rede Social:** uma rede social pode ser definida como um conjunto de atores e seus vínculos. É uma estrutura social na qual os atores podem estar conectados por diversos tipos de relações, por exemplo, interesses profissionais, pessoais, acadêmicos ou políticos.

Com o objetivo de ilustrar os conceitos apresentados anteriormente, a Figura 2.1 apresenta uma rede social fictícia. Considere que esta rede é uma rede de pesquisadores científicos, em que os atores representam os pesquisadores e o vínculo entre pares de pesquisadores é caracterizado por algum trabalho publicado em conjunto por eles. Logo, o conjunto de todos os pesquisadores forma o grupo e o conjunto de todos os vínculos representa a relação de coautoria entre os atores.

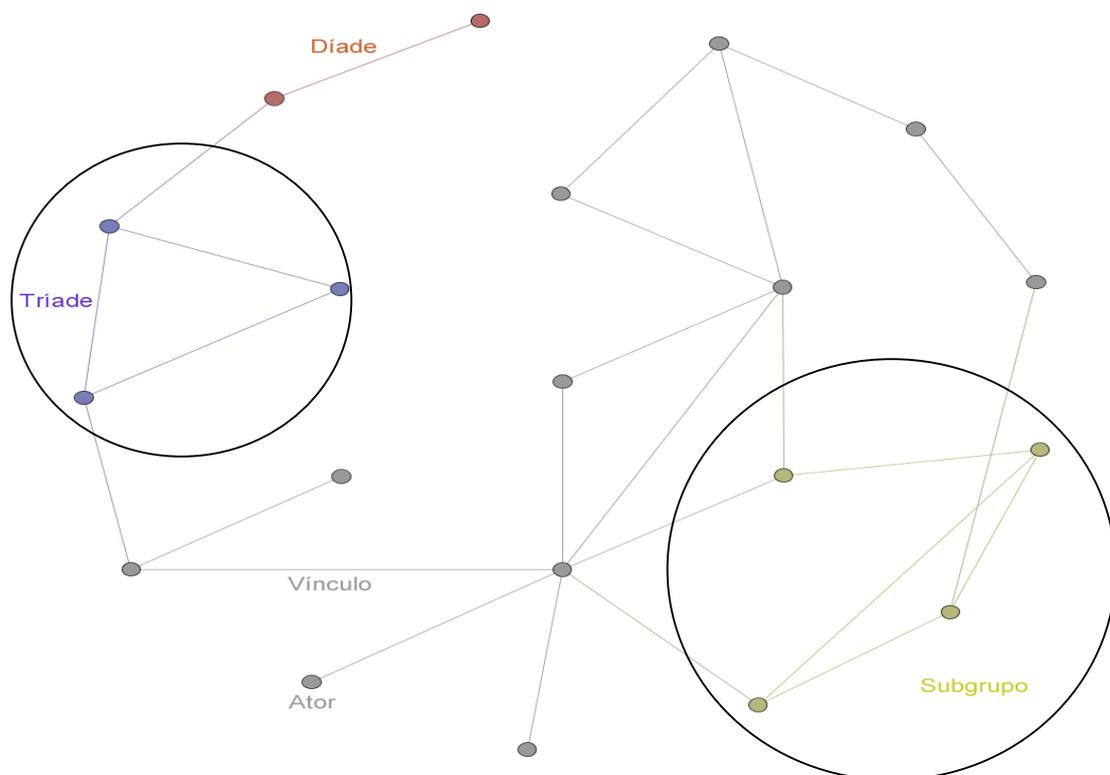


Figura 2.1: Exemplo de uma rede de coautoria com conceitos básicos de redes sociais.

2.2.2 Métricas para Análise de Redes Sociais

Para a análise das redes científicas, diversas métricas de análises de redes sociais podem ser empregadas com o objetivo de explorar técnicas consolidadas para o entendimento dessas redes. Tendo em vista melhorar a compreensão e padronizar a nomenclatura que permeia o restante desta tese, alguns conceitos fundamentais para a análise de redes são descritos a seguir, em conformidade com os trabalhos de Newman (2001a), Newman (2003), Newman (2010) e Lopes (2012).

- **Grau de Centralidade:** o grau de centralidade (*degree centrality*) de um nó v corresponde ao número de arestas incidentes ou ao número de nós adjacentes a ele. Suponha que um nó que possua muitas conexões é considerado importante, enquanto nós com poucas conexões são considerados irrelevantes. Um nó com grau 0 é dito ser um nó isolado. Em uma rede de coautoria, por exemplo, o grau de centralidade indica o total de autores que publicaram trabalhos em colaboração com um determinado autor. Em redes não direcionadas, em que a conexão entre dois nós é não direcionada, a métrica é chamada apenas de grau. Já para redes direcionadas, cujas arestas possuem direcionamento, a métrica é categorizada em grau de entrada (*in-degree*) e grau de saída (*out-degree*) de acordo com a direção dos relacionamentos analisados. A Equação (1) define o grau de centralidade do nó v_i , chamado $dc(v_i)$, em que e é igual a 1, se o nó v_i possui vínculo com o nó v_k , ou 0 caso contrário.

$$dc(v_i) = \sum_{k=1}^n e(v_i, v_k) . \quad (1)$$

- **Grau de Proximidade:** o grau de proximidade (*closeness centrality*) de um nó v mede o quanto o nó que representa o ator está próximo de todos os demais nós da rede. Este índice tenta capturar não só a importância local de um nó em relação aos vizinhos mais próximos, mas em relação a todos os nós da rede. O grau de proximidade é definido pelo inverso da soma das distâncias geodésicas¹ entre um determinado nó e todos os outros nós da rede. Valores do grau de proximidade altos indicam uma maior capacidade dos nós se vincularem a outros nós da rede. Por sua vez, valores baixos indicam que os nós não estão bem posicionados na rede, ou seja, estão distantes dos outros nós.

- **Coefficiente Gigante:** o coeficiente gigante (*giant coefficient*) de uma determinada rede é calculado com base no tamanho do componente gigante da rede. O componente gigante (*giant component*) é aquele componente conectado com o maior número de nós, também conhecido como componente principal (*main component - MC*). O componente principal é um subgrafo do grafo G ou o próprio grafo G , caso o grafo seja totalmente conectado. A Equação (2) apresenta o cálculo do coeficiente gigante, denominado gc , que é calculado como a ordem (número de nós) do componente principal dividido pelo número total de nós da rede (N). Este valor representa o percentual de nós que são parte do componente gigante.

¹ Distância geodésica é o menor caminho entre dois nós, ou seja, a menor quantidade de vínculos que liga dois nós de uma rede.

$$gc(G) = \frac{1}{N} \sum_{i=1}^N mc(v_i), \quad (2)$$

onde:

$$mc(v_i) = \begin{cases} 1, & \text{se } (v_i \in MC); \\ 0, & \text{caso contrário.} \end{cases} \quad (3)$$

- **Diâmetro:** o diâmetro (*diameter*) de uma rede refere-se ao tamanho da maior distância geodésica entre qualquer par de nós. O diâmetro de uma rede pode variar de um mínimo de 1, caso a rede seja completa (todos os nós estão conectados a todos os outros), a um máximo de $V-1$, onde V é o número de nós da rede. Em situações nas quais a rede é não conectada, seu diâmetro refere-se ao maior entre os diâmetros dos componentes conectados.

- **Densidade:** a densidade (*density*) refere-se à razão/proporção entre o número de arestas existentes e o número de arestas possíveis. A densidade de uma rede pode variar de um mínimo de 0 (se não existem arestas na rede) a um máximo de 1 (se cada nó for adjacente a todos os outros). Não é aplicado quando existe a possibilidade de múltiplas arestas entre dois nós, já que não é possível calcular a quantidade máxima de arestas, neste caso. A Equação (4) pode ser usada para calcular a densidade d de um grafo G que represente uma rede. Logo, o número total de arestas existentes é calculado pela soma do grau de centralidade ($dc(v_i)$) de todos os nós da rede, enquanto que o número de todas as possíveis arestas é calculado por $(N(N-1))$, em que N representa o número total de nós.

$$d(G) = \frac{\sum_{i=1}^N dc(v_i)}{N(N-1)}. \quad (4)$$

- **Grau de Intermediação:** o grau de intermediação (*betweenness centrality*) determina a localização global de um nó na rede. Tal medida indica o quanto um nó está no caminho mínimo entre outros dois pares de nós. Possibilita, por exemplo, obter informação como a localização de um nó na rede em senso global, enquanto o grau de centralidade considera apenas os vizinhos diretos. A Equação (5) possibilita obter o grau de intermediação (bc), onde p_{jk} é o número de distâncias geodésicas que ligam os nós j e k , e $p_{jk}(v_i)$ é o número de distâncias geodésicas que passam pelo nó v_i .

$$bc(v_i) = \sum_{j < k} \frac{p_{jk}(v_i)}{p_{jk}}. \quad (5)$$

- **Hub:** a definição de *hub* está associada ao conceito de grau de intermediação, em que os nós que possuem um grau maior de intermediação são os principais conectores da rede, já que possuem um número elevado de vínculos que permitem estabelecer a conexão de diferentes grupos, por exemplo. Para Barabási (2009), na perspectiva dos *hubs*, o mundo pode ser considerado como muito pequeno, já que por meio deles é possível estabelecer contato com um grande número de atores.

- **Caminho Mínimo Médio:** o caminho mínimo médio de uma rede corresponde ao número médio de arestas em todos os caminhos mínimos existentes entre todos os pares de nós da rede.

Para ilustrar o cálculo das métricas, a Figura 2.2 mostra um exemplo para todas as métricas de análise de redes sociais apresentadas nesta seção. Especificamente, a Figura 2.2(a) ilustra uma rede não direcionada, contendo sete nós, que pode ser considerada como uma rede de coautoria, sendo esta caracterizada por nós que representam os autores de trabalhos publicados e os vínculos são estabelecidos quando dois ou mais autores publicaram algum trabalho em conjunto. Já na Figura 2.2(b) são apresentadas as métricas calculadas com seus respectivos valores.

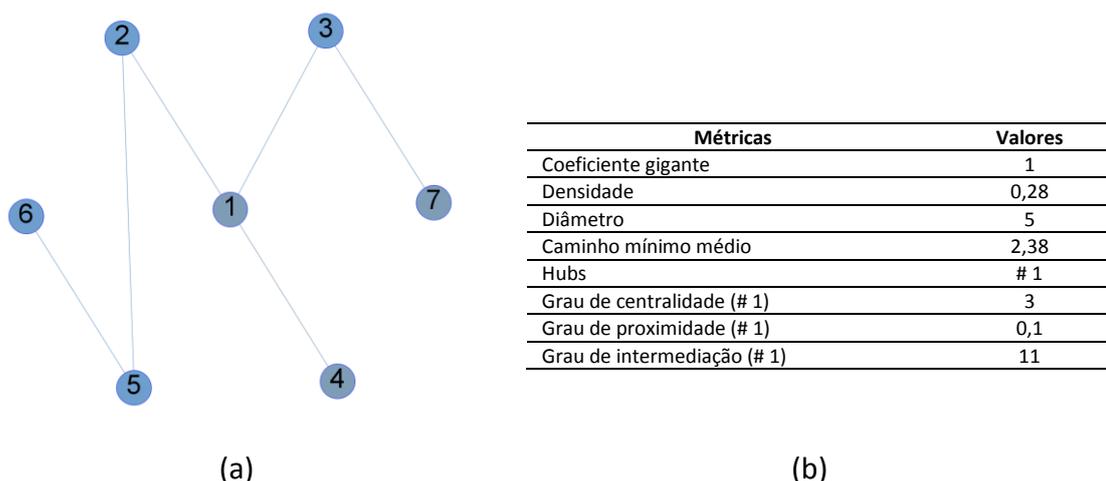


Figura 2.2: Exemplo de uma rede com cálculo das métricas e seus respectivos valores.

Neste exemplo, a rede a ser analisada possui somente um único componente, já que é possível obter um caminho de um determinado nó a qualquer outro da rede. Normalmente, quando existe mais de um componente na rede, o componente principal, componente conectado com o maior número de nós, deve ser utilizado para o cálculo das métricas que avaliam a rede globalmente, ou seja, aquelas que não são calculadas em relação a um único nó.

Como a rede possui um único componente conexo, o coeficiente gigante da rede é 1, tendo em vista que o componente principal é do tamanho da rede. Considerando as métricas que avaliam a rede de forma global, a densidade da rede pode ser calculada pela soma dos graus de cada nó (12), dividida pelo número total de possíveis vínculos (42), resultando em 0,28. O diâmetro da rede é obtido pelo valor máximo obtido entre todos os menores caminhos da rede. No exemplo, o menor caminho que liga os nós 6 e 7, possui 5 arestas, sendo este o menor caminho entre dois nós com o maior valor. Nesse caso, a rede tem como diâmetro o valor 5. Como o cálculo de *Hub* leva em conta os menores caminhos que passam por um determinado nó, o nó 1 pode ser considerado *Hub* da rede em análise.

Para as métricas que são calculadas para nós específicos, foi considerado o nó 1 para análise da rede do exemplo. O grau de um nó considera vínculos diretos que ele possui com os demais nós da rede; neste caso, o nó 1 possui grau 3, já o seu grau de

proximidade é 0,1. Para isto, é calculado o inverso da soma de suas distâncias geodésicas (menor caminho) para todos os outros nós ($1/10 = 0,1$), ou seja, 1 dividido pela soma dos menores caminhos do nó 1 para todos os outros nós da rede (3 caminhos de tamanho 1, 2 caminhos de tamanho 2 e 1 caminho de tamanho 3). Por fim, seu grau de intermediação é 11, esta métrica considera a quantidade de menores caminhos entre quaisquer dois nós que passam por ele.

Tais métricas no contexto de análise de redes sociais possibilitam entender a estrutura e a influência de um determinado nó na rede. Várias outras métricas para análise de redes sociais poderiam ser empregadas objetivando extrair conhecimento sobre os nós que compõem a rede em análise. No entanto, para esta tese, os conceitos anteriormente explicados são suficientes para entendimento dos resultados a serem discutidos. No próximo capítulo, são apresentadas as principais referências que abordam os assuntos que permeiam esta tese.

Capítulo 3

Trabalhos Relacionados

Este capítulo apresenta a síntese de um conjunto de trabalhos relacionados a esta tese. Inicialmente, são apresentados trabalhos relacionados a análises bibliométricas e análises baseadas em redes, com o intuito de compreender como grupos de pesquisadores têm realizado suas pesquisas e como a produção científica pode ser estudada com estas análises (Seção 3.1). A seguir, são apresentados e discutidos os principais trabalhos relacionados que têm utilizado a Plataforma Lattes como principal fonte de dados (Seção 3.2).

3.1 Análises Bibliométricas e Baseadas em Redes

Nesta seção, são apresentados os trabalhos relacionados, que têm utilizado dados de publicações científicas para analisar e compreender como tem ocorrido a colaboração científica, qual o perfil de publicação dos pesquisadores analisados e qual o impacto das publicações dos principais pesquisadores em diversas áreas de atuação. Para isso, tais trabalhos utilizam dados de publicações científicas, possibilitando, desta forma, a adoção de métricas bibliométricas e baseadas em análises de redes sociais.

Morel e Morel (1978) analisam a produção científica brasileira quantitativamente em termos de autores de trabalhos científicos publicados em periódicos indexados pelo *Institute for Scientific Information* (ISI). No trabalho, é avaliada a evolução da produção científica e a relação produção versus população comparativamente a diversos países. Identificou-se que a produção científica brasileira se concentrava na região Sudeste (80,8% do total), em que o estado de São Paulo era responsável por 50,4% da produção nacional, seguido por Rio de Janeiro com 22,9%. Situação esta semelhante à de países também analisados como URSS (50% dos autores em Moscou), Tchecoslováquia (52% dos autores em Praga) e Dinamarca (62% dos autores em Copenhague). Os autores discutem a necessidade e importância de se organizar um Banco de Dados sobre a produção científica brasileira.

No trabalho de Chaimovich (2000), o autor utiliza como referencial para analisar a evolução da ciência no Brasil, dados de trabalhos publicados em periódicos indexados pelo ISI. Ressalta-se que a contribuição brasileira à ciência indexada aumentou de um valor de 0,4% em 1990 para mais de 1,2% em 1999. Tal crescimento é reflexo de maior financiamento à ciência e à pós-graduação. Destaca-se que, no Brasil, a produção de ciência é realizada quase que na sua totalidade nas universidades públicas e que o sistema privado pouco contribui para a produção nacional de ciência e tecnologia. Diante disso, o autor aponta alguns desafios a serem enfrentados pela ciência brasileira, como, por exemplo, o acesso ao ensino superior e à manutenção da estrutura de pesquisa.

Glänzel, Leta e Thijs (2006) apresentam um estudo sobre o perfil da ciência brasileira considerando o período compreendido entre 1991 e 2003, tendo as publicações em periódicos indexados pelo ISI, como principal fonte de dados. Foi identificado que os trabalhos realizados em colaboração aumentaram no período analisado e que a produção científica brasileira aumentava em torno de 8% ao ano, respondendo por, aproximadamente, 45% de toda a produção científica da América Latina. Em outro trabalho (LETA, GLÄNZEL e THIJIS, 2006), os autores analisam com medidas bibliométricas como tem sido conduzidas as pesquisas científicas brasileiras. É ressaltado o investimento na educação científica em cursos de graduação, com o aumento no número de bolsas de iniciação científica, o que favorece a formação dos futuros cientistas, além de prepara-los para a pós-graduação. Também é verificado que mais de 80% da produção científica brasileira indexada na *Web of Science* (Wos) era proveniente de universidades públicas, que concentram a maior parte dos pesquisadores, como também dos estudantes e programas de pós-graduação.

Diversos outros trabalhos têm aplicado indicadores bibliométricos em dados de publicações científicas para melhor compreender a ciência brasileira, dentre os quais Leta e Meis (1996), Meneghini (1996), Zanotto (2002), Jorio et al. (2010) e Leta, Thijs e Glänzel (2013). No entanto, tais trabalhos extremamente relevantes para a compreensão acerca da ciência brasileira, em sua maioria, utilizam apenas dados de trabalhos publicados em periódicos indexados pela ISI ou WoS. Logo, trabalhos publicados em anais de congresso ou de periódicos nacionais que não estão indexadas acabam por serem negligenciados e, dessa forma, não contemplar toda a produção científica brasileira, sendo algumas áreas do conhecimento prejudicadas de forma mais intensa, dada a natureza de suas publicações.

Meadows e Lemos (1999) destacam um aumento no número de trabalhos científicos compartilhados atualmente, impulsionados, principalmente, pela facilidade na forma de distribuir os trabalhos, possibilitando maior velocidade no intercâmbio de informações. Aliado a isso, os autores ressaltam que os trabalhos em colaboração permitem a economia de tempo e de recursos financeiros e materiais, resultando em um processo cada vez mais estimulado por instituições e agências que financiam pesquisas atualmente. Para Sonnenwald (2007), a interação entre dois ou mais cientistas que, em determinado contexto, objetiva compartilhar atividades para obter resultados e atingir objetivos comuns, é definida como colaboração científica e se caracteriza por ser um fenômeno complexo.

Diante do crescente aumento dos trabalhos realizados em colaboração e da facilidade de acesso a estes trabalhos, a análise das redes de coautoria se apresenta como importante ferramenta para compreender, por exemplo, como tem acontecido a colaboração científica, identificando a existência de grupos com colaborações mais intensas e autores com maior grau de colaboração. Redes de colaboração científicas vêm sendo amplamente estudadas, como mostram os trabalhos de Newman (2001a; 2001b; 2001c; 2004) e Kumar (2015). A estrutura das redes de colaboração possibilita revelar características interessantes sobre comunidades científicas (NEWMAN, 2004).

Segundo Stroele, Zimbrão e Souza (2012), a análise de redes de colaboração científica possibilita identificar como os grupos de pesquisadores e centros de estudos estão desenvolvendo seus trabalhos, verificar e entender qual o grau de envolvimento entre os pesquisadores de determinados grupos, de determinadas áreas do conhecimento, de instituições de ensino e pesquisa, e também permitir a indicação de padrões de colaboração que poderiam proporcionar um grande avanço na área, permitindo melhorias na comunicação e colaboração de toda a comunidade científica.

Petersen et al. (2012) analisam fatores relevantes para o sucesso científico em redes de colaboração. Dentre eles, são destacados a quantidade de produção científica que aumenta a atração de novas oportunidades e também o tamanho da equipe de colaboradores, bem como, da própria rede. Diante disso, justifica-se o grande potencial de novos estudos que visam à análise das redes de colaboração, objetivando entender como a colaboração científica acontece e como ela vem evoluindo em cada uma das áreas de pesquisa.

Newman (2001a) apresenta uma análise sobre características das redes de coautoria nas áreas de Computação, Biologia, Física e Medicina entre o período de 1995 a 1999. Em outro trabalho (NEWMAN, 2004), o autor identificou padrões comuns nas redes de coautoria em Biologia, Física e Matemática, analisando propriedades estatísticas destas redes.

Barabási *et al.* (2002) apresentam um estudo das redes de colaboração científica nas áreas de Neurociência e Matemática entre os anos de 1991 e 1998. Os autores concluíram que a rede evolui a partir do momento em que novos nós e vínculos entre os nós existentes são incluídos. E que esta evolução segue um processo denominado *preferential attachment*, em que novos nós tendem a criar seu primeiro vínculo com nós que possuem grande número de vínculos. Como resultado, autores experientes tendem a aumentar seu número de colaboradores, com maior frequência do que autores novatos.

Hou, Kretschmer e Liu (2008) analisam as redes de colaboração entre autores que publicaram no periódico *Scientometrics* entre 1978 e 2004. O conjunto analisado continha um total de 1630 autores e 1927 artigos, sendo que, desses, 875 foram realizados em colaboração. Foram aplicadas métricas de análise de redes como o cálculo da densidade e grau de intermediação dos nós que a compõem. Autores com menos de duas publicações foram extraídos do conjunto a ser analisado, restando 234 autores, e foi identificada a presença de um grande componente, que agregava 59% dos autores, possuindo a rede baixa densidade, ou seja, pouco conectada.

Todos estes trabalhos foram importantes para demonstrar todo o potencial que a análise das redes de colaboração científica pode proporcionar para melhor compreensão sobre como um determinado conjunto de pesquisadores colaboram ao longo do tempo e ainda sobre como ocorre a colaboração nas diversas áreas do conhecimento.

Além das análises baseadas em redes de colaboração, entender a dinâmica dos tópicos de pesquisa e como determinados assuntos estão sendo trabalhados em conjunto surgem como importantes ferramentas para entender como as pesquisas têm sido realizadas nas diversas áreas do conhecimento. Vários desses trabalhos têm explorado o conceito de cliques em grafos para suas análises. Um clique se caracteriza por um subgrafo em que cada vértice está vinculado a todos os outros vértices do grafo (COLIZZA et al., 2006).

Alguns estudos sobre redes de conhecimento começaram a reconhecer a utilidade de palavras-chave e o potencial de análise dessas redes. Lee, Su e Chan (2010) calculam similaridades entre artigos de periódico com base no grau de centralidade das palavras-chave dos artigos analisados. Há o pressuposto de que uma palavra-chave é o portador fundamental mais básico do conhecimento.

As redes de títulos de artigos científicos são formadas utilizando a abordagem de construção de redes de cliques. Vários trabalhos apresentam resultados demonstrando que palavras em uma mesma sentença são inseridas de forma associativa para a apresentação de uma ideia e, com isso, cada palavra pode ter um significado diferente, dependendo do conjunto de palavras à sua volta (FADIGAS et al., 2010; PEREIRA et al., 2011). Conseqüentemente, a dinâmica de construção de uma rede de títulos é dada por justaposição ou sobreposição de cliques. Quando duas cliques são ligadas por apenas um vértice, este processo é denominado justaposição. Quando a ligação acontece com dois ou mais vértices em comum, chama-se de sobreposição (FADIGAS; PEREIRA, 2013).

Kauer e Moreira (2013) aplicam o método de TTM (*Temporal Text Mining*) como forma de identificar automaticamente a evolução dos temas abordados no *Simpósio Brasileiro de Banco de Dados (SBBDD)* ao longo dos anos. O método utilizado consiste na criação de um modelo probabilístico que visa descobrir os temas latentes nas coleções em cada edição e identificar as relações entre estes temas ao longo dos anos.

Cunha et al. (2013) investigam a evolução temporal de vértices e arestas de um grafo formado por palavras impressas nos títulos de artigos científicos publicados no periódico *Nature*, entre janeiro de 1999 a dezembro de 2008. Os vértices são as palavras e as arestas conectam pares de palavras que ocorreram em um mesmo título. Os títulos são agrupados em um documento de texto para, assim, serem construídas as redes. Após a sua construção, uma rede é analisada utilizando métricas clássicas de redes e métricas da abordagem em cliques, conforme proposto por Fadigas e Pereira (2013). Os resultados permitem visualizar a dinâmica do adensamento das redes. Esta abordagem contribui para revelar, ao longo do tempo, a importância dos conceitos expressos nas palavras dos títulos.

elementos para entendimento e avaliação do modo de comunicação, troca de informações e conhecimentos na comunidade estudada.

Análises de palavras-chave têm sido objeto de diversos estudos para compreensão e classificação dos principais elementos de pesquisa, como em Callon, Courtial e Laville (1991), Bhattacharya e Basu (1998), Ding, Chowdhury e Foo (2001), Heckner, Muhlbacher e Wolff (2008), Ke e Chen (2012) e Inomata et al. (2015).

Apesar de ser uma estratégia que tem ganhado cada vez mais atenção devido aos resultados alcançados, a análise baseada em palavras-chave retiradas dos títulos dos artigos, foco de grande parte dos estudos, apresentam limitações, já que nem sempre os títulos dos trabalhos conseguem expressar todo o conteúdo abordado por um determinado trabalho, tendo em vista a preocupação com questões semânticas na descrição dos títulos. Além disso, nem sempre é possível obter o conjunto de palavras-chave de um trabalho, devido ao formato dos repositórios utilizados, que podem não conter estes dados.

3.2 Trabalhos Envolvendo Dados da Plataforma Lattes

Segundo Lane, em artigo publicado na revista *Nature* (LANE, 2010), medir e avaliar o desempenho acadêmico passa a ser um fator crucial para a vida científica. Vários fatores para esta avaliação e medição estão vinculados aos cálculos de métricas. Porém, os sistemas atuais de medição são insuficientes para determinar respostas confiáveis. No trabalho supracitado, são descritos os vários problemas na adoção das métricas normalmente utilizadas para, por exemplo, classificar grupos ou instituições de pesquisa e os desafios nas propostas de métricas eficazes. A autora apresenta esforços empregados para a construção de repositórios confiáveis de dados científicos que poderiam permitir análises com o objetivo de explorar e compreender como a ciência tem evoluído. Embora todos os esforços sejam úteis, alguns apresentam problemas que comprometem o sucesso de tais iniciativas. Exemplo disso são informações dependentes de *softwares* específicos, repositórios dependentes de financiamentos transitórios (e esforços para poder combiná-los) e bancos de dados sem estruturas coesas e confiáveis. Nesse cenário, a Plataforma Lattes é citada como um poderoso exemplo de boas práticas para o fornecimento de dados de alta qualidade, e de como uma plataforma de coleta de dados tem sido incentivada por órgãos federais, instituições e órgãos financiadores. Por fim, é destacado que a Plataforma Lattes é uma das fontes de dados de pesquisadores mais confiáveis existente.

Mugnaini, Leite e Leta (2011) destacam que muito embora não se apresente como uma base de indexação e catalogação de periódicos, a Plataforma Lattes se constitui em uma fonte inesgotável de informações sobre a ciência brasileira, sob diversos aspectos e abordagens. Os autores ressaltam que, apesar de todo o volume de informações, o que se observa é ainda uma baixa frequência de estudos cientométricos, de especialistas brasileiros, que utilizam a Plataforma Lattes. E que isso tem reflexo nas limitações dos mecanismos de recuperação e extração das informações, constituindo-se em obstáculos para os estudos. Porém, a correlação de um grande volume de variáveis pessoais, acadêmicas e de produtividade, somadas a

um acesso livre e gratuito, é um estímulo significativo para uso deste repositório pelos pesquisadores brasileiros da área. Destaca-se, ainda, o fato do repositório de dados reunir toda a produção científica brasileira, viabilizando análises que só seriam possíveis se conduzidas em diversas bases internacionais, o que representaria um custo considerável para tais estudos.

Para Ferraz, Quoniam e Maccari (2014), até o presente momento, não existe no mundo um repositório curricular nacional único semelhante à Plataforma Lattes, sendo que somente repositórios de dados referenciais, de onde se podem extrair referências bibliográficas e fontes de informação secundárias estão disponíveis para livre acesso. Dessa forma, a Plataforma Lattes pode servir como um instrumento de maior importância para o estudo da produção científica brasileira.

Mena-Chalco, Digiampietri e Cesar-Junior (2012) ressaltam alguns motivos que tornam a Plataforma Lattes um interessante estudo de caso:

- Os currículos cadastrados se tornaram um padrão nacional e vêm sendo utilizados na avaliação individual das atividades científicas, acadêmicas e profissionais dos pesquisadores brasileiros;
- Pesquisadores nacionais de diversas áreas do conhecimento estão cadastrados na Plataforma Lattes (atualmente, com, aproximadamente, quatro milhões e oitocentos mil currículos (mais precisamente, 4.828.650 em 26/09/2016)).

Nessa perspectiva, ainda que impulsionada pelas políticas de Ciência e Tecnologia (C&T) nos últimos anos, a ciência brasileira vem apresentando grande crescimento de produção acadêmica que pode ser acompanhado pela análise dos currículos cadastrados.

Cañibano e Bozeman (2009) destacam currículos acadêmicos como uma potencial fonte de informação extremamente abrangente e foco de trabalhos recentes para análises de grupos de pesquisadores. Estudos que utilizam currículos para a análise de redes sociais são ainda menos frequentes, porém deve-se considerar a gama de estudos sobre análise de coautoria e os efeitos das colaborações científicas na carreira do pesquisador (DIGIAMPIETRI; SANTIAGO; ALVES, 2013; LIMA et al., 2013; MENA-CHALCO; CESAR-JUNIOR, 2013).

É importante destacar que, apesar da riqueza de dados registrados na Plataforma Lattes, apenas são disponibilizadas pela plataforma simples consultas que possibilitam recuperar conjuntos de currículos, como, por exemplo, por formação acadêmica ou proficiência em algum idioma. Isto dificulta o processo de obtenção e consequentemente, análise dos dados, sejam de grupos específicos ou de todos os currículos cadastrados. Logo, algumas ferramentas têm sido propostas para auxiliar o processo de extração e análise dos currículos Lattes.

O Lattes Extrator¹ foi desenvolvido pelo próprio CNPq e é uma das ferramentas que compõem a Plataforma Lattes. É acessível via Web e seu acesso é restrito. Atualmente,

¹ <http://lattesextrator.cnpq.br/lattesextrator/>

somente instituições cadastradas podem extrair informações diretamente do repositório de dados de currículos da Plataforma Lattes e, ainda assim, somente informações de seus pesquisadores, docentes, estudantes e colaboradores. As informações extraídas são disponibilizadas em arquivos no formato XML (*eXtensible Markup Language*) e as instituições podem desenvolver aplicações para manipulação e visualização dessas informações, bem como integrar com as suas próprias bases.

Já o ScriptLattes² (MENA-CHALCO; CESAR-JUNIOR, 2009) é uma ferramenta de código-fonte aberto desenvolvida para a extração de dados de um conjunto de currículos cadastrados na Plataforma Lattes, possibilitando a geração de grafos e relatórios a partir dos dados extraídos. Os relatórios apresentam listas com a produção e orientações do conjunto de currículos analisados, e também realiza o tratamento de publicações replicadas, gera mapas de geolocalização dos currículos analisados e identifica as colaborações para a geração dos grafos. Sua primeira versão, lançada em 2005, foi desenvolvida com a finalidade de auxiliar a secretaria do Programa de Pós-Graduação do IME-USP na elaboração de relatórios sobre a produção bibliográfica dos docentes do Departamento de Ciência da Computação. Inicialmente, os relatórios eram baseados apenas nas informações cadastradas nos currículos. Adicionalmente, foram sendo disponibilizadas novas funcionalidades, como mapas de geolocalização baseados nos endereços profissionais de cada currículo e grafos de colaborações. Grafos de colaborações são obtidos a partir de relações entre um grupo de pesquisadores. Tais relações são identificadas considerando publicações com títulos iguais ou similares. Cada grafo é estático, ou seja, não permite qualquer tipo de interação, apenas selecionar os nomes dos pesquisadores que compõem a rede direcionando para uma página que contém o currículo do referido pesquisador.

O XMLattes (FERNANDES; SAMPAIO; SOUZA, 2011) é uma ferramenta que importa os dados curriculares de indivíduos da Plataforma Lattes e os disponibiliza em formato XML. Como os currículos são visualizados em formato HTML, estes são transformados em XML permitindo maior flexibilidade na manipulação dos dados.

O portal CiênciaBrasil³ (LAENDER et al., 2011) permite analisar e visualizar informação sobre pesquisadores brasileiros que participam dos Institutos Nacionais de Ciência e Tecnologia (INCT). O portal tem como objetivo produzir mapas da produção científica dos pesquisadores que integram os INCTs a partir de dados contidos nos currículos da Plataforma Lattes. Diversas técnicas para coleta, integração, análise e visualização de dados da Web são empregadas e, dentre os principais recursos oferecidos pelo portal, estão as visualizações das redes de colaboração entre os pesquisadores de cada instituto. Para tanto, um algoritmo para deduplicação de citações é utilizado especificamente para a identificação dos relacionamentos de coautoria.

Vários trabalhos que visam analisar a colaboração e produção científica de conjuntos de pesquisadores brasileiros têm adotado tais ferramentas para a geração dos resultados, sendo o ScriptLattes a mais utilizada. No entanto, apesar de todo o potencial de análise dessas ferramentas, com a inserção dos captchas em abril de 2015 para acesso a cada um dos currículos cadastrados na Plataforma Lattes, algumas

² <http://scriptlattes.sourceforge.net/>

³ CiênciaBrasil: <http://www.pbct.inWeb.org.br/pbct/>

deixaram de realizar análises de forma automática, sendo necessária a extração dos currículos de interesse em um repositório local, o que limita o potencial de análises de grandes conjuntos de indivíduos.

Mena-Chalco, Digiampietri e Oliveira (2012) analisam os programas de ciência da computação brasileiros identificando o rápido crescimento em termos de produção bibliográfica e formação acadêmica. O trabalho apresenta uma descrição do perfil de produção acadêmica dos programas de ciência da computação avaliados pela CAPES nos triênios 2004-2006 e 2007-2009. O trabalho teve como única fonte de dados as listas de publicações e listas de orientações concluídas disponíveis nos currículos da Plataforma Lattes de cada professor associado aos programas de pós-graduação em Ciência da Computação. A identificação do perfil de produção acadêmica envolve quatro fases: 1) identificação dos programas de pós-graduação avaliados pela CAPES nos triênios 2004-2006 e 2007-2009, 2) identificação dos professores doutores, na modalidade de professor permanente, associados a cada programa de pós-graduação, 3) identificação dos currículos da Plataforma Lattes de cada professor associado a cada programa, e por fim, 4) extração das produções acadêmicas de cada programa (produções bibliográficas e orientações concluídas). Os resultados obtidos pela análise mostram que a área de computação brasileira se caracteriza por: publicar preferencialmente trabalhos completos em anais de congresso (54%), seguido da publicação de artigos completos em periódico (14%). Ainda no trabalho, foi mostrado que, no segundo triênio (2007-2009), houve um aumento na publicação de artigos em periódico.

Perez-Cervantes et al. (2013) introduzem novas medidas para estimar a influência da colaboração em redes científicas. A abordagem é baseada na técnica de predição de *links* e avalia como a presença ou ausência de um pesquisador afeta o processo de predição na rede em análise. Para isto, os cientistas são representados por nós em uma rede de colaboração e, após a remoção de nós, o processo de predição de *links* é realizado de forma iterativa para todos os outros nós. O modelo de aprendizagem supervisionado SVM foi utilizado como preditor e avaliou redes reais de colaboração em vários períodos de tempo. Nesse trabalho, os dados utilizados foram os currículos da Plataforma Lattes. Para os testes, foram selecionados 889 pesquisadores afiliados de 45 departamentos de Ciência da Computação de instituições de ensino brasileiras, sendo as publicações consideradas para a identificação das colaborações entre 2000 e 2012. Os resultados obtidos foram comparados com medidas de análise de redes sociais como grau de proximidade e de centralidade, e se mostraram consistentes para quantificar a influência de um determinado pesquisador em uma rede de colaboração científica.

Digiampietri et al. (2014b) apresentam uma estratégia para caracterizar os programas de pós-graduação em Ciência da Computação brasileiros e a relação entre eles. No trabalho, são apresentadas análises dos programas sob diversas perspectivas. São descritas a classificação sob cada uma destas perspectivas, a correlação entre as métricas de avaliação utilizadas, além de como os programas se relacionam, e quais os aspectos que aumentam a produtividade de pesquisa dos programas. Foram analisados 37 Programas de Pós-graduação acadêmicos em Ciência da Computação avaliados pela CAPES nos triênios 2004-2006 e 2007-2009. Foram identificados 732 professores nos programas avaliados, obtidos pelos relatórios trienais da CAPES, e de

posse dos nomes dos professores foi possível obter o identificador para o currículo de todos os professores, para, posteriormente, serem extraídos da Plataforma Lattes. Após o processo de extração, os currículos passaram por um pré-processamento e seus dados foram inseridos em um banco de dados relacional. Após a inserção dos dados curriculares, estes foram enriquecidos com informações disponíveis em diversos repositórios contendo dados sobre conferências e periódicos de publicações científicas. Essas informações descrevem o nível de cada um dos periódicos ou conferências, a quantidade de citações de um determinado trabalho publicado, dentre outros. Para a identificação de colaborações, foi feita a comparação dos títulos das publicações, dos currículos em análise com todos os demais, visando encontrar títulos iguais ou similares e, desta forma, caracterizar uma colaboração. Para isto, foi calculada a distância de Levenshtein (NAVARRO, 2001).

Já Mena-Chalco et al. (2014) utilizam dados dos currículos da Plataforma Lattes para identificar e caracterizar a rede de colaboração de pesquisadores brasileiros. O trabalho objetivou extrair os dados de currículos cadastrados na Plataforma Lattes, identificar automaticamente a colaboração baseada em informações bibliométricas, produzindo uma rede de colaboração, e aplicar métricas baseadas em análise topológica para compreender como ocorre a interação entre os pesquisadores. Similar ao trabalho de Mena-Chalco, Digiampietri e Cesar-Junior (2012), foi adotado o cálculo da distância de Levenshtein, com o intuito de encontrar títulos iguais ou similares. Com as redes de colaboração geradas, aspectos topológicos foram analisados, considerando apenas informações estruturais das redes, sem considerar pesos nos vínculos entre autores ou qualidade das publicações. Foi possível identificar que, em determinadas áreas, os pesquisadores tendem a realizar suas publicações com um número baixo de coautores em detrimento a outras áreas, em que a quantidade de coautores tende a ser elevada. Outro aspecto importante é que foi possível identificar que poucos pesquisadores possuem um alto grau de conectividade. Além disso, foi possível observar que a quantidade de colaboração entre os pesquisadores tem crescido consideravelmente, identificada pelo grau dos nós das redes analisadas em diferentes períodos de tempo.

Boaventura et al. (2014) caracterizam a evolução das redes de colaboração científica, representadas pelas redes de coautoria das seguintes universidades brasileiras: UFAM, UFMG, UFPE, UFRGS, UFRJ, UNB, UNICAMP e USP. A análise temporal considera os anos entre 2000 e 2013. O foco da análise são pesquisadores (professores) dessas universidades. Métricas tradicionais de redes sociais, tais como densidade e diâmetro da rede, grau de centralidade, grau de proximidade e grau de intermediação, foram calculadas. Os autores apresentam também a análise da endogamia⁴ dessas Instituições, e a sua correlação com a avaliação trienal dos Programas de Pós-Graduação realizada pela CAPES. Por fim, analisam a existência de grupos de pesquisadores que, frequentemente, publicam em conjunto. As principais conclusões da análise apresentada por Boaventura et al. (2014) são as seguintes: (1) a rede de colaboração das universidades analisadas possui a mesma característica de densificação das colaborações entre os seus pesquisadores. O diâmetro relativamente baixo (comparado à cardinalidade da rede) indica que as redes possuem características

⁴ Para Montolio, Dominguez e Larriba-Pey (2013), a endogamia de um conjunto de pesquisadores está relacionada a frequência em que tais pesquisadores colaboram entre si.

de redes de mundo pequeno; (2) universidades que possuem a maior porcentagem de Programas de Pós-Graduação com conceitos acima de 5 possuem valores menores de endogamia.

Lima et al. (2015) fazem uma avaliação do desempenho dos principais pesquisadores que atuam em Ciência da Computação. São utilizados dados extraídos dos currículos da Plataforma Lattes de 406 pesquisadores bolsistas de produtividade em pesquisa do CNPq na área de Ciência da Computação nas cinco modalidades da bolsa (1A, 1B, 1C, 1D e 2). A avaliação considerou três dimensões centrais, sendo: tempo de carreira do pesquisador, em que foi considerada a quantidade de anos após a conclusão do doutorado; quantidade de alunos orientados; e produtividade científica, referente ao volume de publicações e citações. Com relação ao tempo de carreira dos pesquisadores, observa-se que aqueles que possuem os níveis mais elevados de bolsa também possuem maior tempo de carreira. Já com relação às orientações, existe uma distinção entre a quantidade de orientações de mestrado e doutorado, em que há, aproximadamente, uma orientação de doutorado para cada ano de carreira entre todos os bolsistas. Ao se considerar as orientações de mestrado, este valor cai para menos de 0,5 orientação para cada ano após a conclusão do doutorado. Ao se avaliar a produção científica dos pesquisadores, observa-se que o volume de publicações aumenta com o nível das bolsas, exceto no caso do nível 1A, que se assemelha ao 1C, por incluir pesquisadores mais antigos, estes não estavam vinculados à programas com certo nível de maturidade. Estas avaliações demonstram coerência entre os itens avaliados e as modalidades das bolsas em que os pesquisadores se enquadram na área de Ciência da Computação.

Sidone, Haddad e Mena-Chalco (2016) apresentam o papel da geografia na evolução da produção e colaboração científica no Brasil entre 1992 e 2009. Para os estudos, foram utilizados dados dos currículos da Plataforma Lattes de um milhão de pesquisadores. Os autores destacam o processo de desaceleração da produção científica brasileira a partir dos últimos triênios analisados. No trabalho, foi possível identificar que o padrão de crescimento das grandes áreas é semelhante nos triênios iniciais da análise, quando praticamente todas apresentaram produção em taxas duplicadas entre 1992-1994 e 1995-1997. No entanto, percebe-se que o crescimento desacelerado ao longo do período é comum a todas as grandes áreas, porém em intensidades distintas. Como exemplo, verifica-se que a desaceleração no ritmo da produção científica ocorreu de maneira mais forte nos últimos triênios, primordialmente nas grandes áreas de Engenharias, Ciências Exatas e da Terra, Ciências da Saúde e Ciências Biológicas.

Diversos outros trabalhos têm explorado a Plataforma Lattes como principal fonte de dados para análise bibliográfica e de redes de colaboração, dentre os quais os de Pacheco et al. (2007), Cardoso e Machado (2008), Arruda et al. (2009), Alves, Yanasse e Soma (2011), Dias et al. (2013), Digiampietri, Mugnaini e Alves (2013), Roos et al. (2014), Araújo et al. (2014), Furtado et al. (2015), e Brito, Quoniam e Mena-Chalco (2016).

Observa-se que a análise dos dados dos currículos da Plataforma Lattes pode fornecer informações importantes para compreensão do conhecimento científico brasileiro e como ele tem evoluído, tendo em vista a quantidade de trabalhos recentes que têm

considerado os currículos como principal fonte de dados. Percebe-se a importância e interesse em realizar-se análises bibliométricas e baseadas em redes sociais em todo o conjunto de dados. No entanto, devido às restrições impostas pelo CNPq para acesso aos dados, como, por exemplo, a necessidade de se validar os captchas para acesso a cada um dos currículos, limitou-se à utilização de ferramentas já consolidadas conforme já mencionado, reduzindo o potencial de estudos sobre os dados curriculares da Plataforma Lattes. Além disso, particularidades do repositório como ambiguidade entre nome de indivíduos e a falta de vínculos explícitos entre os coautores de trabalhos dificultam muito a análise dos dados, já que a identificação das colaborações passa a ser uma tarefa não trivial, contribuindo para o fato de que a maioria dos trabalhos tem analisado apenas grupos específicos de indivíduos ou pequenos períodos de tempo.

O conteúdo desta tese está inserido no contexto de extração e análise de dados científicos. Entretanto, ao contrário dos trabalhos relacionados anteriormente apresentados nesta seção, o presente trabalho não se restringe a pequenos grupos de pesquisadores ou conjuntos de publicações. A novidade está na extração e análise de todo o conjunto de currículos cadastrados na Plataforma Lattes. Para isso, o LattesDataXplorer é fundamental no processo de coleta, tratamento e análise dos dados, conforme apresentado no Capítulo 4.

Capítulo 4

Coleta e Tratamento dos Dados

Neste capítulo, é descrito o LattesDataXplorer, arcabouço desenvolvido nesta tese e responsável pela coleta e tratamento dos dados extraídos da Plataforma Lattes. São apresentadas algumas motivações que tornam a Plataforma Lattes um importante repositório para a análise de dados científicos e que serviram de motivação para esta tese (Seção 4.1). Após, é apresentada a arquitetura geral do LattesDataXplorer, bem como seu processo de coleta dos dados (Seção 4.2). São ainda detalhadas suas principais funcionalidades como a seleção de grupos (Subseção 4.2.1), a identificação e caracterização de redes de colaboração científica (Subseção 4.2.2) e, por fim, as estatísticas e métricas calculadas (Subseção 4.2.3).

4.1 Motivações

Uma grande parte dos editais de financiamento de projetos de pesquisa realizados por agências de amparo à pesquisa, por instituições de ensino, bem como, pelo próprio CNPq, utiliza dados cadastrados nos currículos da Plataforma Lattes dos proponentes como uma das formas de avaliação das propostas. Isto passou a ser um grande incentivo para que os pesquisadores mantenham seus currículos com informações atualizadas, tornando a Plataforma Lattes uma fonte extremamente rica para análise da produção científica brasileira. Além disso, jovens estudantes que pretendem realizar parte de seus estudos em instituições estrangeiras, financiados por agências governamentais brasileiras, também devem possuir currículos cadastrados na Plataforma Lattes, contribuindo de forma significativa para o aumento na quantidade de currículos atualmente dispostos na Plataforma Lattes. A Figura 4.1 apresenta o crescimento da quantidade de currículos cadastrados durante os meses de maio de 2014 a junho de 2016. Nesse período, houve um aumento de aproximadamente 28% na quantidade de currículos armazenados.

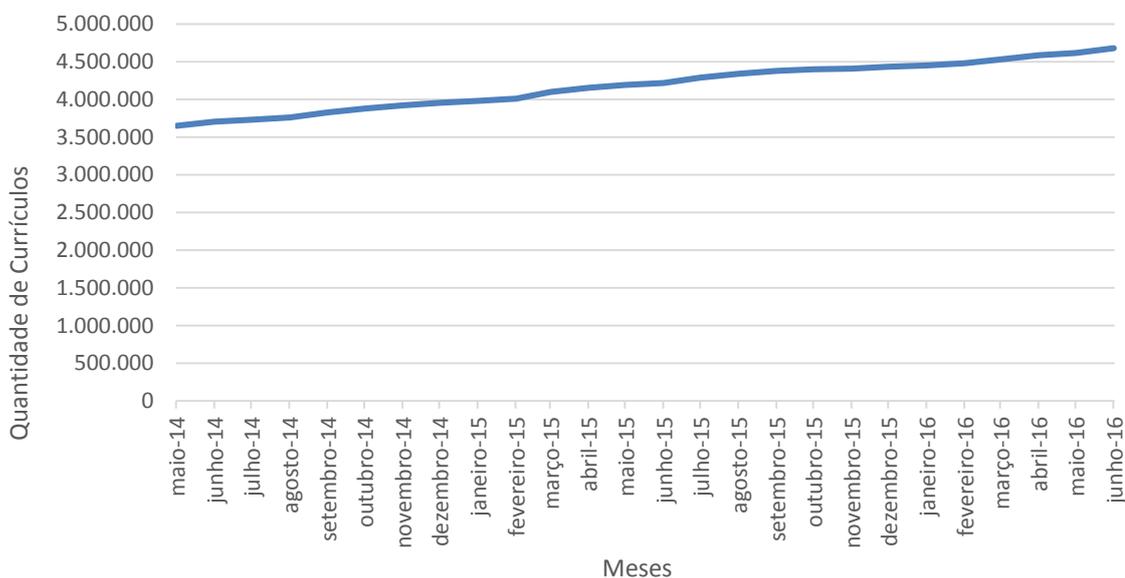


Figura 4.1: Crescimento da Plataforma Lattes entre maio de 2014 e junho de 2016.

É possível identificar que a Plataforma Lattes vem agregando, a cada dia, novos indivíduos que registram dados sobre suas atividades profissionais, acadêmicas e de pesquisa, se apresentando como uma fonte extremamente rica e consistente de informações que podem revelar como têm sido realizadas as pesquisas científicas brasileiras nas diversas áreas do conhecimento. Logo, explorar seus dados com o objetivo de se obter uma visão detalhada de todo o repositório, pode proporcionar uma visão inédita sobre a produção científica brasileira e como tem ocorrido a colaboração entre os pesquisadores brasileiros, fornecendo subsídios para análises e comparações sobre as diversas áreas de atuação dos indivíduos que as compõem.

Uma caracterização geral de todo o repositório de dados curriculares na proporção atual e com grande abrangência de estudos não foi tratada anteriormente, já que os trabalhos correlatos utilizaram apenas grupos específicos de currículos, tendo em vista que a manipulação de todos os currículos da Plataforma Lattes não é uma tarefa trivial. A dificuldade envolve problemas de recuperação de informação e algoritmos eficientes para manipulação de grande volume de dados.

4.2 LattesDataXplorer

Um dos propósitos desta tese é utilizar tecnologias envolvidas no processo de extração de dados da Web para realizar a coleta de todos os currículos da Plataforma Lattes. Com os currículos extraídos, outras consultas são realizadas com o intuito de enriquecer ainda mais os dados obtidos, como, por exemplo, número de citações dos artigos ou informações adicionais como a qualidade dos meios de publicação destes artigos. Para isso, um arcabouço denominado LattesDataXplorer foi desenvolvido (Figura 4.2).

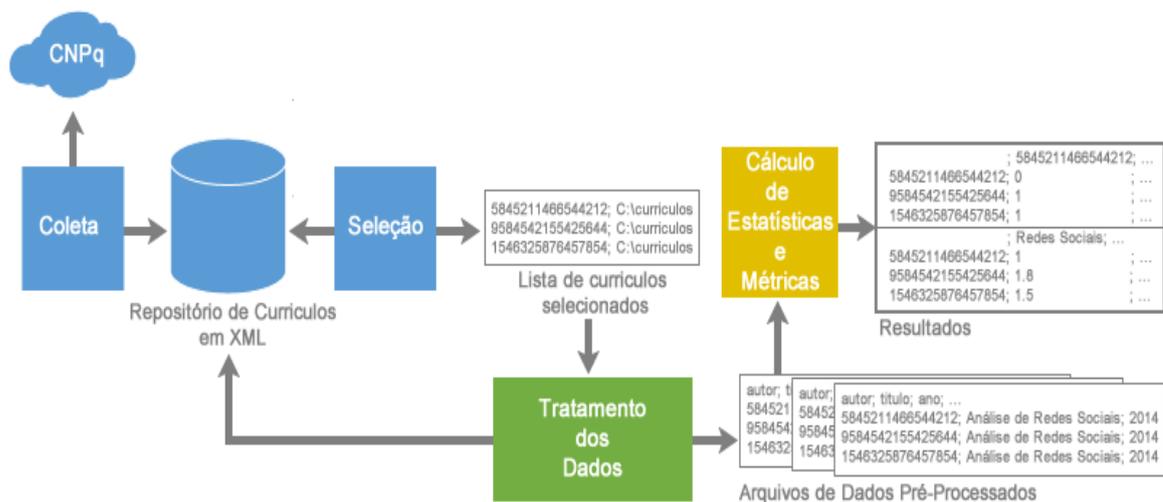


Figura 4.2: Visão geral do LattesDataExplorer.

O LattesDataExplorer é responsável por englobar todo o conjunto de técnicas e métodos para a coleta, tratamento e análise dos dados utilizados nesta tese. Ele é composto por um conjunto de componentes que são responsáveis por todo o processo de coleta e tratamento dos dados. Na Figura 4.3 são apresentados os componentes do módulo de coleta e seleção dos dados. O processo de extração de todos os dados curriculares da Plataforma Lattes é dividido em três componentes que objetivam minimizar o custo computacional: 1) extração de URLs, que é responsável por extrair as referências únicas para todos os currículos cadastrados, e dessa forma, possibilitar o acesso individual a cada currículo, 2) extração de Ids e Data, que visa acessar cada currículo e extrair o seu identificador individual, bem como a data de última atualização, 3) extração de currículos, que é responsável por extrair e armazenar os currículos cuja data de atualização na Plataforma Lattes seja divergente da data de atualização do currículo armazenado localmente.

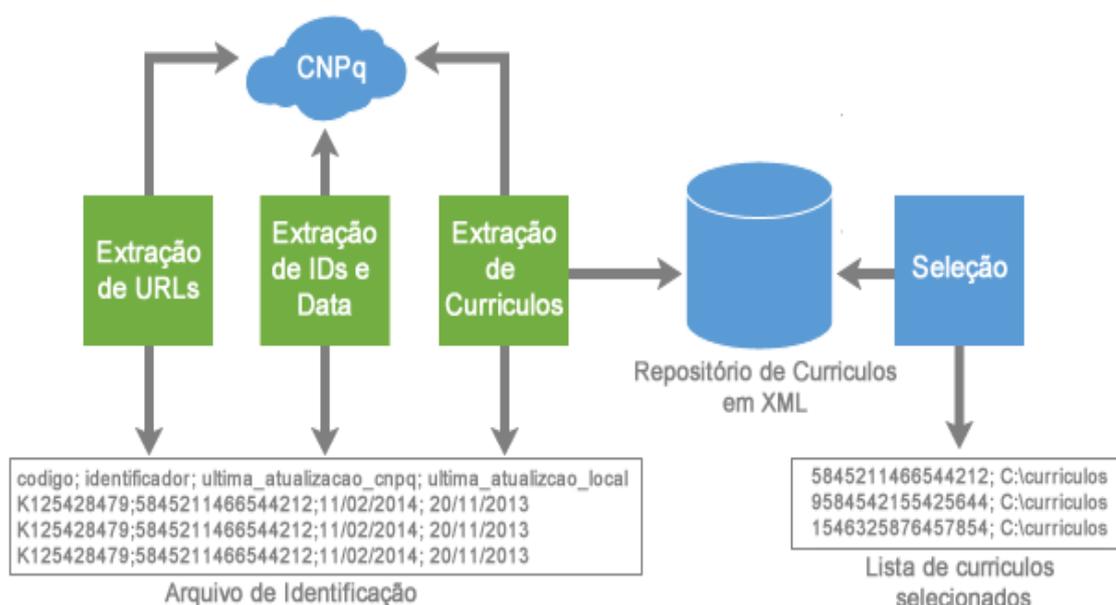


Figura 4.3: Processo de extração e seleção de dados.

Todas essas etapas se fazem necessárias, já que o ideal é manter os dados curriculares atualizados com a maior frequência possível, possibilitando a realização de análises com dados atualizados, tendo em vista que, com a estratégia adotada, não se faz necessário coletar todo o repositório de dados a cada nova extração. É importante, ainda, considerar que os currículos atualizados podem ter novos dados inseridos, bem como, a alteração e exclusão de dados já registrados, o que torna o processo de atualização de campos específicos uma tarefa complexa. Na abordagem adotada nesta tese, todo o currículo atualizado é substituído pelo mais recente.

A Plataforma Lattes possui uma interface de consulta bastante flexível que possibilita a busca por currículos baseada em parâmetros como nome, titulação, idioma, nacionalidade, área de atuação, dentre outros (Figura 4.4). Como resultado das consultas realizadas, são geradas listas de currículos que atendem aos parâmetros informados. Na interface de busca, o parâmetro mais utilizado é o nome do indivíduo (Figura 4.4 (1)), o que limita o universo de consulta, já que se faz necessário saber previamente o nome do indivíduo (ou parte dele) que se deseja consultar, inviabilizando o processo de geração de grupos baseado em determinados critérios. No entanto, pode-se informar parâmetros adicionais para as buscas (Figura 4.4 (2)), o que flexibiliza a consulta, possibilitando listar todos os currículos dos indivíduos de uma determinada área ou que possuem proficiência em um determinado idioma.

The image shows the search interface of the Lattes platform. At the top, there is a search bar with the URL 'buscatextual.cnpq.br/buscatextual/busca.do?metodo=apresentar'. Below the search bar, the CNPq logo and 'Currículo Lattes' are displayed. The main heading is 'Buscar Currículo Lattes (Busca Simples)' with a link to 'Busca Avançada'. The search criteria are set to 'Nome'. Below this, there are filters for 'Nas bases' (Doutores, Demais pesquisadores), 'Nacionalidade' (Brasileira, Estrangeira), and 'País' (Todos). A 'Tipo de filtro' section contains various checkboxes for filtering results, such as 'Bolsistas de Produtividade do CNPq', 'Formação Acadêmica/Titulação', 'Atuação profissional', 'Idioma', 'Atividade Profissional (Instituição)', 'Outros Bolsistas do CNPq', 'Nível do Curso de Pós-graduação onde é Docente', 'Atividade de Orientação', 'Áreas ou Setores da Produção em C&T', and 'Presença no Diretório de Grupos de pesquisa'. A 'Buscar' button is located at the bottom right.

Figura 4.4: Interface de busca da Plataforma Lattes.

Todo o processo de extração do LattesDataXplorer se inicia pela aquisição dos códigos dos currículos da Plataforma Lattes, obtidos por uma requisição na interface de consulta da plataforma, para que, posteriormente, esses códigos possam ser armazenados localmente. A consulta inicial realizada resulta em uma lista contendo todos os códigos de identificação de todos os currículos cadastrados, que permite acessar individualmente cada um dos currículos existentes na Plataforma Lattes (Figura 4.5).



Figura 4.5: Cabeçalho de um currículo Lattes.

A URL (*Uniform Resource Locator*) para acesso a cada um dos currículos é composta pelo endereço de busca na plataforma, concatenado com o código que representa cada um dos indivíduos (Figura 4.5 (1)). Como todos os códigos estão armazenados localmente, é possível acessar todos os currículos de forma sequencial. Posteriormente, o componente de extração de Ids e Data é responsável por acessar cada um dos currículos cujos códigos foram armazenados e extrair de cada currículo o seu identificador único (Figura 4.5 (2)) e a data da última atualização do currículo (Figura 4.5 (3)). Para isso, somente o cabeçalho de cada currículo em que estão presentes estas informações precisam ser carregados. Estes dados são armazenados em um arquivo de identificação que contém código, identificador, data da última atualização no CNPq e data da atualização do currículo armazenado localmente.

Cada currículo possui um código de oito dígitos e um identificador de dezesseis dígitos. Ambos possibilitam o acesso a um determinado currículo da Plataforma Lattes e são únicos para cada indivíduo cadastrado. O componente de extração de URLs extrai os códigos que possibilitam acessar cada um dos currículos. Já os identificadores extraídos pelo componente de extração de Ids e Data também são armazenados, já que são estes identificadores que permitem, por exemplo, vincular coautores de um trabalho publicado em colaboração e registrado no currículo de um determinado indivíduo.

O arquivo de identificação é a base para a extração dos currículos. Toda a vez que é necessário atualizar o repositório local de currículos, o primeiro componente do processo de extração é executado, resultando na extração de todos os códigos cadastrados na plataforma. Códigos já registrados no arquivo de identificação são ignorados por já estarem inclusos, e novos códigos são inseridos ao final do arquivo, já que representam novos currículos que ainda não haviam sido extraídos.

Posteriormente, com o uso dos códigos, são acessados os cabeçalhos de cada um dos currículos e extraídos os respectivos códigos identificadores e as datas de atualização junto à Plataforma Lattes, tanto para currículos já extraídos como para os novos currículos, atualizando o arquivo de identificação a cada nova extração. O acesso ao cabeçalho possibilita maior agilidade em todo o processo aumentando de forma significativa a extração dos dados, já que não se faz necessário esperar todo o currículo ser gerado.

Finalmente, acontece a extração dos currículos. O extrator é responsável por verificar no arquivo de identificação se existem currículos cuja data de atualização no CNPq é divergente da data de atualização local; em caso afirmativo, o currículo é extraído e substitui o que estava no repositório local, atualizando, também, a data de atualização local no arquivo de identificação. Caso contrário, significa que o currículo local já é a última versão atualizada na Plataforma Lattes, tendo em vista que as duas datas são iguais. Nesse caso, não é realizada a extração do currículo, e o próximo identificador é analisado. Finalmente, são extraídos os novos currículos cadastrados, pois foram inseridos ao final do arquivo de identificação os códigos dos novos currículos, seus identificadores e as datas de atualização no CNPq, já que não possuem data de atualização local, tendo em vista que ainda não tinham sido extraídos. Diante disso, esses currículos são extraídos pela primeira vez e sua data de atualização local é registrada no arquivo de identificação.

Todo esse processo possibilita manter um repositório atualizado com baixo custo computacional, já que um percentual considerável de currículos não é atualizado frequentemente. Com exceção da primeira extração, o processo apresentado nesta tese possibilita extrair somente os currículos novos ou os recentemente atualizados. A substituição dos currículos atualizados ocorre, pois o custo para realizar a atualização de uma determinada seção no arquivo armazenado localmente seria maior que simplesmente substituir o currículo existente.

Todos os currículos são armazenados em formato XML. Como o objetivo desta tese é realizar uma análise dos dados curriculares cadastrados na Plataforma Lattes e não gerar um novo repositório com os dados, optou-se por não realizar o armazenamento desses dados em bancos de dados relacionais, tendo em vista que este processo iria resultar em processamento adicional a cada atualização dos dados, já que seriam necessárias atualizações no banco de dados.

Além disso, considerando a necessidade de análises de grupos específicos, como, por exemplo, de uma instituição ou de uma área de pesquisa em particular, um componente de seleção foi criado. Ele é responsável por gerar subgrupos de currículos baseados em critérios de interesse.

4.2.1 Seleção de Grupos

Com todos os currículos armazenados localmente em formato XML, a possibilidade de manipulação dos dados com flexibilidade permite explorar todo o potencial que os dados curriculares da Plataforma Lattes oferecem. Com o intuito de analisar grupos específicos de currículos, como os formados por docentes de um programa de pós-graduação ou de uma instituição em particular, é descrito o componente denominado Seleção, para a composição de subgrupos de currículos baseada em informações presente em seus registros.

Para isso, é usada a linguagem de consulta XPath (*XML Path Language*)¹ para pesquisa e posterior geração dos subgrupos. A linguagem XPath permite construir expressões que vão processar e percorrer um documento XML de forma similar ao uso de expressões regulares. Logo, é possível agrupar um conjunto de currículos com base em parâmetros desejados. Diante disso, em busca pelos parâmetros em cada um dos currículos, independentemente ou não de qual seção ele(s) seja(m) encontrado(s), estes currículos são selecionados e formam um grupo para análises. Os dados destes currículos são organizados em uma lista de currículos selecionados.

Esta lista é responsável por armazenar os identificadores dos currículos e o caminho em que cada um deles está armazenado localmente, para que, dessa forma, seja possível analisar somente os currículos selecionados para um determinado grupo. Isso permite que, quando análises bibliométricas ou de redes forem ser realizadas, basta informar qual o grupo de currículos a ser analisado. Desse modo, será possível analisar a rede composta por todos os indivíduos (rede global com todos os currículos) ou redes específicas que foram geradas seguindo critérios especificados.

É importante destacar que, no momento da consulta, é possível a utilização de operadores lógicos que permitem maior flexibilidade no momento da geração dos subgrupos. O uso dos operadores **e** (*and*) e **ou** (*or*) possibilita a concatenação de parâmetros na consulta XPath. Por exemplo, subgrupos que sigam critérios como 'nível de formação = doutorado completo' e 'ano de conclusão = 2015' poderiam ser gerados, ou, ainda, subgrupos com 'grande área de atuação = ciências exatas e da terra' e 'estado de atuação = Minas Gerais', seriam facilmente caracterizados. Tal facilidade resulta em um mecanismo extremamente importante para a geração de subgrupos específicos que vão possibilitar a análise de outros conjuntos que não podem ser gerados pelas ferramentas de consulta da Plataforma Lattes.

Diante disso, todas as etapas posteriores à geração de grupos específicos utilizam a lista de currículos selecionados no momento de identificar quais serão os currículos a serem analisados. As listas ficam armazenadas em um repositório de listas, em que várias listas podem ser geradas e sendo estas referenciadas no momento de análise. Logo, o componente de seleção se apresenta como uma importante facilidade para a seleção de subgrupos dos currículos a serem analisados. Com a adoção de XPath, as consultas aos arquivos XML são extremamente eficientes e com resultados confiáveis. Após a identificação dos subgrupos, estudos que considerem somente um conjunto particular de indivíduos ou de todo o repositório podem ser realizados.

¹ <https://www.w3.org/TR/xpath/>

4.2.2 Identificação e Caracterização de Redes de Colaboração Científica

Na etapa de identificação e caracterização das redes de colaboração científica, o componente de Tratamento dos Dados é responsável por analisar os currículos dos grupos selecionados e gerar as redes de colaboração científica. O componente utiliza a lista de currículos selecionados, gerada para um determinado grupo, o que possibilita identificar quais são os currículos que vão ser analisados. A lista possui o identificador e o local de armazenamento do referido currículo e, diante desta indicação, apenas os currículos presentes na lista são considerados. Além disso, o componente também é responsável por analisar cada um dos currículos, de tal forma que sejam produzidos arquivos de dados pré-processados que vão possibilitar a realização de análises bibliométricas da produção científica, como também da colaboração científica dos indivíduos que compõem os grupos em análise (Figura 4.6).

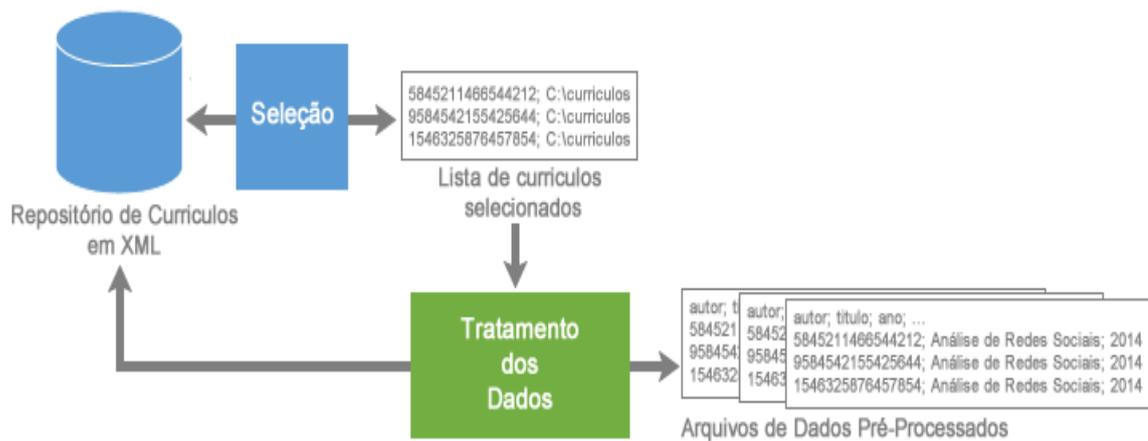


Figura 4.6: Processo de tratamento dos dados.

O maior desafio no tratamento dos dados coletados dos currículos da Plataforma Lattes está relacionado com a maneira com que cada indivíduo preenche os seus dados no currículo: entradas são realizadas manualmente e, portanto, não é uma situação incomum dois indivíduos cadastrarem o mesmo trabalho com informações divergentes, como o título do trabalho ou, até mesmo, nomes de colaboradores incompletos (BOAVENTURA et al., 2014).

Um dos enfoques desta tese é gerar a rede de colaboração dos indivíduos com currículos cadastrados na Plataforma Lattes. O desafio está em identificar as colaborações em um grande volume de dados em tempo hábil, tendo em vista os diversos problemas envolvidos como duplicatas nos nomes de autores e divergência no título da mesma publicação quando cadastrada por coautores. O currículo a ser analisado pode possuir várias seções que vão permitir a identificação das colaborações com outros currículos, como, por exemplo, trabalhos realizados em conjunto. No entanto, por ausência de um processo automático de identificação de colaborações, a relação de colaboradores de um indivíduo não é identificada automaticamente pela Plataforma Lattes, (como também acontece em outros repositórios de publicações científicas). Logo, técnicas para a identificação das colaborações de forma automática se fazem necessárias para análise de grandes volumes de dados.

Para a identificação de colaborações, podem ser utilizados vários elementos que permitam relacionar pares de indivíduos, como: publicação de trabalhos em conjunto (ou seja, coautorias), orientações, participação em bancas de avaliação, organização de eventos, participação em projetos de pesquisa, dentre outros. Diante disso, especificamente na identificação de colaboração científica em publicações, um pré-processamento nos títulos dos trabalhos para a identificação é realizado. Este pré-processamento consiste em três etapas principais, aqui denominadas de remoção de *stop words*, normalização e padronização.

Em um documento textual, podem ser encontrados diversos *tokens* que não possuem valor semântico. Estes *tokens* são importantes para melhor entendimento do texto de forma geral e também são classificados como *stopwords*. As *stopwords* compõem uma lista denominada *stoplist*, a qual é frequentemente utilizada em sistemas de mineração de textos. Palavras que compõem a *stoplist* são removidas do documento original, reduzindo o texto a ser analisado, de modo a não provocar problemas no conteúdo semântico do documento. Uma *stoplist* bem elaborada permite a eliminação de muitos termos irrelevantes, tornando mais eficiente o resultado obtido pelo processo de mineração de textos (CARRILHO-JUNIOR, 2007). A normalização e a padronização consistem em remover caracteres especiais, letras com acentuação e espaços em branco, e deixar a *string* resultante com todas as letras em minúsculo.

É importante destacar que esta tese não possui ênfase na desambiguação de autores, sendo esta tarefa foco de diversos outros trabalhos que tentam tratar a melhor forma de desambiguar nome de autores (FERREIRA; GONÇALVES; LAENDER, 2012). Nesta tese, são considerados os identificadores de cada autor e não o nome dos autores para o processo de identificação de colaborações. Ou seja, cada indivíduo é referenciado pelo seu identificador único na Plataforma Lattes.

De posse de todos os currículos armazenados e utilizando a lista de currículos previamente selecionados, que especifica o subgrupo de indivíduos a serem analisados, é proposto um método para a identificação de colaborações científicas. Para isso, todos os títulos dos artigos cadastrados no currículo de cada um dos autores são analisados e, conseqüentemente, se tornam a base para construção da rede de colaboração. Para auxiliar o processo de caracterização, é utilizado um dicionário que possibilita vincular os artigos (chaves do dicionário) a seus autores (identificadores). Todas as etapas do método de identificação podem ser visualizadas no Algoritmo 4.1.

1. $n \leftarrow$ número de artigos do autor
2. **for** $i \leftarrow 1$ **to** n
3. $x \leftarrow$ *string*[i] // x recebe o título do artigo [i]
4. $x \leftarrow$ *stopword*[x] // remove *tokens* sem valores semânticos
5. $x \leftarrow$ *normalization*[x] // remove espaços e acentuação
6. $x \leftarrow$ *lowercase*[x] // padroniza em minúsculo
7. $x \leftarrow$ *concatenat*[x],*year* // acrescenta o ano no final da string
8. **if** *hash*[x] **in** *dictionary* // verifica se x está no dicionário
9. $\text{dictionary}[x] \leftarrow$ *id_author*
10. **else** *dictionary* $\leftarrow x, \text{id_author}$

Algoritmo 4.1: Etapas do método para caracterização de redes de colaboração.

Como pode ser observado pelo Algoritmo 4.1, cada título de um trabalho cadastrado em um determinado currículo passa por uma transformação que tem como objetivo obter o título sem as palavras que não tenham valor semântico, sem acentuação e sem os espaços. Essa estratégia tem como intuito minimizar os erros ortográficos que podem estar inseridos nos títulos dos artigos e estabelecer uma chave única para cada trabalho. Conseqüentemente, todo o texto é padronizado em letras minúsculas e a *string* resultante é concatenada com o ano da publicação para, posteriormente, ser transformada em uma chave que representa o trabalho em análise, conforme mostram as linhas 2 a 7 do algoritmo. Exemplo desta transformação pode ser visualizado na Tabela 4.1.

Tabela 4.1: Exemplo de transformação dos títulos em chaves para o dicionário de identificação.

Etapa do Algoritmo	Resultado
3	Modelagem e Caracterização de Redes Científicas: Um Estudo Sobre a Plataforma Lattes
4	Modelagem Caracterização Redes Científicas Estudo Sobre Plataforma Lattes
5	ModelagemCaracterizacaoredescientificasestudosobreplataformalattes
6	modelagemcaracterizacaoredescientificasestudosobreplataformalattes
7	modelagemcaracterizacaoredescientificasestudosobreplataformalattes2013

A razão da inserção do ano da publicação no final da *string* tem como objetivo minimizar a identificação de falsas colaborações. Pelo método descrito, caso existissem dois títulos idênticos de pesquisadores distintos que nunca trabalharam em conjunto, seria identificada uma colaboração, tendo em vista que a chave resultante seria a mesma. Com a inclusão do ano na *string*, falsas colaborações somente seriam consideradas se existirem dois artigos distintos com o mesmo título e publicados no mesmo ano.

Após a transformação do título da publicação em uma chave, verifica-se no dicionário utilizado para a caracterização da rede de colaboração se a chave já está presente. Caso a chave já exista no dicionário, o identificador do autor do currículo em análise é vinculado a essa chave, concatenando o seu identificador à lista de identificadores já vinculados a ela. Caso contrário, a chave é inserida no dicionário, sendo vinculado a ela o identificador do indivíduo. Um exemplo de dicionário pode ser visualizado na Tabela 4.2.

Tabela 4.2: Exemplo de dicionário de identificação.

Chave do Dicionário	Autores
modelagemcaracterizacaoredescientificasestudosobreplataformalattes2013	Id01, Id25
studyaboutinfluenceacademicperformancestudentssocialnetworks2013	Id25, Id175, Id98
analysiscollaborationnetworksscificpublications2013	Id01, Id98, Id67
...	
...	
processoparaidentificacaocolaboradoresredescientificas2013	Id01, Id28, Id174

A utilização dos identificadores é confiável já que cada um dos currículos cadastrados na Plataforma Lattes possui um identificador único. Este identificador é utilizado tanto para caracterizar o indivíduo na Plataforma Lattes, como também para permitir acesso ao seu currículo. Como resultado, isso significa que identificadores que estejam vinculados a uma mesma chave possuem colaboração entre si.

Uma dificuldade encontrada pelos autores no cadastro de uma publicação está no momento da citação de seus colaboradores. Devido à falta de padronização e possibilidade de ambiguidade no nome de citação dos colaboradores de um trabalho, não é possível garantir que o nome informado na lista de colaboradores pertença a somente um indivíduo da Plataforma Lattes. Isso dificulta a caracterização das redes de colaboração apenas pelos nomes de citação informados em cada uma das publicações cadastradas. Diante disso, a Plataforma Lattes permite que um autor vincule seus coautores pelos seus identificadores. Isso possibilita que, ao contrário de simplesmente informar um nome de citação de um colaborador, seja informado o seu nome vinculado ao identificador único desse colaborador. Logo, quando há a citação deste colaborador que foi previamente vinculado, além do nome de citação, o seu código de identificação também é incorporado à publicação cadastrada.

A Figura 4.7 apresenta um fragmento de um currículo em formato XML, em que constam dados de uma publicação. Além do indivíduo do currículo em análise (Figura 4.7 (1)), é possível obter os identificadores dos colaboradores que foram vinculados (Figura 4.7 (2)), diferentemente dos autores cujos vínculos não foram identificados (Figura 4.7 (3)). Nesse último caso, só é possível obter os nomes.

2 Clique para visualizar o currículo

1 **DIAS, Thiago. M. R. ; Moita, Gray Farias ; Patrícia Mascarenhas Dias ; T. Moreira ; FERREIRA, L. R. .**
 Modelagem e Caracterização de Redes Científicas: Um Estudo Sobre a Plataforma Lattes. In: BRASNAM - II Brazilian Workshop on Social Network Analysis and Mining, 2013, Maceio. Anais do BRASNAM - II Brazilian Workshop on Social Network Analysis and Mining. Porto Alegre: SBC, 2013. v. 01. p. 10-20.

```

<TRABALHO-EM-EVENTOS>
  <DADOS-BASICOS-DO-TRABALHO NATUREZA="COMPLETO"
    TITULO-DO-TRABALHO="Modelagem e Caracterização de Redes
    Científicas: Um Estudo Sobre a Plataforma Lattes"
    ANO-DO-TRABALHO="2013"
    PAIS-DO-EVENTO="Brasil"/>
  <DETALHAMENTO-DO-TRABALHO CLASSIFICACAO-DO-EVENTO="INTERNACIONAL"
    NOME-DO-EVENTO="BRASNAM - II Brazilian Workshop on Social Network
    Analysis and Mining"
    CIDADE-DO-EVENTO="Maceio"
    ANO-DE-REALIZACAO="2013"/>
  1 <AUTORES NOME-COMPLETO-DO-AUTOR="Thiago Magela Rodrigues Dias"
    NOME-PARA-CITACAO="DIAS, Thiago. M. R."
    ORDEM-DE-AUTORIA="1"/>
  2 <AUTORES NOME-COMPLETO-DO-AUTOR="Gray Farias Moita"
    NOME-PARA-CITACAO="Moita, Gray Farias"
    ORDEM-DE-AUTORIA="2"
    NRO-ID-CNPQ="2550201329788172"/>
  2 <AUTORES NOME-COMPLETO-DO-AUTOR="Patrícia Mascarenhas Dias"
    NOME-PARA-CITACAO="Patrícia Mascarenhas Dias"
    ORDEM-DE-AUTORIA="3"
    NRO-ID-CNPQ="6871965805554986"/>
  3 <AUTORES NOME-COMPLETO-DO-AUTOR="Tales Moreira"
    NOME-PARA-CITACAO="T. Moreira"
    ORDEM-DE-AUTORIA="4"
    NRO-ID-CNPQ=""/>
  3 <AUTORES NOME-COMPLETO-DO-AUTOR="Leandro Roberto Ferreira"
    NOME-PARA-CITACAO="FERREIRA, L. R."
    ORDEM-DE-AUTORIA="5"
    NRO-ID-CNPQ=""/>
</TRABALHO-EM-EVENTOS>
  
```

Figura 4.7: Exemplo de lista de colaboradores vinculados a seus identificadores.

Muitos indivíduos nem sempre realizam a vinculação do identificador único de um colaborador ao seu nome. Isso ocorre em grande maioria por não ser um processo automático na atualização dos currículos, ou seja, o autor de um trabalho publicado deve vincular seus colaboradores a seus identificadores manualmente, o que gera trabalho extra e desestimula uma grande parte dos indivíduos a realizar estes vínculos. Porém, quando tais vínculos são realizados, isso possibilita ao método de identificação de colaborações descrito nesta tese obter melhores resultados do que outros métodos geralmente utilizados para identificação de colaborações. Tal situação possibilita que, mesmo quando apenas um dos autores tenha cadastrado um determinado trabalho, seja possível identificar a colaboração de seus coautores, diferentemente dos métodos que trabalham, por exemplo, com comparação de títulos, em que, obrigatoriamente, só é identificada a colaboração entre dois coautores se ambos tenham registrado a mesma publicação.

Logo, quando o nome de um colaborador que tenha sido vinculado a seu identificador está inserido em uma publicação de um determinado indivíduo que esteja sendo analisada, no momento da inserção desta publicação no dicionário de chaves e identificadores, não apenas o identificador do pesquisador cujo currículo esteja sendo analisado é inserido, mas também os de todos os coautores que estiverem sido vinculados. Isto faz com que o algoritmo melhore seus resultados, já que é possível identificar colaborações mesmo que apenas um dos autores tenha cadastrado a publicação em seu currículo.

Além disso, como cada título de uma publicação é analisado uma única vez, o algoritmo tem custo linear $\theta(n)$, tendo em vista que não são feitas comparações com outros títulos, diferentemente de outras técnicas, como o cálculo da distância de Levenshtein, em que são comparados todos os títulos uns com os outros, visando encontrar títulos iguais ou semelhantes e, dessa forma, caracterizar um trabalho em colaboração. Tais técnicas possuem custo polinomial ($O(n^2)$), o que inviabiliza a caracterização de redes com grande quantidade de nós.

Mena-Chalco, Digiampietri e Cesar-Junior (2012) apresentam o custo computacional do ScriptLattes, que faz a comparação de títulos com o cálculo da distância de Levenshtein. Com o uso do ScriptLattes, o número de comparações requeridas para identificar coautorias em uma lista de n produções bibliográficas está na ordem de $\theta(1/2(n^2 + n)) = O(n^2)$. Na prática, se $n = 10.000$, e cada comparação for realizada em 0.001 segundos, então, serão requeridas em torno de 13,89 horas para processar todas as comparações do grupo.

Tal custo inviabiliza a caracterização de redes com um grande volume de publicações, como, por exemplo, o conjunto de todos os artigos cadastrados nos currículos da Plataforma Lattes. Considerando a estratégia proposta nesta tese, com custo na ordem de $\theta(n)$ e, seguindo os mesmos critérios descritos anteriormente, seriam necessários para $n = 10.000$, em torno de 10 segundos para a realização do processo de identificação.

Como prova de conceito, foram escolhidos grupos de indivíduos que possuem currículos cadastrados na Plataforma Lattes com o objetivo de verificar a eficiência do método descrito para o processo de identificação de colaborações, bem como, a

qualidade de seus resultados. Os resultados obtidos demonstram a viabilidade do método proposto (DIAS; MOITA, 2015).

Ao considerar o custo computacional, a adoção de métodos que trabalham com comparação de títulos entre as publicações torna-se inviável quando é necessário o processamento de uma grande quantidade de dados, como é o caso de todos os currículos inseridos na Plataforma Lattes, tendo em vista o tempo de resposta para avaliar redes de grande porte com centenas de milhares de títulos. Já o método proposto nesta tese possui um custo computacional para n produções da ordem de $\theta(n)$ comparações, já que a única comparação realizada é a da existência da chave no dicionário, ou seja, a única comparação realizada com cada chave (título) transformada é se ela existe ou não no dicionário de identificação utilizado para a caracterização das redes.

Se considerados os resultados obtidos e o custo para a análise de grandes quantidades de trabalhos, o método proposto se apresenta como uma alternativa interessante para a análise de grandes volumes de dados com resultados satisfatórios. Após a análise de todos os títulos, a rede é gerada com a abordagem de construção de redes de cliques. Um conjunto de nós é dito como uma clique quando todos eles estão interligados entre si. Diante disso, para cada chave do dicionário que corresponde a um artigo, os elementos que representam os autores desse artigo são inseridos na rede como uma clique. Como um nó a ser inserido já pode estar na rede de colaboração, os cliques vão sendo unidos por justaposição do nó em comum.

Após a modelagem da rede, é possível identificar os nós que possuem maior grau de colaboração, caracterizados pelas arestas mais densas, classificar os nós por áreas de atuação, além de diversas outras características que podem ser extraídas dos indivíduos e de seus vínculos. Com a rede caracterizada, é possível aplicar diversas métricas de análise de redes sociais para melhor entendimento das características particulares de cada um dos nós, bem como, das características topológicas da rede.

Além da identificação das colaborações, são gerados arquivos de dados pré-processados que contêm todas as informações dos nós que vão compor a rede. Estes arquivos possuem informações como titulação, proficiência, afiliação, áreas de pesquisa, dentre outras. Estes arquivos são importantes, pois, diante das informações que eles agregam, é possível realizar diversas outras análises bibliométricas sem a necessidade de consultar o conjunto de currículos novamente.

4.2.3 Cálculo de Estatísticas e Métricas

Como resultados das etapas de tratamentos dos dados, são produzidos arquivos de dados pré-processados que contêm todas as informações necessárias para o cálculo de diversas métricas bibliométricas e baseadas em análise de redes sociais. O cálculo das estatísticas e métricas utilizando esses arquivos é facilitado, já que eles sumarizam todos os dados contidos nos currículos. Consequentemente, esses arquivos se tornam a base para o componente de Cálculo de Estatísticas e Métricas (Figura 4.8).

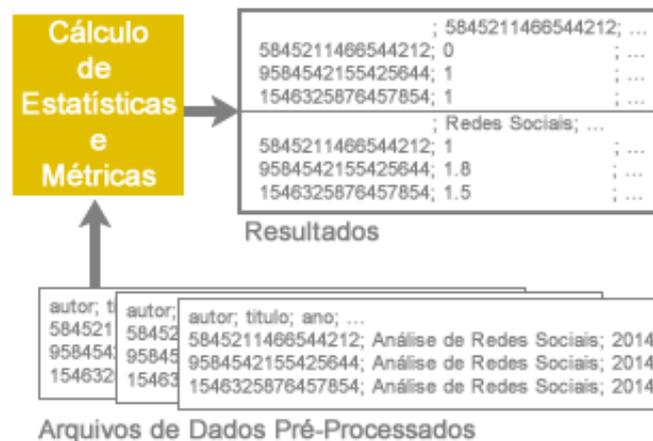


Figura 4.8: Componente para cálculo de estatísticas e métricas.

O componente de cálculo de métricas analisa arquivos de dados pré-processados que possuem somente os dados relevantes para os cálculos das métricas implementadas. Diante disso, não se fazem necessárias novas consultas aos currículos, o que possibilita realizar os cálculos de forma eficiente tendo em vista que todos os dados a serem utilizados no cálculo de uma determinada métrica estão agrupados em um único arquivo. Logo, o componente explora os dados produzidos pelas etapas anteriores e, dessa forma, extrai informações sobre como as redes estão estruturadas e qual o comportamento dos elementos em cada uma das redes em que se encontram. Também estão implementadas métricas bibliométricas que possibilitam verificar dados quantitativos e qualitativos sobre a produção científica dos indivíduos analisados. Exemplos de estatísticas e métricas que estão implementadas são: média de publicação por indivíduo, média de autores por trabalho publicado, frequência de publicação de uma determinada área, tópicos de pesquisa mais frequentes, pesquisadores mais relevantes em uma rede considerando seus graus, densidade das redes, caminhos mínimos entre nós, coeficientes de intermediação, h-index dos indivíduos, média de citações dos trabalhos, dentre outras.

Todas essas métricas possibilitam explorar os dados extraídos da Plataforma Lattes, proporcionando uma visão ampla da produção científica brasileira, tendo como fonte de dados principal os artigos publicados e cadastrados nos currículos, e ainda, analisar como ocorre a colaboração científica nacional e como a ciência brasileira tem evoluído. No próximo capítulo, é apresentada uma caracterização geral de todos os indivíduos com currículos cadastrados na Plataforma Lattes.

Capítulo 5

Caracterização dos Dados da Plataforma Lattes

Neste capítulo é apresentada uma caracterização geral dos dados contidos na Plataforma Lattes, a partir dos currículos coletados em abril de 2015. São apresentadas estatísticas gerais dos dados extraídos (Seção 5.1) e estatísticas por áreas e grandes áreas do conhecimento (Seção 5.2). Os resultados aqui apresentados podem servir como base para diversos outros trabalhos, em diferentes áreas de pesquisa.

5.1 Estatísticas Gerais

Os dados utilizados nesta caracterização foram coletados em abril de 2015 e correspondem a 4.156.635 currículos, distribuídos nas diversas áreas do conhecimento. Esses dados foram coletados conforme descrito no Capítulo 4, possibilitando realizar uma caracterização geral dos dados contidos nos currículos da Plataforma Lattes. As publicações consideradas para esta caracterização cobrem o período de 1961 (trabalho mais antigo cadastrado) até dezembro de 2014. Logo, apesar de a coleta ter sido realizada no início de abril de 2015, somente publicações até o fim do ano de 2014 foram consideradas. Dessa forma, espera-se que os diversos trabalhos finalizados no final de 2014 já estejam registrados nos currículos de seus respectivos autores, tendo em vista que tais trabalhos são incorporados geralmente após a sua conclusão, como no caso de dissertações e teses, ou publicação, como artigos em anais de congresso ou periódico.

É importante destacar a diversidade dos dados que estão registrados neste conjunto de currículos, como artigos publicados em anais de congresso e periódico, apresentação de trabalhos, participação em eventos, nível de formação, orientações concluídas, dentre outros. Ressalta-se que um determinado trabalho pode estar inserido em currículos distintos, já que este pode ter sido realizado em colaboração. Logo, no repositório da Plataforma Lattes, um trabalho pode aparecer várias vezes,

tendo em vista que ele pode ser registrado por cada um de seus autores. A Tabela 5.1 apresenta o quantitativo geral de todos os trabalhos registrados na Plataforma Lattes contabilizando os trabalhos de todos os currículos, bem como os totais únicos, em que são desconsiderados títulos de trabalhos iguais. Na coluna “Único”, são considerados apenas trabalhos com títulos distintos, para que, dessa forma, um mesmo trabalho registrado em vários currículos seja contabilizado uma única vez. No caso específico de orientações, tal situação ocorre quando existe o papel de coorientador em um trabalho, em que o trabalho orientado está registrado na seção de orientações concluídas tanto no currículo do orientador como também no currículo do coorientador.

Tabela 5.1: Quantitativo dos dados da Plataforma Lattes em abril de 2015.

Tipo de Trabalho	Geral	Único
Artigos em Anais de Congresso	11.591.142	7.437.979
Artigos em Periódico	4.560.921	2.806.910
Capítulos de Livro	1.055.388	820.732
Demais Trabalhos	667.944	610.778
Livros	455.447	360.902
Textos em Jornais e Revistas	1.353.645	1.220.309
Trabalhos Técnicos	1.538.111	1.278.378
Outras Produções Bibliográficas	882.443	799.036
Orientações de Doutorado	224.303	224.078
Orientações de Especialização	981.886	979.922
Orientações de Graduação	3.005.267	3.002.261
Orientações de Mestrado	737.572	736.096
Supervisões de Pós-Doutorado	24.662	24.662

Os dados apresentados correspondem aos registros de publicações de todos os currículos, independentemente de sua área de atuação ou de sua data de atualização. A quantidade de dados registrada corrobora a importância da Plataforma Lattes, confirmando ser um dos maiores repositórios de dados científicos mundiais da atualidade (LANE, 2010), caracterizando-se como uma fonte extremamente rica para análise da produção científica brasileira. É possível observar a tendência de publicação de artigos em anais de congresso, seguida em menor número pela publicação de artigos em periódico. Além dos artigos, estão registrados outros tipos de publicação, tais como capítulos de livro, livros, textos em jornais e revistas, trabalhos técnicos¹, outras produções bibliográficas que não se enquadrem nas categorias anteriores e demais trabalhos que um indivíduo julgar serem relevantes e tenha interesse em cadastrar, bem como, dados de orientações nos mais diversos níveis de formação.

É importante ressaltar que cada indivíduo possui um único currículo cadastrado, não existindo a possibilidade de dados duplicados sobre a produção de um mesmo indivíduo. Além disso, em cada currículo, é possível identificar a data de sua última atualização. Tendo em vista que a atualização dos currículos é realizada pelos próprios indivíduos, frequentemente são encontrados currículos que estão há algum tempo sem atualização. A Figura 5.1 apresenta a distribuição dos currículos pela sua última data de atualização.

¹ Correspondem à produção técnica de um pesquisador como assessoria e consultoria.

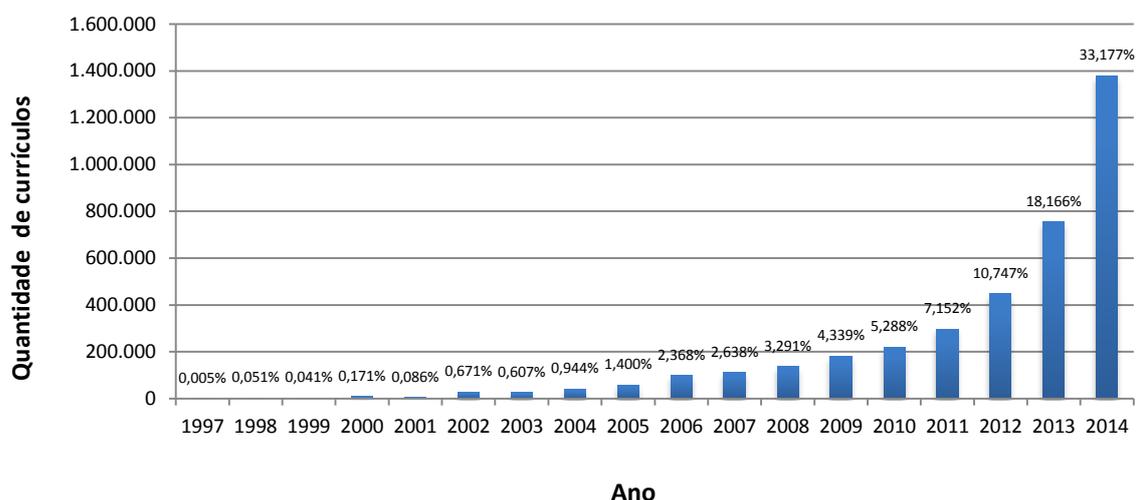


Figura 5.1: Distribuição dos currículos pela data da última atualização.

É possível verificar que, apesar de existirem currículos com data da última atualização em 1997, a quantidade é baixa (201 currículos) se comparada com o total de currículos atualizados nos anos posteriores, principalmente a partir de 2004 (39.248). A não atualização desses currículos pode ter motivos diversos e de difícil identificação. Porém, a grande maioria dos currículos possui data de atualização recente. Cerca de 33% dos currículos (1.379.057) possuem data de última atualização em 2014 e 51,34% (2.134.154) foram atualizados nos últimos dois anos. Tendo em vista que, como a coleta dos dados foi realizada em abril de 2015, 368.177 currículos possuem data de atualização já em 2015. Ou seja, além da quantidade e diversidade de dados presentes em cada um dos currículos, estes possuem, em geral, data de atualização recente, justificando todo o seu potencial de análise.

Uma informação importante para compreender o perfil dos indivíduos que possuem currículos cadastrados diz respeito ao seu nível de formação. Nos currículos da Plataforma Lattes, é possível informar o histórico acadêmico que se inicia no ensino básico e vai até o pós-doutorado². A Figura 5.2 apresenta o nível mais alto de formação dos indivíduos.

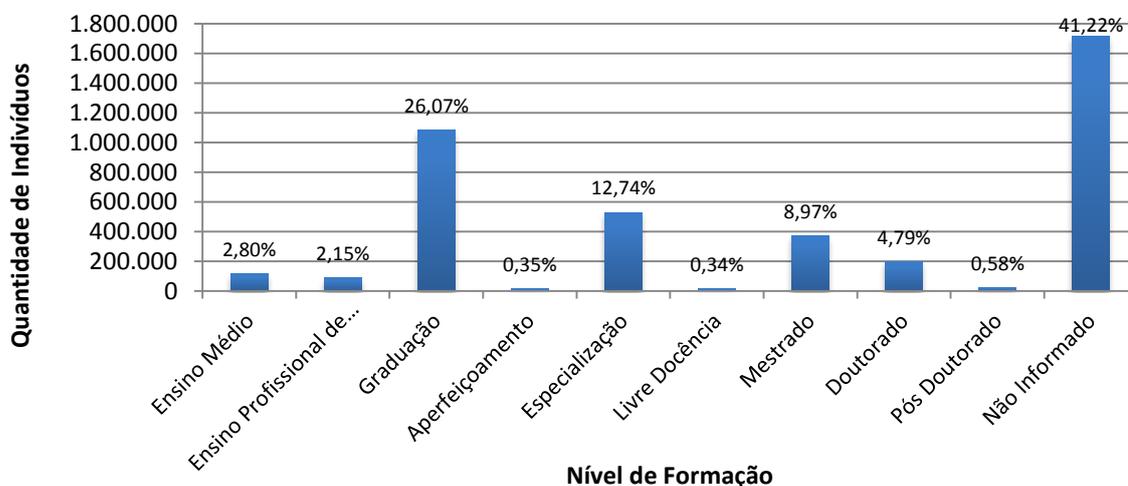


Figura 5.2: Nível mais alto de formação concluída dos indivíduos.

² Embora pós-doutorado não seja uma formação acadêmica formal, foi assim considerado para fins estatísticos.

Com relação às informações correspondentes aos mais altos níveis de formação concluída informadas nos currículos da Plataforma Lattes, aproximadamente 26% dos indivíduos têm como maior titulação a graduação. Este grande número de indivíduos com graduação concluída tem relação com as exigências de ingresso em grande parte dos cursos de pós-graduação e recentes programas de incentivo à capacitação no exterior por órgãos governamentais, que exigem que os candidatos tenham seus currículos cadastrados na Plataforma Lattes. Existe uma queda natural entre os níveis de Mestrado, Doutorado e Pós-Doutorado, no entanto, é importante destacar que, como grande parte dos pesquisadores com esses níveis de formação estão vinculados a instituições de ensino ou pesquisa, por exigência, eles também devem ter seus currículos cadastrados na Plataforma Lattes, o que significa que representam quase a totalidade dos indivíduos que possuem esse nível de formação. Diferentemente, os níveis inferiores de capacitação como Ensino Médio e Ensino Profissional de Nível Técnico possuem pequena quantidade de indivíduos cadastrados, pois, em geral, esses indivíduos ainda estão em formação e não cadastram seus currículos. Ressalta-se que 41,22% dos currículos não possuem nenhum registro sobre algum nível de formação concluído, logo, foram categorizados como Não Informado.

Grande parte dos indivíduos informa em seus currículos seus dados profissionais que descrevem o histórico de seus vínculos profissionais. Um dado importante diz respeito ao seu endereço profissional atual. Com ele, é possível analisar como esses indivíduos estão distribuídos geograficamente (Tabela 5.2). Entretanto, do total de currículos cadastrados na Plataforma Lattes, apenas 29,89% possuem endereço profissional registrado. Considerando os currículos que possuem endereço profissional informado, é possível verificar que as regiões Sudeste e Sul do país abrigam o maior número de indivíduos, podendo ser justificado pela concentração de instituições de ensino e centros de pesquisa que, em sua maioria, estão nessas regiões. Este cenário, não por acaso, também coincide com índices nacionais que apresentam desigualdades semelhantes, como níveis de escolaridade e distribuição de renda. A região Sudeste-Sul concentra 62,6% dos indivíduos com endereço profissional cadastrado em seus currículos e que foram considerados para esta análise. Já a quantidade de indivíduos dos estados do Amapá, Acre e Roraima juntos, não correspondem a 1%, confirmando a distribuição irregular quando considerada a localidade de atuação dos indivíduos.

Tabela 5.2: Distribuição geográfica dos indivíduos.

Nível de Formação	Estado													
	AC	AM	PA	RR	RO	AP	TO	BA	MA	PI	CE	RN	PB	PE
Não Graduado	97	735	465	135	205	116	209	988	451	197	732	573	509	132
Graduado	2.580	15.923	20.422	2.201	5.835	2.871	6.564	43.009	11.144	9.781	27.780	12.398	14.502	30.975
Mestre	672	4.304	6.184	730	1.291	761	1.486	11.657	2.883	2.510	8.044	4.211	4.991	9.342
Doutor	491	2.562	4.246	450	607	387	961	8.243	1.872	1.729	5.596	4.022	4.874	7.550
Total	3.840	23.524	31.317	3.516	7.938	4.135	9.220	63.897	16.350	14.217	42.152	21.204	24.876	48.999
	0,31%	1,89%	2,52%	0,29%	0,65%	0,32%	0,74%	5,14%	1,32%	1,14%	3,39%	1,71%	2,00%	3,94%
Nível de Formação	Estado													
	AL	SE	MT	MS	DF	GO	MG	SP	ES	RJ	PR	SC	RS	Total
Não Graduado	269	353	505	283	343	697	1.841	2.607	351	1.630	1.070	1.087	1.644	19.224
Graduado	11.740	9.501	13.952	13.267	20.518	29.402	77.560	135.555	11.714	55.114	55.642	38.931	58.582	737.463
Mestre	2.375	2.335	3.995	3.612	9.219	7.207	27.110	59.710	4.954	29.409	19.225	13.882	23.467	265.566
Doutor	1.589	1.820	2.637	2.451	7.059	4.473	21.698	62.306	2.929	28.428	14.335	8.558	18.393	220.266
Total	15.973	14.009	21.089	19.613	37.139	41.779	128.209	260.178	19.948	114.581	90.272	62.458	102.086	1.242.519
	1,29%	1,13%	1,70%	1,58%	2,99%	3,36%	10,32%	20,94%	1,61%	9,22%	7,27%	5,03%	8,22%	100,00%

Há um certo equilíbrio na proporção de mestres e doutores em praticamente todos os estados, com exceção de São Paulo, que possui mais doutores do que mestres. Outro fator a ser considerado é a baixa quantidade de currículos de não graduados, que, em sua maioria, são de profissionais de nível técnico atuando em instituições de ensino ou de graduandos que estão realizando atividades de iniciação científica e, por receberem bolsas de estudo, informam as instituições a que estão vinculados como seus endereços profissionais.

Indivíduos com níveis de formação mais altos tendem a ter maior produção científica e, cada vez mais, necessitam divulgar seus resultados de pesquisa. Motivados ainda pela competitividade na distribuição de recursos que, em geral, leva em consideração a produtividade dos indivíduos, houve um aumento significativo do volume de publicações a partir da década de 1980. Entretanto, como será analisado a seguir (Figura 5.3), nos últimos anos houve uma diminuição considerável do volume de publicações, que retorna a patamares da década de 2000.

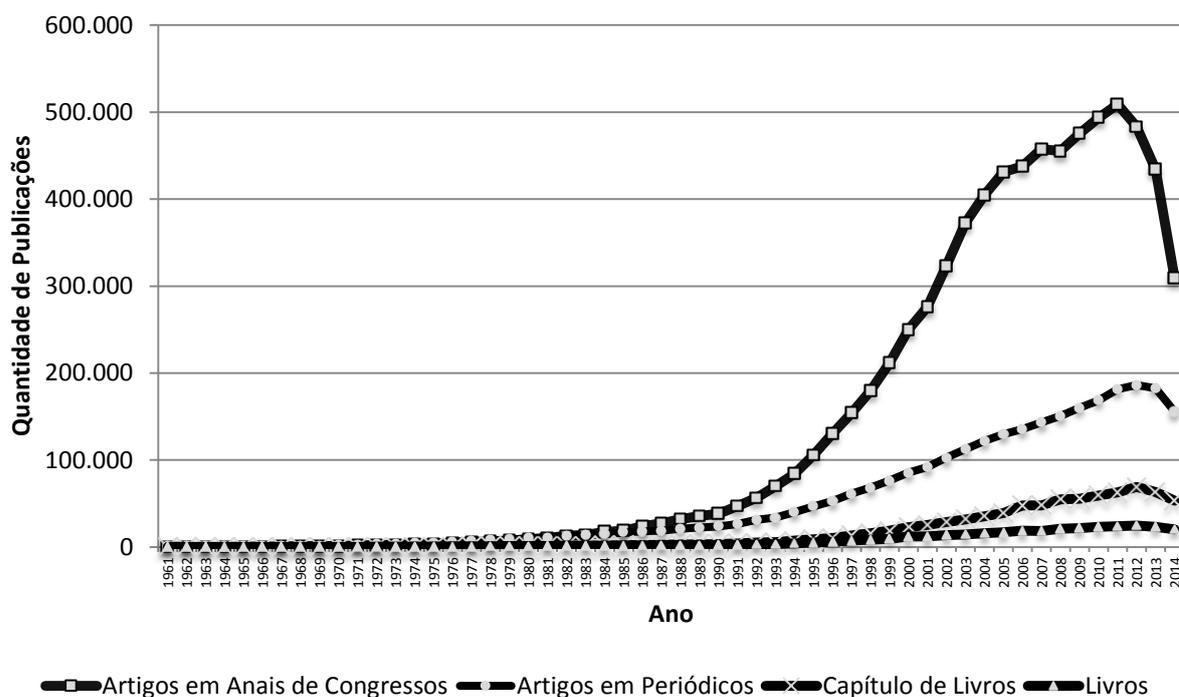


Figura 5.3: Total de publicações por ano.

Para esta análise, foram considerados trabalhos únicos registrados na Plataforma Lattes, ou seja, o mesmo trabalho registrado em currículos distintos foi considerado uma única vez. A publicação de livros e capítulos de livro teve um aumento considerável, principalmente após o ano de 1999, atingindo seu ápice em 2012, representando um crescimento no período de aproximadamente 166% na publicação de livros e 267% na publicação de capítulos de livro. Percebe-se, ainda, que o volume de publicações, tanto em periódico como em anais de congresso, também teve um aumento significativo, principalmente a partir de meados da década de 1980. O volume de publicações em periódico apresentou crescimento constante até o ano de 2012, com queda nos dois anos seguintes. Já o volume de publicações em anais de congresso, que apresentava crescimento significativo (cerca de 5.005% entre os anos de 1980 e 2007), apresenta queda pela primeira vez em 2008, voltando a crescer até o

ano de 2011, quando novamente volta a apresentar uma queda acentuada. Para melhor visualização dos gráficos da Figura 5.3, a Tabela 5.3 apresenta o total de publicações e a respectiva taxa de crescimento nos últimos 10 anos.

Tabela 5.3: Total de publicações e taxa de crescimento por ano.

Tipo de Publicação	Anos										
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	
Artigos A. Congressos	Quant.	430.969	437.573	457.311	455.091	475.273	493.451	508.729	483.017	434.066	309.058
	Cresc.	6,54%	1,53%	4,51%	-0,48%	4,43%	3,82%	3,09%	-5,05%	-10,13%	-28,79%
Artigos em Periódicos	Quant.	129.686	135.487	143.236	150.448	160.065	168.757	181.442	186.084	182.498	155.702
	Cresc.	6,22%	4,47%	5,71%	5,03%	6,39%	5,43%	7,51%	2,55%	-1,92%	-14,68%
Capítulo de Livros	Quant.	39.609	47.771	47.940	54.835	55.739	59.321	62.812	69.201	63.385	53.446
	Cresc.	11,62%	20,60%	0,35%	14,38%	1,64%	6,42%	5,88%	10,17%	-8,40%	-15,68%
Livros	Quant.	16.960	18.750	18.512	20.946	21.972	23.173	23.943	24.826	23.399	20.488
	Cresc.	7,65%	10,55%	-1,26%	13,14%	4,89%	5,46%	3,32%	3,68%	-5,74%	-12,44%

Como é possível observar, há uma queda na produção de todos os tipos de publicação nos últimos dois anos, com destaque para as publicações em anais de congresso, onde essa queda entre os anos de 2012 a 2014 é de aproximadamente 40%. Os dados apresentados poderiam indicar uma considerável queda na produção científica brasileira a partir de 2011. No entanto, tal fenômeno parece estar ocorrendo também em outros países, como pode ser observado no gráfico da Figura 5.4 com dados da produção de artigos em periódico dos 20 países mais produtivos, extraídos da SCImago Journal & Country Rank³ em março de 2016.

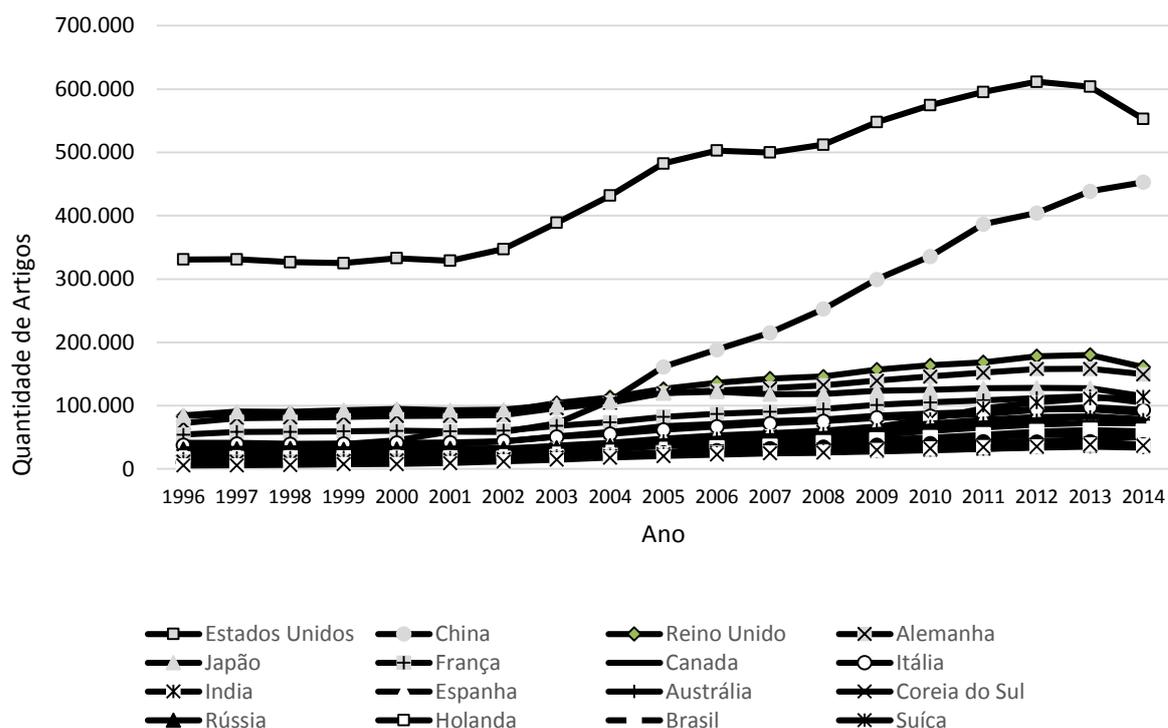


Figura 5.4: Artigos em periódico dos 20 países mais produtivos (SCImago (2016)).

³ SCImago Journal & Country Rank: <http://www.scimagojr.com/>

O SCImago é um portal que apresenta indicadores científicos de vários países, tendo como fonte de dados o portal da Scopus. Nesse portal, é possível obter a quantidade de publicações de um determinado país a partir de 1996. Além da quantidade de publicações, também é possível verificar, ano a ano, o número total de citações, dentre outras métricas bibliométricas. Para a análise aqui realizada, foi utilizada a quantidade total de publicações dos 20 países mais bem colocados, tendo como critério de classificação o número total de publicações. Como pode ser observado, em geral, existe uma tendência de crescimento no número de artigos em periódico principalmente a partir de 2001 para todos os países. Esta tendência de crescimento se estende até o ano de 2012, a exemplo da queda observada nos diversos tipos de publicação analisados nos currículos da Plataforma Lattes. Países que ocupam as primeiras posições em volume de produção como os Estados Unidos (1º) e Japão (5º) apresentam queda acentuada a partir de 2012, a exemplo de Reino Unido (3º) e da Alemanha (4º), que deixam de aumentar a sua produção a partir de 2013 apresentando queda em 2014.

A China (2º) surge como exceção dentre os países mais produtivos, já que, diferentemente dos outros países, mantém crescimento constante. Em 2004, sua produção representava apenas 25,05% do total de artigos publicados nos Estados Unidos, e passou a representar 54,64% e 81,94% nos anos de 2009 e 2014 respectivamente. A queda da produção americana nos últimos dois anos corresponde a 9,61%. A exemplo da China, apenas outros dois países dos vinte mais produtivos não apresentaram queda em sua produção nos últimos anos, Índia (9º) e Rússia (13º). Todos os outros países analisados apresentam a mesma tendência de queda em sua produção científica, incluindo o Brasil (15º), corroborando com os dados extraídos dos currículos da Plataforma Lattes.

Conforme descrito no Capítulo 2, vários trabalhos têm dedicado atenção ao estudo das redes de colaboração (BARABÁSI et al., 2002; NEWMAN, 2004; LOPES et al., 2012; BENEVENUTO; LAENDER; ALVES, 2015; KUMAR, 2015; BRANDÃO; DINIZ; MORO, 2016). Tais estudos apontam como os trabalhos vêm sendo realizados cada vez mais em parcerias. Analisando as características das publicações dos indivíduos brasileiros, é possível verificar que a pesquisa nacional vem, ao longo dos anos, sendo realizada cada vez mais em colaboração (Figura 5.5). Um outro ponto importante observado é que a média de autores por artigo vem aumentando de forma geral, principalmente em artigos publicados em anais de congresso e em periódico. Esses resultados confirmam que, cada vez mais, os grupos de pesquisa tendem a trabalhar coletivamente, resultando em maior produtividade.

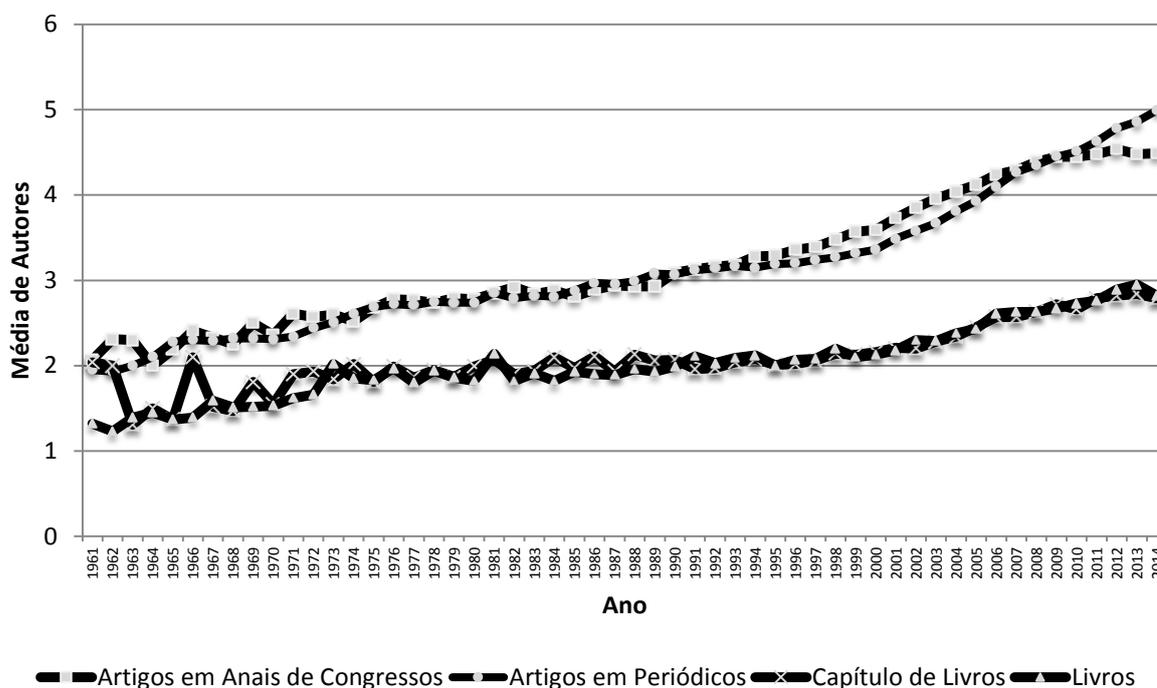


Figura 5.5: Média de autores por trabalho.

As médias de autores em artigos de periódico e em anais de congresso seguiam um mesmo padrão de crescimento até 1993, quando os artigos publicados em anais de congresso passaram a ter maior média de autores do que os artigos em periódico. Esta proporção se reduz entre 2006 e 2009, quando artigos em periódico passam a ser publicados com maior número de colaboradores do que os artigos em anais de congresso. Da mesma forma a publicação de livros e de capítulos de livro passou a ser realizada com um número maior de colaboradores a partir de 2003.

É importante destacar, neste momento, que o crescente aumento no número de colaborações, principalmente nos artigos publicados em anais de congresso e em periódico, coincide com a queda no número de publicações de ambos os tipos a partir de 2011. Conforme já mencionado, o trabalho em colaboração facilita o processo de publicação, tendo em vista que a troca de conhecimento é maior e favorece o desenvolvimento das pesquisas, já que é esperado um menor esforço a ser empregado individualmente em um determinado trabalho, pois envolve um número maior de colaboradores. Como resultado, seria esperado que, quanto maior a colaboração, maior a quantidade de trabalhos publicados, o que não tem acontecido nos últimos anos. Entretanto, como será avaliado mais à frente, essa redução do volume de publicações veio acompanhada de uma melhoria dos veículos de publicação.

A capacidade de publicação, principalmente em bons periódicos, está intimamente vinculada à habilidade de ler e escrever trabalhos em outros idiomas, tendo em vista que, para a publicação de bons trabalhos, se faz necessária a leitura de referencial que, frequentemente, é escrito em outros idiomas (notadamente em Inglês). Além disso, para a publicação e consequente apresentação de trabalhos em bons congressos internacionais, se faz necessária a fluência em línguas estrangeiras. Assim, a seguir é realizada uma breve análise da proficiência em língua estrangeira (fala, compreensão, escrita e leitura) dos indivíduos com currículos cadastrados na Plataforma Lattes (Figura 5.6).

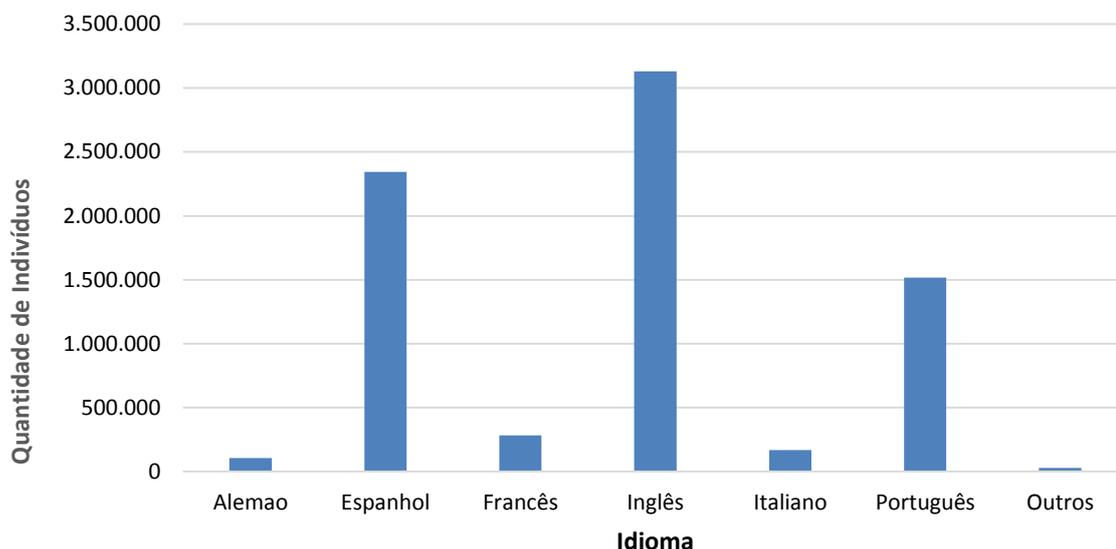


Figura 5.6: Proficiência dos indivíduos em idiomas.

Para a análise da proficiência dos indivíduos em idiomas, foram considerados somente indivíduos que indicaram algum tipo de habilidade em algum idioma. Ou seja, nesta análise, foram desconsiderados os currículos que não tinham nenhuma informação sobre o nível de conhecimento para qualquer idioma. Diante disso, foram identificados seis idiomas principais (Alemão, Espanhol, Francês, Inglês, Italiano e Português), já que menos de 1% dos currículos continha informação sobre o conhecimento de outros idiomas, como Japonês, Latim e Chinês, que foram agrupados na categoria Outros.

O idioma estrangeiro com maior popularidade entre os currículos analisados é o Inglês (75,3%). Um dos fatores desta proficiência diz respeito a esta língua dominar a grande maioria dos periódicos de relevância e ter se tornado um padrão para a comunicação científica. A seguir, surge o Espanhol, com 56,3% de indivíduos com algum tipo de fluência no idioma. O motivo deste percentual parece estar relacionado ao fato de que a maioria dos currículos cadastrados na Plataforma Lattes seja de brasileiros e estes, pela similaridade com a Língua Portuguesa, consideram ter conhecimento do Espanhol, principalmente para compreensão e leitura.

Curiosamente, o Português surge apenas como o terceiro idioma mais citado. O principal motivo deste aparente resultado deve-se à nacionalidade dos indivíduos do repositório. Como a grande maioria dos indivíduos cadastrados na Plataforma Lattes é de brasileiros, muitos deles negligenciam o Português, supondo que não é necessário registrar este dado, já que, intrinsecamente, possuem conhecimento do idioma ou que esta informação se refere somente à proficiência em idiomas estrangeiros. Francês, Italiano e Alemão aparecem como os outros idiomas mais citados.

Análises específicas possibilitam identificar o nível de proficiência em cada um dos idiomas mais citados. Nos currículos da Plataforma Lattes, deve ser informada a proficiência nas quatro habilidades: Compreensão, Fala, Leitura e Escrita. O grau de proficiência segue três níveis: Pouco, Razoavelmente e Bem. Considerando esses dados, é apresentado o nível de proficiência dos indivíduos nos idiomas considerados (ver Figura 5.7).

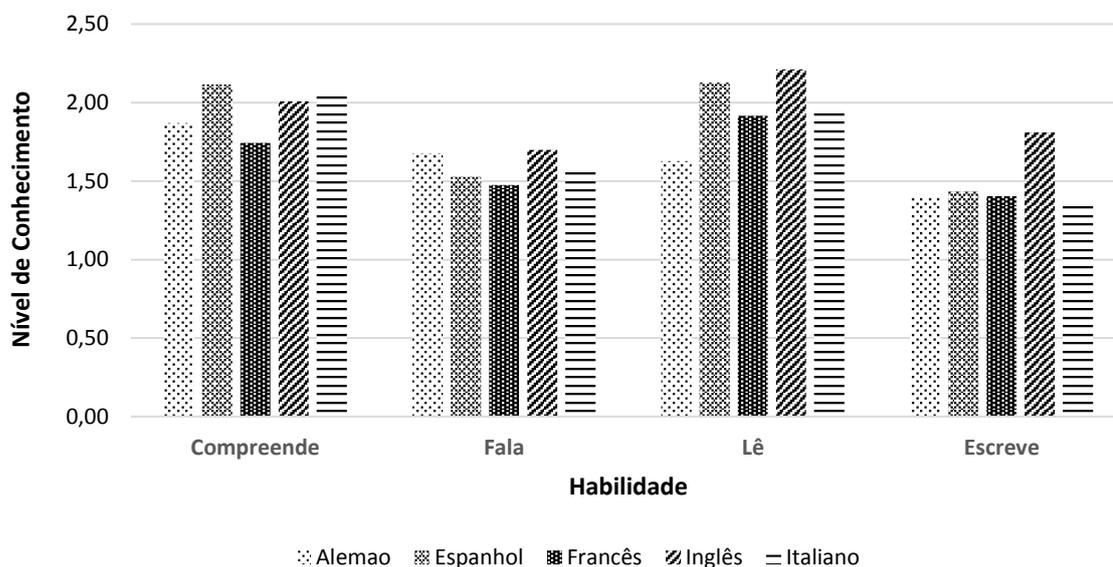


Figura 5.7: Nível de proficiência dos indivíduos.

Os níveis de proficiência em cada idioma foram ponderados para as habilidades da seguinte maneira: 3 para o nível Bem; 2 para Razoavelmente; 1 para Pouco; e 0 para situações nas quais o nível não foi informado. Nesta análise, o Português foi negligenciado, pois atingiu o nível Bem em todas as habilidades. Logo, o inglês e espanhol surgem com os maiores níveis de proficiência, na leitura e escrita, e em seguida, o Espanhol, com valores superiores de proficiência na habilidade compreensão. Em geral, percebe-se que os maiores níveis de habilidade estão na leitura e compreensão e, os piores, na escrita e fala.

5.2 Estatísticas por Áreas e Grandes Áreas

Uma informação presente nos currículos diz respeito à grande área, área, subárea e especialidade de atuação do indivíduo. Este dado opcional visa possibilitar que um determinado indivíduo informe suas áreas de atuação. A classificação das áreas de conhecimento disponíveis para indicação nos currículos da Plataforma Lattes segue, em geral, a mesma classificação da CAPES⁴ com algumas pequenas distinções. Na Plataforma Lattes, a grande área Multidisciplinar não existe e em seu lugar, tem-se a opção Outra, que engloba as seguintes áreas: Defesa; Bioética; Ciências Ambientais; Divulgação Científica; Robótica, Mecatrônica e Automação; e Microeletrônica.

Com a indicação da grande área, área, subárea e especialidade de atuação de cada indivíduo, é possível realizar agrupamentos e verificar como os currículos estão distribuídos nas diversas áreas do conhecimento. Como é possível inserir mais de uma grande área e, conseqüentemente, áreas e subáreas, para a classificação adotada nesta tese, o primeiro registro contendo esta informação foi utilizado, tendo em vista que, neste primeiro momento, os currículos foram classificados em apenas uma grande área. A distribuição dos indivíduos por grande área de atuação pode ser observada na Figura 5.8.

⁴ <http://www.capes.gov.br/avaliacao/instrumentos-de-apoio/tabela-de-areas-do-conhecimento-avaliacao>

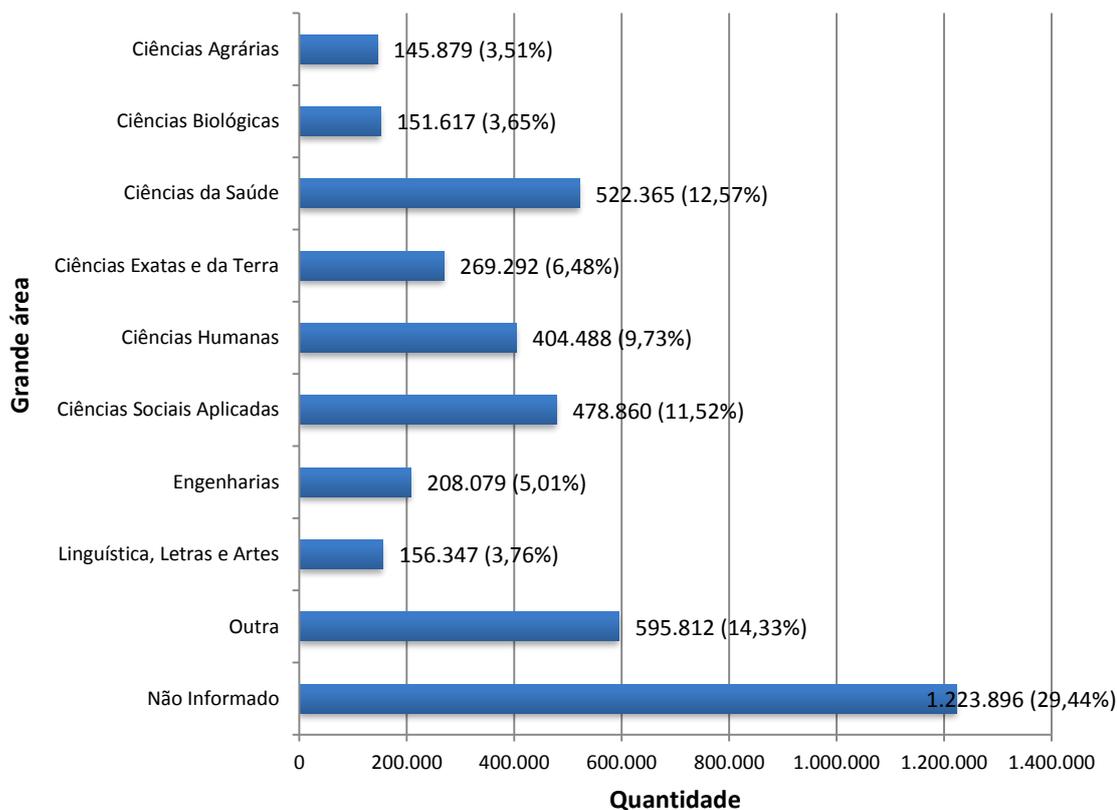


Figura 5.8: Distribuição dos indivíduos por grande área de atuação.

Aproximadamente, 29% dos currículos não indicam uma grande área e, conseqüentemente, a área específica de atuação não é informada. A exemplo dos currículos sem níveis de formação informados, em sua maioria, os currículos sem grande área informada não possuem data de atualização recente e, em geral, possuem apenas dados básicos como dados de identificação e endereço eletrônico para acesso ao currículo. Ou seja, tais currículos não apresentam dados relevantes além da identificação do indivíduo, já que várias outras seções não foram preenchidas. Aproximadamente, 71% dos currículos sem grande área informada não são atualizados há mais de dois anos.

É possível observar, ainda, que em 14,33% dos currículos cadastrados na Plataforma Lattes foi selecionada a opção Outra. Além disso, é possível notar também grande diversidade na distribuição dos currículos nas grandes áreas, ressaltando, ainda, que a quantidade de áreas em cada uma das grandes áreas não é uniforme, podendo contribuir para a distribuição irregular da quantidade de indivíduos atuando nas diversas grandes áreas do conhecimento. Logo, uma análise baseada nas áreas que compõem cada uma das grandes áreas se faz necessária (ver Figura 5.9).

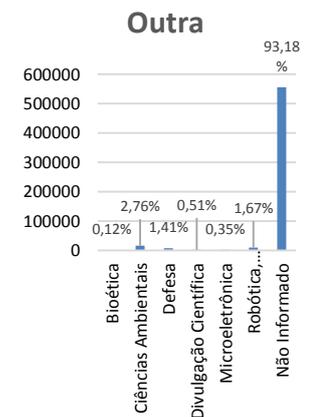
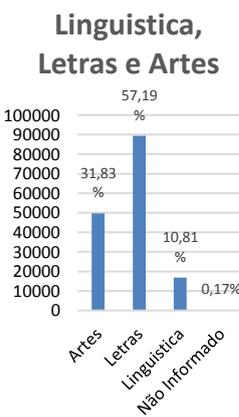
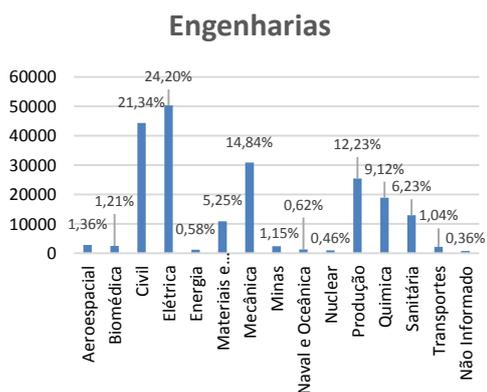
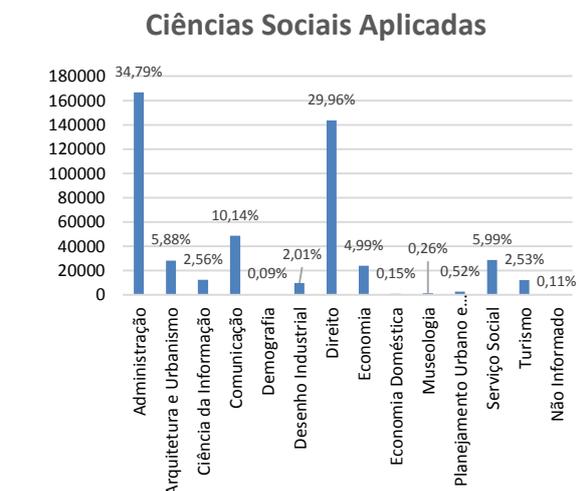
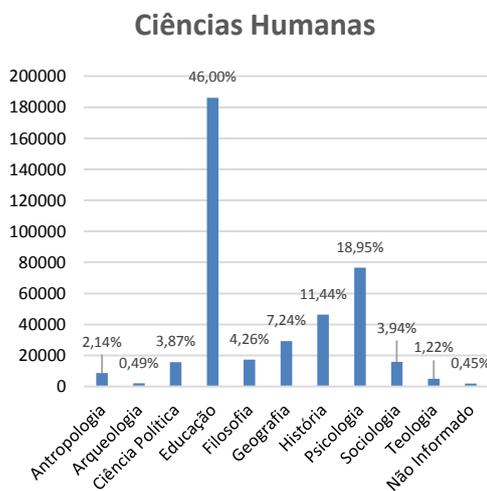
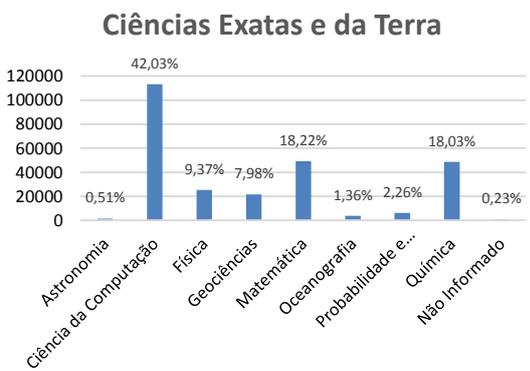
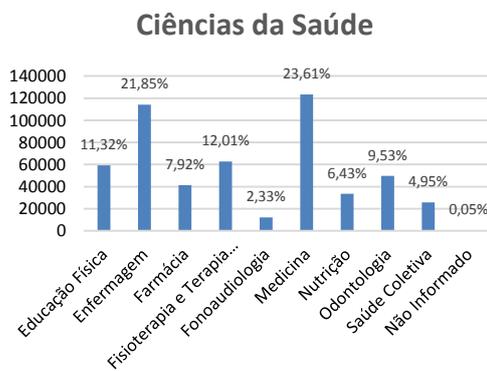
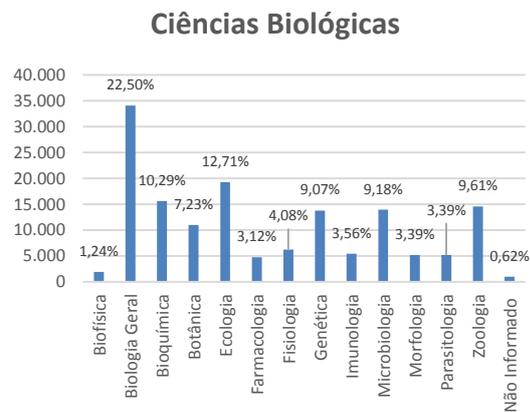
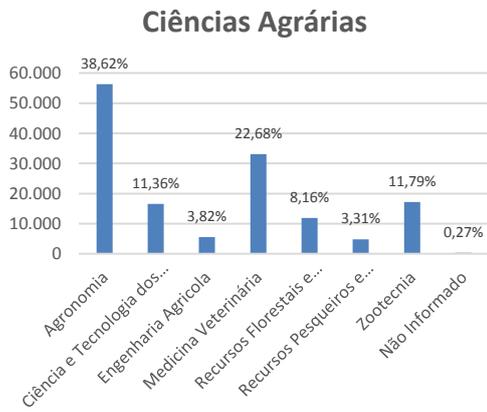


Figura 5.9: Distribuição de indivíduos por área de atuação.

A grande área de Ciências da Saúde e a opção Outra se caracterizam como aquelas com maior quantidade de currículos cadastrados na Plataforma Lattes, totalizando 522.365 (12,57%) e 595.812 (14,33%), respectivamente. Já a grande área de Engenharias possui a maior quantidade de áreas (quatorze), contendo, no entanto, apenas 208.079 (5,01%) currículos cadastrados, o que mostra que a quantidade de áreas não está relacionada com a quantidade de indivíduos nas grandes áreas. A área de pesquisa com a maior quantidade de currículos cadastrados é a área de Educação, com 186.045 currículos, o que corresponde a 46% do total de sua grande área, comprovando a inexistência de um padrão de distribuição nas grandes áreas. A Tabela 5.4 apresenta as áreas com as maiores quantidades de currículos.

Tabela 5.4: Áreas de pesquisa com a maior quantidade de currículos.

Grande Área	Área	Quantidade
Ciências Humanas	Educação	186.045
Ciências Sociais Aplicadas	Administração	166.597
Ciências Sociais Aplicadas	Direito	143.475
Ciências da Saúde	Medicina	123.355
Ciências da Saúde	Enfermagem	114.118
Ciências Exatas e da Terra	Ciência da Computação	113.188
Total		846.778

As seis maiores áreas são responsáveis por 20,3% de todos os currículos da Plataforma Lattes. Essas áreas indicadas na Tabela 5.4 apresentam quantidades muito superiores, se comparadas com todas as outras áreas. Além disso, um total de 670.951 (16,14%) indivíduos informaram a sua grande área, porém sem definir a sua área de pesquisa. Deste total, aproximadamente, 56,6% selecionaram a opção Outra como grande área.

Cada uma das grandes áreas possui indivíduos com perfis distintos com relação ao maior nível de formação informado nos currículos, como pode ser observado na Tabela 5.5. Currículos com a opção Outra, são aqueles com maior percentual sem área informada (93,18%), e também com o maior percentual sem formação informada (79,87%), além de corresponderem aos indivíduos com o maior nível de formação em cursos técnico-profissionalizantes. Ressaltam-se, ainda, as grandes áreas de Ciências Exatas e da Terra e Engenharias, com grande quantidade de currículos neste nível de formação, sendo uma característica das áreas que as compõem a formação de pessoal de nível técnico. A grande área de Ciências da Saúde possui a maior quantidade de currículos com o maior nível de formação em graduação e, apesar de ser a segunda grande área com o maior número de currículos cadastrados, não possui a mesma representatividade nos níveis mais altos de formação. Neste aspecto, destaca-se a grande área de Ciências Humanas, que possui a maior quantidade de indivíduos com mestrado e doutorado concluídos, mesmo sendo apenas a quarta grande área com maior quantidade de currículos.

Tabela 5.5: Distribuição dos currículos por grande área e maior nível de formação concluída.

Grande Área / Maior Nível de Formação	Total de Currículos	Sem Formação Declarada	Curso Técnico / Profissionalizante	Graduação	Especialização	Mestrado	Doutorado
Ciências Agrárias	145.879	54.108	6.512	35.484	7.478	22.362	19.935
Ciências Biológicas	151.617	43.966	2.606	38.820	12.118	27.399	26.708
Ciências da Saúde	522.365	169.120	7.786	173.058	88.183	50.390	33.828
Ciências Exatas e da Terra	269.292	86.754	10.584	74.939	30.373	36.717	29.925
Ciências Humanas	404.488	105.751	3.677	120.314	78.520	62.240	33.986
Ciências Sociais Aplicadas	478.860	138.765	6.050	162.496	92.073	59.488	19.988
Engenharias	208.079	79.807	9.826	56.818	15.529	27.144	18.955
Linguística, Letras e Artes	156.347	46.493	1.812	48.436	25.580	22.432	11.594
Outra	595.812	475.904	12.092	87.858	13.502	4.688	1.768
Total	2.932.739	1.200.668	60.945	798.223	363.356	312.860	196.687

Outra informação importante a ser analisada diz respeito ao perfil de publicação dos indivíduos cadastrados na Plataforma Lattes. Nas várias áreas do conhecimento, a média de trabalhos publicados sofre variações consideráveis, o que demonstra perfis distintos entre essas áreas (Tabela 5.6). Ao se fazer uma análise comparativa entre as grandes áreas, Ciências da Saúde possui a maior quantidade de artigos em anais de congresso (2.688.365). No entanto, essa grande área possui uma média de apenas 5,14 artigos por indivíduo, bem abaixo da grande área de Ciências Agrárias, que tem média de 12,27 artigos por indivíduo. Tal fato também ocorre com artigos em periódico, pois, apesar de a grande área de Ciências da Saúde possuir o maior número de publicações, possivelmente relacionado ao fato de possuir o maior número de currículos, tem média de publicação por indivíduos inferior à da grande área de Ciências Biológicas. Já as grandes áreas Ciências Humanas e Engenharias possuem a maior média de trabalhos técnicos.

Tabela 5.6: Publicações por grande área de pesquisa e média por indivíduo.

Grande Área	Artigos em Anais de Congresso	Artigos em Periódico	Livros	Capítulos de Livro	Textos em Jornais e Revistas	Trabalhos Técnicos
C. Agrárias	1.790.278 (12,27)	577.376 (3,95)	23.560 (0,16)	67.759 (0,46)	89.619 (0,61)	99.584 (0,68)
C. Biológicas	1.577.271 (10,40)	662.714 (4,37)	18.104 (0,11)	74.178 (0,48)	51.081 (0,33)	111.891 (0,73)
C. da Saúde	2.688.365 (5,14)	1.147.751 (2,19)	49.761 (0,09)	257.609 (0,49)	173.692 (0,33)	185.931 (0,35)
C. Exatas e da Terra	1.291.901 (4,79)	589.394 (2,18)	27.920 (0,10)	52.429 (0,19)	49.079 (0,18)	118.503 (0,44)
C. Humanas	1.307.838 (3,23)	484.958 (1,19)	133.751 (0,33)	262.993 (0,65)	310.166 (0,76)	342.656 (0,84)
C. Sociais Aplicadas	690.660 (1,44)	409.488 (0,85)	96.595 (0,20)	164.379 (0,34)	382.514 (0,79)	344.079 (0,71)
Engenharias	958.200 (4,60)	267.256 (1,28)	15.898 (0,07)	33.297 (0,16)	41.124 (0,19)	152.954 (0,74)
Linguística, Letras e Artes	336.027 (2,14)	146.526 (0,93)	53.145 (0,33)	79.344 (0,50)	128.243 (0,82)	98.002 (0,62)
Outra	98.253 (0,16)	24.578 (0,04)	3.197 (0,005)	5.699 (0,01)	12.802 (0,02)	15.470 (0,02)

A menor média de publicações está a cargo dos indivíduos que selecionaram a opção Outra como grande área em seus currículos. Isto se deve, provavelmente, a fatores já apresentados anteriormente, como baixa quantidade de currículos nos maiores níveis de formação, mas, principalmente, por conter uma grande quantidade de currículos sem nenhum tipo de publicação registrada, influenciando diretamente para que a média de publicações associada à opção Outra seja insignificante. Para demonstrar o quanto os currículos sem produção influenciam na média de publicações em cada grande área, a Tabela 5.7 apresenta o quantitativo de currículos sem nenhum tipo de produção cadastrada, o que impacta nos cálculos sobre produtividade média das grandes áreas analisadas.

Tabela 5.7: Quantidade e percentual de currículos sem produção registrada.

Grande Área	Quantidade de currículos sem produção
C. Agrárias	61.786 (42,35%)
C. Biológicas	55.399 (36,54%)
C. da Saúde	264.973 (50,73%)
C. Exatas e da Terra	152.526 (56,64%)
C. Humanas	203.687 (50,36%)
C. Sociais Aplicadas	298.684 (62,37%)
Engenharias	130.185 (62,57%)
Linguística, Letras e Artes	82.736 (52,92%)
Outra	543.162 (91,16%)

Considerando que 91,16% dos indivíduos que indicaram Outra como sua grande área não possuem produção informada em seus currículos, para o restante dos estudos apresentados nesta tese, tais indivíduos serão desconsiderados.

Logo, com o intuito de verificar a variabilidade dos dados de cada uma das grandes áreas, foram analisados os artigos publicados em anais de congresso e os artigos em periódico, principais tipos de publicação informados nos currículos e objeto de estudo

de diversos trabalhos relacionados. Para cada grande área, foi gerado um *boxplot*⁵ com todos os seus currículos, a fim de apresentar a distribuição das publicações e mostrar como essa distribuição é dispersa (Figuras 5.10 e 5.11).

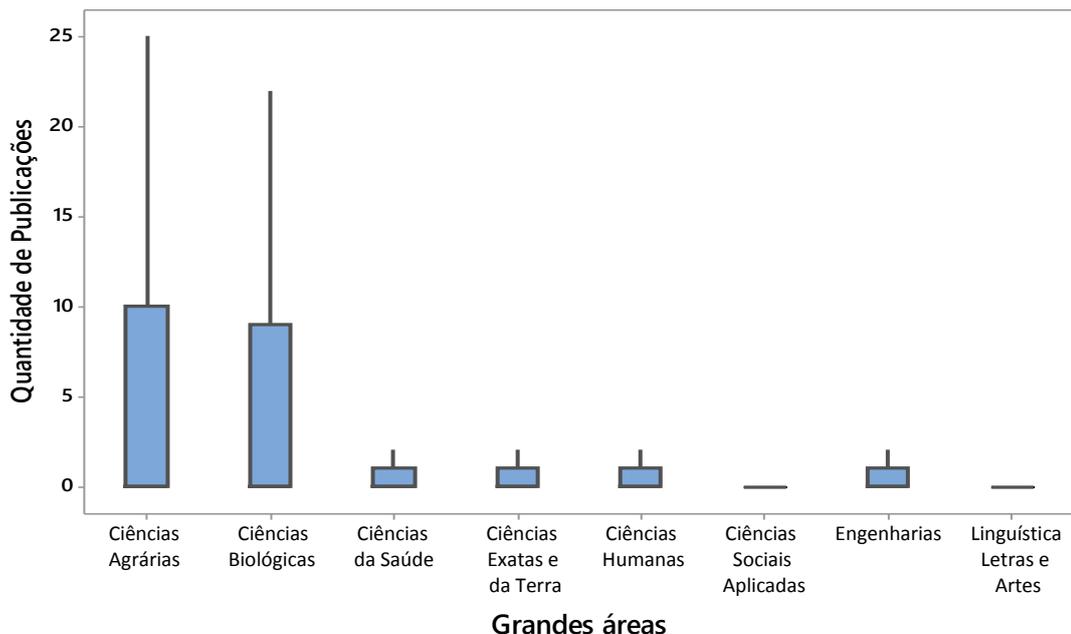


Figura 5.10: Distribuição do número de publicações em anais de congresso.

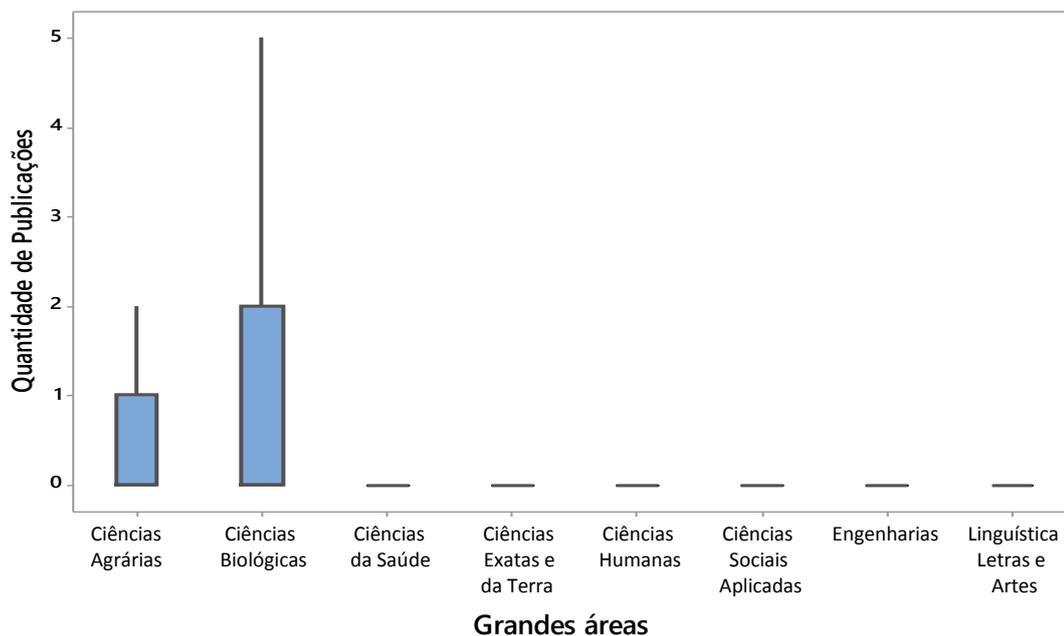


Figura 5.11: Distribuição do número de publicações em periódico.

É possível identificar que, independentemente do tipo de publicação, os currículos sem produção influenciam significativamente a análise dos dados. Em todas as grandes áreas a mediana tanto da publicação de artigos em anais de congresso como em periódico é igual a 0 (zero), já que pelo menos a metade dos currículos em cada grande área não tem nenhum registro para cada tipo de publicação. No caso de artigos

⁵ Para uma melhor visualização dos gráficos *boxplot* apresentados nesta tese, tendo em vista a grande variação entre as taxas de cada amostra, os *outliers* de alguns deles não foram incluídos.

publicados em anais de congresso, as grandes áreas de Ciências Sociais Aplicadas e Linguística, Letras e Artes possuem, pelo menos, 75% dos seus currículos sem publicações, tornando os gráficos concentrados no limite inferior, com exceção dos *outliers*. Situação semelhante ocorre quando são analisados os artigos em periódico, em que somente as grandes áreas de Ciências Agrárias e Ciências Biológicas possuem valores acima de zero no cálculo do seu terceiro quartil. Esse grande número de currículos sem registro de produção nos dois tipos de publicação analisados faz com que as suas medianas sejam iguais a zero em todas as grandes áreas. Para melhor visualizar, a distribuição do número de publicações por autor, histogramas para cada tipo de publicação, com os currículos de todas as grandes áreas, foram gerados (Figuras 5.12 e 5.13).

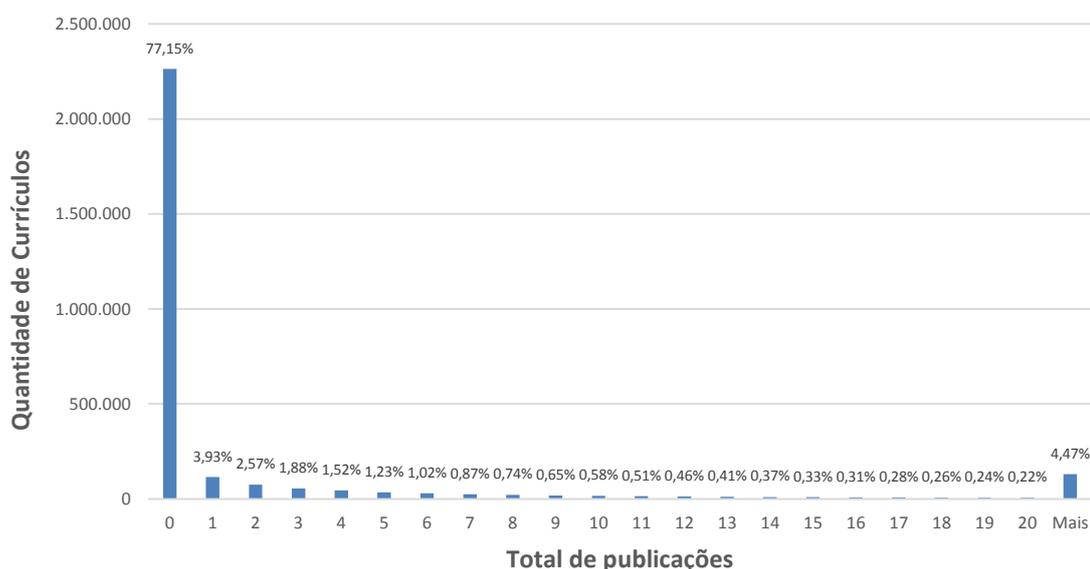


Figura 5.12: Distribuição da quantidade de currículos pelo total de artigos em anais de congresso.

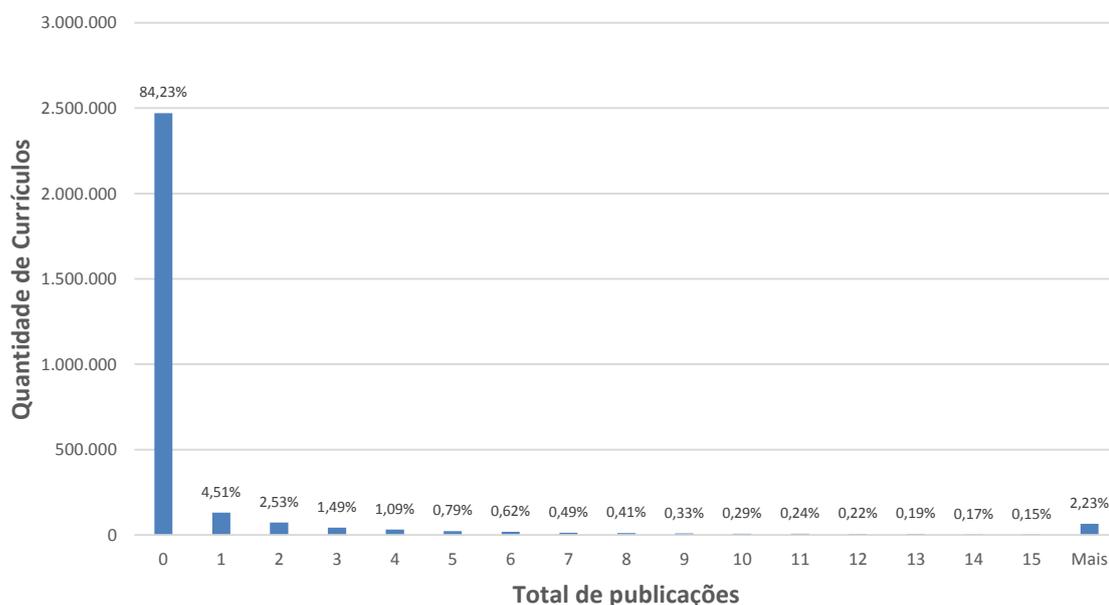


Figura 5.13: Distribuição da quantidade de currículos pelo total de artigos em periódico.

Destaca-se a grande quantidade de currículos sem qualquer registro de produção para ambos os tipos de publicação. Ao analisar a publicação de artigos em anais de congresso, percebe-se uma grande concentração de indivíduos com baixa produção, sendo que 77,15% dos currículos não possuem nenhuma publicação registrada e apenas 4,47% dos currículos possuem mais do que vinte publicações. Situação semelhante ocorre com artigos em periódico, em que 84,23% dos currículos não possuem nenhum registro nesta categoria e, apenas 2,23%, ou seja, 65.437 currículos, possuem mais de 15 publicações, corroborando com o fato de que poucos currículos concentram um grande número de publicações e a grande maioria dos indivíduos possui uma produtividade muito baixa ou nula. Diante disso, fica evidente o quanto os currículos sem produção influenciam na média da produção científica em todas as grandes áreas (Tabela 5.8). Logo, se faz necessária a identificação de subconjuntos de indivíduos que representem de forma efetiva a produção científica em suas áreas de atuação para novos estudos.

Tabela 5.8: Cálculo da média, desvio padrão e mediana.

Grande Área	Artigos em Anais de Congresso				Artigos em Periódico			
	Total	Média	Desvio Padrão	Mediana	Total	Média	Desvio Padrão	Mediana
Ciências Agrárias	1.790.278	12,27	33,34	0	577.376	3,95	15,04	0
Ciências Biológicas	1.577.354	10,40	29,03	0	662.714	4,37	16,21	0
Ciências da Saúde	2.688.365	5,14	21,86	0	1.147.751	2,19	11,95	0
Ciências Exatas e da Terra	1.291.901	4,79	19,88	0	589.394	2,18	12,29	0
Ciências Humanas	1.307.838	3,23	12,16	0	484.958	1,19	5,84	0
Ciências Sociais Aplicadas	690.660	1,44	7,71	0	409.488	0,85	5,61	0
Engenharias	958.200	4,60	20,28	0	267.256	1,28	7,85	0
Linguística, Letras e Artes	336.027	2,14	8,16	0	146.526	0,93	4,74	0

No caso dos dois tipos de publicação analisados, a quantidade de currículos sem produção cadastrada é superior a 50% em todas as grandes áreas, resultando em uma mediana igual a 0 (zero). As grandes áreas de Ciências Agrárias e Ciências Biológicas possuem as maiores médias dentre todas as grandes áreas e também os maiores desvio padrão. Apesar de a grande área de Ciências da Saúde possuir a maior quantidade de publicações, ela possui média inferior à das duas últimas citadas, isso devido ao grande percentual de currículos sem publicações registradas de ambos os tipos (ver Tabela 5.9).

Tabela 5.9: Currículos sem produção registrada.

Grande Área	Total de currículos	Artigos em Anais de Congresso		Artigos em Periódicos	
		Sem Produção	%	Sem Produção	%
Ciências Agrárias	145.879	77.837	53,36	101.805	69,79
Ciências Biológicas	151.617	79.127	52,19	97.576	64,36
Ciências da Saúde	522.365	371.319	71,08	411.311	78,74
Ciências Exatas e da Terra	269.292	193.185	71,74	218.285	81,06
Ciências Humanas	404.488	290.123	71,73	330.285	81,66
Ciências Sociais Aplicadas	478.860	400.645	83,67	412.419	86,13
Engenharias	208.079	153.833	73,93	177.538	85,32
Linguística, Letras e Artes	156.347	119.310	76,31	131.970	84,41

A grande área de Ciências Biológicas se destaca com os menores percentuais em ambos os tipos de publicação. Conforme mencionado, currículos sem publicações não são úteis para trabalhos que visam analisar dados sobre produção científica, como é o caso do estudo sobre as redes de colaboração. Logo, optou-se por excluir os currículos sem produção. Com isso, os valores dos quartis e também das medianas alteraram significativamente, conforme mostrado nas Figuras 5.14 e 5.15.

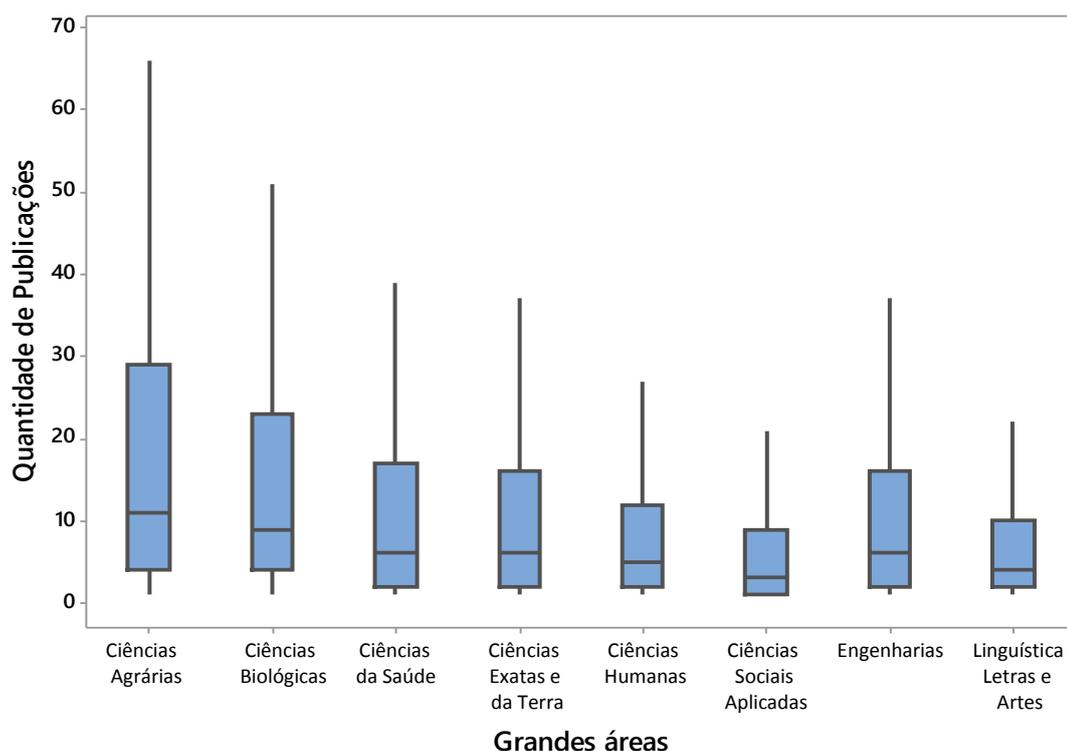


Figura 5.14: Distribuição dos artigos em anais de congresso nos currículos com produção.

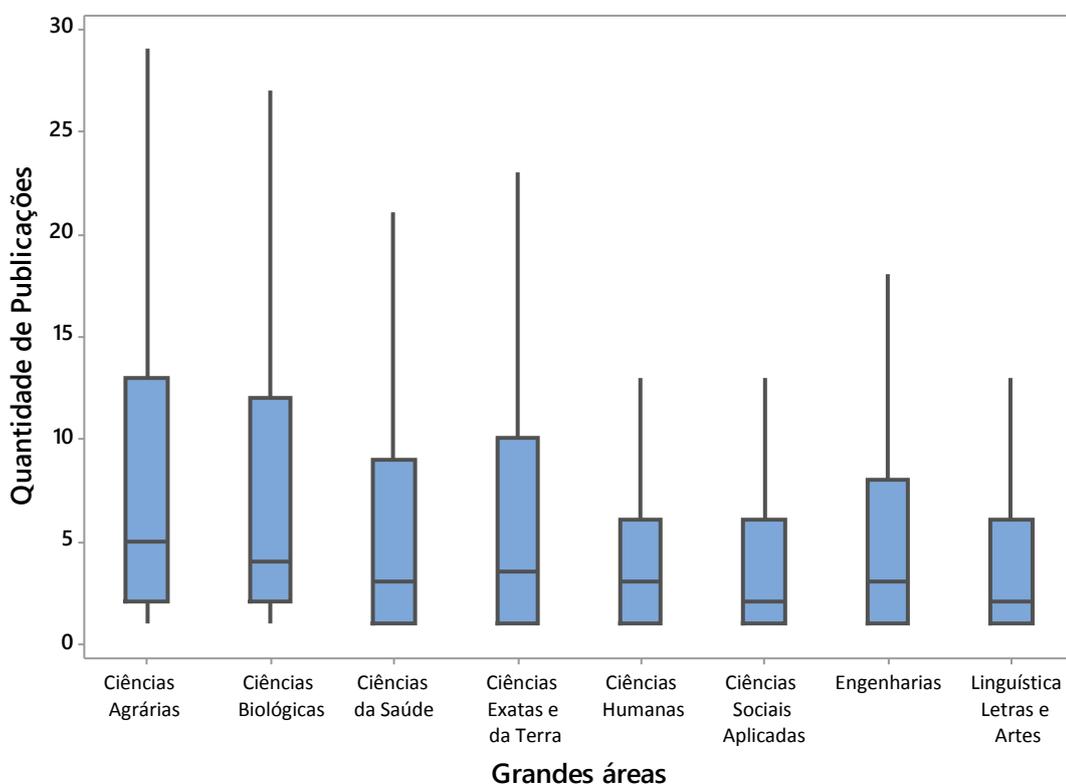


Figura 5.15: Distribuição dos artigos em periódico nos currículos com produção.

É possível verificar que as medianas passam a ter valores não nulos, já que não existe mais currículos com produção nula em cada uma das grandes áreas. Destacam-se com as melhores medianas, tanto para artigos em anais de congresso como também para artigos em periódico, as grandes áreas de Ciências Agrárias (11 e 5) e Ciências Biológicas (9 e 4), sendo as únicas com primeiro quartil maior do que 3 (três) no caso de artigos publicados em anais de congresso. As grandes áreas de Ciências Sociais e Linguística, Letras e Artes, a exemplo do que ocorre com a análise incluindo os currículos sem produção, continuam a ter as medianas com valores baixos, já que possuem, em sua maioria, currículos com pouca produção. Mesmo excluindo os currículos sem produção, ainda é possível identificar a ocorrência de *outliers* em todas as grandes áreas, que corroboram a disparidade do conjunto em análise. Alguns desses *outliers* possuem valores extremamente altos, se comparados com a maioria dos indivíduos em todas as grandes áreas, no entanto, eles representam uma minoria.

Diante disso, pode-se confirmar que o conjunto de currículos verificados corresponde a uma grande parcela de indivíduos com pouca ou nenhuma produção, existindo, por outro lado, um pequeno grupo com produção extremamente alta. A Tabela 5.10 lista os quantitativos correspondentes aos currículos referentes aos indivíduos mais produtivos em cada uma das grandes áreas, apresentando a quantidade de trabalhos publicados considerando artigos em anais de congresso, artigos em periódico e livros. Para cada tipo de publicação, são considerados os cinco currículos com os maiores quantitativos, que não correspondem, necessariamente, a currículos de um mesmo indivíduo.

Tabela 5.10: Indivíduos mais produtivos em cada grande área.

Grande Área	Posição	Artigos em Anais de Congressos	Artigos em Periódicos	Livros
C. Agrárias	1º	876	744	96
	2º	866	518	65
	3º	827	476	62
	4º	807	436	60
	5º	771	431	49
C. Biológicas	1º	1.281	922	185
	2º	953	631	75
	3º	942	612	63
	4º	940	506	54
	5º	909	503	48
C. da Saúde	1º	2.107	702	134
	2º	1.770	671	125
	3º	1.400	603	112
	4º	1.157	600	79
	5º	1.134	578	62
C. Exatas e da Terra	1º	1.502	887	144
	2º	1.289	852	109
	3º	1.181	836	98
	4º	782	744	92
	5º	781	635	76
C. Humanas	1º	578	570	172
	2º	557	348	157
	3º	535	327	143
	4º	529	316	141
	5º	499	311	128
C. Sociais Aplicadas	1º	766	1.076	367
	2º	567	1.043	150
	3º	560	697	138
	4º	477	532	118
	5º	445	370	117
Engenharias	1º	982	814	57
	2º	839	459	51
	3º	773	452	43
	4º	646	434	40
	5º	626	388	39
Linguística, Letras e Artes	1º	378	309	134
	2º	331	262	131
	3º	296	252	125
	4º	246	231	107
	5º	225	208	106

Os dados apresentados mostram que existe realmente uma parcela de indivíduos com produtividade muito acima da média. Além disso, existe uma grande distinção entre as grandes áreas, destacando-se a grande área de Ciências da Saúde com os indivíduos com maior produção de artigos em anais de congresso, e as grandes áreas de Ciências Sociais Aplicadas e Ciências Exatas e da Terra com os indivíduos com maior produção de artigos em periódico. Para ressaltar a distinção entre as grandes áreas, o indivíduo com o maior número de trabalhos publicados em anais de congresso na grande área de Ciências da Saúde possui 2.107 publicações, produção bem maior do que a dos cinco indivíduos mais produtivos da grande área de Linguística, Letras e Artes, que têm os menores valores nesta categoria. É importante destacar que esses indivíduos altamente produtivos são, em sua maioria, bolsistas de produtividade em pesquisa do CNPq. Dentre o conjunto apresentado na Tabela 5.10, apenas 5 não recebem bolsa atualmente.

Diante do exposto, é possível identificar que, apesar da grande quantidade de currículos cadastrados na Plataforma Lattes, somente uma parcela deles possui dados que possibilitam realizar análises específicas sobre a produção científica brasileira, como as análises baseadas em redes de colaboração científica ou sobre a qualidade e o impacto das publicações. A identificação de uma parcela que represente os indivíduos mais produtivos não é uma tarefa trivial, tendo em vista a dificuldade para se definir os critérios de seleção. Para isso, uma tabulação considerando o maior nível de formação informado pelos indivíduos em seus respectivos currículos foi realizada (Tabela 5.11).

Tabela 5.11: Número de publicações por maior nível de formação.

Produção / Maior Nível de Formação	Sem Formação Declarada / Ensino Médio	Curso Técnico / Profissionalizante	Graduação	Especialização	Mestrado	Doutorado	Total
Total de Currículos	1.857.980 44,70%	89.280 2,15%	1.083.439 26,07%	529.440 12,73%	373.019 8,97%	223.477 5,38%	4.156.635
Artigos em Anais de Congressos	283.367 2,44%	34.534 0,31%	802.993 6,93%	433.959 3,74%	2.540.123 21,91%	7.496.166 64,67%	11.591.142
Artigos em Periódicos	32.285 0,71%	3.692 0,08%	162.556 3,56%	130.813 2,87%	833.275 18,27%	3.398.300 74,51%	4.560.921
Capítulos de Livros	7.191 0,68%	587 0,06%	36.145 3,42%	37.406 3,54%	163.144 15,46%	810.915 76,84%	1.055.388
Livros	3.675 0,80%	315 0,07%	19.288 4,23%	26.357 5,79%	85.776 18,84%	320.036 70,27%	455.447
Textos em Jornais e Revistas	25.357 1,87%	2.297 0,17%	107.834 7,97%	132.890 9,82%	310.819 22,96%	774.448 57,21%	1.353.645
Média de Artigos em Periódicos	0,02	0,04	0,15	0,25	2,23	15,20	1,10

Como pode ser observado, os indivíduos com maior nível de formação concluída são, consequentemente, os mais produtivos. Existe um crescimento gradual em praticamente todos os tipos de publicação entre os níveis de formação mais baixos para os mais elevados, em que o conjunto que possui titulação de doutor se apresenta bem mais produtivo que todos os outros. Apesar de ser o grupo com o segundo menor número de currículos, os doutores possuem o maior percentual de publicações dentre todos os grupos considerados. Importante destacar a quantidade de publicações de artigos em periódico dos doutores, que detêm 74,51% de todos os artigos deste tipo de publicação cadastrados na Plataforma Lattes. Esta quantidade de artigos faz com que a média de artigos em periódico para cada doutor (16,55) seja bem superior à dos outros grupos. Logo, diante do exposto na Tabela 5.11, o restante desta tese irá se concentrar no conjunto de indivíduos com doutorado concluído.

Capítulo 6

Análise da Produção Científica dos Doutores

Neste capítulo, é apresentada uma análise da produção científica do conjunto de doutores encontrados na Plataforma Lattes. Este conjunto de pesquisadores se caracteriza por possuir o nível mais elevado de formação acadêmica e por ser responsável pela maioria dos trabalhos cadastrados nessa plataforma. Inicialmente, é apresentada uma visão geral desse conjunto de pesquisadores que possibilita traçar o perfil de publicação dos doutores brasileiros (Seção 6.1). Após, é realizado um estudo da produção científica do conjunto de doutores (Seção 6.2) e, ainda, uma análise baseada em redes de colaboração científica que possibilita compreender como ocorre a evolução da colaboração científica entre os doutores brasileiros ao longo dos anos (Seção 6.3). Posteriormente, são realizadas análises tendo como objeto de estudo as principais áreas do conhecimento, para que, dessa forma, possam ser identificadas aquelas mais representativas em cada uma das grandes áreas consideradas na Plataforma Lattes (Seção 6.4). Finalmente, são apresentadas as considerações finais (Seção 6.5).

6.1 Caracterização Geral

Para análises detalhadas da produção científica brasileira e ainda para mostrar o potencial do LattesDataXplorer na coleta e tratamento dos dados, considerou-se o conjunto de doutores com currículos cadastrados na Plataforma Lattes. O motivo da seleção desse conjunto de indivíduos para análises específicas tem como motivação os resultados apresentados no Capítulo 5, no qual foi possível observar que, apesar desse conjunto corresponder a somente 5,38% do total de indivíduos, ele é responsável por 74,51% dos artigos publicados em periódico e 64,67% dos artigos publicados em anais de congresso. Além disso, em geral esse conjunto de currículos possui data de atualização recente e praticamente em sua totalidade possui algum tipo de publicação registrada. Este grupo de indivíduos que, em sua maioria, tem atuado em pesquisas,

seja em instituições de ensino seja em institutos de ciência e tecnologia, ainda é responsável pela formação dos alunos nos principais programas de pós-graduação *stricto sensu* no Brasil, e vários deles são reconhecidos por sua elevada produção científica. Com isso, ressalta-se que o conjunto de indivíduos analisado neste capítulo compreende grande parte dos docentes de pós-graduação e todos os bolsistas de produtividade em pesquisa do CNPq que serão analisados no próximo capítulo (ver Figura 6.1). Ressalta-se que nem todos os bolsistas estão vinculados a programas de pós-graduação, estando alguns deles alocados em centros de pesquisa, como, por exemplo, Embrapa e Fiocruz.

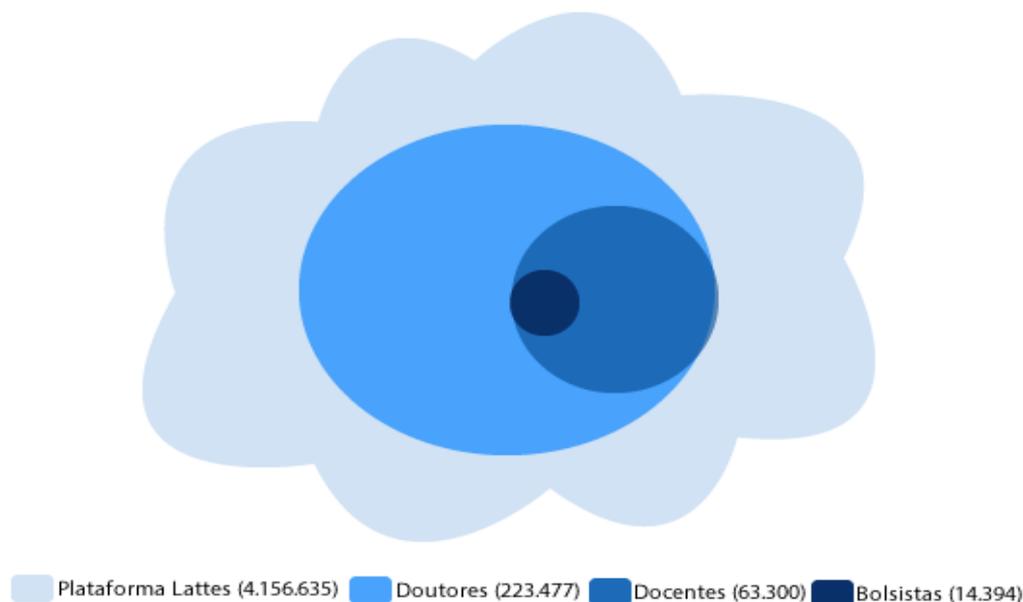


Figura 6.1: Representatividade dos Doutores na Plataforma Lattes.

É possível observar que, do total de currículos cadastrados na Plataforma Lattes, 223.477 correspondem a indivíduos que possuem como titulação máxima concluída o doutorado. Como vários desses doutores estão vinculados a programas de pós-graduação ou possuem bolsas de produtividade em pesquisa do CNPq, eles são motivados a manter seus currículos atualizados. Outro fator a ser considerado é que, como grande parte dos doutores está vinculada a grupos de pesquisa, grupos estes vinculados ao Diretório de Grupos de Pesquisa no Brasil do CNPq¹, devido ao *Censo 2014*, realizado no final do ano de 2014, todos os pesquisadores vinculados a algum desses grupos foram recomendados a atualizar seus currículos. Com isso, é possível observar que estes currículos que fazem parte das análises aqui apresentadas possuem, em geral, dados recentemente atualizados (Figura 6.2).

Aproximadamente 70% dos currículos dos doutores foram atualizados há menos de um ano, destes, 39% já nos primeiros meses de 2015. Os currículos não atualizados recentemente, em geral, são de indivíduos que informaram apenas informações básicas e possuem poucos registros sobre publicações ou dados sobre atuações profissionais. A grande vantagem em se analisar um conjunto de currículos recentemente atualizado como o dos doutores é que trabalhos recém-publicados provavelmente já estarão registrados nesses currículos, proporcionando uma visão atual da produção científica brasileira.

¹ <http://lattes.cnpq.br/web/dgp>

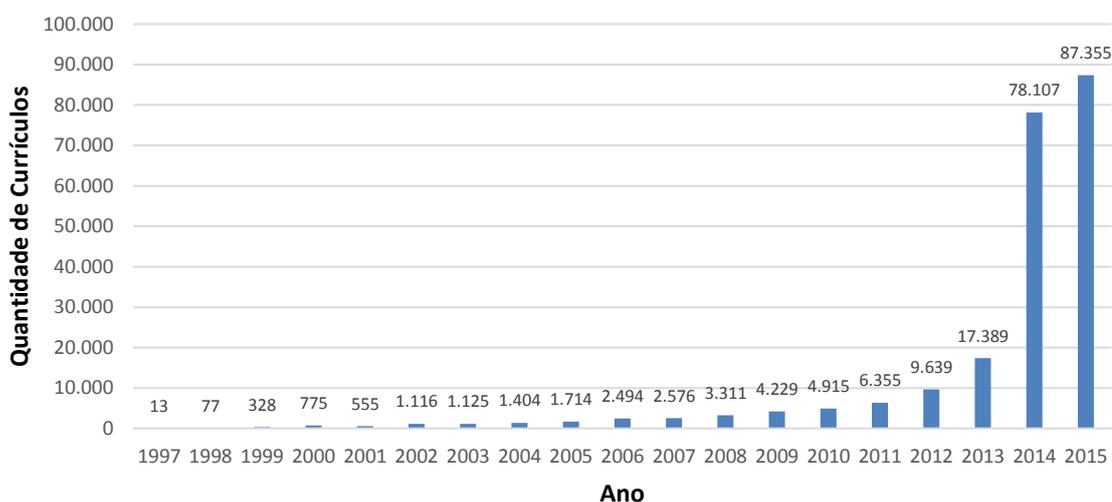


Figura 6.2: Data da última atualização dos currículos dos doutores.

A análise dos doutores que possuem currículos cadastrados na Plataforma Lattes tem como objetivo avaliar o conjunto de pesquisadores que possuem o nível mais alto de qualificação acadêmica. Em geral, ter doutorado concluído se faz necessário para atuar em programas de pós-graduação como também para participar de diversos editais de fomento à pesquisa no Brasil. Tendo em vista traçar a tendência de publicação desse conjunto de indivíduos, ao se analisar os principais tipos de publicação, percebe-se que a maioria possui grande quantidade de artigos publicados em anais de congresso, seguido por artigos em periódico (ver Figura 6.3).

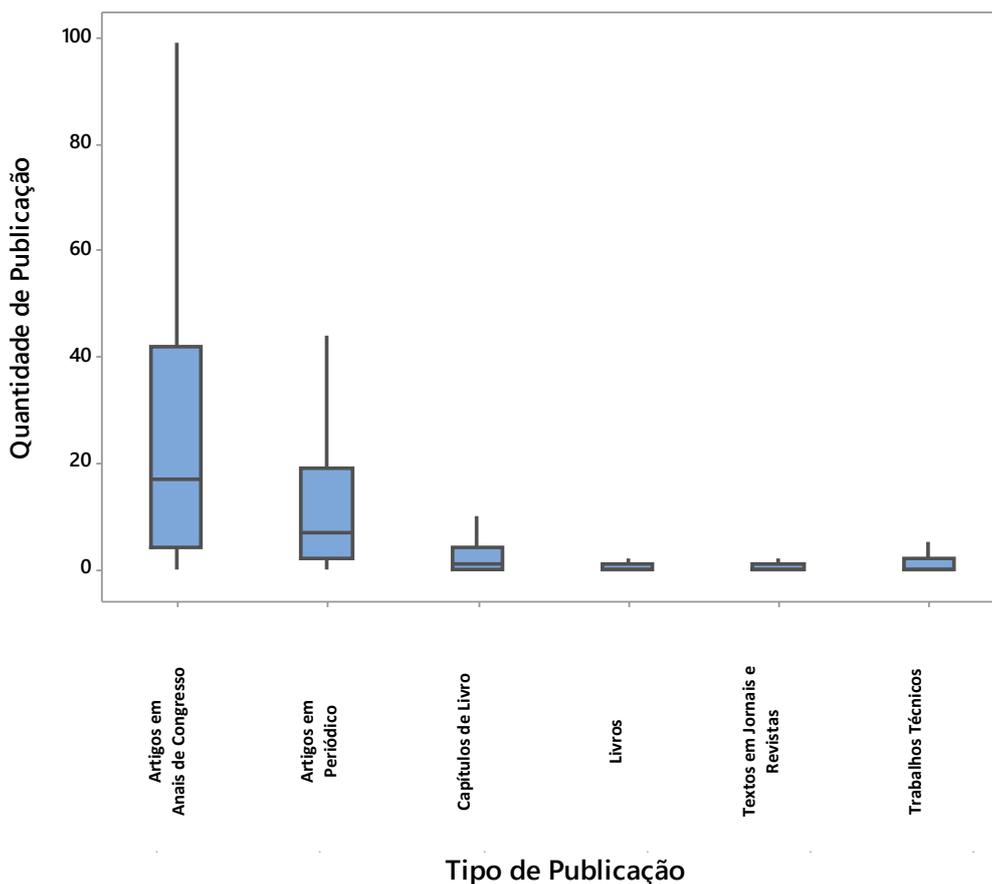


Figura 6.3: Taxa de publicação dos doutores por tipo de produção.

Mesmo com a queda acentuada no número de artigos em anais de congresso evidenciada nos últimos anos, em geral esse tipo de publicação é o que aparece com mais frequência nesse conjunto. Vale ressaltar que, em algumas áreas, a publicação em conferências é tão relevante quanto a publicação em periódicos, como, por exemplo, na área de Ciência da Computação (LAENDER et al., 2008).

Além do exposto, diferentemente do que ocorre quando considera-se todo o conjunto de indivíduos cadastrados na Plataforma Lattes, no qual o valor da mediana do total de artigos em periódico é igual a zero, no caso do conjunto dos doutores, este valor é igual a sete. O motivo das baixas médias quando se analisa o todo conjunto de indivíduos está diretamente relacionado ao fato de que a maioria dos currículos não inclui qualquer publicação (ver Figuras 5.12 e 5.13). No entanto, com o conjunto de doutores, é possível identificar apenas uma pequena parcela do conjunto que não possui publicações cadastradas (ver Figuras 6.4 e 6.5).

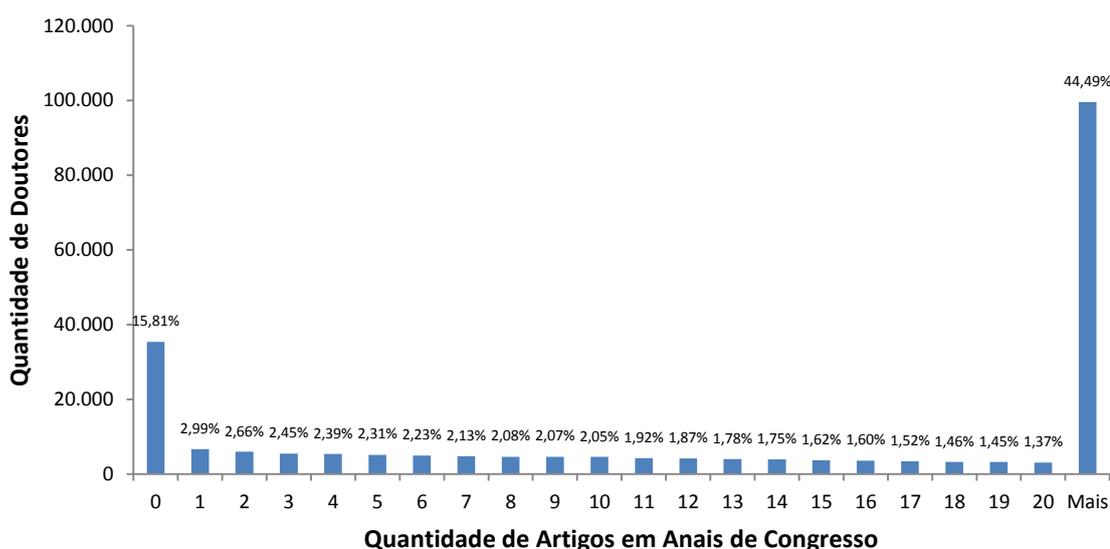


Figura 6.4: Distribuição da quantidade de artigos em anais de congresso por doutor.

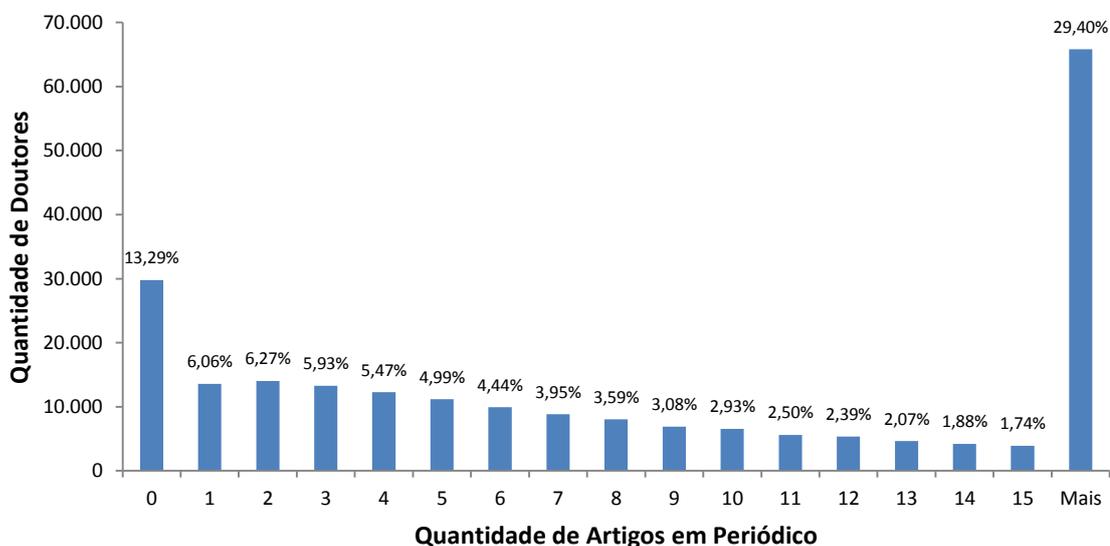


Figura 6.5: Distribuição da quantidade de artigos em periódico por doutor.

Como pode ser observado, apenas 15,81% e 13,29% dos currículos não possuem registro de publicações em anais de congresso e em periódico respectivamente. Ao verificar tais currículos sem registro de publicações, foi possível identificar que em sua maioria, são de doutores em início de carreira que, em geral, ainda não possuem publicações, ou, ainda, de currículos que possuem somente dados básicos registrados, como resumo biográfico e formação acadêmica, estando há anos sem atualização. Tais currículos, em geral, são de indivíduos que não estão vinculados a instituições de ensino.

Outra informação importante para compreensão do perfil científico nacional está relacionada à quantidade de doutores que se formam a cada ano. A partir da Figura 6.6, é possível verificar que a quantidade de doutores tem aumentado consideravelmente a cada ano. O doutorado concluído mais antigo identificado no conjunto de indivíduos analisado é de 1932, seguido por outro concluído em 1939 e, posteriormente, um terceiro em 1944. Após, o número de conclusões aumenta de forma muito modesta até meados de 1970 com algumas dezenas de doutores formados por ano. De 1975 até 1996, ocorreu um crescimento constante no número de doutores formados e, já em 1997, houve um aumento significativo nesse número que não se manteve no ano seguinte, mas que voltou a se manifestar de forma constante a partir de 1998. Entre 2011 e 2013, a taxa de crescimento aumenta novamente de forma significativa, sofrendo, no entanto, uma queda em 2014. Esse período coincide exatamente com a expansão dos cursos de doutorado ocorrido nos últimos anos no país, conforme destacado nas avaliações trienais da CAPES de 2007 e 2010.

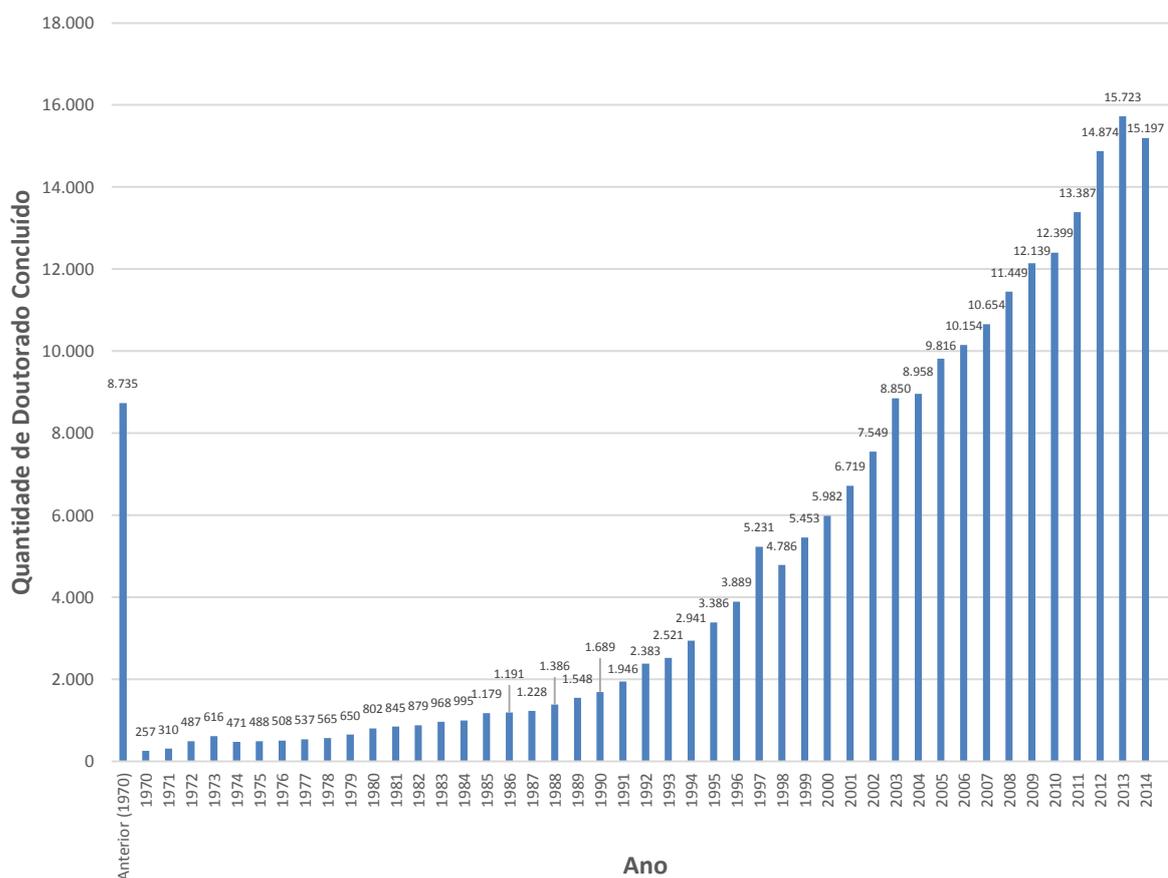


Figura 6.6: Quantidade de doutores formados por ano.

6.2 Análise da Produção Científica dos Doutores

Tendo em vista a expansão dos cursos de doutorado ocorrida a partir do final da década de 1990, nesta seção é apresentada uma análise detalhada da produção científica dos doutores cadastrados na Plataforma Lattes, com ênfase nos artigos publicados em anais de congresso e em periódico. A exemplo do que foi apresentado no Capítulo 5, a produção científica desse conjunto de indivíduos também tem sofrido queda em todos os tipos de publicação analisados, sendo mais significativa no caso dos artigos em anais de congresso (ver Figura 6.7 e Tabela 6.1).

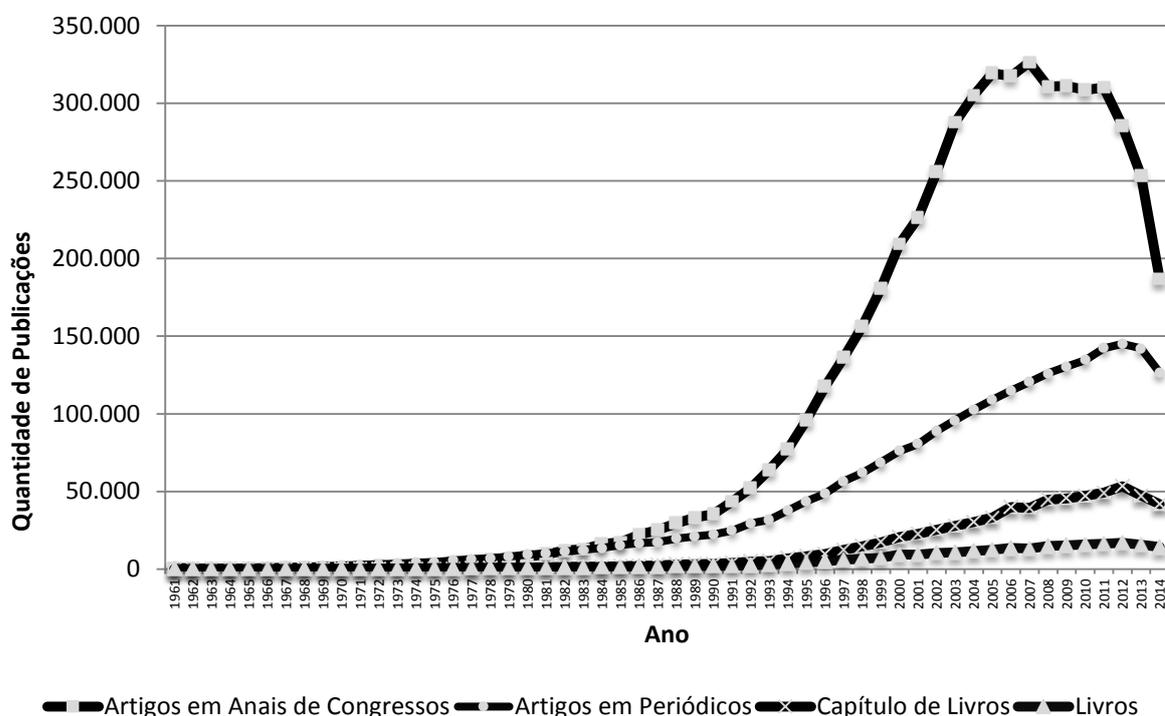


Figura 6.7: Produção científica dos doutores por ano.

Tabela 6.1: Produção científica dos doutores e taxa de crescimento por ano.

Tipo de Publicação	Anos										
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	
Artigos A. Congressos	Quant.	319.463	317.478	326.395	310.743	311.300	308.469	310.198	285.612	253.399	186.801
	Cresc.	1,04%	-0,63%	2,80%	-4,80%	0,18%	-0,91%	0,56%	-7,93%	-11,28%	-26,28%
Artigos em Periódicos	Quant.	108.966	114.971	120.437	126.069	130.449	134.614	142.180	145.148	142.126	126.480
	Cresc.	1,06%	5,51%	4,75%	4,68%	3,47%	3,19%	5,62%	2,09%	-2,08%	-11,00%

Como os doutores são responsáveis pela grande maioria dos trabalhos publicados no país, era esperado que esse comportamento fosse mesmo semelhante ao do conjunto de todos os indivíduos com currículos cadastrados na Plataforma Lattes. Assim, em praticamente todos os tipos de publicação, há uma queda acentuada a partir de 2012, ano em que se encerra um período de crescimento constante, iniciado na década de 1990. Importante destacar que, para todos os tipos de publicação, só foram considerados trabalhos únicos, ou seja, trabalhos em colaboração que estão cadastrados em mais de um currículo são contabilizados uma única vez.

Os artigos em periódico, cuja produção teve crescimento considerável a partir de 1992, atingem seu ápice em 2012, com 145.148 publicações. No entanto, a partir de 2013 inicia-se uma queda, sendo que em 2014 foram registradas apenas 126.480 publicações, queda de 11% em relação a 2013, ano com maior número de doutorados concluídos. Já os artigos em anais de congresso possuem características distintas dos demais tipos de publicação, tendo seu crescimento acentuado a partir de 1990, mantendo-se até 2005, ano em que há uma pequena queda e, conseqüentemente, um comportamento irregular em suas taxas de crescimento, com variações positivas e negativas bem próximas de 0. No entanto, a partir de 2011, percebe-se uma queda acentuada, reduzindo o número de publicações em, aproximadamente, 40% em um período de apenas três anos.

Uma hipótese para a queda acentuada no número de artigos em anais de congresso a partir de 2011 pode estar relacionada a exigências da CAPES, que passa a considerar apenas artigos em periódico em suas avaliações. Conforme descrito pelos Instrumentos de Apoio para a Classificação da Produção Intelectual², o Qualis-Periódicos³ é o conjunto de procedimentos utilizados pela CAPES para estratificar a qualidade da produção intelectual dos programas de pós-graduação. O Qualis afere a qualidade das publicações a partir da análise da qualidade dos veículos de divulgação, ou seja, periódicos científicos. Como um grande percentual dos doutores está vinculado a programas de pós-graduação, estes passaram a dedicar seus esforços na publicação de artigos em periódico, em detrimento aos artigos em anais de congresso, já que não seriam mais relevantes para as avaliações e, portanto, pouco influenciariam na melhoria dos conceitos de seus programas. Porém, considerando-se a produção científica dos anos seguintes, não se percebe um crescimento no número de artigos em periódico como era de se esperar.

Uma outra hipótese que poderia justificar de forma geral a queda na produção científica brasileira está relacionada ao REUNI⁴ (Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais), que visa a ampliação do acesso e permanência na educação superior, no nível de graduação presencial, para melhor aproveitamento da estrutura física e de recursos humanos existentes nas universidades federais, bem como, a criação de novos campi no interior do país. Houve uma expansão no número de cursos e vagas ofertadas pelas instituições de ensino superior do Brasil, expansão esta que não aconteceu com a mesma intensidade com relação ao número de docentes contratados, impactando diretamente no aumento da carga de trabalho dos docentes dessas universidades (LÉDA; MANCEBO, 2009; NISHIMURA, 2014). Com isso, doutores que estavam exclusivamente realizando pesquisas ingressaram em universidades, em sua maioria localizadas fora dos grandes centros, resultado da interiorização do REUNI, o que impactaria diretamente na capacidade desses doutores em dar continuidade às suas pesquisas. Além disso, professores doutores, vinculados às instituições afetadas pelo REUNI, tiveram que dedicar grande parte de seus esforços em atividades administrativas desses novos cursos e, também, com a preparação de um número cada vez maior de aulas. Logo, com menos tempo para a pesquisa, sua produtividade na publicação de artigos é comprometida de forma significativa.

² www.capes.gov.br/avaliacao/instrumentos-de-apoio/classificacao-da-producao-intelectual

³ <https://qualis.capes.gov.br/>

⁴ <http://reuni.mec.gov.br/>

Com o intuito de verificar esta hipótese, foram selecionados cinco grupos de doutores: (1) de uma instituição que foi criada com o REUNI, (2) de uma instituição que expandiu a oferta de cursos com o REUNI na região Sudeste, (3) de uma instituição que expandiu a oferta de cursos com o REUNI na região Norte, e (4 e 5) de centros de pesquisa sem atividades de ensino e que, portanto, não sofreriam interferência do REUNI. A análise da produção científica desses grupos mostrou, entretanto, que todos possuem o mesmo comportamento, comportamento este, idêntico ao do conjunto total de doutores, incluindo os dois grupos compostos por pesquisadores que não estão envolvidos com atividades de ensino (ver Figura 6.8). Dessa forma, não se pode alegar que o REUNI seja o principal responsável pela queda na produção científica, tendo em vista que mesmo doutores vinculados a instituições não afetadas pelo programa também tiveram queda em suas publicações.

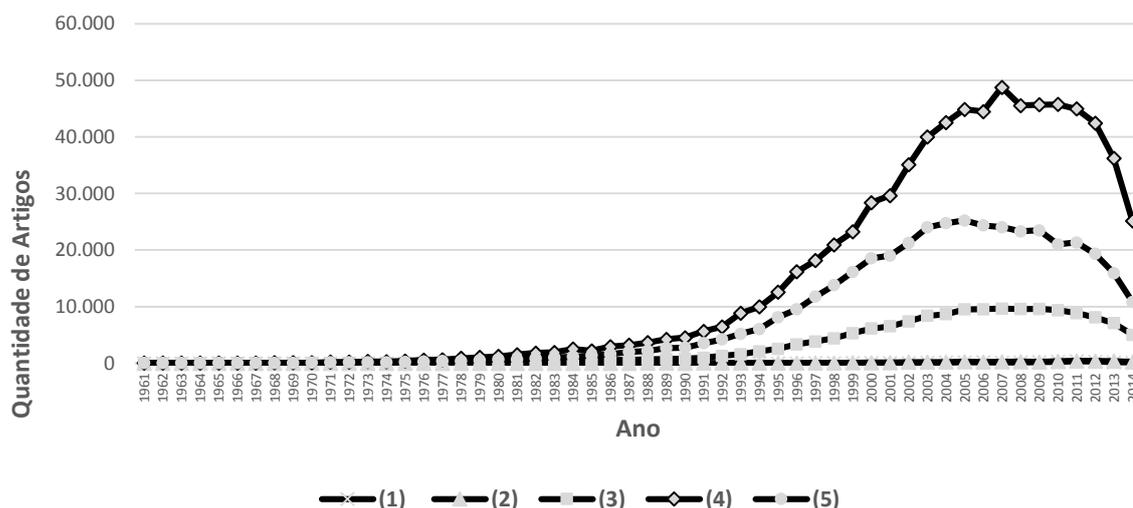


Figura 6.8: Produção de artigos em anais de congresso dos grupos considerados.

Para melhor compreensão e, ainda, para tentar encontrar justificativas que expliquem as razões da recente queda em todos os tipos de publicação, foi calculada a produção per capita do conjunto de doutores (ver Figura 6.9).

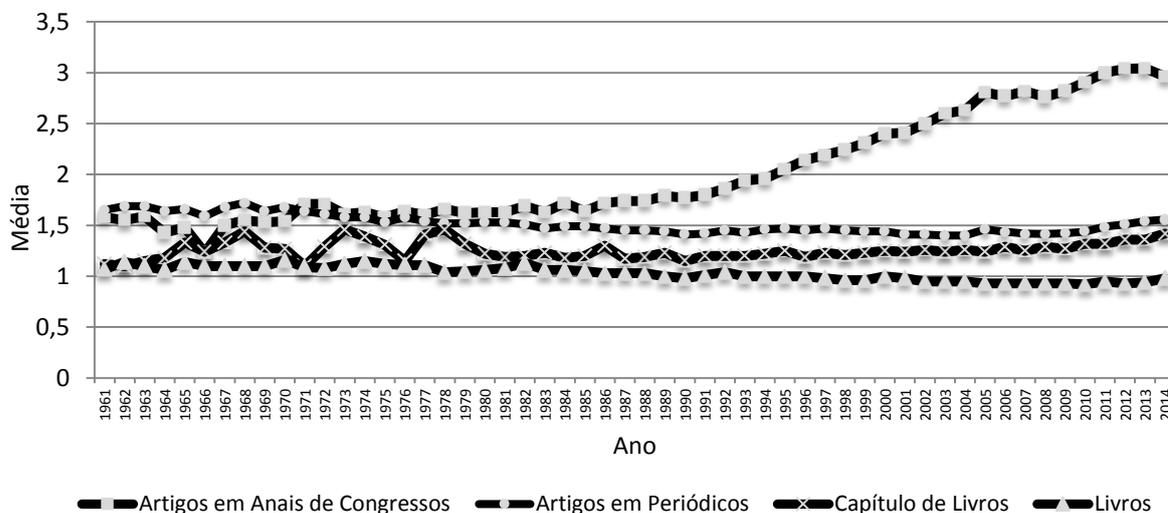


Figura 6.9: Produção per capita dos doutores por ano.

Para o cálculo da produção per capita foi considerado o número de publicações em cada um dos tipos em cada ano, tendo como universo de pesquisadores aqueles que tiveram alguma publicação daquele tipo no ano em questão. A divisão desses valores aqui denominada de produção per capita representa a taxa de publicação, desconsiderando todos os pesquisadores que não publicaram em um determinado ano. Como pode ser observado, há um crescimento nos últimos anos em todos os tipos de publicação, exceto no caso dos artigos em anais de congresso no último ano (2014). Baseado na quantidade de publicações (Tabela 6.1) e considerando o conjunto dos doutores analisados, era de se esperar que a tendência dos valores da produção per capita reduzisse ao longo dos últimos anos, devido à queda geral na produção científica. Porém, não é o que ocorre, ou seja, apesar de recentemente ter havido uma queda geral na produção científica, a produção per capita aumentou nesse mesmo período, exceto em 2014 no caso de artigos em anais de congresso, resultado de uma queda na quantidade de doutores que têm publicado a cada ano, superior à de trabalhos publicados.

Com o intuito de verificar a hipótese sobre a queda do número de autores, foi realizada uma comparação entre o número de publicações e a quantidade de autores que publicaram no ano em questão (ver Figura 6.10). Como pode ser observado, a quantidade de autores que publicaram nos últimos anos também sofre uma redução considerável a exemplo do que acontece com a quantidade de publicações. Neste cálculo, um autor somente é contabilizado se teve, pelo menos, uma publicação naquele referido ano. Percebe-se que a quantidade de autores que publicaram em anais de congresso atinge seu ápice em 2007 (115.809) e, a partir deste ano, sofre uma queda anual constante. Já com relação aos autores que publicaram em periódicos, a queda se inicia após 2012, ano com a maior quantidade de autores (96.079), sofrendo queda considerável nos dois próximos anos. Assim, com o passar do tempo, o número de doutores com produção científica tem diminuído. Para melhor visualizar tal fenômeno, a Tabela 6.2 apresenta uma comparação das taxas de queda da produção científica e da quantidade de autores.

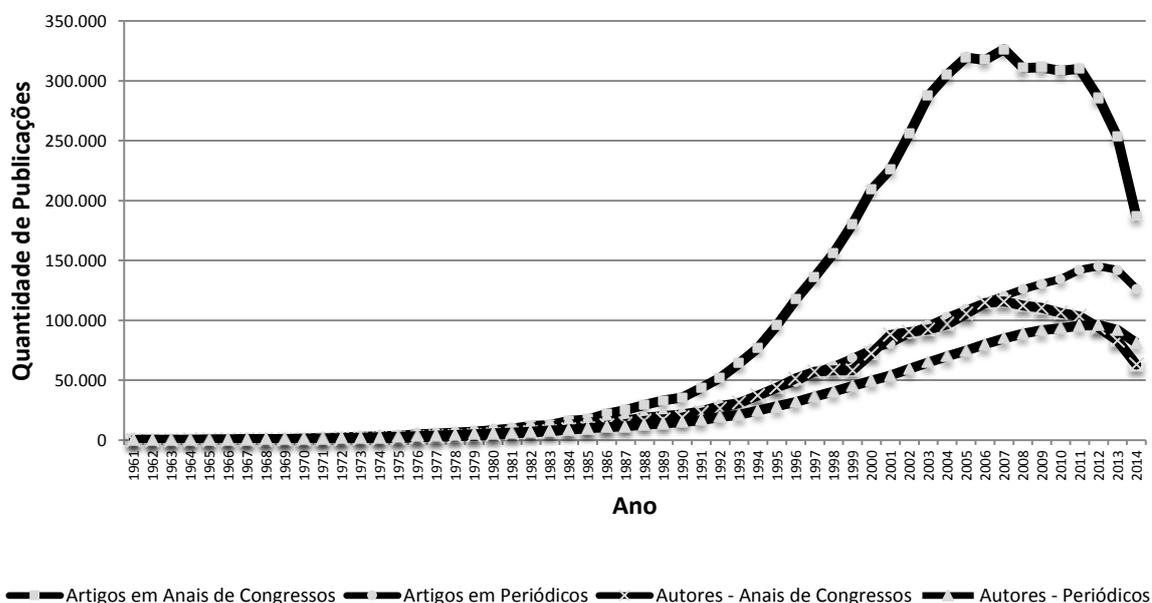


Figura 6.10: Quantidade de publicações e quantidade de autores por ano.

Tabela 6.2: Taxa de crescimento dos trabalhos publicados e dos autores.

		Anos									
		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Artigos A. Congressos	Quant.	319.463	317.478	326.395	310.743	311.300	308.469	310.198	285.612	253.399	186.801
	Cresc.	1,04%	-0,63%	2,80%	-4,80%	0,18%	-0,91%	0,56%	-7,93%	-11,28%	-26,28%
Autores A. Congressos	Quant.	105.348	114.622	115.809	112.493	110.406	106.179	103.405	94.003	83.252	63.035
	Cresc.	8,90%	0,73%	1,04%	-2,86%	-1,86%	-3,82%	-2,61%	-9,09%	-11,44%	-24,28%

		Anos									
		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Artigos em Periódicos	Quant.	108.966	114.971	120.437	126.069	130.449	134.614	142.180	145.148	142.126	126.480
	Cresc.	1,06%	5,51%	4,75%	4,68%	3,47%	3,19%	5,62%	2,09%	-2,08%	-11,00%
Autores em Periódicos	Quant.	74.651	80.041	84.928	88.982	91.741	93.537	95.582	96.079	92.112	81.266
	Cresc.	6,53%	7,22%	6,11%	4,77%	3,10%	1,96%	2,19%	0,52%	-4,13%	-11,77%

Outro esforço considerado, ainda, com o intuito de verificar as razões da queda acentuada da produção científica brasileira, foi comparar a produção científica dos pesquisadores que realizaram seus doutorados em instituições brasileiras com a produção de quem fez doutorado em instituições estrangeiras (Figura 6.11). Do conjunto de doutores analisados, 23.584 (10,55%) informaram em seus currículos, terem concluído seus cursos de doutorado em instituições estrangeiras. Conseqüentemente, a produção científica dos pesquisadores que se formaram no Brasil é bem superior. No entanto, o objetivo da análise é verificar se a produção científica dos dois grupos nos últimos anos possui comportamento semelhante. Observa-se que, ao se comparar as quedas na quantidade de publicações, percebe-se que tais quedas em ambos os tipos de publicação são equivalentes. Embora haja uma pequena distinção na produção científica entre os doutores formados em instituições brasileiras e estrangeiras, ela é mínima, não podendo ser considerada uma justificativa para a queda na produção dos doutores.

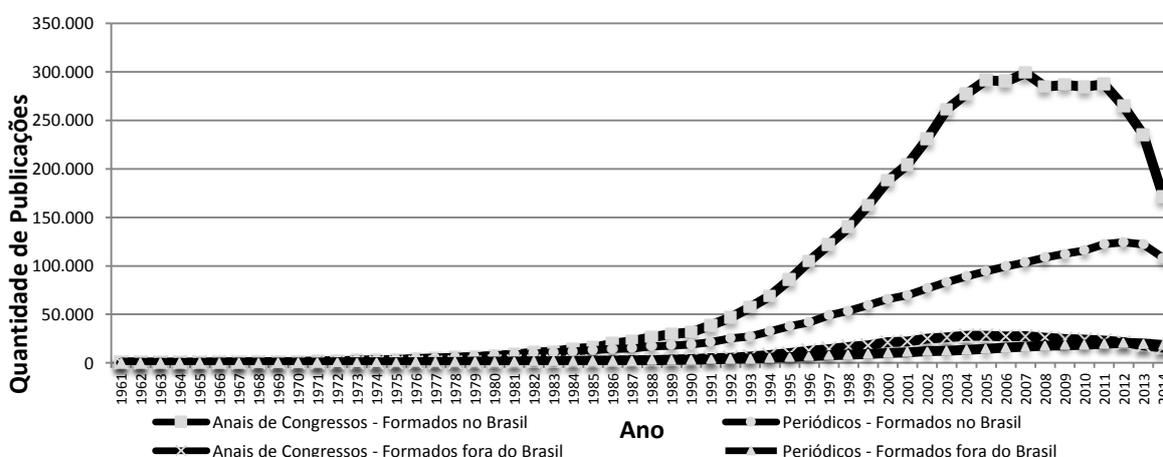


Figura 6.11: Produção de doutores formados no Brasil e em instituições estrangeiras.

Considerando o ano de conclusão dos doutorados, é possível identificar o período da carreira em que os doutores são mais produtivos. Tal informação é importante pois possibilita mensurar quando um determinado pesquisador vai contribuir de forma mais significativa com as pesquisas em sua área de atuação (ver Figuras 6.12 e 6.13).

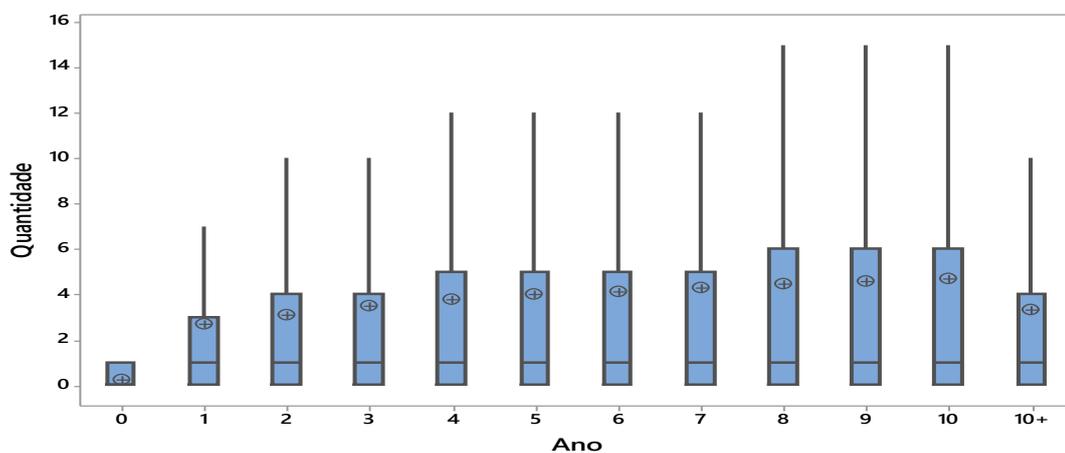


Figura 6.12: Distribuição da quantidade de artigos em anais de congresso pós-doutorado.

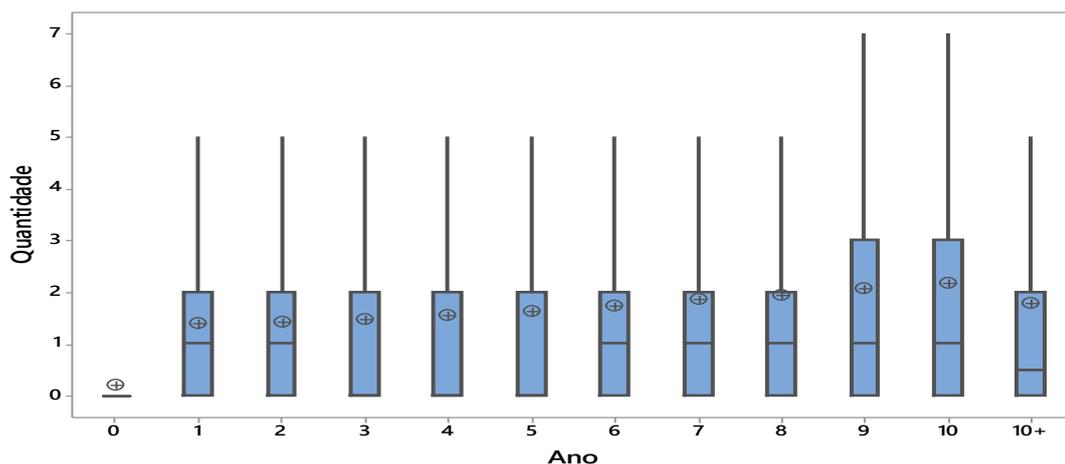


Figura 6.13: Distribuição da quantidade de artigos em periódico pós-doutorado.

Os dados apresentados correspondem à produção científica dos doutores a partir do ano em que concluíram os seus cursos de doutorado (ano 0). Para cada doutor, é analisada sua produção científica em cada ano posterior à conclusão do seu doutorado, ressaltando que só foram considerados os pesquisadores que já possuíam tempo de conclusão em cada ano analisado. Tal análise é realizada para os primeiros 10 anos e, posteriormente, agrupada para os anos seguintes para aqueles com mais de uma década de doutorado concluído (10+). Para o agrupamento em 10+, foi utilizada a mediana de sua produção científica dos anos que sucederam os seus primeiros 10 anos após o doutorado.

Ao verificar a produção pós-doutorado de artigos em anais de congresso, é possível identificar que, no ano de conclusão de seus cursos, a taxa de publicação dos doutores é igual a 1 e a mediana é 0, apresentando pouco volume de publicação. Consequentemente, em todos os outros anos analisados, a mediana é igual a 1. Porém, ao analisar as taxas de publicação, há uma tendência de aumento no número de artigos publicados já no primeiro ano, intensificando-se nos anos seguintes e alcançando as melhores taxas entre o oitavo e décimo ano, sofrendo queda após esse período. Tal fenômeno está relacionado ao fato de que, nos primeiros anos, o pesquisador publica resultados de seus esforços em pesquisas relacionadas ao trabalho desenvolvido em seu doutorado e, com o passar dos anos, quando já possui alunos orientados, geralmente em níveis de capacitação inferiores (iniciação científica,

graduação e mestrado), começa a participar da produção científica de tais alunos que, geralmente, acontece em anais de congresso. Em geral, a média (\oplus) tem acréscimo constante até o décimo ano.

Já para os artigos em periódico, a taxa de publicação e a mediana no ano de conclusão, são ambas iguais a zero. Após, as taxas de publicação até o oitavo ano são iguais a 2. No entanto, nos dois primeiros anos após a conclusão, a mediana é 1. Posteriormente, a mediana volta a ser zero entre o terceiro e quinto ano, retornando a ter valores próximos aos seus primeiros anos como doutor somente a partir do sexto ano de doutoramento. Uma explicação para tal fenômeno está relacionada ao fato de que, após a divulgação dos resultados de sua pesquisa no doutorado, o pesquisador não consegue ser tão produtivo, já que, em geral, se vincula a instituições de pesquisa e precisa dedicar-se a outras funções. Somente após algum período é que esses pesquisadores conseguem orientar alunos em níveis de capacitação mais elevados, como, por exemplo, doutorado, e dessa forma, começam a usufruir dos resultados dessas orientações, voltando a ter publicações em periódicos, alvo de interesse dos alunos concluintes deste nível de capacitação. A exemplo do que ocorre com os artigos em anais de congresso, a média também aumenta de forma constante até o décimo ano após o doutorado.

6.3 Análise das Redes de Colaboração Científica

Um tipo de análise que comumente vem sendo realizada é a análise sobre como a comunidade científica brasileira tem colaborado (MENA-CHALCO; DIGIAMPIETRI; CESAR-JUNIOR, 2012; BOAVENTURA et al., 2014; MENA-CHALCO et al. 2014; DIGIAMPIETRI, 2015). Nesta seção, são analisadas as redes de colaboração científica dos doutores brasileiros, considerando suas grandes áreas de atuação no período de 1960 a 2014, período este que compreende o ano de publicação da produção mais antiga cadastrada na Plataforma Lattes ao último ano do período de análise considerado. É possível observar que uma grande quantidade (35,85%) dos artigos publicados em anais de congresso e em periódico são de autoria única e, conseqüentemente, não caracterizam colaborações (ver Figura 6.14).

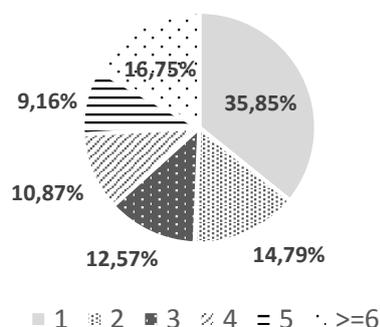


Figura 6.14. Quantidade de coautores em artigos publicados em anais de congresso e em periódico do conjunto de doutores.

Observa-se ainda que 14,79% dos artigos possuem dois coautores e 12,57% têm três coautores, mostrando que a maioria dos trabalhos são realizados com a participação de poucos autores. Além disso, foi identificada a existência de artigos com centenas de

coautores, fato este influenciado por algumas áreas do conhecimento que tendem a realizar seus trabalhos de forma mais colaborativa. Para melhor visualização desta distinção no perfil de colaboração dos pesquisadores brasileiros, é apresentada a quantidade de coautores em cada grande área (ver Figura 6.15).

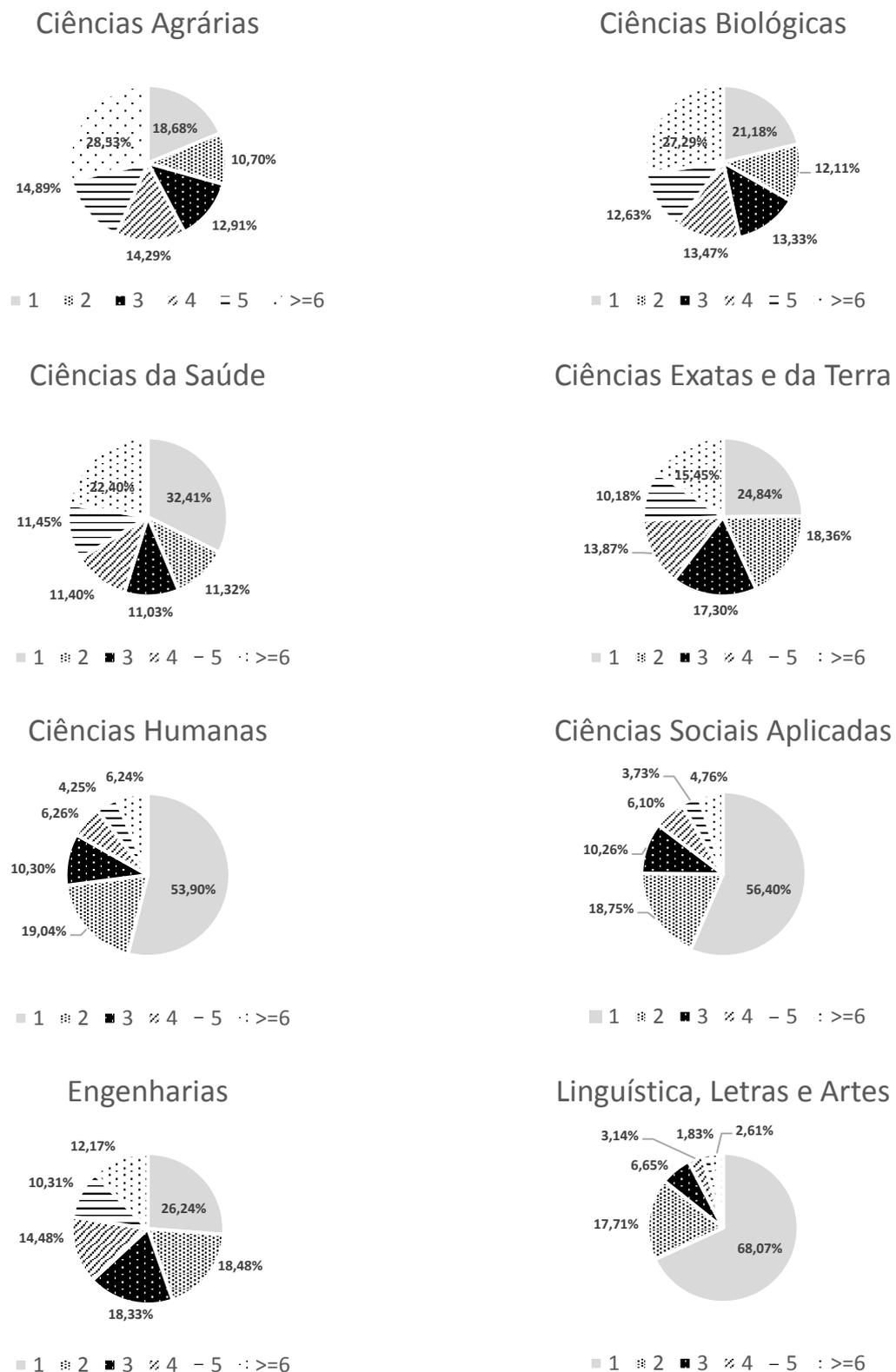


Figura 6.15. Quantidade de coautores em artigos do conjunto de doutores nas grandes áreas.

Como pode ser observado, os perfis de publicação dos pesquisadores em cada uma das grandes áreas são distintos. Pesquisadores das grandes áreas de Ciências Agrárias e Ciências Biológicas tendem a realizar seus trabalhos de forma mais colaborativa, em que uma quantidade considerável de seus trabalhos foi realizada com seis ou mais coautores. Outras grandes áreas como as de Ciências Humanas, Ciências Sociais Aplicadas e Linguística, Letras e Artes possuem pesquisadores que, em sua maioria, publicam de forma isolada, ou seja, sem colaboradores. Em especial, destaca-se a grande área de Linguística, Letras e Artes, em que 68,07% dos seus artigos foram realizados por um único autor e 17,71% com dois coautores.

No entanto, de maneira geral, a colaboração científica entre os doutores brasileiros tem aumentado de forma significativa nos últimos anos conforme mostrado na Figura 6.16. A publicação de livros e a de capítulos de livro têm comportamento similar quando se analisa o número de autores envolvidos, tendo havido um tímido crescimento na quantidade média de autores nos últimos anos. Já os artigos publicados em anais de congresso e em periódico tinham médias de autores similares até 2007, quando os artigos em periódico passam a ser produzidos de forma mais colaborativa. Atualmente, a média de autores dos artigos em periódico é de 5,22, já a dos artigos em anais de congresso é de 4,49. É importante destacar que, como as publicações estão sendo realizadas cada vez mais de forma colaborativa, esperava-se que esta colaboração influenciasse no número de trabalhos publicados de forma mais significativa, já que grupos mais colaborativos tendem a ser bem mais produtivos, o que não foi efetivamente observado.

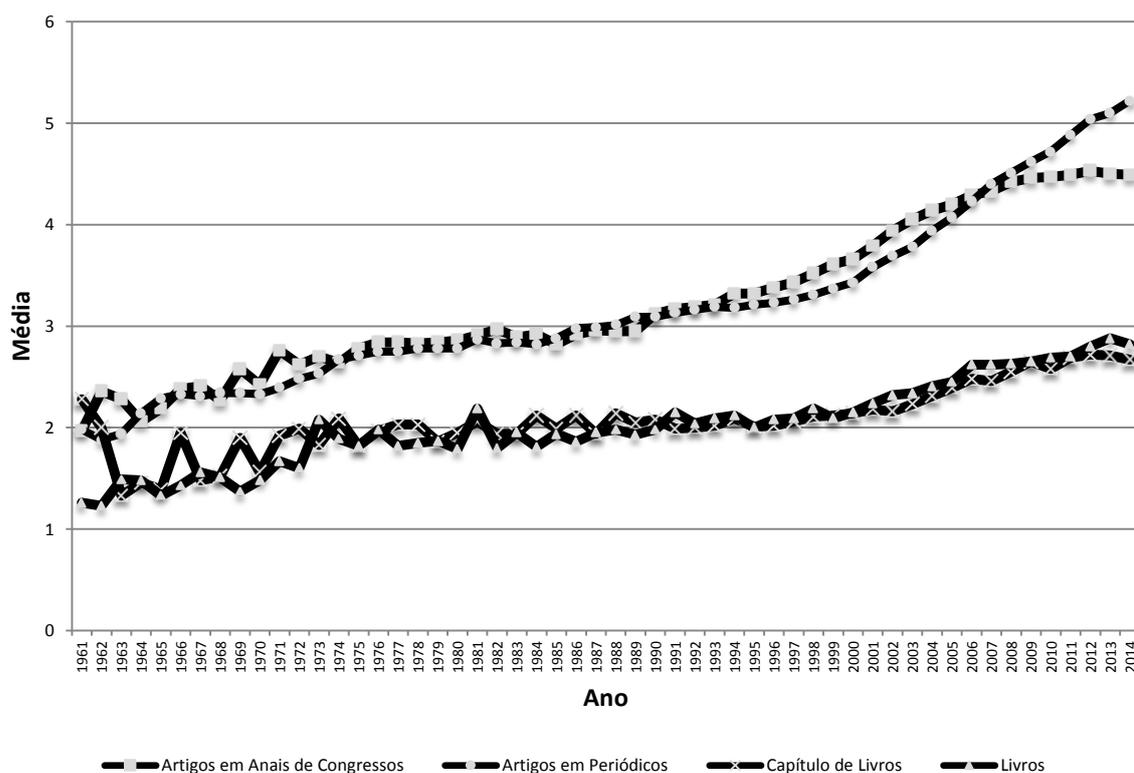


Figura 6.16: Média de autores por trabalho com pelo menos um doutor.

A análise da colaboração científica, entretanto, pode ser mais bem evidenciada com a caracterização das respectivas redes de colaboração. Tais redes possibilitam analisar, com a adoção de métricas específicas, como a colaboração tem evoluído ao longo dos anos. Nesta tese, foram caracterizadas redes temporais e, para suas análises, as métricas apresentadas na Subseção 2.2.2 foram aplicadas.

As métricas adotadas (grau médio dos nós, total de nós no componente gigante, densidade da rede, diâmetro da rede e caminho mínimo médio) são métricas clássicas, usualmente adotadas por diversos trabalhos que analisam redes de colaboração (SZWARCFITER, 1986; NEWMAN, 2003; LEMIEUX; OUIMET, 2008; SCOTT, 2009; WASSERMAN; FAUST, 2009). De forma que se possa observar como a colaboração científica evolui ao longo dos anos, tendo em vista a realização de uma análise temporal dessas redes, foram definidas janelas temporais. Tais janelas cobrem um período de cinco anos e se iniciam em 1960, ano que engloba as primeiras publicações identificadas na Plataforma Lattes, e terminam em 2014, último ano considerado em todas as análises apresentadas nesta tese. Para cada janela de tempo, foram caracterizadas as redes contendo todos os indivíduos que possuíam doutorado concluído no período em questão, independentemente de sua área de atuação ou tempo de titulação como doutor. Somente a produção científica do conjunto de doutores publicada no período foi considerada para a caracterização das redes de colaboração. Além disso, para melhor visualização de como a colaboração tem ocorrido nas grandes áreas e nas áreas específicas, também são apresentadas redes contendo somente o componente gigante. As cores dos nós que compõem as redes estão relacionadas às suas grandes áreas de atuação (ver Figura 6.17 e 6.18).

Para caracterização das redes que contêm todos os elementos, foram adotados métodos que visam concentrar os nós mais conectados no centro da rede e, conseqüentemente, aqueles nós com menor quantidade de ligações ou isolados são deslocados para as extremidades das redes (Figura 6.17). Já para a visualização dos componentes gigantes, componentes estes que contêm a maior quantidade de nós conectados, a ênfase está em agrupar os nós com maior colaboração em suas grandes áreas, possibilitando verificar como tem ocorrido a colaboração em cada uma delas (Figura 6.18). Como pode ser observado, as colaborações entre os doutores se iniciam no início da década de 1960 e são intensificadas a partir da década de 1990. É possível observar como as colaborações em algumas grandes áreas se estabilizaram ao longo do tempo, por exemplo, nas grandes áreas de Ciências da Saúde e Ciências Biológicas, bem como outras que passaram a ter um maior grau de colaboração nos últimos anos, como por exemplo, as grandes áreas de Ciências Exatas e da Terra e Ciências Agrárias.

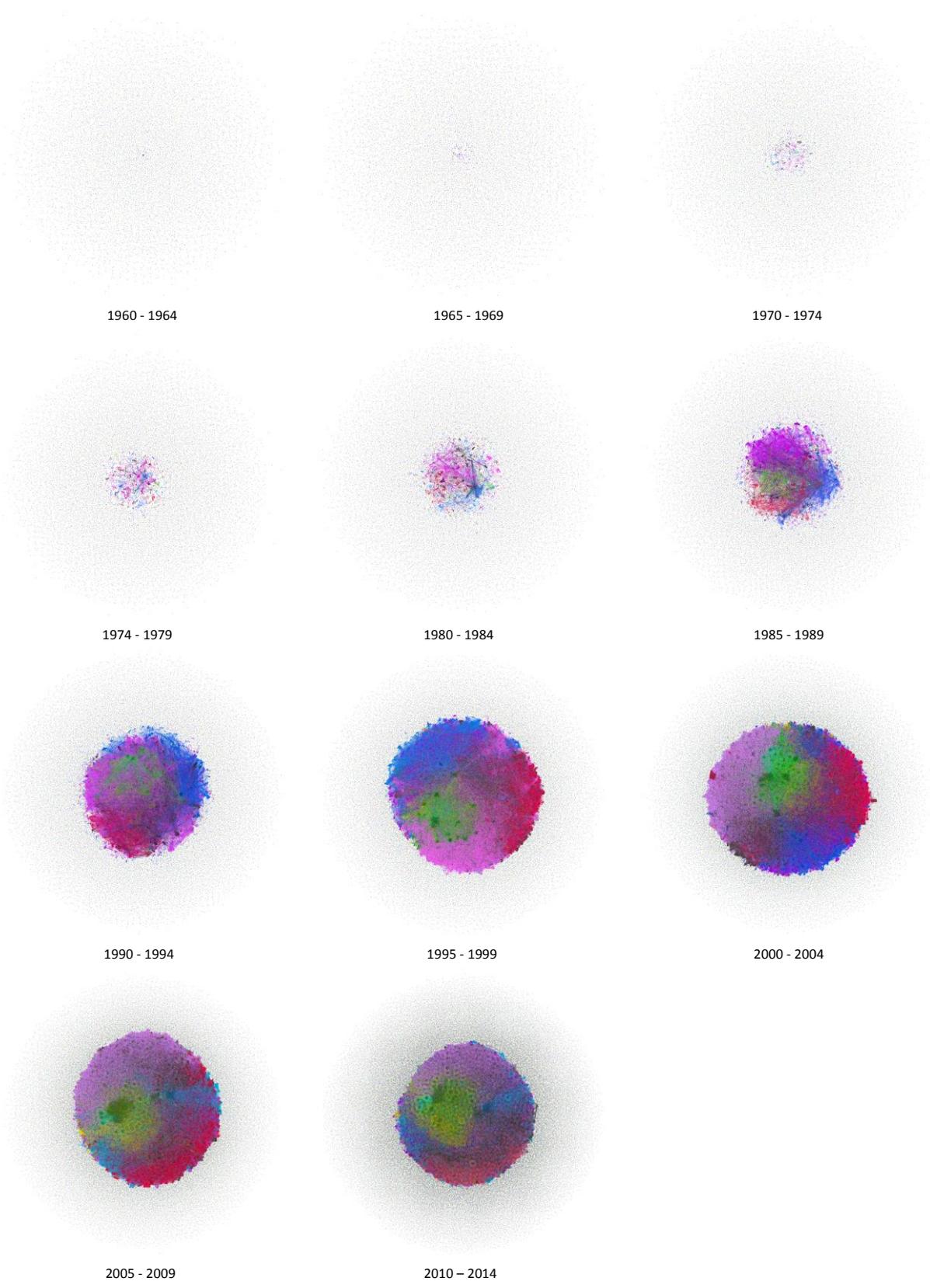


Figura 6.17: Redes gerais de colaboração científica contendo todos os nós.

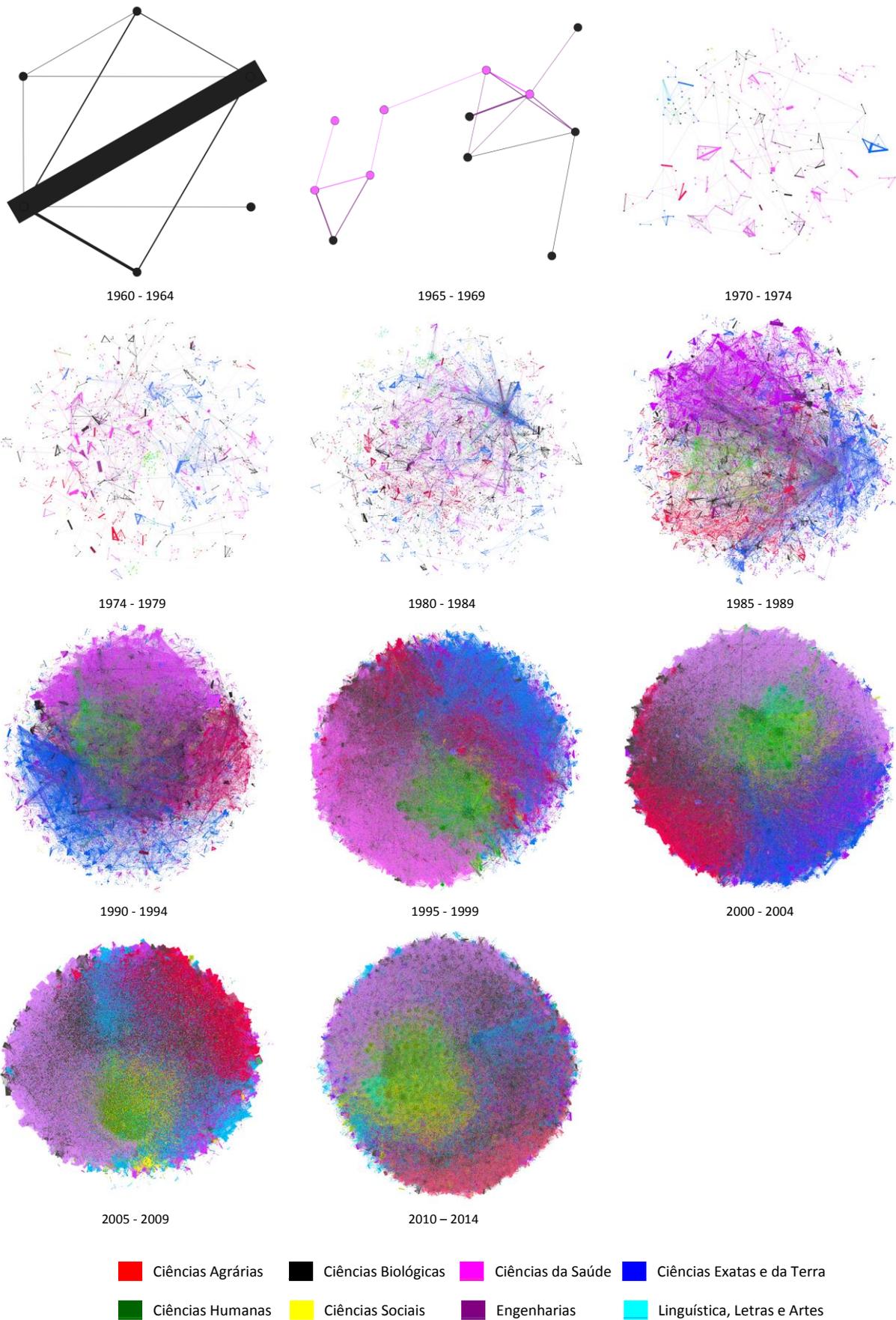


Figura 6.18: Componentes gigantes.

A Figura 6.19 apresenta a evolução temporal do tamanho do componente gigante da rede de doutores. É possível observar um crescimento mais intenso desse componente a partir do início da década de 1990, que se estende até o ano de 2011, período coincidente com o crescimento da produção científica brasileira. Esse componente cresce ao longo dos anos, já que outros componentes vão se conectando a ele. Verifica-se, ainda, que, em 2011, o componente gigante representa cerca de 61% da rede. Após, o componente gigante tem redução em todos os anos seguintes, período este também coincidente com a queda da produção científica brasileira.

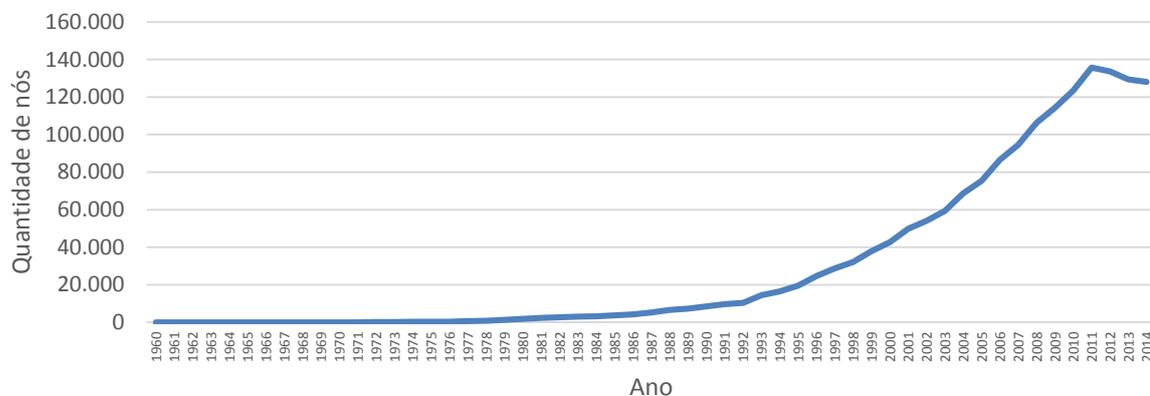


Figura 6.19: Evolução temporal do componente gigante.

Na Figura 6.20, é possível observar a evolução temporal do diâmetro da rede, o qual é influenciado pela evolução do componente gigante. Como o componente gigante vai se tornando mais conectado, o diâmetro tende a diminuir.

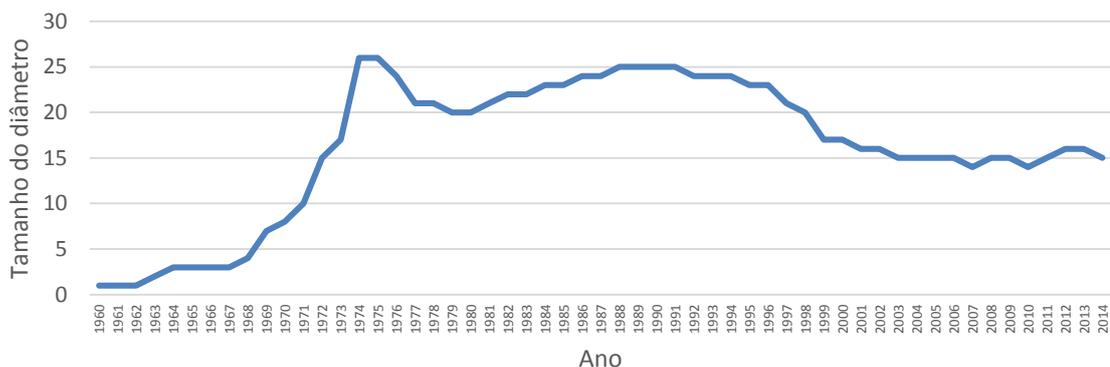


Figura 6.20: Evolução do diâmetro da rede.

A Figura 6.21 apresenta a evolução temporal do caminho mínimo médio da rede. Observa-se que, inicialmente, esta medida segue um padrão semelhante ao do diâmetro da rede. No entanto, após atingir seu maior valor (11,02) em 1975, tem redução constante até 1995. Posteriormente, volta a ter aumento no final da década de 1990 e, a partir do ano 2000, apresenta uma redução mais estável. Tal redução é reflexo de o componente gigante estar mais conectado. É importante destacar o aumento do caminho mínimo médio no quinquênio compreendido entre 1995 a 1999, período este, que também coincide com significativo aumento na quantidade de arestas e, ainda, com o lançamento e padronização do Currículo Lattes⁵ (agosto de 1999) como sendo o formulário de currículo a ser utilizado no âmbito do Ministério da Ciência e Tecnologia e CNPq.

⁵ <http://lattes.cnpq.br/>

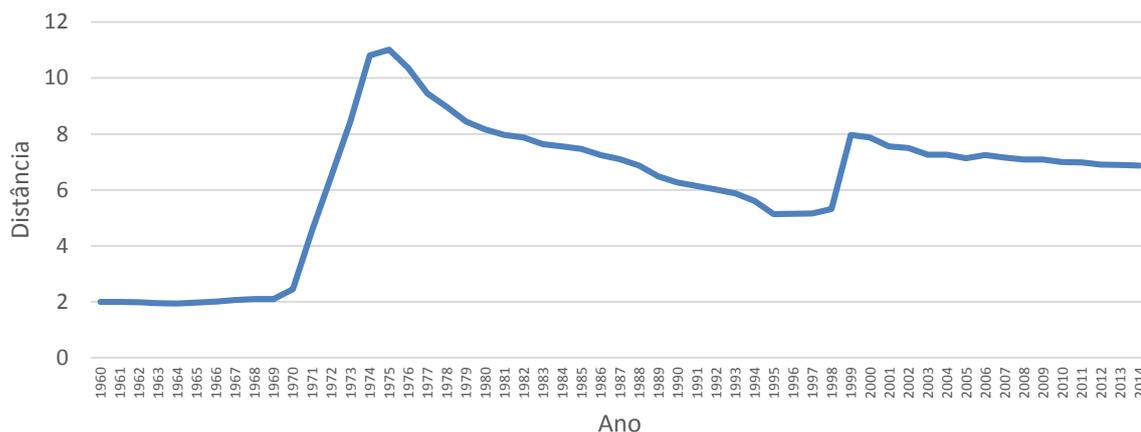


Figura 6.21: Evolução do caminho mínimo médio.

A Tabela 6.3 apresenta uma sumarização das redes caracterizadas e das métricas adotadas em todos os períodos analisados. É possível observar a evolução das redes, sendo que a rede inicial possuía apenas 8.186 nós e um componente gigante composto por poucos nós. Ressalta-se ainda a quantidade de arestas que, a exemplo da produção científica, tem crescimento em todos os períodos, com exceção do quinquênio compreendido entre 2010 a 2014, em que a produção científica coincidentemente apresenta significativa redução.

Tabela 6.3: Resultado das métricas adotadas.

Período	Métrica									
	Total de Nós	Total de Arestas	Grau Médio dos Nós	Total de Nós no Componente Gigante	% de Nós no Componente Gigante	Total de Arestas no Componente Gigante	Densidade da Rede	Diâmetro da Rede	Caminho Mínimo Médio	Total de Componentes Isolados
1960 - 1964	8.186	33	0,008	6	0,07%	9	0,00000098	3	1,939	8.140
1965 - 1969	8.735	129	0,028	12	0,13%	14	0,00000338	7	2,101	8.576
1970 - 1974	10.876	939	0,167	300	2,76%	455	0,00001587	26	10,804	10.036
1975 - 1979	13.624	3.176	0,455	1.333	9,78%	2.618	0,00003422	20	8,449	11.553
1980 - 1984	18.113	9.200	0,997	3.200	17,67%	8.283	0,00005608	23	7,551	13.711
1985 - 1989	24.645	26.781	2,145	7.310	29,66%	25.817	0,00008818	25	6,484	15.948
1990 - 1994	36.125	104.560	5,740	16.402	45,40%	103.691	0,00016024	24	5,601	18.409
1995 - 1999	58.870	688.180	5,970	37.970	64,50%	687.757	0,00039714	17	7,965	20.167
2000 - 2004	96.928	1.011.010	6,769	68.677	70,85%	1.009.655	0,00021522	15	7,256	27.443
2005 - 2009	151.140	1.225.113	7,380	114.292	75,62%	1.218.245	0,00010726	15	7,089	36.170
2010 - 2014	222.720	997.751	6,725	128.059	57,50%	992.195	0,00004022	15	6,875	67.240

A rede correspondente ao primeiro quinquênio analisado possui apenas 33 arestas, sendo que a aresta mais densa tem peso 58, indicando que somente uma dupla de doutores realizou no período 58 trabalhos em colaboração. O grau médio da rede é de apenas 0,008 e o componente gigante possui apenas seis nós, interligados por nove arestas. O diâmetro da rede é igual a 3 e seu caminho mínimo médio é pequeno (1,939), significando que com menos de duas arestas, em média, é possível chegar a um determinado nó a partir de qualquer outro que compõe o componente gigante. O componente gigante é composto por doutores que estão vinculados à Universidade de São Paulo (USP) e ao Instituto Butantã, todos atuando nas grandes áreas de Ciências da Saúde e Ciências Biológicas.

O próximo período analisado (1965 - 1969) possui 129 arestas, sendo a mais densa com peso 90. O grau médio dos nós é baixo (0,028), possuindo o componente gigante 12 nós, o que corresponde a 0,13% de todo o conjunto. É possível identificar que tanto o diâmetro como o caminho mínimo médio aumentaram se comparados ao período anterior. Doutores das áreas de Medicina e Farmacologia são a grande maioria dos que compõem o componente gigante, mas também existem nós correspondentes a doutores de outras áreas como Química e Bioquímica. Estes doutores em quase sua totalidade estão vinculados a instituições localizadas no estado de São Paulo, sendo as principais USP, Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Instituto Butantã e Universidade Estadual de Campinas (UNICAMP), nessa ordem.

O terceiro quinquênio analisado caracteriza-se pelo crescimento da quantidade de nós no componente gigante. Com isso, o diâmetro da rede aumenta de forma substancial (26), bem como o caminho mínimo médio. Tais fatores são justificados pela presença de doutores de todas as grandes áreas do conhecimento e um grau médio seis vezes maior que no período anterior. Há também uma maior presença de doutores da grande área de Ciências da Saúde, com destaque para o surgimento de grande quantidade de doutores vinculados à Fundação Oswaldo Cruz (FIOCRUZ).

A próxima rede analisada, a exemplo da anterior, também apresenta crescimento no número de arestas, indicando um aumento significativo na colaboração e também no grau médio dos nós. Aproximadamente 10% dos nós e 82% das arestas estão no componente gigante. O diâmetro da rede diminui e o caminho mínimo médio tem redução. Na rede, as grandes áreas de Ciências da Saúde e Ciências Biológicas têm a maior frequência de colaboração entre seus doutores, mas é possível identificar crescimento na colaboração entre pesquisadores das grandes áreas de Ciências da Saúde e Ciências Exatas e da Terra.

A rede correspondente ao período de 1980 a 1984 possui grau médio de 0,99 e diâmetro superior ao do período anterior. Aproximadamente 90% das arestas identificadas e aproximadamente 18% dos nós estão no componente gigante. A colaboração entre pesquisadores das áreas de Medicina e Química se consolida e um núcleo formado por doutores de Ciências Agrárias é responsável por um aumento significativo no grau de colaboração da rede. Apesar de o componente gigante possuir doutores de todos os estados, destaca-se São Paulo com 39,3%, Rio de Janeiro com 12,52% e Minas Gerais com 6,99%, como os estados mais representativos.

No próximo quinquênio (1985 - 1989), a colaboração entre os pesquisadores aumenta de forma significativa e, pela primeira vez, o grau médio é maior do que 1 (2,14). Aproximadamente, 30% dos nós estão no componente gigante e o diâmetro da rede (25) aumenta, ao contrário do caminho mínimo médio (6,48) que tem redução. Há um aumento considerável nas colaborações entre as grandes áreas de Ciências Humanas e Ciências Sociais Aplicadas. Ademais, Química (4,27%) segue como a terceira área mais representativa no componente gigante, atrás de Agronomia (5,92%) e Medicina (14,29%).

O próximo período analisado (1990 - 1994) destaca-se pelo crescimento na quantidade de arestas (104.560) e grau médio (5,74) dos nós que compõem a rede. Além do caminho mínimo médio (5,60), o diâmetro da rede também tem redução (24). Aproximadamente, 45% dos nós estão no componente gigante, englobando 99% das arestas. É possível identificar a ocorrência de uma grande colaboração entre pesquisadores das grandes áreas de Ciências da Saúde e Ciências Biológicas, e um aumento na colaboração entre pesquisadores da grande área de Ciências Biológicas, como também da grande área de Ciências Exatas e da Terra.

As redes caracterizadas com as publicações dos anos de 1995 a 1999 têm aumento significativo no número de arestas (688.180) e também no grau médio dos nós (5,97). Além disso, o diâmetro da rede tem redução (17), havendo um aumento no caminho mínimo médio dos nós (7,96). O componente gigante possui, aproximadamente, 64% dos nós da rede. Percebe-se uma maior colaboração entre pesquisadores das grandes áreas de Ciências Exatas e da Terra e Engenharias, e um baixo grau de colaboração entre pesquisadores da grande área de Linguística, Letras e Artes e das demais grandes áreas.

O quinquênio correspondente aos anos de 2000 a 2004 apresenta um grande aumento na quantidade de colaborações identificadas, o que resultou em um aumento do número de arestas, e conseqüentemente, do grau dos nós, chegando a uma média de 6,76 colaboradores para cada pesquisador. Além disso, é possível identificar que, aproximadamente, 71% dos nós estão no componente gigante e, a exemplo dos outros períodos, esses nós estão vinculados por 99% das arestas de toda a rede. Com relação às colaborações entre as grandes áreas, percebe-se um aumento na colaboração entre pesquisadores das grandes áreas de Ciências Humanas e Ciências Sociais, bem como entre pesquisadores das grandes áreas de Ciências Agrárias e Ciências Exatas e da Terra.

O próximo período analisado (2005-2009) apresenta crescimento no número de arestas identificadas e também no grau médio dos nós que compõem a rede. O diâmetro da rede continua o mesmo do período anterior e o caminho mínimo médio tem queda sensível (7,08). Já a quantidade de nós que integram o componente principal é de, aproximadamente 76%, englobando, praticamente, 99% das arestas. Nesse período, há um aumento no grau de colaboração entre pesquisadores das grandes áreas de Ciências da Saúde e Ciências Exatas e da Terra, como também entre pesquisadores das grandes áreas de Ciências Agrárias e Ciências Biológicas. A grande área de Ciências da Saúde ainda mantém o maior percentual de pesquisadores no componente gigante (17,88%), seguido das grandes áreas de Ciências Humanas (14,93%) e Ciências Exatas e da Terra (14,14%). Com relação às maiores áreas, a área

de Medicina ainda detém a maior quantidade de nós (6,99%), mas com percentual inferior à dos outros períodos, estando à frente das grandes áreas de Educação com 5,28% e Agronomia com 4,45%.

O último quinquênio (2010-2014), que engloba as publicações mais recentes, sofre queda no número de colaborações identificadas, sendo esta queda consequência direta da redução da produção científica nos últimos anos, conforme já apresentado. São identificadas apenas 997.751 arestas, uma queda de, aproximadamente, 20%, se comparado com o período anterior. Como consequência desta queda, há uma redução no grau médio dos nós, tendo em vista que, pela primeira vez, o número de componentes isolados, que indica nós com grau 0, aumenta de forma considerável. Esses nós que não possuem colaborações correspondem a pesquisadores que publicaram de forma isolada ou não tiveram nenhuma publicação no período analisado, reforçando o fato que, nos últimos anos, houve uma redução geral da produção científica brasileira.

Para sumarizar todo o exposto nesta seção, pode-se afirmar que a análise temporal das redes de colaboração científica do conjunto de doutores com currículos cadastrados na Plataforma Lattes possibilita obter uma visão de como ocorreu o processo de coautoria ao longo dos anos. Tendo como base as publicações científicas desses doutores, a primeira colaboração se remete ao início dos anos 1960 e, como pôde ser observado, a cada quinquênio, o número de colaborações tem aumentado, tendo como base a quantidade de arestas e o grau médio das redes que tiveram crescimento significativo, com exceção do último período analisado, em que a quantidade de componentes isolados cresceu de forma substancial. Destaca-se o período compreendido entre 2005 a 2009, que possui o maior valor do grau médio e também a maior quantidade de arestas, podendo ser considerado o período de maior colaboração da comunidade científica brasileira. Foi possível ainda identificar que, apesar de as redes possuírem baixa densidade, ao longo dos anos, a quantidade de componentes isolados, que correspondem a autores que publicaram sem nenhuma colaboração ou que não publicaram no período, teve crescimento bem inferior ao crescimento dos nós que compõem as redes, exceto no último quinquênio (2010-2014).

Também foi possível verificar como alguns pesquisadores têm trabalhado em colaboração de forma muito intensa. Analisando as arestas mais densas, desde o primeiro período analisado, percebe-se que alguns pesquisadores trabalham em colaboração com uma frequência muito alta. Em geral, tais colaborações são entre pesquisadores que atuam na mesma área ou em áreas correlatadas, que possuem a mesma formação acadêmica e, em determinadas situações, orientados pelos mesmos orientadores.

As análises apresentadas possibilitam compreender como ocorreu a evolução da colaboração científica brasileira de forma temporal, com quinquênios que compreendem todas as publicações encontradas na Plataforma Lattes considerando o conjunto de doutores analisados.

6.4 Análise por Áreas de Atuação

Para melhor compreensão do conjunto de doutores, é possível, ainda, identificar como eles estão distribuídos pelas suas grandes áreas de atuação (ver Figura 6.22). Esta identificação possibilita verificar quais as grandes áreas com o maior número de doutores e, dessa forma, viabilizar análises específicas com o intuito de compreender o comportamento e o padrão de publicação de cada uma delas.

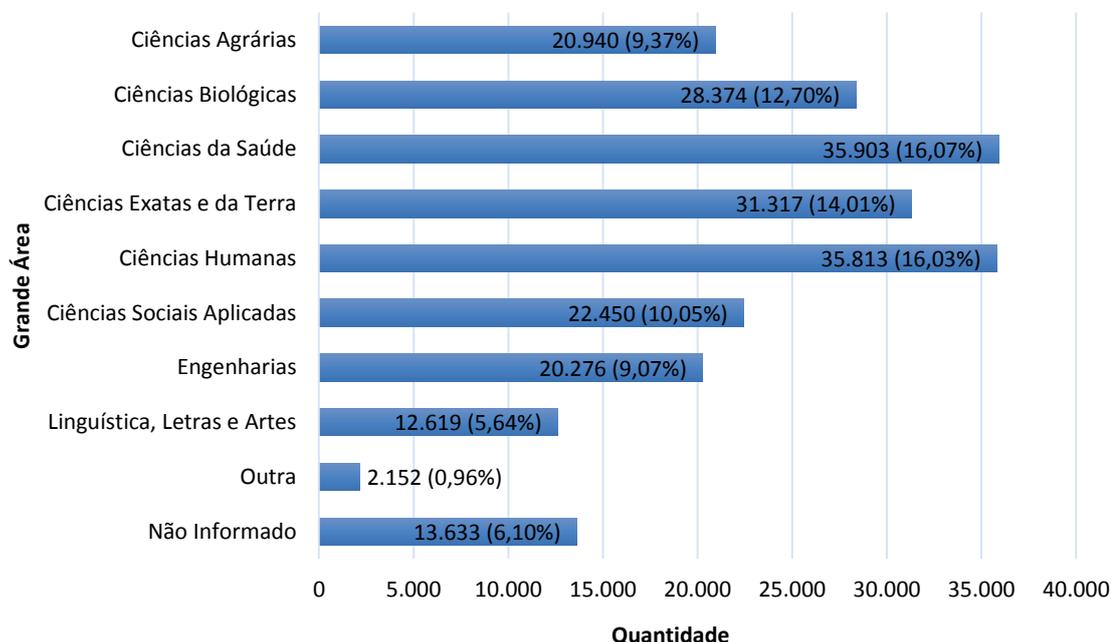


Figura 6.22: Distribuição dos doutores por grande área de atuação.

Como pode ser observado na Figura 6.22, as grandes áreas com o maior número de doutores são Ciências da Saúde e Ciências Humanas, respectivamente, com quantidades muito próximas. Especificamente no caso da grande área de Ciências Sociais Aplicadas, que é a segunda grande área com maior número de currículos na Plataforma Lattes (ver Figura 5.8), esta surge como a quinta grande área com o maior número de doutores, mostrando não haver qualquer correlação entre a quantidade de indivíduos e a quantidade de doutores dentro de uma mesma grande área. Além disso, apenas cerca de 6% dos doutores não informaram sua grande área em seus currículos, mostrando que a grande maioria dos doutores teve o cuidado de especificar suas áreas de atuação, possibilitando análises mais confiáveis. Tendo em vista a pouca quantidade de currículos com a opção Outra, optou-se por desconsiderá-la para as análises específicas aqui apresentadas.

Conforme mencionado no Capítulo 5, além das grandes áreas, cada currículo ainda pode indicar as áreas, subáreas e especialidades do respectivo indivíduo. A exemplo das análises anteriores, no caso de mais de uma área de atuação em um currículo, foi considerada a primeira como sendo a sua principal. As análises da distribuição por áreas possibilitam compreender como estão distribuídos os pesquisadores em uma determinada grande área (ver Figura 6.23).

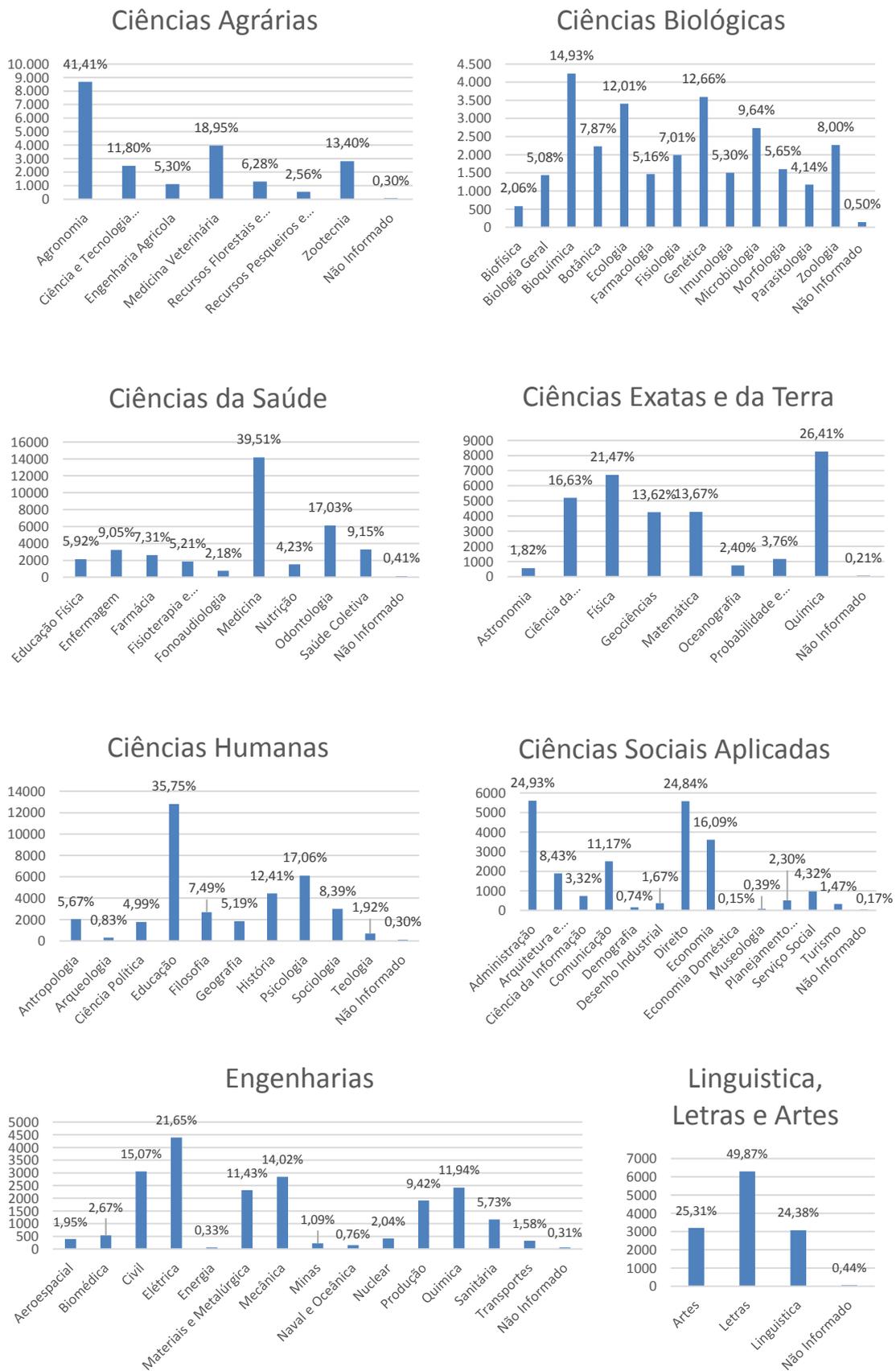


Figura 6.23: Distribuição dos doutores por área de atuação.

Observa-se que não há uma distribuição uniforme do número de doutores de cada área em todas as grandes áreas. Algumas dessas grandes áreas possuem uma quantidade maior de áreas, como Ciências Biológicas e Engenharias, o que faz com que a distribuição entre tais áreas seja um pouco mais homogênea, não destacando-se uma determinada área com percentuais muito elevados de pesquisadores. Por outro lado, determinadas áreas apresentam percentuais muito superiores se comparados com os de outras dentro de sua própria grande área, como Letras, por exemplo, que possui 49,87% dos doutores da grande área de Linguística, Letras e Artes. Destacam-se, ainda, as áreas de Agronomia, Medicina e Educação, com percentuais bem mais elevados do que os de outras áreas de sua grande área.

Ao se comparar a distribuição de todos os indivíduos pelas suas grandes áreas e com a distribuição dos doutores, apenas as grandes áreas de Ciências Biológicas e Ciências Exatas e da Terra divergem nas suas maiores áreas. Inicialmente, foi observado que, na grande área de Ciências Biológicas, Biologia Geral possui a maior quantidade de indivíduos, porém, Bioquímica possui o maior número de doutores. Já na grande área de Ciências Exatas e da Terra, a área de Ciência da Computação possui o maior número de indivíduos, porém, Química é a área detentora da maior quantidade de doutores. Tais dados indicam uma grande quantidade de indivíduos nessas áreas que estão se capacitando ou que poderão se capacitar em níveis mais elevados de formação, contribuindo significativamente para o desenvolvimento delas. Assim, em razão dessas diferenças, para análises comparativas entre as áreas de cada uma das grandes áreas, considerou-se a área mais relevante como aquela com a maior quantidade de doutores (ver Tabela 6.4).

Tabela 6.4: Áreas com a maior quantidade de doutores.

Grande Área	Área	Total de Doutores	% da Grande Área	% de Doutores	% na Plataforma Lattes
Ciências Agrárias	Agronomia	8.672	41,41%	3,88%	0,21%
Ciências Biológicas	Bioquímica	4.236	14,93%	1,90%	0,10%
Ciências da Saúde	Medicina	14.186	39,51%	6,35%	0,34%
Ciências Exatas	Química	8.270	26,41%	3,70%	0,20%
Ciências Humanas	Educação	12.804	35,75%	5,73%	0,31%
Ciências Sociais	Administração	5.596	24,93%	2,50%	0,13%
Engenharias	Engenharia Elétrica	4.390	21,65%	1,96%	0,11%
Linguística, Letras e Artes	Letras	6.293	49,87%	2,82%	0,15%

Como pode ser observado, há uma grande variação nos percentuais de cada grande área, tendo em vista a distribuição irregular na quantidade de áreas de cada uma delas. Destaca-se a área de Letras com praticamente metade dos doutores de sua grande área e apenas 2,82% do conjunto analisado, a exemplo da área de Agronomia com apenas 3,88% de todo o conjunto de doutores, mas com 41,41% de sua grande área, composta por outras seis áreas. Bioquímica é a área com o menor número de

doutores dentre todas as áreas selecionadas. Isto porque a grande área de Ciências Biológicas possui treze áreas, sendo aquela que possui a distribuição mais uniforme se comparada com todas as demais. Situação semelhante ocorre na área de Engenharia Elétrica, que possui apenas 4.390 doutores, mas que está inserida na grande área de Engenharias, que possui a maior quantidade de áreas (14). Importante ainda destacar a Medicina como a área que possui a maior quantidade de doutores, englobando 6,35% de todo conjunto analisado. Ao se analisar a produção científica de cada uma dessas áreas, é possível identificar qual a tendência de publicação de cada uma delas e, ainda, quais são as que possuem maiores taxas de produção (ver Figuras 6.24 e 6.25).

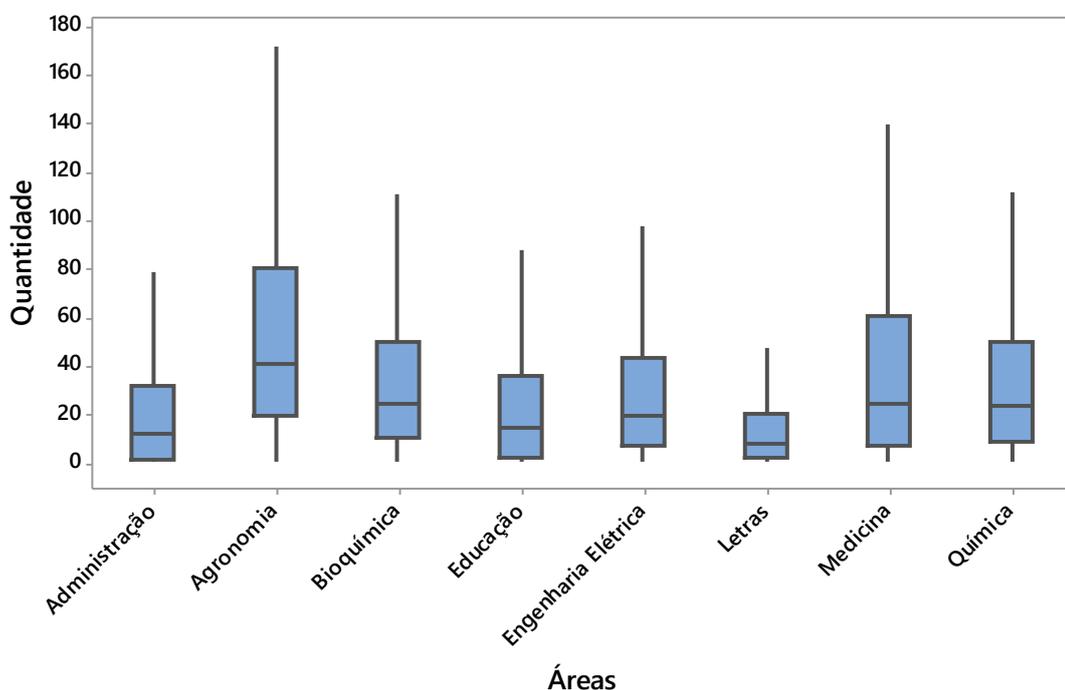


Figura 6.24: Taxa de publicação de artigos em anais de congresso por área.

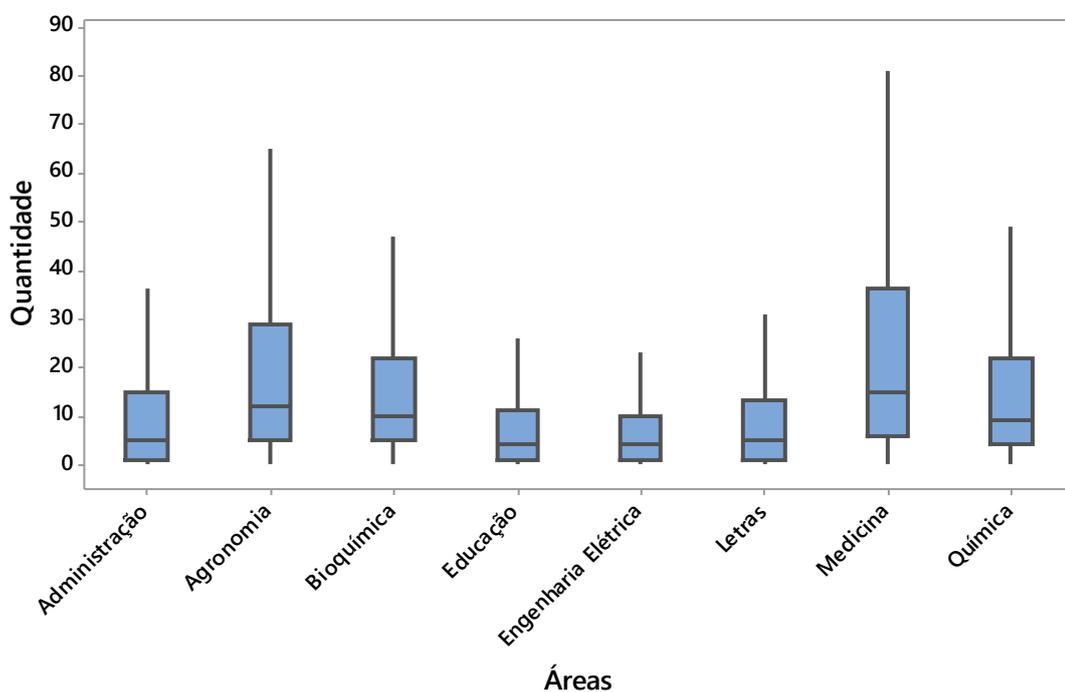


Figura 6.25: Taxa de publicação de artigos em periódico por área.

O perfil de publicação de cada uma das áreas é distinto e, a exemplo do que já foi apresentado para o conjunto de todos os doutores, a média de publicação de artigos em anais de congresso é de modo geral bem superior à de artigos em periódico. Ao se analisar a produção de artigos em anais de congresso, a área de Agronomia se destaca, com taxas que variam de 20 a 75 artigos por doutor, taxas estas bem superiores às das demais áreas. Posteriormente, as áreas de Bioquímica, Medicina e Química se equivalem com medianas muito próximas, enquanto a área de Medicina se destaca com taxas mais elevadas. Importante destacar a área de Letras, que possui o menor valor de mediana (8), indicando uma tendência de baixa publicação em conferências. Como esta área representa aproximadamente 50% da produção de sua grande área, ela influencia diretamente o perfil de pouca produção nesse tipo de publicação na grande área de Linguística, Letras e Artes.

Nas análises dos artigos em periódico, a área de Agronomia, que tinha uma taxa de publicação em anais de congresso muito elevada, fica abaixo da área de Medicina, que possui a maior taxa de publicação dentre todas as áreas. As áreas de Bioquímica e Química, a exemplo do que acontece com a quantidade de artigos em anais de congresso, também possuem valores muito aproximados. Além disso, também se equivalem as áreas de Administração e Letras, bem como as áreas de Educação e Engenharia Elétrica, detentoras das menores taxas de publicação. Percebe-se que, apesar de a área de Medicina possuir a maior mediana na produção de artigos em periódico, a diferença para a área de Agronomia é pequena, se comparada com a diferença de suas medianas na publicação de artigos em anais de congresso. Já as áreas de Bioquímica e Química se equivalem em ambos os tipos de publicação, com boas taxas de publicação, mostrando que são áreas com perfis de produção científica muito próximos. Apesar de outras áreas possuírem taxas semelhantes de produção de artigos em periódico, destacam-se as áreas de Engenharia Elétrica, com taxa bem superior às demais na produção de artigos em anais de congresso, e de Letras, com valores muito baixos para ambos os tipos de publicação. Já as áreas de Administração e Educação são aquelas que possuem padrões muito próximos de publicação.

Para compreender o perfil de cada área, a Figura 6.26 apresenta a distribuição de todos os tipos de publicação em cada uma das áreas. Observa-se uma distribuição distinta entre os tipos de publicação em cada uma das áreas analisadas. Em todas elas, o tipo mais representativo é o de artigos em anais de congresso, seguido em menor quantidade dos artigos em periódico. Para os demais tipos de produção, as áreas de Letras e Educação destacam-se por possuírem percentuais consideráveis de suas produções relacionados a capítulos de livro e também a livros. Já em relação a produção de textos em jornais e revistas que possui percentuais baixos em diversas áreas, se destacam as áreas de Letras e Administração como as mais representativas respectivamente. A área de Administração também se destaca na produção de trabalhos técnicos que constituem aproximadamente 20% da produção dessa área.

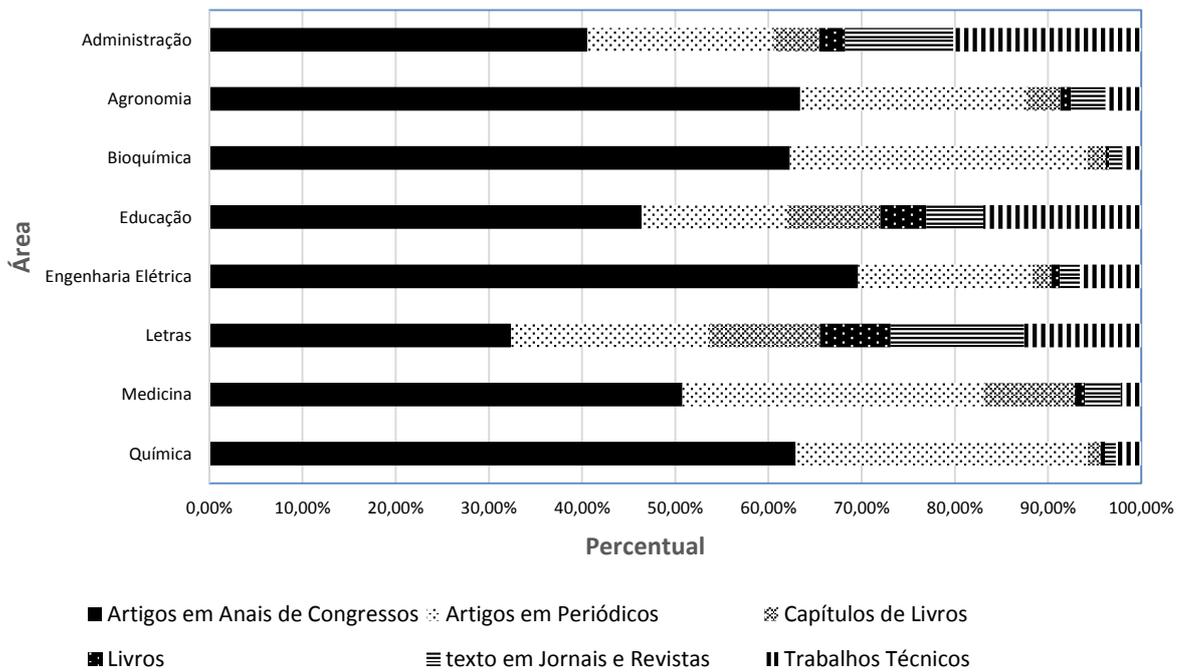


Figura 6.26: Distribuição dos tipos de publicação por área.

Para uma análise comparativa com o objetivo de melhor compreender a evolução da produção científica de cada uma dessas áreas e, ainda, para verificar o impacto dessas áreas em suas grandes áreas, as Figuras 6.27 e 6.28 apresentam o histórico com a evolução da produção científica tanto para artigos em anais de congresso como também para artigos em periódico em cada grande área.

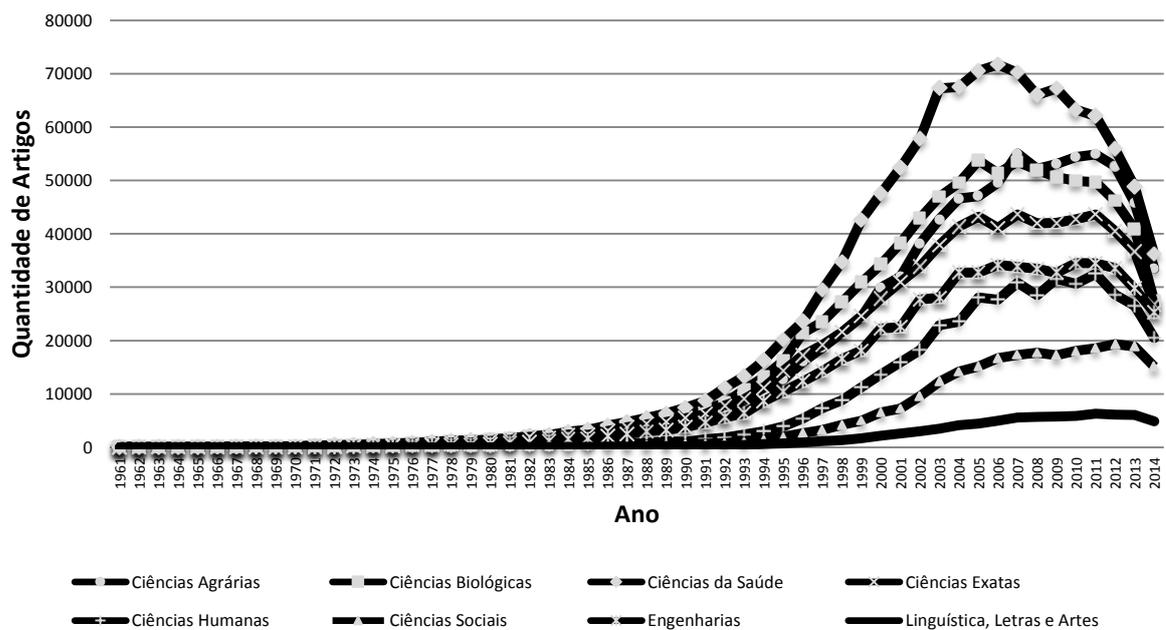


Figura 6.27: Quantidade de artigos em anais de congresso por grande área.

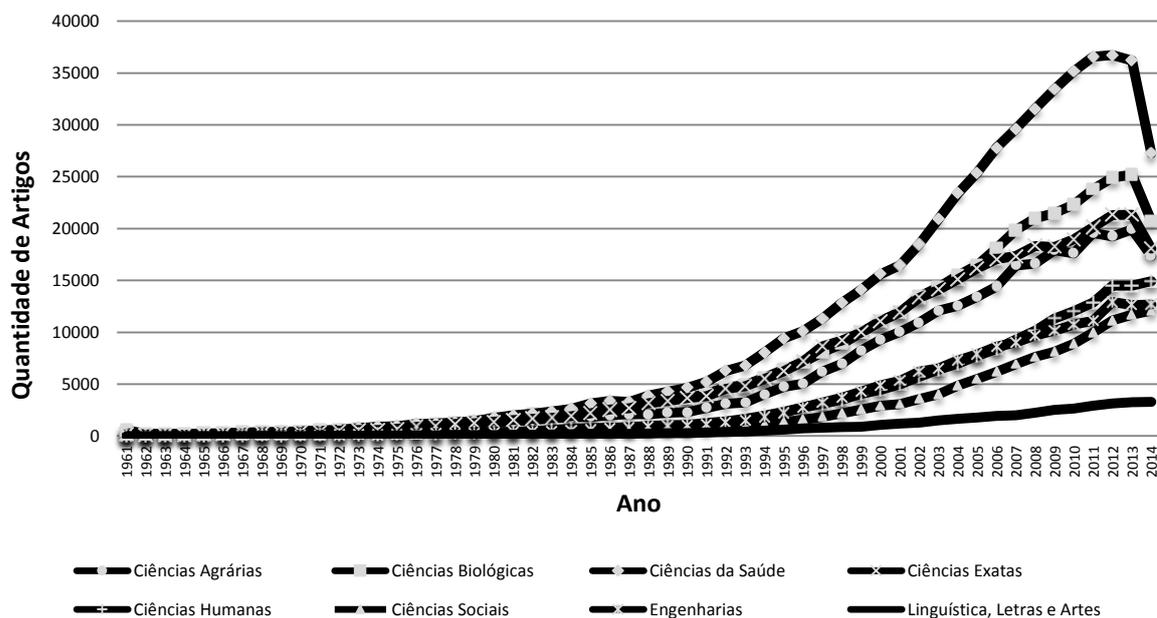


Figura 6.28: Quantidade de artigos em periódico por grande área.

Como pode ser observado, todas as grandes áreas apresentam queda na publicação de artigos em anais de congresso nos últimos anos, sendo que, algumas delas, como Ciências da Saúde, Ciências Biológicas e Ciências Agrárias, que estão entre as mais representativas, possuem queda bem mais acentuada do que as demais. Nos últimos oito anos, a grande área de Ciências da Saúde teve uma redução de aproximadamente 49%, Ciências Biológicas de 46,4% e Ciências Agrárias de 32,3%. Além disso, percebe-se que, na grande área de Ciências da Saúde, que possui a maior quantidade de trabalhos publicados e a maior quantidade de doutores dentre todas as grandes áreas, sua queda tem início antes das demais (2007). No período de maior redução no número de publicações, compreendido entre 2011 a 2014, houve uma redução de 123.397 artigos em anais de congresso, sendo que somente essas três grandes áreas foram responsáveis por uma redução de 69.581 publicações, aproximadamente 56% do total. Portanto, essas três grandes áreas aparecem como grandes responsáveis pela considerável queda na produção de artigos em anais de congresso.

Já no caso de artigos em periódico, as grandes áreas apresentam comportamento distinto se comparadas em relação aos artigos em anais de congresso. Somente as quatro grandes áreas mais representativas tiveram queda em sua produção. Dessas, a grande área de Ciências da Saúde é a que possui a maior queda, queda esta iniciada em 2012 de forma muito acentuada. Considerando a queda na taxa de publicação de todo o conjunto de doutores entre os anos de 2012 e 2014, que corresponde a 18.668 artigos, somente a grande área de Ciências da Saúde é responsável pela redução de 9.457 artigos, ou seja, aproximadamente, 50% do total. As outras três grandes áreas com os maiores índices de queda são Ciências Biológicas (22,3%), Ciências Exatas e da Terra (17,4%) e Ciências Agrárias (10,1%). Em todas as outras grandes áreas, a quantidade de publicações se manteve, em alguns casos com tímido crescimento. Apesar do crescimento da produção identificado nessas outras grandes áreas, este é pequeno, fazendo com que a análise geral de todo o conjunto tenha como resultado uma redução considerável da produção científica nos dois últimos anos. Neste

contexto, a análise das principais áreas que compõem essas grandes áreas pode fornecer indícios sobre o impacto delas no comportamento das grandes áreas a que estão vinculadas (ver Figuras 6.29 e 6.30).

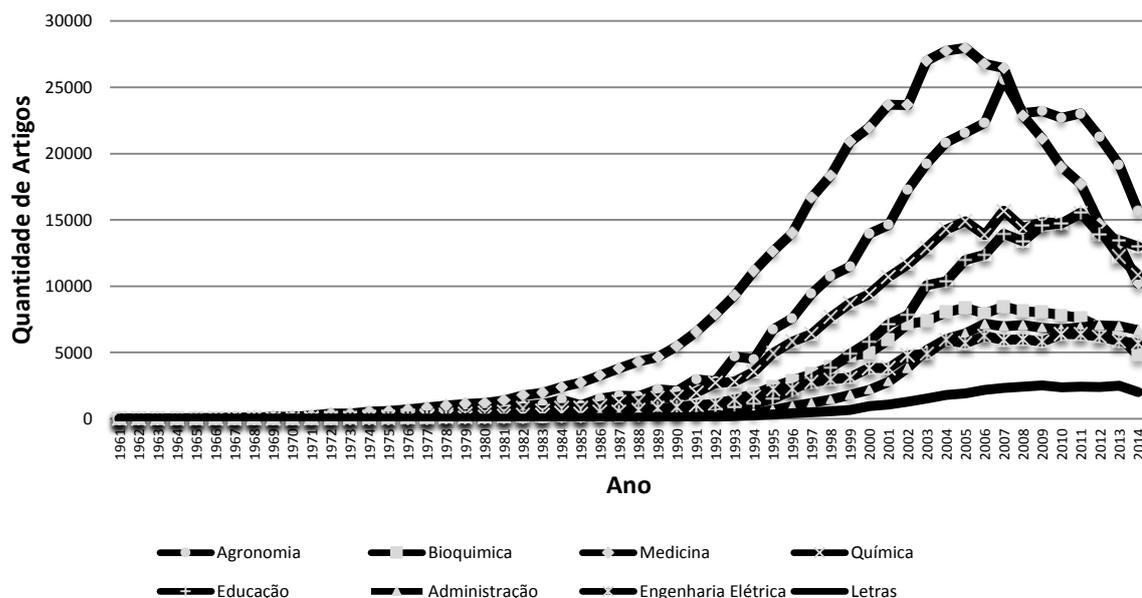


Figura 6.29: Quantidade de artigos em anais de congresso por áreas.

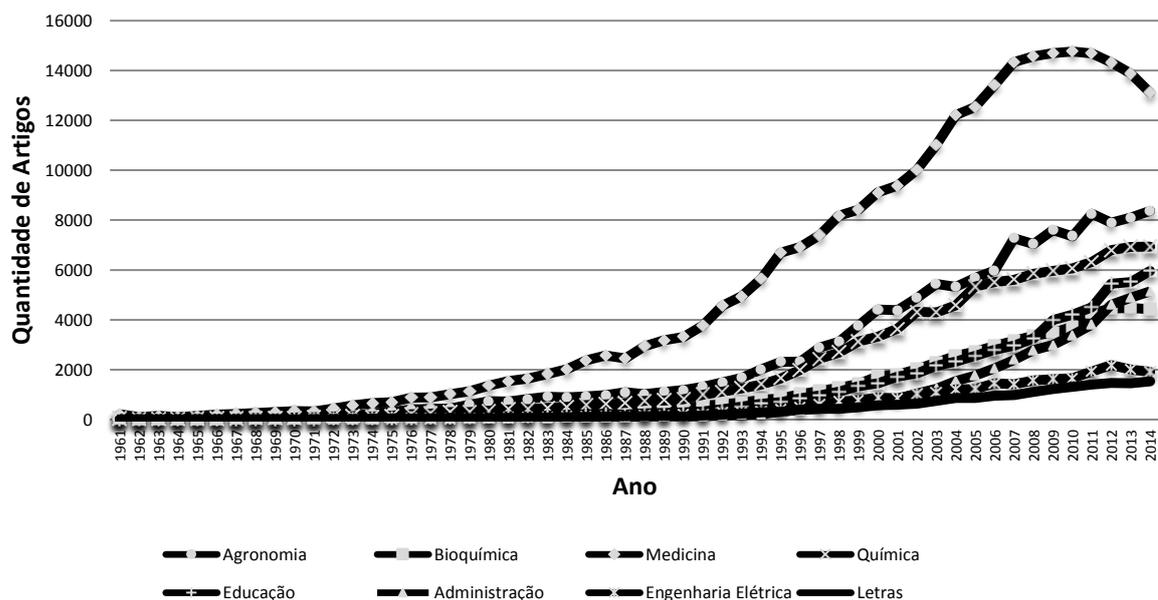


Figura 6.30: Quantidade de artigos em periódico por áreas.

Na análise das áreas em relação à publicação de artigos em anais de congresso, observa-se que a área de Medicina é a mais produtiva e que possui a maior quantidade de doutores dentre todas as outras, sendo detentora de, aproximadamente, 40% dos doutores da grande área de Ciências da Saúde e tendo um comportamento semelhante ao de sua grande área, como era de se esperar. A área de Medicina apresenta uma queda no número de artigos em anais de congresso de,

aproximadamente, 93%, considerando as taxas de queda acumuladas ano a ano, ao longo de 2005 a 2014. Somente nos últimos três anos, sua produção reduziu em, aproximadamente, 50%. Similarmente, a área de Agronomia, que tem uma produção elevada neste tipo de publicação e representa 41,41% de sua grande área, também apresenta uma queda acentuada a partir de 2011. Todas as demais áreas apresentam queda, principalmente a partir de 2011, mas com menor intensidade do que as áreas de Medicina e Agronomia. Logo, como seus percentuais de participação são bem maiores, pode-se afirmar que possuem impacto direto na queda da produção de artigos em anais de congresso de suas respectivas grandes áreas. Entretanto, vale ressaltar que algumas outras áreas vinculadas às mesmas grandes áreas possuem índices de queda superiores, mas com menor impacto (ver Apêndice Tabela A.1).

Com relação aos artigos em periódico, a maioria das áreas tem crescimento tímido ou se mantêm com valores muito próximos ao longo dos anos. Apenas as áreas de Engenharia Elétrica e Medicina apresentam queda, com destaque para Medicina, com queda bem mais acentuada. Diante disso, pode-se afirmar que a área de Medicina, apesar da queda de sua produção científica nos últimos dois anos, influencia de forma tímida a produção de sua grande área neste tipo de publicação. Importante ainda destacar a queda na produção das grandes áreas de Ciências Biológicas, Ciências Exatas e da Terra e Ciências Agrárias, queda não apresentada em suas respectivas áreas, Bioquímica, Química e Agronomia, indicando que outras áreas que compõem essas grandes áreas têm contribuído de forma mais impactante na queda da produção de artigos em periódico (ver Apêndice Tabela A.2).

Diante do exposto, pode-se afirmar que a queda na produção científica não é uniforme em todas as grandes áreas e áreas do conhecimento. Logo, para melhor visualização do impacto de cada área em suas grandes áreas e, a influência dessas áreas na produção científica brasileira de forma geral, o Apêndice apresenta o histórico de publicações de cada área do conhecimento (Figuras A.1 a A.16).

Como pôde ser observado, o comportamento de cada área dentro de sua grande área não é uniforme e várias áreas apresentam taxas de crescimento muito distintas. Tratando-se dos artigos em anais de congresso, praticamente todas as áreas tiveram queda nos últimos anos, mas com uma variação muito grande em relação ao início das quedas. Com relação aos artigos em periódico, o que se observa é que determinadas áreas não apresentam queda e, que em geral, as maiores áreas, ou seja, aquelas mais representativas em cada grande área, não apresentam redução na sua produção científica. Para uma visão geral dos índices de crescimento de ambos os tipos de publicação nos últimos dez anos, as Tabelas A.1 e A.2 do Apêndice apresentam a produção de todas as áreas.

Nos resultados apresentados nas Tabelas A.1 e A.2, para cada ano foram destacados os maiores índices de crescimento e de queda, quando comparados com o ano anterior. De forma geral, os maiores índices positivos são identificados após um ano com grande queda em uma determinada área. Ou seja, após uma queda considerável em um determinado ano de uma área, no outro ano suas taxas de publicação voltam a crescer. Por isto, quando comparadas com o ano anterior, essas áreas passam a ter uma alta taxa de crescimento. Esta variação que, em geral, acontece em áreas com

menor quantidade de doutores, é de difícil explicação, já que vários fatores podem contribuir para tal fenômeno.

No entanto, é importante destacar que tendo em vista a análise temporal da produção científica em cada uma das áreas, e que engloba todo o período compreendido entre 2005 e 2014, esta traz resultados importantes para compreensão da evolução da produção científica brasileira. Para os artigos em anais de congresso, o fato mais relevante são as reduções na quantidade de publicações em várias áreas do conhecimento, destacando-se a queda na área de Medicina (-92,86%), área esta importante dada a sua quantidade de doutores e o volume de publicações, e o crescimento na área de Ciência da Computação (22,14%), o maior dentre todas as demais. Tal constatação vem corroborar com a hipótese anteriormente identificada sobre o impacto da avaliação da CAPES que passa a considerar apenas artigos em periódico em suas avaliações a partir do triênio 2007-2009. Este período coincide com o início da queda na quantidade de publicações, o que pode ter incentivado autores a publicarem os resultados de suas pesquisas em periódicos, em detrimento às conferências, fato este, que não acontece apenas com a área de Ciência da Computação, que continua a considerar os artigos em anais de congresso em suas avaliações. Consequentemente, a área de Ciência da Computação manteve um crescimento constante em suas taxas de publicação, sem queda em nenhum ano analisado, e com crescimento acumulado de 22,14% no período. Já com relação aos artigos em periódico, observa-se que não há um padrão como o que ocorre com os artigos em anais de congresso, sendo que algumas grandes áreas se destacam pela quantidade de áreas com taxas negativas nos últimos dois anos como as de Ciências Biológicas e Ciências da Saúde, que englobam áreas importantes e produtivas, mas apresentam quedas consideráveis em sua produção científica nesses anos.

6.5 Considerações Finais

Após analisar a produção científica dos doutores, que são responsáveis pela maioria das publicações cadastrados na Plataforma Lattes, foi possível apresentar uma caracterização do perfil de publicação e atuação desses pesquisadores. Na análise da produção científica brasileira, foi identificado que, nos últimos anos, a quantidade de trabalhos publicados tem sofrido uma considerável queda dentre os vários tipos de publicação analisados, em especial, artigos em anais de congresso e artigos em periódico. Várias hipóteses foram consideradas para explicar tal comportamento, no entanto, todas de difícil validação.

Ainda foi possível verificar como a colaboração científica do conjunto de doutores evoluiu ao longo dos anos. Para isso, foram considerados quinquênios desde a primeira publicação identificada até os dias atuais. Foi possível observar como a área de Medicina caracterizou-se como uma das principais a realizar os primeiros trabalhos em colaboração e, ainda, como grandes áreas intensificaram seus trabalhos em colaboração em cada período analisado. Além disso, verificando as arestas mais densas, foi identificado como pequenos grupos têm trabalhado em colaboração de forma intensa, aumentando de forma impactante o número de publicações de seus pesquisadores, tendo em vista a participação conjunta dos mesmos coautores, geralmente das mesmas áreas de atuação e das mesmas instituições.

Adicionalmente, nas análises das principais áreas do conhecimento, foi verificado que a área de Agronomia possui a maior quantidade de publicações em sua grande área e, que, como a área de Medicina, que também é a área mais representativa em sua grande área, sofreu uma redução muito considerável na publicação de artigos em anais de congresso, o que tem contribuído para a redução da produção de suas grandes áreas. Vale destacar ainda que outras áreas menos representativas também vêm apresentando índices de queda consideráveis. Além disso, foi possível constatar que a área de Ciência da Computação, diferentemente das demais áreas, possui um crescimento constante no número de artigos em anais de congresso, sendo a única área que ainda considera este tipo de publicação nas avaliações da CAPES após o triênio 2007-2009. Observa-se, ainda, a tendência de algumas áreas em valorizar outros tipos de publicação, como, por exemplo, livros e trabalhos técnicos, contribuindo para a queda no número de artigos em anais de congresso.

No Capítulo 7, é analisada a produção científica do conjunto de docentes dos programas de pós-graduação e dos bolsistas de produtividade em pesquisa do CNPq. Essa análise visa apresentar como os responsáveis pela formação de novos pesquisadores têm produzido ao longo dos anos e, ainda, como os pesquisadores de excelência do Brasil têm realizado suas pesquisas considerando para isso fatores qualitativos de suas publicações.

Capítulo 7

Análise da Produção Científica dos Docentes de Pós-Graduação e dos Bolsistas de Produtividade em Pesquisa do CNPq

Neste capítulo, é apresentada uma análise do conjunto de docentes que atuam nos Programas de Pós-Graduação *Strictu Sensu* reconhecidos pela CAPES e, também, do conjunto de bolsistas de Produtividade em Pesquisa do CNPq. São apresentados dados sobre a caracterização desses conjuntos, como sua distribuição geográfica, instituições mais representativas a que estão vinculados, orientações, bem como, suas grandes áreas e áreas de atuação (Seção 7.1). Posteriormente, é realizada uma análise quantitativa de sua produção científica, que visa demonstrar como suas pesquisas têm sido realizadas (Seção 7.2). Após, são analisados os artigos publicados em periódico pelos bolsistas de produtividade em pesquisa, buscando identificar a qualidade das publicações. Para isso, são utilizados dados referentes ao número de citações das publicações, fator de impacto dos periódicos e nível das bolsas dos pesquisadores (Seção 7.3). Finalmente são apresentadas as considerações finais (Seção 7.4).

7.1 Caracterização Geral dos Conjuntos

Para as análises apresentadas neste capítulo, um conjunto contendo 63.300 docentes que atuam em programas de pós-graduação *strictu sensu* reconhecidos pela CAPES foi selecionado. Além disso, um segundo conjunto, contendo 14.394 bolsistas de Produtividade em Pesquisa do CNPq, também foi considerado, sendo que, do total de bolsistas, 98,5% também são docentes em programas de pós-graduação, e conseqüentemente, também estão presentes no conjunto de docentes. A escolha desses conjuntos visa possibilitar obter uma visão sobre o perfil dos docentes, que, atualmente, são responsáveis pela formação dos mestres e doutores no Brasil, bem

como, comparar a produção científica desses docentes com a dos principais pesquisadores brasileiros que recebem auxílio pela excelência de suas pesquisas em suas áreas de atuação.

7.1.1 Docentes de Pós-Graduação

O primeiro conjunto foi selecionado a partir da relação dos docentes de programas de pós-graduação, obtida a partir dos dados do *Coleta CAPES 2013*¹. Apesar de tal listagem ser referente aos docentes que estavam atuando nos programas no final de 2013, todos os dados a serem analisados foram extraídos dos currículos coletados em maio de 2015.

Diferentemente da análise realizada com o conjunto de doutores, em que é possível identificar a data de conclusão do doutorado e, dessa forma, considerar a produção somente após essa data, as informações sobre quando um determinado indivíduo iniciou sua atuação como docente em um determinado programa de pós-graduação nem sempre é possível de ser identificada nos currículos da Plataforma Lattes. O mesmo ocorre em relação à falta de informação nos currículos sobre o período de início da bolsa de produtividade em pesquisa de um determinado pesquisador. Diante disso, as análises aqui apresentadas irão considerar todo o histórico de produção científica registrado nos currículos dos indivíduos considerados.

Atualmente, não existe um critério único para que um determinado docente possa ser credenciado em cursos de pós-graduação *strictu sensu*, já que cada programa define suas regras, mas, em geral, exige-se o título de doutor. Apesar de o conjunto analisado considerar somente os docentes que estão vinculados a programas brasileiros, foi possível identificar que alguns dos docentes possuem endereço profissional em outros países, sendo os principais, Estados Unidos, Portugal, Chile, Colômbia, Inglaterra, França e Espanha. Observou-se que a grande maioria destes docentes é estrangeira e que tem colaborado com os programas brasileiros, fruto de parcerias internacionais em que estes têm atuado como professores colaboradores. Outras situações encontradas, mas em menor quantidade, são de docentes que, mesmo estando vinculados a instituições brasileiras, registraram em seus currículos da Plataforma Lattes o endereço profissional de onde estão se capacitando, realizando, por exemplo, pós-doutorado.

Outra informação importante para compreensão do perfil deste conjunto de docentes é com relação às instituições as quais eles estão vinculados. Tal informação possibilita identificar onde estão alocados os docentes responsáveis pela formação dos níveis mais altos de capacitação e, conseqüentemente, como é a distribuição geográfica deles. A Tabela 7.1 apresenta os estados e as instituições mais representativas.

¹ <http://www.capes.gov.br/avaliacao/plataforma-sucupira/modulo-coleta-de-dados>

Tabela 7.1: Distribuição dos docentes de pós-graduação por estados e nas principais instituições.

Posição	Estado	% de docentes	Posição	Instituição	% de docentes
1º	SP	25,37	1º	USP	8,44
2º	RJ	12,31	2º	UNESP	3,72
3º	MG	9,33	3º	UFRJ	3,47
4º	RS	8,08	4º	UNICAMP	2,98
5º	PR	6,38	5º	UFMG	2,89
6º	SC	3,52	6º	UFRGS	2,64
7º	BA	3,49	7º	UNB	2,29
8º	PE	3,15	8º	UFSC	2,14
9º	DF	2,75	9º	UFF	1,92
10º	CE	2,35	10º	UFPE	1,86
11º	PB	2,03	11º	UFPR	1,81
12º	GO	1,97	12º	UFBA	1,70
13º	RN	1,89	13º	UERJ	1,61
14º	PA	1,74	14º	UFC	1,48
15º	ES	1,45	15º	UFG	1,48
16º	MS	1,11	16º	UFRN	1,47
17º	MT	1,06	17º	UNIFESP	1,45
18º	AM	1,05	18º	UFPB	1,36
19º	SE	0,94	19º	UFSCAR	1,31
20º	AL	0,71	20º	UFES	1,22
21º	MA	0,69	21º	FIOCRUZ	1,21
22º	PI	0,64	22º	UFSM	1,19
23º	TO	0,25	23º	UFPA	1,13
24º	RR	0,23	24º	UFU	1,08
25º	RO	0,21	25º	UEM	0,95
26º	AC	0,15	26º	UFV	0,93
27º	AP	0,10	27º	UEL	0,87

A exemplo do que ocorre com o conjunto de doutores, os estados de São Paulo, Rio de Janeiro e Minas Gerais possuem a maior quantidade de docentes, resultado este influenciado diretamente por estes estados abrigarem as principais instituições de ensino do país. Estas instituições de excelência possuem uma grande quantidade de cursos de pós-graduação. Dentre as quatro instituições com a maior quantidade de docentes de pós-graduação, três estão localizadas em São Paulo, sendo a USP a instituição com o maior percentual de docentes, cerca de 8,5% do conjunto estão vinculados a ela. Rio de Janeiro se destaca, principalmente, pela quantidade de docentes vinculados à UFRJ, a exemplo de Minas Gerais, com a UFMG. Destaca-se, ainda, a Fiocruz, que se caracteriza por ser uma instituição de ciência e tecnologia em saúde, referência em pesquisas na área de Saúde Pública, e que também possui programas de pós-graduação, justificando a quantidade de docentes a ela vinculados. Por fim, a região Norte, que possui a menor quantidade de indivíduos e doutores cadastrados na Plataforma Lattes, também se caracteriza como a região com o menor número de docentes atuando na pós-graduação. Com exceção do Amazonas, que surge na 18ª posição, com um total de 663 docentes, destacam-se, negativamente, os estados do Acre e Amapá, com apenas algumas dezenas de docentes.

Os atuais docentes dos programas de pós-graduação *strictu sensu* são responsáveis pela maioria das orientações concluídas nos níveis mais altos de formação identificada em todo o conjunto de currículos da Plataforma Lattes (Tabela 7.2).

Tabela 7.2: Quantitativo de orientações concluídas identificadas em todos os currículos da Plataforma Lattes e as orientadas pelo conjunto de docentes.

Tipo de Orientação	Geral	Docentes
Orientações de Doutorado	224.303	200.261 (89,28%)
Orientações de Mestrado	737.572	653.620 (88,61%)
Supervisões de Pós-Doutorado	24.662	23.732 (96,22%)

Percebe-se que nem todas as orientações foram realizadas pelo conjunto de docentes. Aproximadamente 11% dos doutores e mestres foram formados fora do país ou orientados por docentes que, atualmente, não estão vinculados aos programas de pós-graduação. O mesmo ocorre com aproximadamente 3% das supervisões de pós-doutorado. Ao se verificar as orientações que não foram realizadas sob supervisão dos docentes que compõem o conjunto considerado, foi possível observar, que em geral, são de doutores que já foram orientadores em programas de pós-graduação e que, por algum motivo, não estão mais atuando nos respectivos programas, já que a relação utilizada considera apenas quem estava atuando em algum programa no final do ano de 2013. Mesmo sendo orientadores dos níveis mais altos de formação, alguns docentes também têm realizado orientações de outras naturezas (ver Tabela 7.3).

Tabela 7.3: Quantitativo de orientações em andamento identificadas em todos os currículos da Plataforma Lattes e as orientadas pelo conjunto de docentes.

Tipo de Orientação	Geral	Docentes
Supervisão de Pós-Doutorado	7.688	7.445 (96,97%)
Orientações de Doutorado	101.960	101.221 (99,28%)
Orientações de Mestrado	160.475	139.653 (87,02%)
Orientações de Especialização	218.381	5.492 (2,51%)
Orientações de Graduação	1.464.544	22.386 (1,52%)

Diferentemente do esperado, nem todas as orientações em andamento, identificadas nos currículos da Plataforma Lattes, são conduzidas pelos atuais docentes. Um problema identificado para tal situação está relacionado à falta de atualização de alguns currículos, onde são informados cursos em andamento cujos orientadores já não estão mais vinculados aos atuais programas de pós-graduação. Percebe-se ainda, que apenas um pequeno número desses docentes de pós-graduação tem realizado orientações nos níveis de Especialização e Graduação.

Importante destacar que é de se esperar que todos os docentes que atuam na pós-graduação tenham o título de doutor. No entanto, existe um pequeno percentual deste conjunto que não possui doutorado informado em seus currículos. Dentre todos os docentes, um total de 562 não possui tal titulação. Ao verificar a formação desses docentes, foi possível identificar que os seus currículos são, em sua maioria, currículos que não incluem informação sobre os níveis de formação do indivíduo ou que incluem a formação “Livres Docência”.

A Figura 7.1 apresenta a distribuição dos docentes nas grandes áreas do conhecimento, o que possibilita identificar aquelas que possuem maior potencial para

formação de pesquisadores. Ressalta-se que tal distribuição levou em consideração a principal grande área informada no currículo de cada docente e não a grande área na qual o programa em que ele está vinculado se enquadra. Observa-se que a grande área de Ciências Humanas possui a maior quantidade de docentes, quantidade esta pouco superior à das grandes áreas de Ciências Exatas e da Terra e Ciências da Saúde. A grande área de Ciências da Saúde possui a maior quantidade de doutores, dos quais cerca de 27% estão atuando como docentes em programas de pós-graduação, sendo a terceira grande área mais representativa desse conjunto de docentes. Tal situação difere do que acontece na grande área de Ciências Humanas, que possui o 4º maior percentual de doutores, dos quais 31,33% são docentes. Esse percentual, é ligeiramente menor que o das grandes áreas de Engenharias, Linguística, Letras e Artes, e Ciências Exatas e da Terra, todas com aproximadamente 32% dos seus doutores atuando como docentes em programas de pós-graduação.

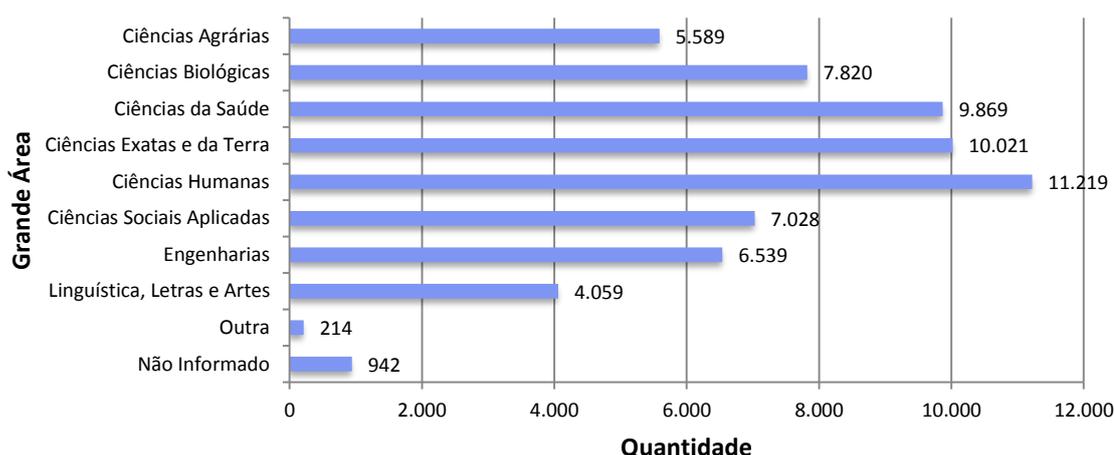


Figura 7.1: Distribuição dos docentes pelas suas grandes áreas de atuação.

Além das grandes áreas, também é possível verificar as principais áreas de atuação dos docentes (ver Tabela 7.4). Como pode ser observado, a área de Medicina é detentora da maior quantidade de docentes, sendo também responsável pela maioria dos doutores com currículos cadastrados na Plataforma Lattes. Ao se considerar as quatro primeiras áreas com a maior quantidade de docentes, estas também são as principais áreas de suas respectivas grandes áreas com relação à quantidade total de doutores. De todas as grandes áreas, apenas as de Engenharias e Ciências Biológicas não possuem representantes dentre as dez áreas com a maior quantidade de docentes.

Tabela 7.4: Áreas com a maior quantidade de docentes.

Posição	Área	Quantidade
1º	Medicina	3.680
2º	Educação	3.588
3º	Química	2.388
4º	Agronomia	2.077
5º	Física	2.045
6º	Ciência da Computação	1.873
7º	Letras	1.851
8º	Direito	1.701
9º	Odontologia	1.660
10º	Administração	1.644

É importante destacar a área de Ciência da Computação, que, apesar de ser apenas a terceira em quantidade de doutores na grande área de Ciências Exatas e da Terra, possui a sexta maior quantidade de docentes, à frente de áreas consolidadas como Letras, Direito e Administração. Conforme já apresentado no Capítulo 5, ela surge como uma das áreas com a maior quantidade de indivíduos com currículos cadastrados na Plataforma Lattes. Uma possível explicação está relacionada à expansão dos programas de pós-graduação da área e, ainda, à presença de docentes formados em Ciência da Computação e que atuam em disciplinas relacionadas à computação em programas diversos, como os das áreas de Engenharia Elétrica e Interdisciplinar.

Considerando ainda a grande área de atuação dos docentes, é possível identificar a distribuição das orientações por eles concluídas (ver Figura 7.2). Os dados apresentados correspondem à quantidade de orientações concluídas em cada grande área, considerando aquela de atuação do orientador. O total de orientações concluídas deste conjunto corresponde a 877.613, considerando todas as orientações independente do período em que foram realizadas. A grande área de Ciências Humanas, que inclui o maior número de docentes, é também responsável pelo maior número de orientados, cerca de 17% de todas as orientações concluídas. Logo, após, surgem as grandes áreas de Ciências da Saúde com 13,97%, Ciências Exatas e da Terra com 13,34%, e Ciências Biológicas com 13,14%. A grande área de Linguística, Letras e Artes possui o menor número de orientados, apenas 5,65%, percentual próximo à sua representatividade na quantidade total de docentes, cerca de 6% do conjunto.

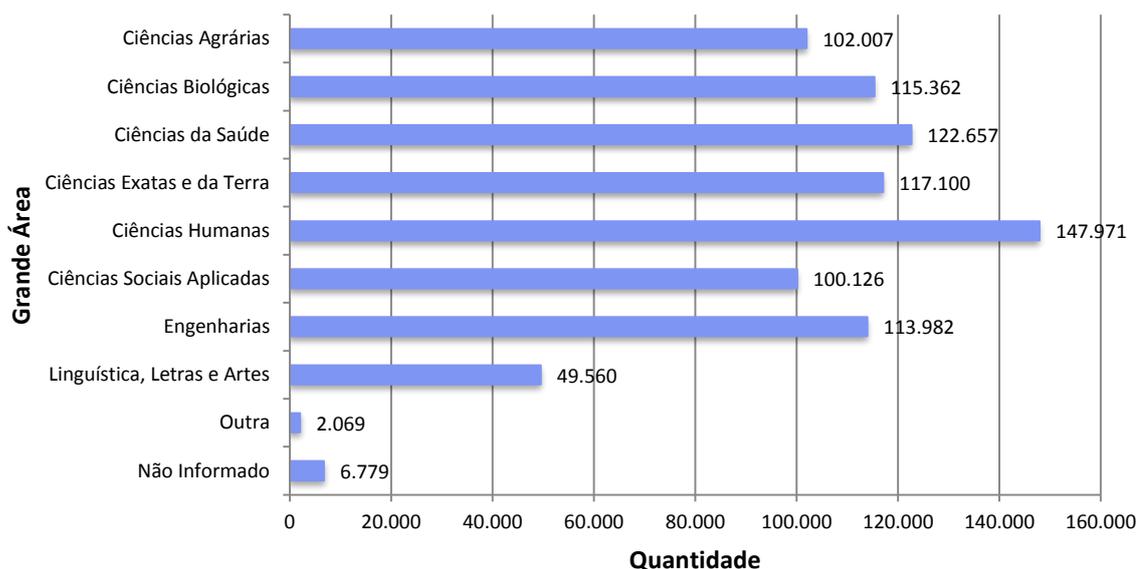


Figura 7.2: Distribuição das orientações concluídas considerando a grande área do orientador.

Diante do exposto, é possível verificar uma grande relação entre o número de orientadores e o número de orientações concluídas em cada grande área. Para ilustrar tal relação, foi calculado o coeficiente de correlação de Spearman para os dois conjuntos e o resultado apresenta uma correlação positiva significativa de 0,90, que corresponde a uma grande similaridade entre as duas amostras (Figura 7.3).

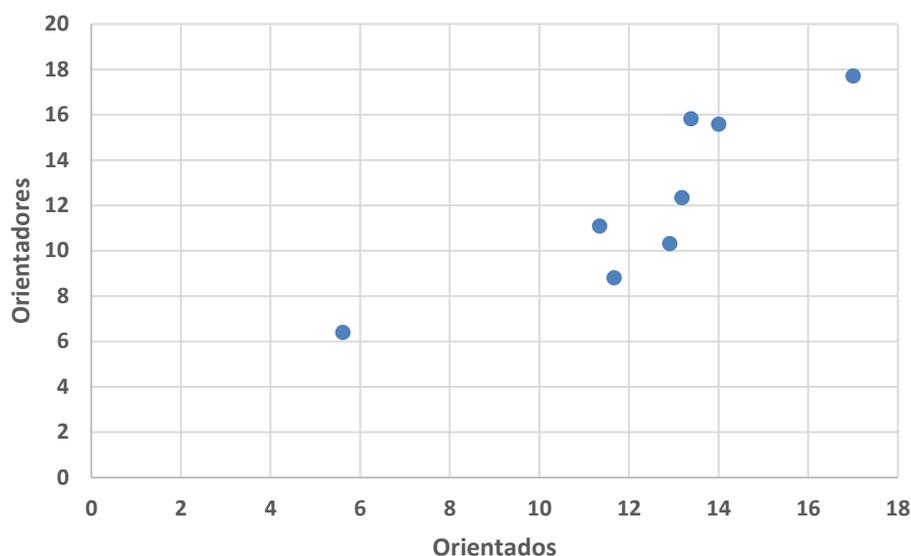


Figura 7.3: Correlação entre número de orientadores e orientações concluídas nas grandes áreas.

Apesar de a relação direta entre a quantidade de docentes e a quantidade de orientações concluídas, foi observada, uma distribuição irregular quando verificado o total de orientações de cada docente. Existem diversos docentes sem orientações concluídas e docentes com um número muito elevado de orientações, conforme mostrado na Tabela 7.5.

Tabela 7.5: Docentes com o maior número de orientações concluídas.

Posição	Orientações	Áreas
1º	389	Filosofia
2º	344	Direito
3º	302	Probabilidade e Estatística
4º	256	Zootecnia
5º	255	Engenharia Civil
6º	246	Engenharia de Produção
7º	219	Economia
8º	219	Artes
9º	215	Direito
10º	211	Engenharia Civil

Na Tabela 7.5 foram consideradas somente orientações concluídas de mestrado e doutorado. É possível observar que, dentre os dez orientadores com maior número de orientações concluídas, existe uma grande diversidade em suas áreas de atuação. Somente as grandes áreas de Ciências da Saúde e Ciências Biológicas não possuem representantes. A partir da análise dos docentes com maior número de orientações, foi possível identificar que todos foram ou são pesquisadores bolsistas de produtividade em pesquisa do CNPq e que estão há décadas atuando de forma significativa na orientação de alunos.

Em contraposição a esses orientadores muito representativos, dado o grande número de orientações ao longo de suas carreiras, foram identificados vários docentes sem nenhuma orientação concluída. Verificando mais cuidadosamente estes docentes,

detectou-se que, em sua maioria, eles estavam em início de carreira e ainda não tinham concluído nenhuma de suas orientações. Há também, mesmo atuando na pós-graduação, um pequeno número de docentes que possuem currículos somente com informações básicas, em geral, com data de atualização antiga.

7.1.2 Bolsistas de Produtividade em Pesquisa do CNPq

Já o segundo conjunto analisado considera todos os atuais pesquisadores que recebem bolsas de produtividade em pesquisa do CNPq. Eles foram selecionados a partir da informação inserida automaticamente nos currículos pelo CNPq. A bolsa de produtividade em pesquisa² é destinada aos pesquisadores que se destacam entre seus pares, valorizando sua produção científica segundo critérios normativos definidos pelo CNPq e pelos seus Comitês de Assessoramentos (CAs). Dentre os principais critérios para a concessão da bolsa, o solicitante deve ter o título de doutor ou perfil científico equivalente e dedicar-se a atividades de pesquisa. As bolsas são divididas em modalidades, sendo a modalidade 1 subdividida nos níveis, A, B, C e D. Já na modalidade 2 não há subdivisão de nível. O acesso às bolsas por um determinado pesquisador se inicia pela modalidade 2 e, ao longo do tempo, pode ser promovido por progressões até chegar à modalidade 1A, mais elevado nível, dedicado a pesquisadores que tenham demonstrado excelência continuada na sua produção científica e na formação de recursos humanos, e que liderem grupos de pesquisa consolidados. Logo, o grupo de pesquisadores bolsistas de produtividade em pesquisa se caracteriza por ser um segmento significativo da excelência científica brasileira.

Tais bolsistas podem ser agrupados por diversos critérios, como, por exemplo, nível da bolsa, áreas de atuação ou instituição à qual estão vinculados. A Tabela 7.6 apresenta as instituições com o maior número de bolsistas vinculados. Do total de bolsistas de produtividade em pesquisa, aproximadamente 64% deles estão vinculados a 20 instituições. Como pode ser observado, dentre as instituições com o maior número de bolsistas, destaca-se a USP. Como é a instituição detentora do maior número de doutores e, também, de docentes de pós-graduação, era de se esperar que fosse responsável por uma grande parcela dos bolsistas, destacando seu percentual que supera por mais de duas vezes o da segunda instituição. As outras posições são ocupadas, em geral, por instituições com grande número de docentes de pós-graduação, fato este que também era esperado, tendo em vista, que muitos dos bolsistas são também docentes de pós-graduação. Ressalta-se, entretanto, alguns casos específicos como, por exemplo, a UFRJ, que possui o terceiro maior número de docentes e o segundo de bolsistas, e a UFMG, que possui o quinto maior número de docentes e o quarto maior número de bolsistas.

² Cnpq.br/web/guest/view/-/jornal_content/56_INSTANCE_OoED/10157/100343#16061

Tabela 7.6: Instituições com o maior número de bolsistas de produtividade em pesquisa.

Instituição	Posição	Nº de Bolsistas	% das Bolsas	Modalidade da Bolsa					UF
				1A	1B	1C	1D	2	
USP	1º	2.015	14,09	282	236	245	349	903	SP
UFRJ	2º	898	6,28	131	127	111	151	378	RJ
UNESP	3º	695	4,86	43	49	65	139	399	SP
UFMG	4º	693	4,85	58	66	78	138	353	MG
UNICAMP	5º	691	4,83	106	100	90	105	290	SP
UFRGS	6º	658	4,60	78	84	70	124	302	RS
UFSC	7º	418	2,92	32	43	39	73	231	SC
UNB	8º	317	2,22	30	34	30	51	172	DF
UFPE	9º	312	2,18	28	30	20	61	173	PE
UFPR	10º	289	2,02	13	33	19	51	173	PR
UFF	11º	273	1,91	17	18	32	44	162	RJ
UERJ	12º	269	1,88	18	22	27	58	144	RJ
UFV	13º	242	1,69	46	28	18	47	103	MG
UFC		242	1,69	13	12	18	54	145	CE
UNIFESP	15º	223	1,56	28	20	26	40	109	SP
UFSCAR	16º	215	1,50	23	36	20	21	115	SP
FIOCRUZ	17º	214	1,50	24	25	31	27	107	RJ
UFSM	18º	205	1,43	11	7	13	45	129	RS
UFBA	19º	198	1,38	17	11	24	27	119	BA
PUC-RJ	20º	187	1,31	32	24	26	29	76	RJ

A exemplo do que também ocorre com a análise dos docentes, destaca-se a Fiocruz, com grande representatividade dentre os bolsistas, e ainda, a PUC-RJ, única instituição de ensino privada a figurar dentre as mais representativas. Também, a exemplo do que ocorre com a distribuição dos doutores e docentes de pós-graduação, a região Sudeste se caracteriza como a região de atuação da grande maioria dos bolsistas, sendo que o estado de São Paulo possui, aproximadamente, 32% de todo o conjunto, seguido do Rio de Janeiro (15,96%) e Minas Gerais (10,39%). Conseqüentemente, os estados menos representativos estão localizados na região Norte, sendo eles, Rondônia com dois bolsistas, Acre e Roraima com três e Amapá com quatro, reforçando a grande desigualdade encontrada na distribuição geográfica também dos bolsistas.

Com relação às modalidades de bolsas, existe grande disparidade entre as instituições. De maneira geral, a grande maioria dos bolsistas está enquadrada na modalidade 2. Em todas as instituições, esta modalidade de bolsa é a que detém o maior número de bolsistas, seguido pela modalidade 1D, com exceção para a Fiocruz, que caracteriza-se como a única a ter mais bolsistas 1C do que 1D. Porém, dentre os demais níveis, não existe um padrão específico. USP, UFRJ e UNICAMP destacam-se com grande número de bolsistas na modalidade 1A.

A Figura 7.4 apresenta a distribuição de todos os bolsistas entre as cinco modalidades de bolsa. A modalidade 2 concentra 55,73% dos bolsistas, enquanto 16,80% estão na modalidade 1D. Já nas demais modalidades (1A, 1B e 1C), o número de bolsistas gira em torno de 9% em cada uma. Esse salto da modalidade 2 para as demais modalidades significa que há uma grande barreira de entrada em tais modalidades, em que a progressão entre a modalidade 2 e a modalidade 1D torna-se bem mais competitiva. Outro aspecto que deve ser considerado são os critérios adotados para as modalidades 1 e 2. De uma forma geral, a modalidade 2 é a porta de entrada dos pesquisadores bolsistas. Já a mudança para as modalidades 1 requer como principais critérios ter uma elevada produção científica e maior tempo de carreira.

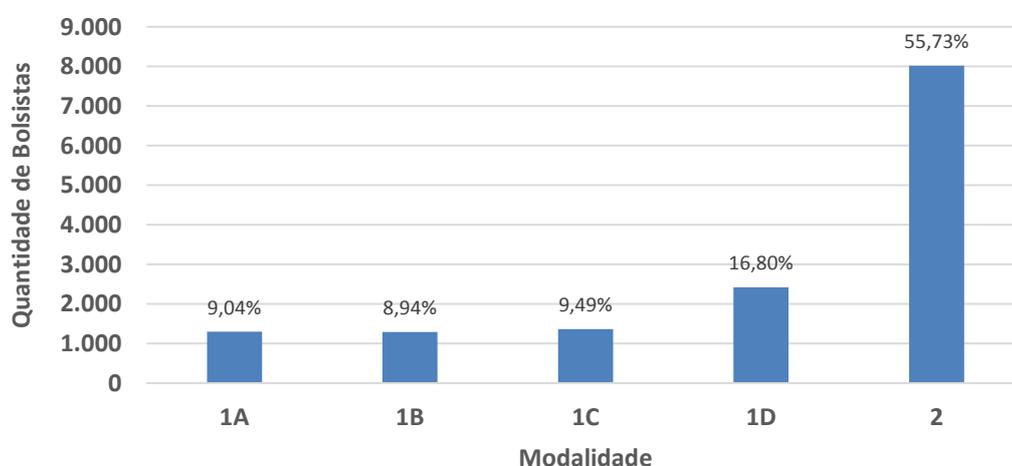


Figura 7.4: Distribuição dos bolsistas de produtividade em pesquisa por modalidade de bolsa.

Considerando que os bolsistas de produtividade em pesquisa são avaliados sob diversos aspectos relativos à sua carreira como pesquisadores, uma variável

especificamente importante é o tempo de carreira. A carreira de um bolsista é aqui considerada como sendo iniciada após o término de seu doutorado, ressaltando que um dos critérios para a concessão da bolsa é que o pesquisador tenha, no mínimo, três anos de doutorado concluído. Diante disso, é importante destacar o tempo de carreira em cada uma das modalidades de bolsa.

É possível observar a partir da Figura 7.5 a distribuição de tempo desde que o pesquisador obteve o grau de doutor, agrupados pela modalidade da bolsa. Verifica-se que o valor da mediana aumenta gradativamente da modalidade 2 em direção à modalidade 1A, formando um caminho de evolução bem consistente. Percebe-se que, não existem *outliers* inferiores com relação à quantidade de anos pós-doutorado, o que comprova a exigência de o candidato a bolsista ter um tempo mínimo de formação como doutor. É importante salientar a grande quantidade de *outliers* superiores na modalidade 2, evidenciando que o tempo de carreira por si só não é suficiente para alcançar níveis mais altos.

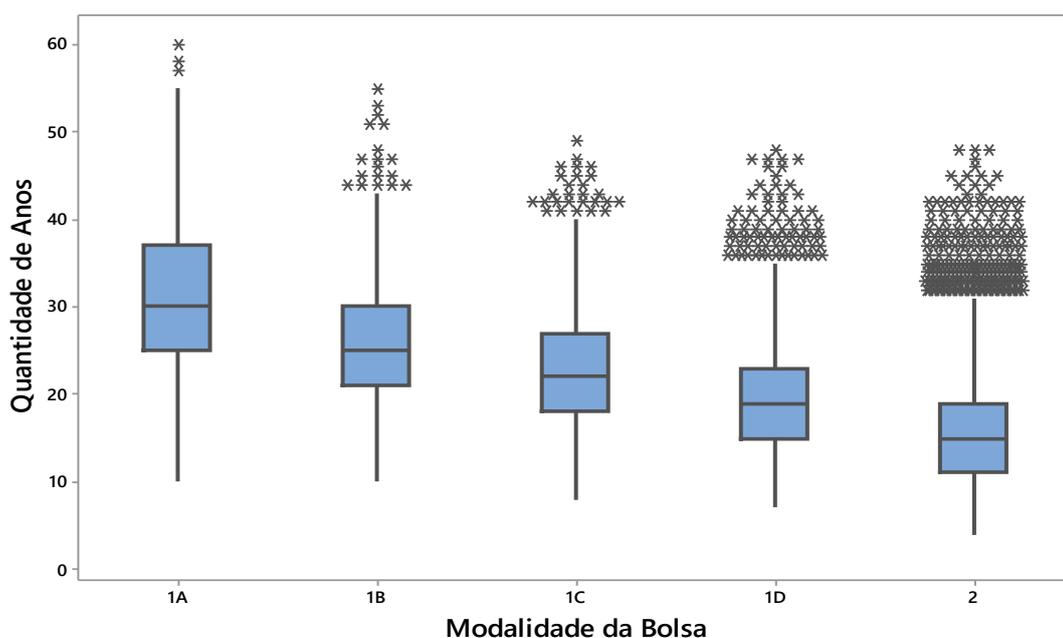


Figura 7.5: Tempo de doutorado em anos dos bolsistas de produtividade em pesquisa.

Outra informação importante ao se caracterizar os bolsistas refere-se à relação com as grandes áreas e áreas de atuação. A distribuição geral dos bolsistas pelas oito grandes áreas conforme informadas em seus currículos pode ser observada na Figura 7.6. Percebe-se que a grande área de Ciências Exatas e da Terra concentra, aproximadamente, 23% do total de bolsistas, tendo a grande área de Ciências Biológicas o segundo maior percentual com 18,09%. Estas duas grandes áreas, responsáveis por 40,98% dos bolsistas, são detentoras de apenas 26,71% do conjunto de doutores com currículos cadastrados na Plataforma Lattes. Já a grande área de Linguística, Letras e Artes é a de menor representatividade, com apenas 3,8% dos bolsistas. Note que um pequeno número de currículos (1,36%) inclui a opção Outra ou simplesmente nada informa sobre a grande área de atuação do bolsista.

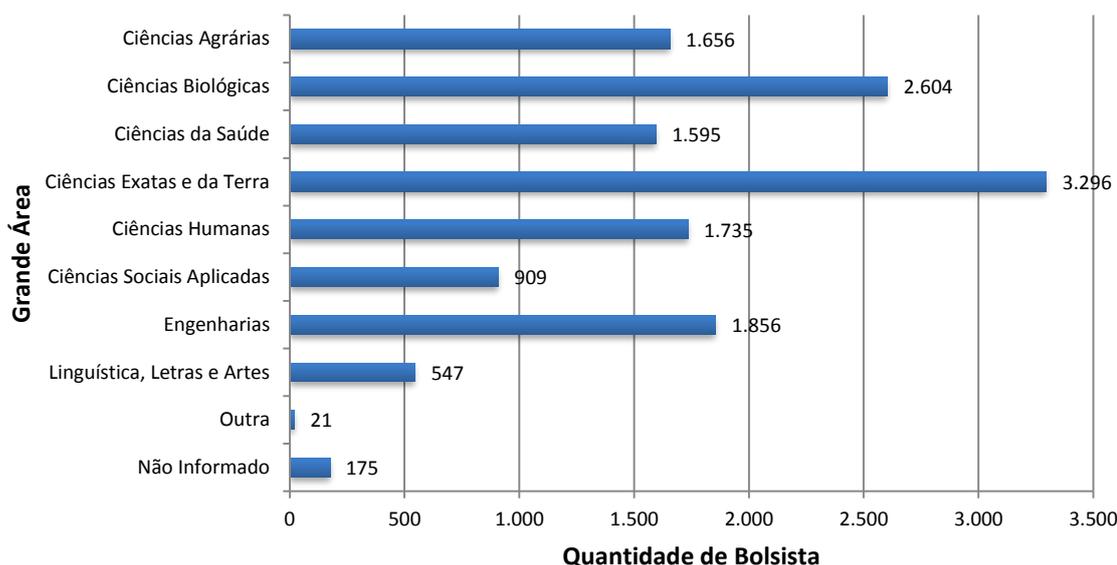


Figura 7.6: Distribuição dos bolsistas de produtividade em pesquisa por grandes áreas.

Aqui, é importante destacar que o intuito é identificar as áreas de atuação dos bolsistas e não necessariamente os Comitês de Assessoramento a que estão vinculadas suas bolsas. As grandes áreas de cada bolsista, bem como a sua área de atuação, foram extraídas de seus currículos e correspondem à sua principal grande área de atuação informada. Como principal grande área, foi considerada a primeira registrada na seção “Áreas de Atuação” em cada um dos currículos, a exemplo das análises anteriores. Foi possível, ainda, verificar que as grandes áreas com maior quantidade de bolsistas sofrem grande influência de suas áreas mais representativas (ver Tabela 7.7).

Tabela 7.7: Áreas de atuação com a maior quantidade de bolsistas de produtividade em pesquisa.

Posição	Área	Quantidade
1º	Física	978
2º	Química	808
3º	Agronomia	703
4º	Medicina	609
5º	Bioquímica	444
	Geociências	444
7º	Ciência da Computação	442
8º	Engenharia Elétrica	364
9º	Educação	352
10º	Matemática	343

Dentre as dez áreas mais relevantes com relação à quantidade de bolsistas, cinco são da grande área Ciências Exatas e da Terra, grande área esta que possui oito áreas, impulsionando-a como a grande área mais representativa. Medicina, que se caracteriza como uma das áreas com a maior quantidade de doutores (6,35%), possui 4,23% dos bolsistas, proporção bem inferior à Física, que possui apenas 3% dos doutores e 6,79% dos bolsistas. Uma explicação para a baixa representatividade da área de Medicina pode estar relacionada ao fato de trata-se de uma área com diversificado perfil profissional. Portanto, apesar de haver muitos doutores que atuam

como docentes em programas de pós-graduação, estes mantêm sua atuação profissional como médicos e, portanto, não se interessam ou não alcançam, por vezes, o perfil necessário para obter uma bolsa de produtividade em pesquisa. Este fato também está relacionado às áreas de Engenharia, que possuem um perfil bem mais tecnológico. Já com relação a Ciências Exatas e da Terra, embora a grande área não apresente a maior quantidade de doutores, a maioria deles dedica-se a atividades de pesquisa quase que exclusivamente e, portanto, possuem perfil e produção científica que os qualificam para as bolsas. Por exemplo, doutores de áreas como Matemática e Física, que atuam em programas de pós-graduação e orientam alunos, possuem, em sua maioria, dedicação integral às suas instituições e, portanto, de forma geral, detêm produção científica que os qualificam a uma bolsa de produtividade em pesquisa. Situação semelhante ocorre também na grande área de Ciências Biológicas.

A Tabela 7.8 apresenta a distribuição dos bolsistas por modalidade em cada uma das grandes áreas. Nota-se que a grande área de Ciências Exatas e da Terra possui a maioria dos bolsistas em todas as modalidades. Com valores próximos, surge a grande área de Ciências Biológicas. Conforme já mencionado, uma das razões para estes percentuais está relacionada ao perfil de seus doutores, que, em geral, dedicam-se exclusivamente às atividades de docência e pesquisa. Destaca-se, ainda, a pequena quantidade de bolsistas que declararam como grande área a opção Outra ou que não informaram em seus currículos suas respectivas grandes áreas de atuação, diferentemente do que acontece quando se analisa todo o conjunto de indivíduos com currículos cadastrados na Plataforma Lattes, em que este percentual é muito elevado.

Diante do exposto, observa-se como o conjunto de docentes de pós-graduação e, conseqüentemente, o conjunto dos bolsistas de produtividade em pesquisa são subconjuntos específicos dentre o conjunto total de doutores, cujos indivíduos apresentam reconhecida capacidade de pesquisa e orientação. As áreas de atuação destes indivíduos possuem perfis distintos e, para compreensão de cada uma delas, estudos que considerem as particularidades destas áreas se fazem necessário.

Tabela 7.8: Quantidade de bolsistas de produtividade em pesquisa por grande área em cada modalidade.

Grande Área	1A	1B	1C	1D	2	Total
Ciências Agrárias	134	119	161	311	931	1.656
Ciências Biológicas	268	250	301	441	1.344	2.604
Ciências da Saúde	144	137	137	279	898	1.595
Ciências Exatas e da Terra	289	303	336	499	1.869	3.296
Ciências Humanas	157	172	144	285	977	1.735
Ciências Sociais Aplicadas	72	61	71	150	555	909
Engenharias	180	185	150	337	1.004	1.856
Linguística, Letras e Artes	53	55	59	102	278	547
Outra	0	1	2	2	16	21
Não Informado	7	4	5	13	146	175
Total	1.304	1.287	1.366	2.419	8.018	14.394

Considerando a produção científica dos doutores que, apresenta queda considerável nos últimos anos (conforme mostrado na Seção 6.1), é importante verificar o comportamento ocorrido com os docentes de pós-graduação e os bolsistas de

produtividade em pesquisa, considerando sua representatividade no conjunto de doutores com currículos cadastrados na Plataforma Lattes. Tal análise é apresentada na próxima seção.

7.2 Análise Quantitativa da Produção Científica dos Docentes de Pós-Graduação e dos Bolsistas de Produtividade em Pesquisa

Ao se analisar a produção científica dos dois conjuntos estudados neste capítulo, destaca-se que, com relação aos docentes de pós-graduação, após o início das orientações, grande parte de seus trabalhos passa a ser realizada, cada vez mais, de forma colaborativa com alunos ou com seus pares que compõem seus grupos de pesquisa. Logo, uma análise da produção científica desses dois conjuntos de pesquisadores se faz importante. A Figura 7.7 apresenta a produção geral dos docentes de pós-graduação ao longo dos anos.

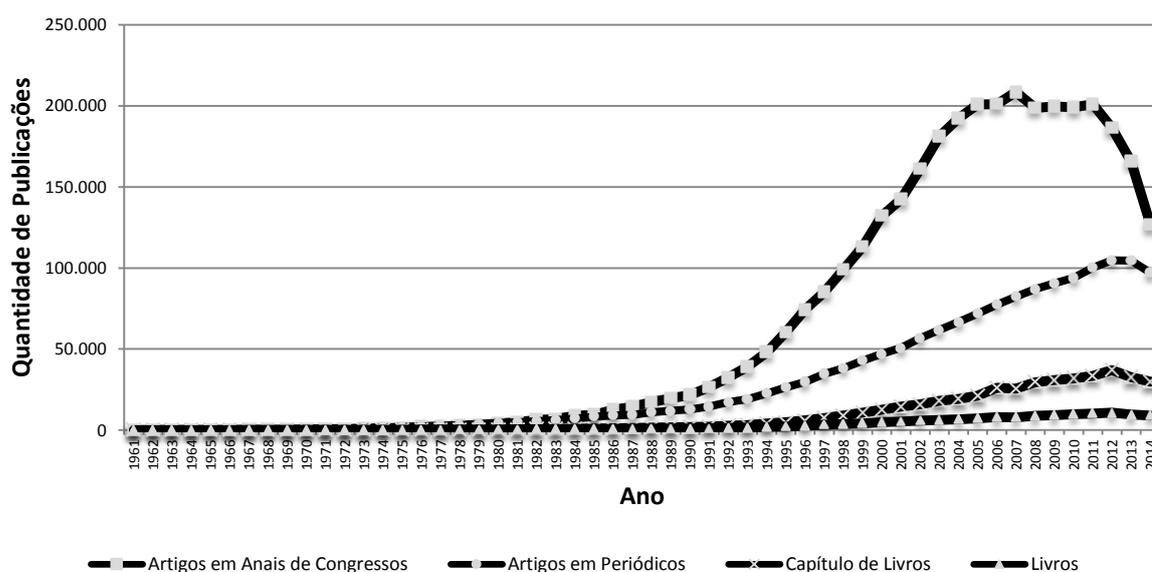


Figura 7.7: Produção científica dos docentes de pós-graduação.

A exemplo do que ocorre com a análise da produção dos doutores, e tendo em vista que os docentes de pós-graduação são um subconjunto relevante dos doutores, observa-se também uma queda na produção científica desse subconjunto em todos os tipos de publicação nos últimos anos. No entanto, a queda mais acentuada manifesta-se nos artigos em anais de congresso. Esta categoria de publicação tem seu ápice em 2007. Em seguida, sofre uma queda acentuada a partir de 2011 e, após três anos, apresenta uma redução de aproximadamente 37%. Dentre as hipóteses já mencionadas, a que melhor se aplica para os docentes dos programas de pós-graduação é a que diz respeito às avaliações da CAPES, que deixam de considerar os artigos em anais de congresso, dando ênfase aos artigos em periódico. Enfatiza-se, aqui, que a avaliação dos programas tem dependência direta com a produção científica dos docentes. Com relação aos artigos em periódico, o ano de 2012 apresenta a maior produção, sofrendo uma sensível queda em 2013, que se mostra ainda mais acentuada em 2014. Porém, é importante destacar que o ano de 2014 se caracteriza como aquele em que a diferença entre a quantidade de artigos em anais de congresso e artigos em

periódico é a menor dos últimos anos (29.742), caracterizando uma tendência dos docentes de pós-graduação em concentrar a publicação de seus artigos em periódico.

Ao analisar a produção científica dos bolsistas de produtividade em pesquisa, que, conforme já mencionado, são os pesquisadores de excelência em suas áreas de atuação e se caracterizam por publicações de maior impacto, nota-se comportamento semelhante à produção científica dos docentes, mas com queda na publicação de artigos em anais de congresso, ainda mais intensa (ver Figura 7.8).

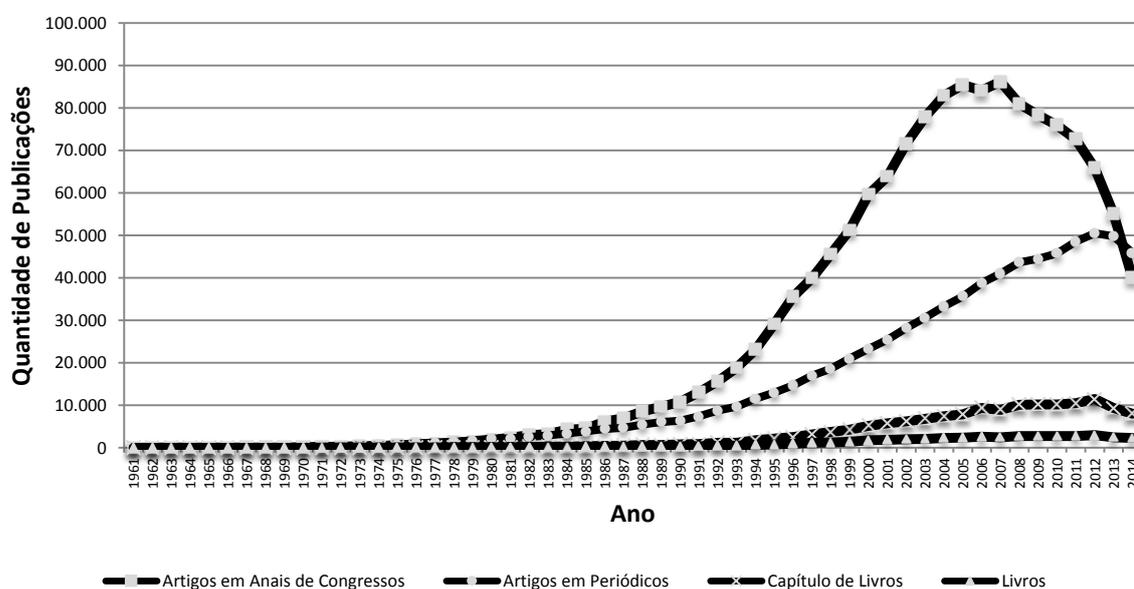


Figura 7.8: Produção científica dos bolsistas de produtividade em pesquisa.

Como o conjunto de bolsistas é praticamente um subconjunto dos docentes de pós-graduação, era de se esperar um comportamento similar ao observado na análise de todo o conjunto de docentes e, conseqüentemente, dos doutores. Porém, vale destacar a queda acentuada na quantidade de artigos em anais de congresso, tornando os artigos em periódico o principal tipo de publicação dos bolsistas de produtividade em pesquisa a partir de 2013. Este fato, que inicialmente ocorre com os bolsistas que passam a publicar mais artigos em periódico do que artigos em anais de congresso, pode representar uma tendência em relação aos outros conjuntos de pesquisadores aqui analisados. Tal fato poderá influenciar diretamente no perfil da produção científica brasileira, tendo em vista que os bolsistas de produtividade em pesquisa são os pesquisadores de excelência responsáveis pela formação de grande parte dos novos pesquisadores e, por isso, também influenciam o perfil de publicações desses novos pesquisadores. Para uma comparação entre a produção científica dos conjuntos analisados, a Tabela 7.9 apresenta os dados de cada conjunto.

É possível observar que o conjunto de docentes de pós-graduação representa apenas 1,52% dos indivíduos com currículos cadastrados na Plataforma Lattes, mas é detentor de 29,47% dos artigos em anais de congresso e 34,01% dos artigos em periódico. Logo, apesar de os docentes de pós-graduação corresponderem a apenas 28,33% dos doutores, eles são responsáveis por 45,57% dos artigos em anais de congresso e 45,65% dos artigos em periódico publicados pelos doutores. Dessa forma, mesmo sendo uma parcela relativamente pequena dos indivíduos que possuem doutorado

concluído, os docentes de pós-graduação, se caracterizam por serem responsáveis por boa parcela da produção científica do conjunto de doutores, com taxas de publicação bem superiores às dos que não lecionam na pós-graduação. Prova disso, são as médias de artigos em periódico dos dois conjuntos.

Já os bolsistas de produtividade em pesquisa representam apenas 0,35% de todos os indivíduos com currículos cadastrados na Plataforma Lattes, mas detêm 16,67% dos artigos em periódico. Isso faz com que esse conjunto apresente uma média de 52,81 artigos publicados em periódico por bolsista, média esta, bem superior à de todos os outros conjuntos analisados. Observa-se, ainda, que tais pesquisadores, de reconhecida capacidade em pesquisa, têm publicado, preferencialmente, em periódicos.

Tabela 7.9: Distribuição da produção científica dos indivíduos dos principais conjuntos analisados.

Produção / Conjuntos analisados	Plataforma Lattes	Doutores	Docentes de Pós-Graduação	Bolsistas de Produtividade em Pesquisa
Total de Currículos	4.156.635	223.477 (5,38%)	63.300 (1,52%)	14.394 (0,35%)
Artigos em Anais de Congressos	11.591.142	7.496.166 (64,67%)	3.416.031 (29,47%)	1.420.402 (1,23%)
Artigos em Periódicos	4.560.921	3.398.300 (74,51%)	1.551.243 (34,01%)	760.175 (16,67%)
Capítulos de Livro	1.055.388	810.915 (76,84%)	441.671 (41,85%)	153.585 (14,55%)
Livros	455.447	320.036 (70,27%)	151.354 (33,23%)	47.413 (10,41%)
Média de Artigos em Periódicos	1,10	15,21	24,51	52,81

Ao se comparar a produção de artigos dos docentes de pós-graduação e dos bolsistas de produtividade em pesquisa (Figura 7.9), levando em consideração todo o histórico de publicações, é possível verificar que a produção dos bolsistas é claramente superior à dos docentes. Para melhor visualização, os *outliers* foram excluídos, já que existem valores muito discrepantes, o que dificulta a apresentação gráfica dos resultados. Enquanto a mediana dos artigos em anais de congresso dos docentes é 46, a dos bolsistas é 91, com taxas de publicação que variam de 43 a 158 artigos. Com relação aos artigos em periódico, a mediana dos docentes é igual a 22, já a dos bolsistas é 54, com taxas de publicação que variam entre 32 e 90 artigos. Destaca-se a mediana tanto dos docentes quanto também a dos bolsistas, que são muito próximas às suas médias (\oplus), demonstrando certa homogeneidade entre os conjuntos analisados.

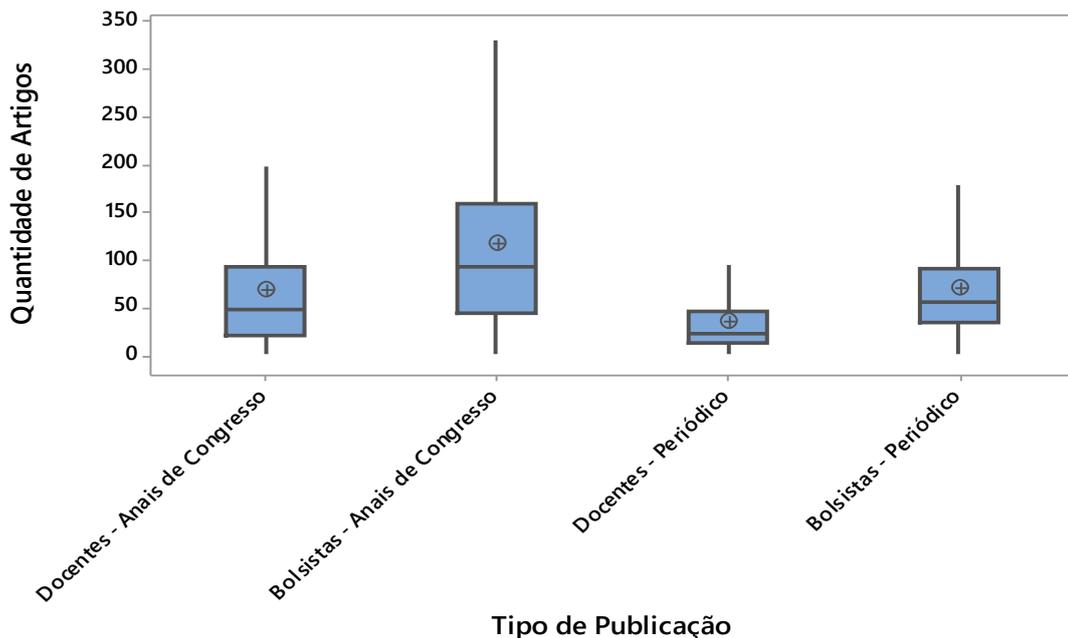


Figura 7.9: Produção científica dos docentes de pós-graduação e bolsistas de produtividade em pesquisa.

Importante ressaltar que a diferença dos valores apresentados tem relação direta com o tempo de carreira dos docentes e bolsistas. Considerando que as análises aqui apresentadas levam em consideração todo o histórico de produção científica registrado nos currículos analisados, observa-se que, de forma geral, os bolsistas possuem maior tempo de carreira (ver Figura 7.10). Ressalta-se que um dos critérios para a solicitação de bolsas de produtividade em pesquisa do CNPq é possuir um determinado tempo como doutor, dependendo da modalidade pleiteada. O tempo de carreira após o doutorado dos docentes de pós-graduação varia entre 7 a 18 anos com mediana igual a 12, enquanto que no caso dos bolsistas esse tempo varia entre 13 a 23 anos com mediana igual a 17. Logo, tais dados mostram que, em geral, os bolsistas de produtividade em pesquisa possuem maior tempo de carreira, contribuindo com taxas superiores de publicação quando comparadas com os docentes de pós-graduação.

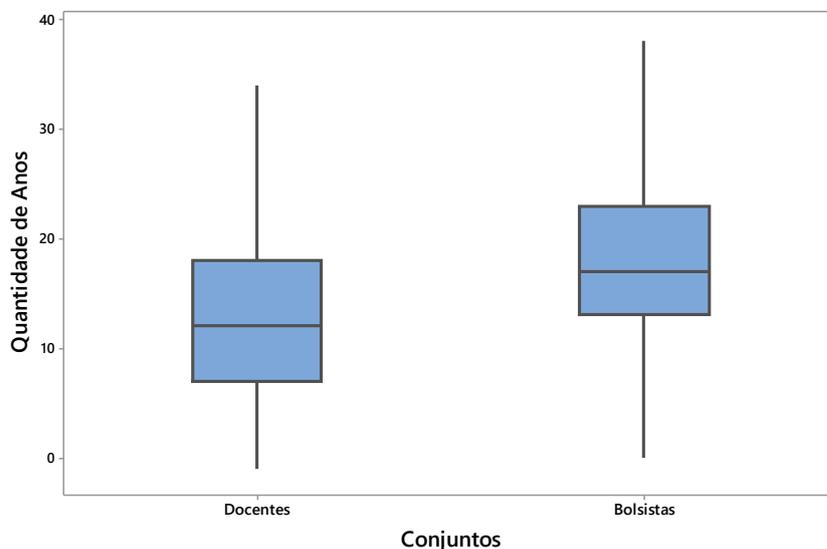


Figura 7.10: Tempo de carreira em anos após a conclusão do doutorado.

Ao investigar o impacto das orientações na produção científica dos docentes de pós-graduação, foi possível verificar que existe uma relação direta entre o número de orientações concluídas e a quantidade de trabalhos publicados. Apesar disso, percebe-se grande distinção no perfil de publicação desses docentes em relação à quantidade de orientações. Ao se comparar os dez casos mais representativos, verifica-se, por exemplo, uma grande diferença entre os dois primeiros colocados, com produções científicas muito distintas (ver Tabela 7.10).

Tabela 7.10: Docentes de pós-graduação com o maior número de orientações concluídas.

Posição	Orientações	Áreas	Artigos em Anais de Congresso	Artigos em Periódico
1º	389	Filosofia	219	130
2º	344	Direito	16	93
3º	302	Probabilidade e Estatística	259	247
4º	256	Zootecnia	876	519
5º	255	Engenharia Civil	350	112
6º	246	Engenharia de Produção	72	46
7º	219	Economia	129	143
8º	219	Artes	48	163
9º	215	Direito	0	43
10º	211	Engenharia Civil	503	64

Uma possível explicação para tal distinção entre a produção científica desses docentes pode ser as suas áreas de atuação, já que algumas áreas tendem a publicar mais do que outras. Além disso, também não existe um padrão de publicação, já que alguns docentes publicam mais artigos em anais de congresso e outros em periódico, dependendo de suas áreas de atuação.

Para melhor compreender a relação entre o número de orientações e a quantidade de trabalhos publicados, uma análise de correlação foi realizada. Para tal análise, foi calculado o coeficiente de correlação de Spearman tanto para os artigos em anais de congresso como também para os artigos em periódico, visando verificar a existência de uma relação direta entre o número de alunos orientados e a quantidade de trabalhos publicados por um determinado docente de pós-graduação (ver Figuras 7.11 e 7.12).

Os resultados apresentam um coeficiente de correlação positivo de 0,443 para os artigos em anais de congresso e de 0,528 para os artigos em periódico. Pelos valores obtidos, é possível comprovar uma relação direta entre o número de publicações e a quantidade de orientações de um determinado orientador.

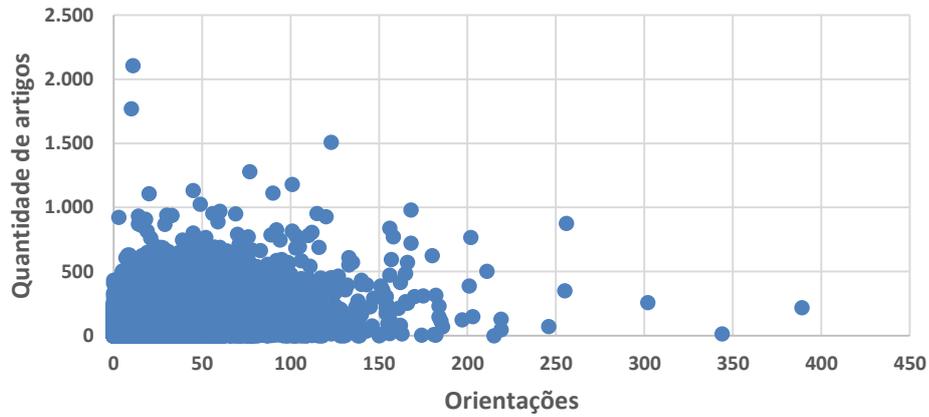


Figura 7.11: Correlação entre número de artigos em anais de congresso e orientações concluídas.

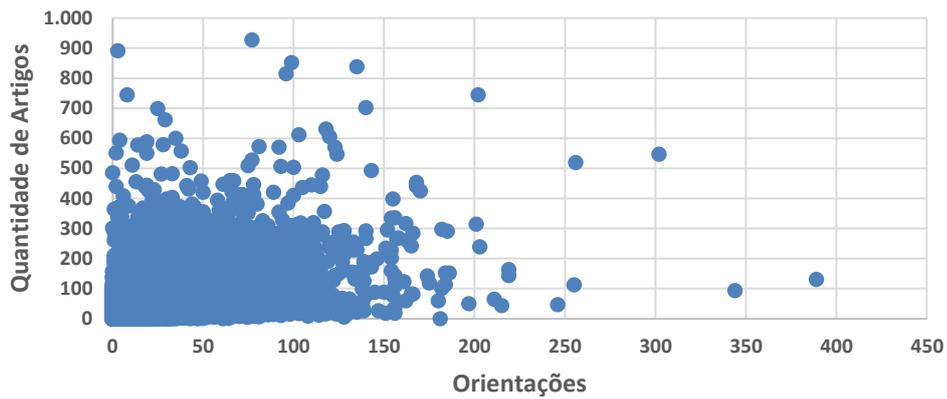


Figura 7.12: Correlação entre número de artigos em periódico e orientações concluídas.

Já com relação à produção científica dos bolsistas de produtividade em pesquisa, é importante verificar a distribuição de suas publicações em relação às modalidades das bolsas. Tendo em vista que a entrada de um doutor não bolsista é na modalidade 2 e os pesquisadores de excelência estão na modalidade 1A, é de se esperar que o perfil de publicação entre os níveis tenha crescimento gradativo (ver Figuras 7.13 e 7.14).

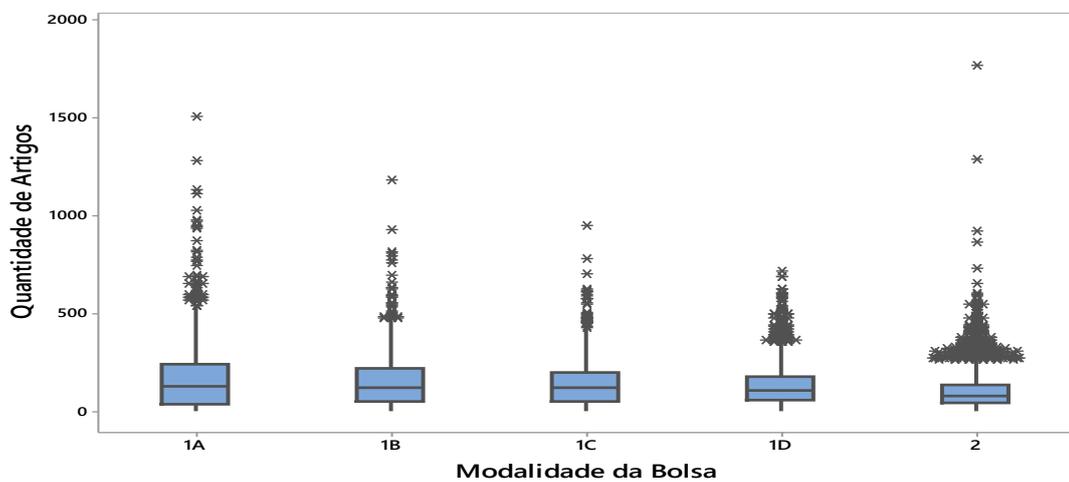


Figura 7.13: Distribuição dos artigos em anais de congresso por modalidade de bolsa de pesquisa.

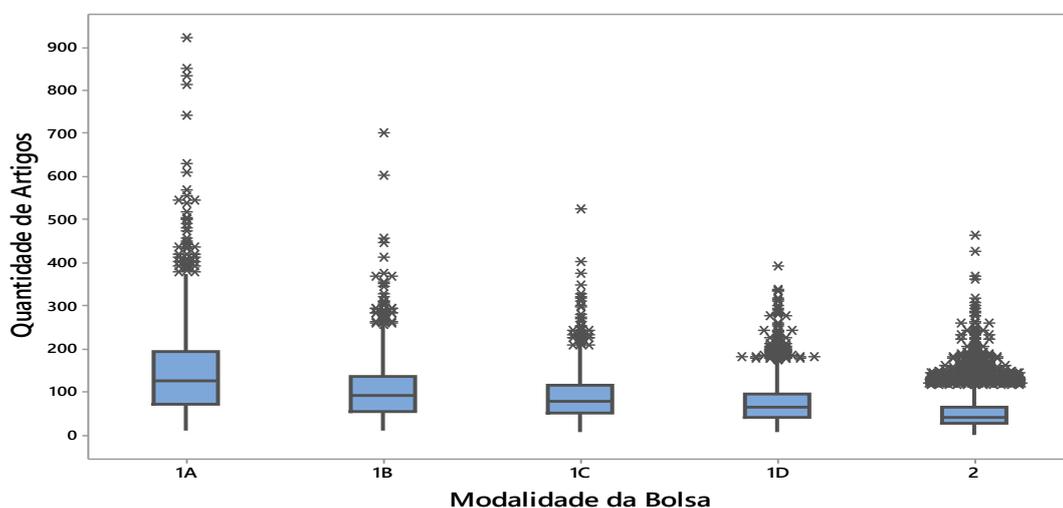


Figura 7.14: Distribuição dos artigos em periódico por modalidade de bolsa de pesquisa.

Ao se analisar a produção de artigos em anais de congresso, observa-se que o valor das medianas está bem próximo entre as modalidades do nível 1. Sendo o maior valor para a modalidade 1A que possui mediana igual a 125. E de forma hierárquica, o valor da mediana diminui até a modalidade 2 (78). Dessa forma, é importante notar que a maior diferença entre as modalidades está nos valores máximos (3º quartil), que alteram de forma mais intensa, entre as modalidades (modalidade 1A igual a 238 e modalidade 2 igual a 121). Também é importante destacar a quantidade de *outliers* que aumentam gradativamente das modalidades mais altas para as mais baixas. A maior parte dos *outliers* está próxima dos valores máximos em cada modalidade subsequente, demonstrando que, em se tratando de produção de artigos em anais de congresso, esses *outliers* poderiam estar no nível superior ao que se encontram.

Com relação à produção de artigos em periódico, nota-se que a diferença da modalidade 1A para as demais fica evidenciada, já que a modalidade 1A possui mediana igual a 126, enquanto a modalidade 1B possui mediana igual a 91, valores estes bem superiores aos da modalidade 2, com mediana igual a 41. O valor da mediana da modalidade 1A está bem próximo do terceiro quartil da modalidade 1B, a exemplo do que acontece com as modalidades 1D e 2. É importante destacar que, a exemplo dos artigos em anais de congresso, os *outliers* das modalidades inferiores estão próximos dos valores máximos das modalidades posteriores, *outliers* estes que aparecem em maiores quantidades nos níveis mais baixos.

Destaca-se ainda que, pelos dados apresentados, ao se considerar a produção científica dos bolsistas de produtividade em pesquisa, levando em consideração a modalidade de suas bolsas, fica evidente uma maior contribuição dos bolsistas das modalidades mais altas. Também pode-se destacar os *outliers* que possuem produção científica equivalente aos valores máximos dos níveis superiores, o que possibilitaria uma mudança de nível se o critério adotado fosse apenas produção científica. Entretanto, apenas uma análise quantitativa não é suficiente para avaliar a produção científica dos bolsistas de produtividade em pesquisa nas cinco modalidades da bolsa. Logo, a Seção 7.3 avalia qualitativamente a produção em periódicos desses bolsistas.

7.3 Análise Qualitativa da Produção Científica dos Bolsistas de Produtividade em Pesquisa

Os resultados desta seção visam apresentar uma caracterização sobre a qualidade dos trabalhos realizados pelos bolsistas de produtividade em pesquisa. Suas publicações são aqui avaliadas de forma qualitativa, possibilitando, dessa forma, compreender como têm sido publicados os resultados das pesquisas brasileiras ao longo dos últimos anos, nas diversas áreas do conhecimento. Para tanto, na análise apresentada, serão considerados apenas os artigos em periódico, tendo em vista a disponibilidade de dados adicionais sobre esse tipo de publicação, como número de citações recebidas e fator de impacto do respectivo periódico, que possibilitam a realização de tais análises.

Esses dados adicionais são geralmente obtidos a partir de fontes externas. Para isso, o ISSN (*International Standard Serial Number*) e o DOI (*Digital Object Identifier*) associados aos artigos são de fundamental importância. Do total de artigos em periódico produzidos pelos bolsistas de produtividade em pesquisa, apenas 8,79% não possuem registrado o respectivo ISSN. Ao verificar tais artigos, foi possível observar que, de forma geral, são publicações para as quais somente dados básicos, como título e lista de autores foram registrados, não incluindo outros dados como DOI, número da edição e paginação, imprescindíveis, por exemplo, para o processo de obtenção das citações.

Uma informação importante para análises qualitativas refere-se à abrangência do periódico calculada pelo seu Fator de Impacto (FI). O FI, proposto por Garfield (1955), começou a ser utilizado como instrumento para indicação da qualidade das publicações nos anos de 1960. Desde então, o FI tem sido usado como meio de avaliação de periódicos nas diversas áreas do conhecimento. Atualmente, o cálculo do FI é realizado pela Thomson Reuters e divulgado anualmente pelo *Journal of Citation Report – JCR*³. Diversos pesquisadores consideram o FI como um dos critérios para a escolha do periódico, para que seu trabalho possa ter maior visibilidade. Além disso, diversas agências de fomento à pesquisa têm adotado o FI como um dos critérios para a identificação de pesquisadores com maior mérito, dada a sua produção científica. Para o cálculo do FI de um determinado periódico em um certo ano, é levado em consideração o número de citações recebidas naquele ano pelos artigos publicados no periódico nos dois anos precedentes, dividido pelo número de artigos publicados pelo periódico no mesmo período. Logo, o FI é um importante mecanismo para mediar a qualidade do periódico, não sendo, entretanto, relevante para medir a qualidade de um artigo em particular.

A Figura 7.15 apresenta a distribuição da mediana dos FIs referentes aos periódicos dos artigos publicados pelos bolsistas de produtividade em pesquisa ao longo dos últimos anos. Como pode ser observado, a partir do início dos anos 2000 existe uma pequena variação na mediana dos FIs, que estabiliza-se em 1,71 a partir de 2007. Após três anos estáveis, a mediana passa a ter crescimento constante, apresentando valores maiores a cada ano. O ápice é atingido em 2014, com o valor de 1,99. É importante ressaltar que, para fins de visualização, os *outliers* não foram apresentados, já que, para todos os anos, existem FIs com valores muito altos, o que faz com que suas médias sejam bem superiores às medianas. Logo, tendo em vista que o FI está relacionado à qualidade dos periódicos em que os resultados de pesquisa dos bolsistas

têm sido divulgados e, considerando ainda que esses bolsistas representam o conjunto de pesquisadores com as maiores médias de publicações em periódicos, pode-se afirmar que a qualidade dos veículos em que os artigos têm sido publicados tem melhorado a cada ano.

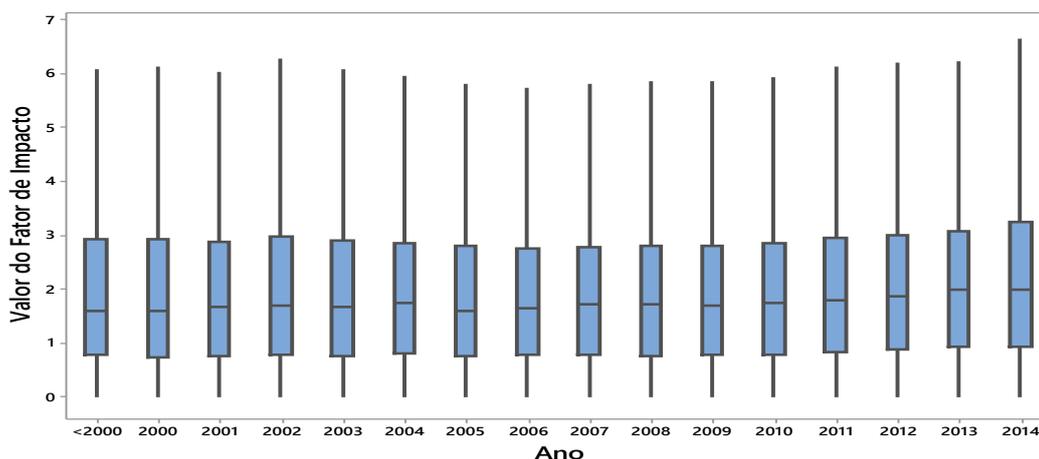


Figura 7.15: Distribuição da mediana do Fator de Impacto das publicações dos bolsistas de produtividade em pesquisa.

Além do FI, útil para verificar a qualidade de um periódico como veículo de publicação, a análise de citações surge como um importante mecanismo para verificar a qualidade de uma determinada publicação. Para Abbasi, Altmann e Hossain (2011), um pesquisador possui grande visibilidade quando os seus trabalhos são citados por outros pesquisadores. Logo, para analisar a qualidade individual dos trabalhos, o número de citações de cada artigo em periódico publicado pelos bolsistas foi estimado a partir de dados extraídos pelos sítios *Web of Science (WOS)*⁴ e *Scopus*⁵.

É importante destacar que a qualidade da produção científica de um determinado pesquisador não pode ser medida simplesmente pelo número de citações que seus trabalhos receberam. Isso se deve a vários fatores, como, por exemplo, o seu tempo de carreira ou a rede de colaboração na qual ele está inserido. Para Podlubny (2005), a qualidade da produção científica de um pesquisador não pode estar associada apenas ao número de citações recebidas, principalmente porque isso depende muito de sua área de atuação. O número de referências por artigo pode ser bastante diferente em cada uma das diversas áreas do conhecimento. Por exemplo, artigos de Ciências Exatas geralmente têm menor densidade de citações do que aqueles ligados às Ciências da Saúde, o que explica, em parte, porque os FIs dos periódicos dessa área são, em média, muito maiores em comparação àqueles das Ciências Exatas, como no caso da Matemática (GARFIELD, 2000). Para Thomaz, Assad e Moreira (2011), o FI publicado pelo JCR, ao levar em consideração somente as citações dos dois últimos anos, tende a beneficiar aqueles periódicos de áreas cujo ritmo de atualização do conhecimento é muito acelerado. Nesses casos, a citação dos artigos ocorre imediatamente após a sua publicação, favorecendo áreas como Ciências Exatas e Ciências Biológicas em detrimento de áreas com crescimento menos acelerado, como Ciências Sociais e Humanas. Diante disso, a Tabela 7.11 apresenta uma classificação dos periódicos de acordo com as grandes áreas com base em seus respectivos FI.

⁴ <https://webofknowledge.com/>

⁵ <https://www.scopus.com/>

Tabela 7.11: Média, Mediana e Desvio Padrão das citações por estrato do Fator de Impacto dos periódicos de cada grande área.

Grande Área		MÉDIA						MEDIANA						DESVIO PADRÃO				
		Muito Baixo	Baixo	Médio	Alto	Muito Alto		Muito Baixo	Baixo	Médio	Alto	Muito Alto		Muito Baixo	Baixo	Médio	Alto	Muito Alto
Ciências Agrárias	WOS	0,18	0,03	0,62	3,27	16,34	WOS	0	0	0	0	7	WOS	3,31	1,13	2,23	10,69	37,34
	Scopus	1,18	1,01	0,91	3,42	18,82	Scopus	0	0	0	0	9	Scopus	4,74	3,65	2,80	11,89	41,02
Ciências Biológicas	WOS	1,81	0,42	2,59	9,80	22,21	WOS	0	0	0	5	10	WOS	12,94	4,28	7,00	18,13	53,48
	Scopus	3,12	2,27	3,31	10,49	24,23	Scopus	0	0	0	5	12	Scopus	14,68	8,68	8,41	18,33	54,71
Ciências da Saúde	WOS	0,40	0,06	0,65	5,26	19,13	WOS	0	0	0	1	7	WOS	7,15	1,56	2,53	11,07	59,68
	Scopus	1,57	0,91	1,73	6,67	22,77	Scopus	0	0	0	2	9	Scopus	11,26	3,87	5,73	12,86	72,89
Ciências Exatas e da Terra	WOS	2,19	0,27	3,33	8,80	20,36	WOS	0	0	0	4	10	WOS	14,34	3,13	10,11	18,97	48,65
	Scopus	2,70	1,16	3,54	9,48	19,51	Scopus	0	0	1	4	9	Scopus	14,36	5,03	9,43	19,28	55,32
Ciências Humanas	WOS	0,03	0,01	0,01	0,01	3,91	WOS	0	0	0	0	0	WOS	1,77	0,62	0,49	0,25	15,67
	Scopus	0,41	0,30	0,15	0,30	5,06	Scopus	0	0	0	0	0	Scopus	2,72	1,63	0,92	1,54	18,89
Ciências Sociais Aplicadas	WOS	0,01	0,02	0,01	0,01	3,85	WOS	0	0	0	0	0	WOS	0,20	0,75	0,03	0,08	13,86
	Scopus	0,43	0,22	0,14	0,11	5,40	Scopus	0	0	0	0	0	Scopus	2,36	2,76	0,94	0,91	17,38
Engenharias	WOS	0,43	0,14	0,41	4,60	15,12	WOS	0	0	0	1	8	WOS	5,28	2,02	1,96	11,12	40,94
	Scopus	1,20	0,73	0,94	6,16	18,35	Scopus	0	0	0	2	9	Scopus	7,17	3,21	3,05	14,88	46,29
Linguística, Letras e Artes	WOS	0,01	0,01	0,01	0,00	0,38	WOS	0	0	0	0	0	WOS	0,25	0,44	0,03	0,00	4,19
	Scopus	0,14	0,05	0,07	0,04	0,73	Scopus	0	0	0	0	0	Scopus	3,63	0,53	0,63	0,40	6,45

Dada a disparidade entre o valor do FI dos diversos periódicos, a classificação foi dividida em cinco estratos (Muito Baixo, Baixo, Médio, Alto e Muito Alto), cada um contendo 20% dos periódicos identificados conforme ordenação crescente do valor de seus FIs, onde o estrato Muito Baixo contém os 20% dos periódicos com os menores FIs e Muito Alto contém os 20% dos periódicos com os maiores FI. Logo, os artigos foram classificados em cada um desses estratos, de acordo com o periódico em que eles foram publicados.

Tendo em vista a grande discrepância na quantidade de citações recebidas pelos artigos de periódico nas diversas áreas do conhecimento, a Tabela 7.11 apresenta, a média, a mediana e o desvio padrão desse indicador em cada um dos estratos definidos para o FI e para cada uma das grandes áreas, considerando tanto os valores obtidos na WOS como também na Scopus. Como pode ser observado, a distribuição das citações em cada grande área é bem distinta em cada um dos estratos analisados. Essa distinção também é identificada quando é analisada a quantidade de citações registradas na WOS e Scopus, sendo que a média de citações da Scopus é superior em todos os estratos e em todas as grandes áreas.

De forma geral, percebe-se que a média de citações é maior nos estratos mais altos, com os valores do estrato Muito Alto sendo bem superiores em relação a todos os outros. Já os três estratos mais baixos, recebem um número muito baixo de citações. Um fato identificado é que, com exceção das grandes áreas de Ciências Biológicas, Ciências da Saúde e Ciências Exatas e da Terra, todas as demais possuem a maioria de suas publicações no estrato Muito Baixo, tornando suas médias nesse estrato superior à dos estratos Baixo e Médio.

A grande área com a melhor média de citações é a de Ciências Biológicas, seguida pela grande área de Ciências Exatas e da Terra. Já a grande área de Linguística, Letras e Artes possui a menor média de citações dentre todas as grandes áreas. Tais valores, extremamente baixos, podem ser justificados pelas características de suas áreas, cujos pesquisadores geralmente publicam seus trabalhos em periódicos nacionais que tendem a receber uma quantidade menor de citações devido à sua menor visibilidade internacional. É importante destacar os valores do desvio padrão das amostras analisadas nos casos em que as grandes áreas com a maior média de citações também possuem os maiores valores do desvio padrão, devido a uma grande variação entre os valores das amostras. Tal fato pode ser explicado pela presença de artigos altamente citados em relação a outros que não são tão referenciados. Esses artigos com um número de citações muito elevado acabam por influenciar a média em suas grandes áreas. Especificamente, a grande área de Ciências da Saúde possui os maiores valores para o desvio padrão no estrato Muito Alto, reflexo de artigos muito citados e de uma distribuição irregular das citações de seus trabalhos nos periódicos de maior FI.

Logo, ao se verificar a mediana das citações, é possível identificar que, nos três estratos inferiores, que, conseqüentemente, concentram 60% dos periódicos com os menores FI, percebe-se que, em praticamente sua totalidade, o valor da mediana é zero. Mesmo nos estratos superiores, algumas grandes áreas, como as de Ciências Humanas, Ciências Sociais Aplicadas e a de Linguística, Letras e Artes, também possuem medianas iguais a zero. As razões disso estão relacionadas ao fato de que tais áreas têm, geralmente, uma grande quantidade de seus artigos publicados em

periódicos com baixo FI e, portanto, de baixa visibilidade, de modo que a maioria de seus artigos acaba por não receber citações, resultando em medianas com valores nulos.

Tabela 7.12: Percentual de artigos que não possuem citações em cada grande área.

Grande Área	WOS	Scopus	WOS + Scopus
Ciências Agrárias	69,37%	60,57%	49,97%
Ciências Biológicas	38,93%	32,26%	27,31%
Ciências da Saúde	60,62%	48,98%	43,85%
Ciências Exatas e da Terra	33,85%	32,15%	26,29%
Ciências Humanas	95,12%	85,80%	84,53%
Ciências Sociais Aplicadas	94,77%	86,56%	86,00%
Engenharias	53,69%	44,86%	42,37%
Linguística, Letras e Artes	99,24%	95,71%	95,54%

A Tabela 7.12 apresenta o percentual de artigos sem citações em cada grande área. Observa-se que, em todas as grandes áreas, a quantidade de artigos sem citações é maior na WOS, sendo que a maior diferença entre os artigos sem citações na WOS e Scopus ocorre na grande área de Ciências da Saúde (11,64%) e a menor na grande área de Ciências Exatas e da Terra (1,70%). Os valores apresentados demonstram que uma grande quantidade de artigos não possui qualquer citação, corroborando com os valores das medianas apresentados na Tabela 7.11, com destaque para as grandes áreas de Ciências Humanas, Ciências Sociais Aplicadas e Linguística, Letras e Artes. Logo, fica evidente a influência dos artigos altamente citados no cálculo das médias. Tais artigos são atípicos, fazendo com que as médias sejam bem maiores que as medianas encontradas. A Tabela 7.13 apresenta o total de citações recebidas pelos artigos mais citados, conforme dados da WOS como da Scopus, bem como o FI do respectivo periódico, a área de atuação de seu(s) autor(es) e o ano de publicação.

Tabela 7.13: Artigos com maior quantidade de citações.

WOS			
	Citações	FI	Área
1º	4.212	6,131	Física
2º	3.731	45,217	Medicina
3º	3.205	6,131	Engenharia Elétrica
4º	3.174	55,873	Medicina
5º	2.981	46,568	Química
6º	2.934	55,873	Saúde Coletiva
7º	2.799	8,665	Fisiologia
8º	2.667	4,774	Botânica
9º	2.658	55,873	Medicina
10º	2.561	41,456	Genética
Scopus			
	Citações	FI	Área
1º	5.265	2,183	Física
2º	4.847	45,217	Medicina
3º	4.344	55,873	Medicina
4º	4.328	6,131	Física
5º	3.484	55,873	Saúde Coletiva
6º	3.209	55,873	Medicina
7º	3.138	46,568	Química
8º	3.096	2,183	Química
9º	2.944	55,873	Medicina
10º	2.776	4,774	Botânica

É possível observar uma pequena distinção entre os dados dos dois repositórios consultados. Em geral, os artigos consultados possuem uma quantidade maior de citações na Scopus quando comparados com as citações da WOS. Apesar desta distinção, foi possível verificar que sete dos 10 (dez) artigos mais citados são os mesmos nos dois repositórios. Vale ressaltar que as áreas dos artigos mais citados foram identificadas a partir da principal área de atuação de seus autores e que apenas dois artigos em ambos os repositórios tinham entre seus autores mais de um bolsista de produtividade em pesquisa, sendo que, em ambos os casos, a principal área desses autores era a mesma. Além disso, corroborando com o que foi apresentado, a maioria dos artigos mais citados foi publicada em periódicos com alto FI, destacando-se o *The New England Journal of Medicine*⁶ criado em 1812 e cujo fator de impacto é 55,873, o maior dentre os periódicos das áreas médicas. No entanto, existem periódicos que mesmo não possuindo um FI muito alto, incluem artigos extremamente citados.

Logo, fica evidente pelas constatações anteriores que, como as publicações dos bolsistas nos últimos anos têm sido realizadas cada vez mais em periódicos com maior FI, a tendência é que os novos artigos irão receber uma quantidade maior de citações dada a sua maior visibilidade.

Outra medida comumente utilizada para avaliar a qualidade da produção científica de um determinado pesquisador é o índice-h. Criado por Hirsch (2005), o índice-h de um pesquisador é calculado com base no número de citações recebidas pelas suas publicações, mas sem estabelecer um limite temporal. O índice-h de um pesquisador é definido como sendo o número de suas publicações que possuem uma quantidade de citações igual ou maior a esse número. Considerando que o índice-h de um pesquisador seja igual a oito, isto significa que ele possui, pelo menos, oito publicações, e que cada um deles tenha, pelo menos, oito citações. Diante disso, o índice-h busca o equilíbrio entre a quantidade de trabalhos publicados e a quantidade de citações recebidas pelas publicações de um determinado pesquisador. Diante disso, tendo em vista uma análise qualitativa dos bolsistas de produtividade em pesquisa, foi calculado o índice-h para cada um deles. Para este cálculo, foram utilizadas as citações da Scopus, já que, em geral, possuem um maior valor do que as citações da WOS. Neste ponto, é importante destacar que, como a análise de citações é baseada somente nos artigos publicados em periódico, conseqüentemente, os índices-h aqui apresentados correspondem somente aos artigos de periódico de um determinado bolsista, o que difere de outras fontes, como, por exemplo, *Google Scholar*⁷ e *Semantic Scholar*⁸, que fazem o cálculo considerando também artigos publicados em anais de congresso e outros tipos de publicação como livros e capítulos de livro. A distribuição dos bolsistas pelos seus índices-h pode ser observada na Figura 7.16.

⁶ <http://www.nejm.org>

⁷ <https://scholar.google.com/>

⁸ <https://www.semanticscholar.org>

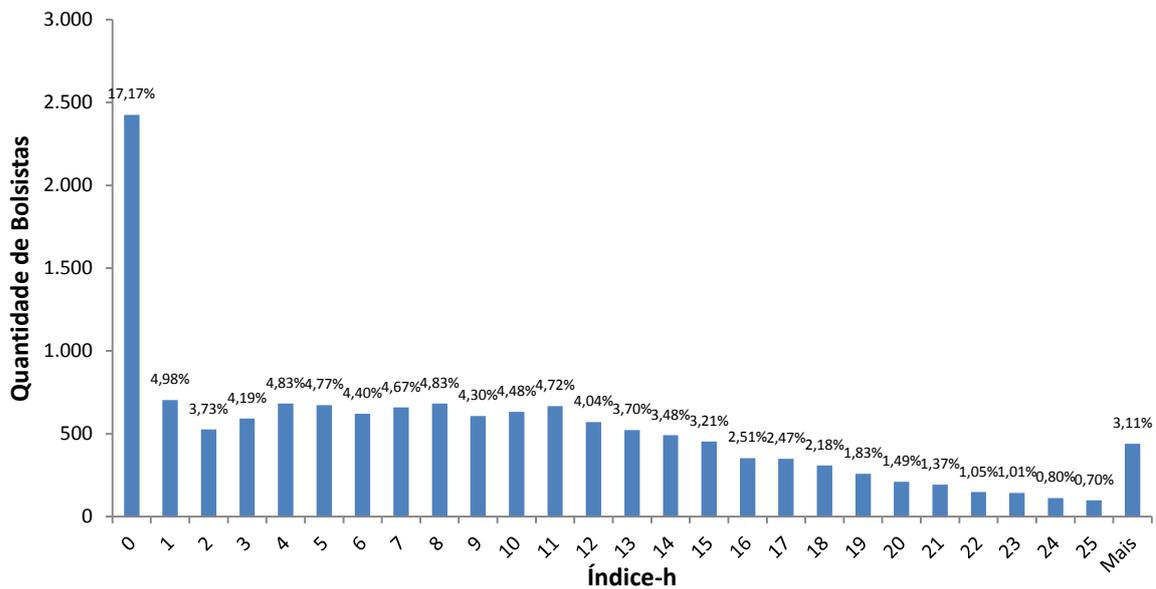


Figura 7.16: Distribuição dos bolsistas de produtividade em pesquisa pelos respectivos valores do índice-h.

É possível verificar que um percentual de 17,17% dos bolsistas não possui citações e, conseqüentemente, seu índice-h é igual a 0. Ressalta-se que, como o cálculo é realizado sobre o número de citações, tal valor não está associado à quantidade de artigos, podendo um bolsista possuir vários artigos e, mesmo assim, ter um índice-h nulo. A distribuição entre os valores mais baixos do índice-h (entre 1 a 10) é irregular, porém, com valores muito próximos a 4%. No entanto, a partir do índice-h 11 é apresentada uma queda constante, ou seja, apenas poucos bolsistas detêm valores maiores do índice-h. Vale ressaltar que apenas cerca de 3% do total de bolsistas possuem índice-h superior a 25 e que o maior índice-h encontrado foi 78. Tendo em vista que o índice-h de um pesquisador pode ser influenciado por vários motivos, como, por exemplo, sua participação em grupos de pesquisa altamente produtivos ou, ainda, por seu tempo de carreira, é de se esperar que exista uma distribuição irregular dos índices-h entre os bolsistas em cada modalidade de bolsa (ver Figura 7.17).

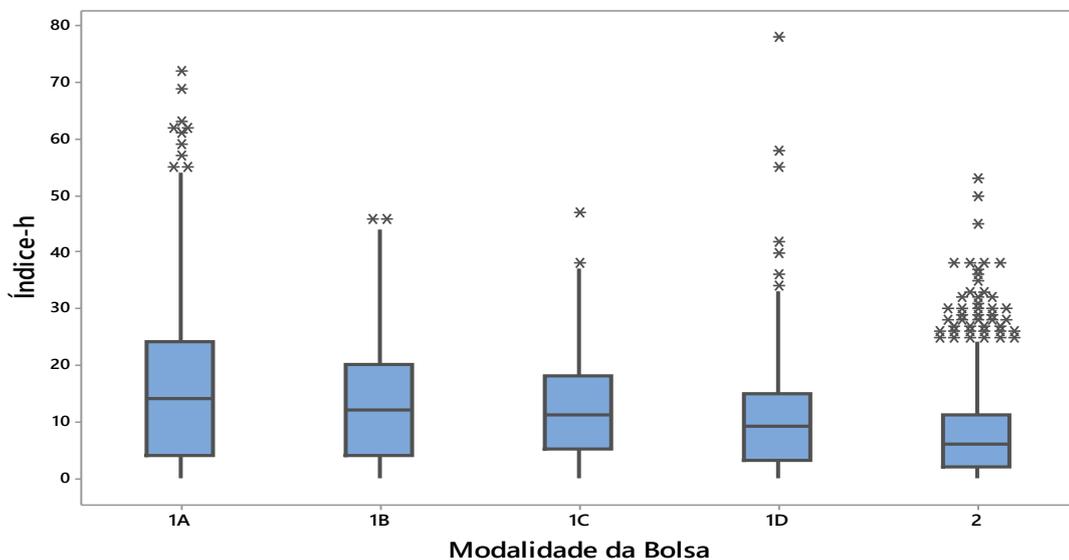


Figura 7.17: Distribuição dos valores dos índices-h pelas modalidades das bolsas.

Observa-se que a distribuição segue uma hierarquia de acordo com a modalidade das bolsas, em que os maiores índices-h estão concentrados na modalidade 1A, modalidade esta em que estão os bolsistas de maior excelência, e que possuem, em geral, carreiras mais longas. Existem alguns *outliers* na modalidade 1A que são bolsistas com índice-h bem acima da média. Já nas modalidades 1B a 1D também existem *outliers*, mas em número menor, diferentemente da modalidade 2 que, por ser a modalidade de entrada dos bolsistas, possui a menor mediana e valores menores do índice-h. Entretanto, a modalidade 2 possui vários bolsistas com índices-h bem superiores aos de muitos bolsistas das outras modalidades, o que pode indicar a existência de uma certa retenção nessa modalidade.

Ao verificar uma possível correlação entre a modalidade da bolsa e o índice-h dos bolsistas, percebe-se que ela é relativamente baixa, o cálculo do coeficiente de correlação de Spearman é de 0,349, considerando pesos de 1 a 5 para cada modalidade. Foi possível observar bolsistas nos níveis mais baixos com alto índice-h e bolsistas nos níveis mais altos com índice-h igual a zero. Logo, a Tabela 7.14 apresenta, para cada modalidade de bolsa, os dados dos cinco bolsistas com maiores valores do índice-h em cada modalidade de bolsa e suas respectivas áreas de atuação.

Tabela 7.14: Bolsistas de produtividade em pesquisa com os maiores valores de índice-h em cada modalidade de bolsa.

Modalidade	Posição	Índice-h	Área
1A	1º	72	Medicina
1A	2º	69	Saúde Coletiva
1A	3º	62	Bioquímica
1A		62	Farmacologia
1A	5º	61	Química
1B	1º	46	Física
1B		46	Física
1B	3º	42	Farmacologia
1B	4º	41	Medicina
1B	5º	40	Engenharia Elétrica
1C	1º	38	Física
1C	2º	37	Química
1C	3º	36	Odontologia
1C	4º	35	Química
1C	5º	34	Física
1D	1º	78	Física
1D	2º	58	Física
1D	3º	55	Física
1D	4º	42	Medicina
1D	5º	40	Agronomia
2	1º	53	Física
2	2º	50	Medicina
2	3º	38	Física
2	4º	37	Medicina
2	5º	36	Astronomia

Como pode ser observado, a modalidade 1A é a detentora dos maiores valores, como era de se esperar. No entanto, a modalidade 1D também se destaca com o maior índice-h dentre todos os bolsistas (78). Tal modalidade também apresenta média

superior à da modalidade 1B ao se comparar os bolsistas com os maiores valores do índice-h. Considerando as áreas de atuação desses bolsistas, a área de Física se destaca, uma vez que, com exceção da modalidade 1A, em todas as demais detém o maior valor do índice-h. Além disso, há uma grande representatividade das áreas de Ciências da Saúde, em particular da área de Medicina. Foi possível identificar ainda que, dentre os bolsistas com maiores valores do índice-h, independente da modalidade, todos possuem mais de 14 anos de doutorado concluído.

Como os valores do índice-h variam de uma área para outra, para verificar esta distinção, foi realizada uma análise do índice-h dos bolsistas em função de suas grandes áreas de atuação (ver Figura 7.18). Como pode ser observado, as maiores medianas correspondem às grandes áreas de Ciências Biológicas e Ciências da Saúde (14), e posteriormente a grande área de Ciências Exatas e da Terra (12). Observa-se ainda que as grandes áreas de Ciências Agrárias e Engenharias possuem taxas muito semelhantes, enquanto as grandes áreas de Ciências Humanas, Ciências Sociais Aplicadas e Linguística, Letras e Artes possuem medianas nulas. Dessas três últimas, a grande área de Ciências Sociais Aplicadas apresenta um número maior de *outliers*, o que eleva a sua média em detrimento das demais. Logo, percebe-se que os bolsistas da grande área de Ciências Biológicas possuem valores superiores aos das demais grandes áreas e que a grande área de Linguística, Letras e Artes possui os menores valores bem como a menor média dentre todas as demais. Com isso, aliado ao que já foi exposto, ao se avaliar a qualidade da produção dos pesquisadores, deve-se considerar as suas respectivas áreas de atuação, tendo em vista os perfis e as particularidades de cada uma delas.

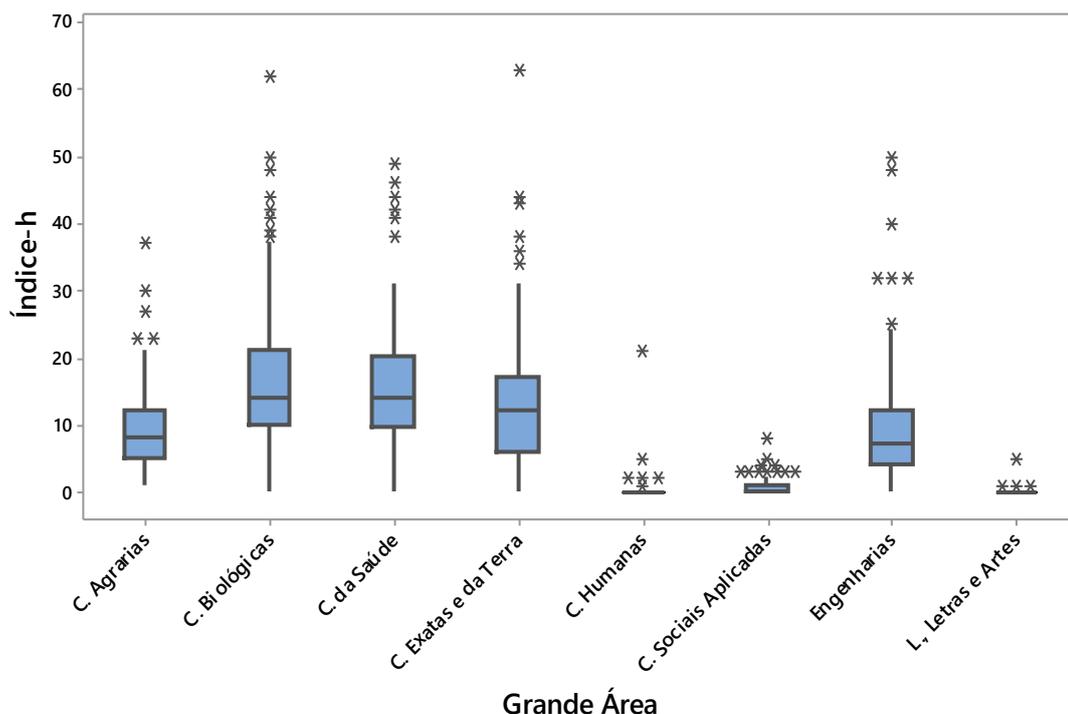


Figura 7.18: Análise dos valores dos índices-h dos bolsistas de produtividade em pesquisa de acordo com as respectivas grandes áreas.

7.4 Considerações Finais

Neste capítulo, pôde ser observado que o conjunto de docentes de pós-graduação e bolsistas de produtividade em pesquisa do CNPq correspondem a uma parcela considerável dos pesquisadores brasileiros responsáveis pela formação nos níveis mais altos de capacitação e que contribuem de forma significativa para a produção científica brasileira. Foi possível identificar como os docentes estão distribuídos pelas principais instituições de ensino do país, concentrados, em sua maioria, nas regiões Sudeste e Sul. Também foram observadas as principais áreas de atuação desses docentes, sendo Ciências Humanas, Ciências Exatas e da Terra e Ciências da Saúde as principais grandes áreas de atuação de um grande percentual dos docentes que atuam na pós-graduação. Como vários desses docentes orientarem trabalhos nos programas de pós-graduação em que atuam, foi constatado que existe uma correlação entre a quantidade de orientações e o número de publicações dos docentes.

Já considerando os bolsistas de produtividade em pesquisa, estes também estão, em sua maioria, concentrados nas regiões Sudeste e Sul, sendo a USP a instituição com a maior quantidade de bolsistas em todas as modalidades de bolsa. Além disso, como era de se esperar, o tempo de carreira desses bolsistas tende a aumentar conforme o nível de enquadramento de suas respectivas bolsas. Esses bolsistas atuam, em grande parte, nas grandes áreas de Ciências Exatas e da Terra ou Ciências Biológicas, sendo a grande área de Ciências Exatas e da Terra detentora da maior quantidade de bolsistas entre todas as modalidades.

A exemplo do que ocorre com a produção científica do conjunto de todos os doutores, há uma queda, também, na produção do conjunto de docentes de pós-graduação e bolsistas de produtividade em pesquisa, queda esta que mostra-se mais acentuada nos artigos em anais de congresso. Neste aspecto, destacam-se os bolsistas que, a partir de 2013, com a queda acentuada na publicação de artigos em anais de congresso, passam a ter como principal tipo de publicação artigos em periódico. Tal resultado parece ter relação direta com os critérios de avaliação da CAPES, conforme já discutido, que influenciam diretamente os grupos analisados neste capítulo, tendo em vista que estão vinculados a programas de pós-graduação e a produção científica de tais programas depende, diretamente, da produção de seus docentes.

Considerando as análises qualitativas realizadas com base nos artigos em periódico dos bolsistas, foi possível perceber que o FI dessas publicações tem aumentado nos últimos anos e que as publicações em periódicos com maior FI tendem a receber mais citações. Este fato está relacionado à maior visibilidade desses periódicos. O resultado disso é que, apesar de a quantidade de publicações ter diminuído nos últimos anos, a qualidade dos respectivos veículos tem aumentado, o que é uma indicação da melhoria da qualidade dos trabalhos produzidos, tendo em vista a dificuldade natural de se publicar em periódicos com maior fator de impacto. Entretanto, observou-se uma grande quantidade de artigos sem citações, dos quais um percentual significativo foi publicado em periódicos com baixo FI, comprovando que a qualidade do veículo de publicação contribui para a maior visibilidade dos trabalhos.

A quantidade de citações recebidas pelas publicações de um pesquisador é também uma forma efetiva de avaliar qualitativamente o nível de suas pesquisas. Verificando o índice-h dos bolsistas de produtividade em pesquisa, percebe-se que aqueles enquadrados nas modalidades mais altas da bolsa possuem índices-h mais altos, sendo que entre os bolsistas enquadrados na modalidade 2 há uma grande quantidade de *outliers*. Esses *outliers* possuem índices-h superiores às médias de outras modalidades, o que pode ser explicado pelo fato de ser esta a modalidade de enquadramento inicial dos bolsistas e de o índice-h não ser um critério explícito para a mudança do nível da bolsa. De forma geral, as grandes áreas que possuem os pesquisadores com os maiores valores do índice-h são as de Ciências Biológicas e Ciências da Saúde. Considerando as áreas específicas de atuação, destaca-se a Física cujos pesquisadores possuem os maiores valores do índice-h em quase todas as modalidades da bolsa. Por outro lado, pesquisadores de áreas como as que compõem a grande área de Linguística, Letras e Artes recebem poucas citações como reflexo direto do fato de seus trabalhos serem geralmente publicados em periódicos com baixo FI, principalmente em veículos nacionais. Diante disso, há de se ter critérios distintos quando se compara a produção científica de pesquisadores com atuações em áreas diferentes, para que, dessa forma, cada particularidade possa ser devidamente considerada.

Por fim, percebe-se que o perfil de publicação em cada área do conhecimento é distinto e isto influencia as análises quantitativas e qualitativas da produção científica dos respectivos pesquisadores. Entretanto, o importante é que, ao se constatar a melhoria geral da qualidade da produção científica brasileira, a queda do volume dessa produção é plenamente justificável pela melhoria do nível dos trabalhos produzidos.

Capítulo 8

Conclusões

Neste capítulo, são apresentadas as considerações finais desta tese, destacando-se a sumarização de suas principais contribuições (Seção 8.1). Diante dessas contribuições, são apresentadas as publicações resultantes do trabalho realizado, as submissões em andamento e os projetos que têm utilizado o arcabouço desenvolvido para extração e tratamento de dados da Plataforma Lattes (Seção 8.2). Por fim, são discutidas as possíveis extensões para trabalhos futuros (Seção 8.3).

8.1 Contribuições

O principal objetivo desta tese foi realizar uma análise da produção científica brasileira a partir de dados curriculares da Plataforma Lattes e, dessa forma, apresentar um retrato dessa produção nos últimos anos. Para isso, foi desenvolvido um arcabouço denominado LattesDataXplorer para extração, processamento e análise dos dados curriculares cadastrados na Plataforma Lattes. Pela riqueza dos dados que compõem os currículos da Plataforma Lattes, diversos trabalhos têm dedicado atenção em analisá-los com o objetivo de avaliar a produção científica de grupos específicos de pesquisadores brasileiros (MENA-CHALCO; CESAR-JUNIOR, 2013; DIGIAMPIETRI et al., 2014b; LIMA et al., 2015; TUESTA et al., 2015). Diante disso, utilizando o arcabouço desenvolvido, foi possível realizar uma caracterização geral dos dados contidos na Plataforma Lattes, com estatísticas gerais e por áreas/grandes áreas de atuação dos pesquisadores. Mais especificamente, também foram analisados os dados do conjunto de doutores com currículos cadastrados na Plataforma Lattes, bem como, dos docentes de pós-graduação e bolsistas de produtividade em pesquisa do CNPq.

Os resultados apresentados nesta tese mostram todo o potencial dos dados da Plataforma Lattes para a realização de análises baseadas tanto em métricas bibliométricas como em métricas de redes sociais. Essas análises possibilitaram extrair informações em diversas vertentes, tendo em vista a quantidade e diversidade dos dados coletados. Logo, foi possível visualizar um retrato da produção científica

brasileira considerando todo o conjunto de indivíduos com currículos cadastrados na Plataforma Lattes, apresentando informações que possibilitam analisar e compreender como a ciência brasileira é produzida e tem evoluído.

Particularmente, foi identificado que, apesar do crescimento constante da quantidade de currículos na Plataforma Lattes, um grande percentual desses currículos não possui dados que contribuam para um amplo estudo sobre a produção científica brasileira. Isso porque esses currículos contêm apenas dados básicos, como identificação pessoal dos indivíduos, em sua grande maioria alunos de graduação que, por exigências de editais para concessão de bolsas de iniciação científica ou intercâmbio acadêmico, cadastram-se apenas para fins de elegibilidade. Diante disso, apesar de a Plataforma Lattes possuir aproximadamente 4,8 milhões de currículos cadastrados, uma grande quantidade desses currículos é irrelevante para análises bibliométricas e análises baseadas em redes sociais, já que não incluem quaisquer dados sobre produção científica. Assim sendo, considerou-se apenas o conjunto de indivíduos com doutorado concluído como relevante para o estudo realizado sobre a produção científica brasileira, mesmo representando, apenas, 5% de todos os indivíduos com currículo cadastrado na plataforma.

Ao analisar a produção científica do conjunto de doutores, observou-se uma queda na quantidade de trabalhos publicados ao longo dos últimos anos, em todos os tipos de publicação. As maiores taxas de redução foram as de artigos publicados em anais de congresso e artigos publicados em periódico. A redução da quantidade de artigos publicados em periódico acompanha uma tendência mundial, quando comparada com a de países mais produtivos cientificamente. Com relação à redução da quantidade de artigos publicados em anais de congresso, com o intuito de identificar uma justificativa para a grande redução dos trabalhos publicados, várias hipóteses foram consideradas, sendo que a que mostrou-se mais consistente foi a influência dos critérios de avaliação dos programas de pós-graduação brasileiros utilizados pela CAPES, em particular as modificações introduzidas no Qualis. Nas análises realizadas, observou-se uma redução repentina desse tipo de produção após 2007, ano em que a CAPES passa a não mais considerar os artigos em anais de congresso nos indicadores sobre a produção científica. Uma análise considerando áreas representativas de cada grande área do conhecimento apontou que várias dessas áreas, como as áreas de Medicina e Agronomia, detentoras de grande quantidade de trabalhos neste tipo de produção, tiveram redução acentuada após 2007, diferentemente da área de Ciência da Computação, única área que não apresenta tal redução e ainda continua considerando este tipo de publicação na avaliação de seus programas de pós-graduação.

Ressalta-se, ainda, a redução na quantidade de doutores que têm publicado artigos recentemente. Embora a colaboração científica brasileira tenha aumentado de forma significativa, há uma redução de, aproximadamente, 16% na quantidade de autores de artigos em periódico nos últimos dois anos e de 55,96% no caso de autores de artigos em anais de congresso entre 2007 a 2014.

Ao avaliar a colaboração científica brasileira, destacam-se os pesquisadores das grandes áreas de Ciências da Saúde e Ciências Biológicas com os primeiros trabalhos em colaboração, com data de publicação no início da década de 1960. Tais trabalhos, de autoria de pesquisadores em sua maioria vinculados à USP, UNESP e UNICAMP,

fazem com que estas instituições sejam consideradas mais colaborativas, dada a quantidade de trabalhos realizados em colaboração por seus pesquisadores. Uma das principais razões para isso é a proximidade geográfica entre elas, bem como, a relação orientador-orientado, já que vários pesquisadores da UNESP e UNICAMP capacitaram-se na USP.

Tendo em vista caracterizar os principais pesquisadores do país, foi apresentado o perfil do conjunto de docentes dos programas de pós-graduação e dos bolsistas de produtividade em pesquisa do CNPq. Além da caracterização geral desses pesquisadores, sua produção científica também foi analisada, mostrando que, por serem parte do conjunto de indivíduos com doutorado concluído, suas produções científicas também apresentam redução, embora com elevada média de publicações em periódicos (52,81 no caso dos bolsistas). Consequentemente, com o intuito de avaliar a qualidade da produção científica brasileira, foram analisados os artigos em periódico dos bolsistas de produtividade em pesquisa do CNPq, já que representam uma parcela de pesquisadores com reconhecida qualidade em suas pesquisas. Os resultados mostram que, embora tem havido uma redução na quantidade de trabalhos publicados, a qualidade dos periódicos em que esses trabalhos têm sido publicados tem aumentado, considerando como critério, para isso, os seus respectivos FIs.

Além disso, ao verificar a quantidade de citações recebidas pelos artigos, percebe-se que aqueles publicados em periódicos de maior impacto tendem a receber mais citações. Finalmente, no caso de algumas grandes áreas, como as de Linguística, Letras e Artes e Ciências Sociais Aplicadas, cujos artigos em geral, possuem uma média muito baixa de citações, isso seria resultado da tendência de seus pesquisadores de publicar mais em periódicos nacionais e não indexados, dificultando, assim, a visibilidade de seus trabalhos pela comunidade científica internacional.

Assim sendo, os resultados obtidos nesta tese apontam a validade de suas proposições, além de indicar diversas possibilidades de novos estudos para continuidade e extensão do presente trabalho.

8.2 Publicações

A seguir, são apresentadas as produções científicas elaboradas a partir dos estudos realizados nesta tese.

Como resultado das pesquisas iniciais realizadas a partir dos dados curriculares extraídos da Plataforma Lattes, foram produzidos e publicados os seguintes artigos:

- DIAS, T. M. R. et al. Modelagem e Caracterização de Redes Científicas: Um Estudo Sobre a Plataforma Lattes. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 2., 2013. **Anais...** Maceió, 2013.
- _____; MOITA, G. F. Análise da Rede de Colaboração Científica dos Congressos Ibero Latino Americano de Métodos Computacionais em Engenharia Extraídos da Plataforma Lattes. In: IBERIAN LATIN AMERICAN CONGRESS ON

COMPUTATIONAL METHODS IN ENGINEERING, 34., 2013. **Anais...** Pirenópolis, 2013.

- DIAS, T. M. R.; DIAS, P. M.; MOITA, G. F. Analysis of the Collaboration Networks of Scientific Publications. In: CONTECSI - INTERNATIONAL CONFERENCE OF INFORMATION SYSTEMS AND TECHNOLOGY MANAGEMENTT – CONTECSI, 10., 2013. **Anais...** São Paulo, 2013.
- _____; MOITA, G. F.; DIAS, P. M. Um processo de extração de dados da web para identificação de colaboração científica em grandes bases de dados. In: IBERIAN LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, 35., 2014. **Anais...** Fortaleza, 2014.

Com a consolidação do processo de extração dos dados e objetivando a análise de grandes redes de colaboração, foi proposto um método de identificação de colaborações científicas com baixo custo computacional que gerou os seguintes artigos:

- DIAS, T. M. R.; MOITA, G. F. An efficient process for extraction and identification in scientific collaboration networks. In: WORLD CONGRESS ON COMPUTATIONAL MECHANICS, 11., 2014. **Anais...** Barcelona, 2014.
- _____; _____. A method for the identification of collaboration in large scientific databases. **Em Questão**, v. 21, p. 140-161, 2015.

Adicionalmente, relativos à proposta de construção e análise de redes de palavras-chave, foram publicados os seguintes artigos:

- DIAS, T. M. R.; MOITA, G. F. Caracterização e Análise de Redes de Palavras-chave em Repositórios de Publicações Científicas. In: SIMPÓSIO DE MECÂNICA COMPUTACIONAL, 11., 2014. **Anais...** Juiz de Fora, 2014.
- _____; _____. Identifying relevant keywords in scientific collaboration networks. In: WORLD CONGRESS ON COMPUTATIONAL MECHANICS, 11., 2014. **Anais...** Barcelona, 2014.

Como consequência do artigo publicado no BraSNAM 2013 (II Brazilian Workshop on Social Network Analysis and Mining), os autores receberam um convite para envio de uma versão estendida do artigo para o periódico iSys (Revista Brasileira de Sistemas de Informação). O artigo foi publicado na terceira edição, volume 7 de 2014:

- DIAS, T. M. R.; MOITA, G. F. Identificação e Caracterização de Redes Científicas de Dados Curriculares. **iSys: Revista Brasileira de Sistemas de Informação**, v. 7, n. 3, 2014.

Após alterações realizadas no arcabouço de coleta e tratamento dos dados, com a incorporação de elementos que tornaram o processo de coleta dos currículos e caracterização das redes de colaboração científica mais eficiente, um novo artigo

detalhando todos os passos e métodos utilizados foi publicado no periódico *Encontros Bibli* (Revista Eletrônica de Biblioteconomia e Ciência da Informação).

- DIAS, T. M. R.; MOITA, G. F. Adoção da Plataforma Lattes como fonte de dados para caracterização de redes científicas. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 21, n. 47, p. 16-26, 2016.

Além dos trabalhos que apresentaram o arcabouço de extração e a caracterização de dados da Plataforma Lattes, um artigo baseado nos resultados apresentados no Capítulo 5 encontra-se em elaboração e será submetido a um periódico de circulação internacional.

- DIAS, T. M. R.; MOITA, G. F.; LAENDER, A. H. F. An Overview of the Brazilian Scientific Production based on Data from the Lattes Platform.

Outros artigos descrevendo os resultados apresentados nos Capítulos 6 e 7 estão em fase de elaboração para submissão a periódicos de relevância na área, tendo em vista que são resultados importantes para compreensão da produção científica brasileira e, até então, inéditos dada a quantidade de dados analisados.

Objetivando explorar todo o potencial do arcabouço desenvolvido, outros trabalhos realizados por alunos do Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG com base em dados derivados desse arcabouço foram desenvolvidos, o que possibilitou a publicação de artigos apresentando novas linhas de pesquisa:

- AFFONSO, F.; DIAS, T. M. R.; DIAS, P. M. Uma Análise em Redes de Colaboração Científica para Identificação da Intensidade de Colaboração. In: WORKSHOP MINEIRO DE SISTEMAS DE INFORMAÇÃO, 2012. **Anais...**, Juiz de Fora, 2012.
- AFFONSO, F. et al. Identificação de Menores Caminhos em Redes Científicas. In: INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGY MANAGEMENT, 11., 2014. **Anais...**, São Paulo, 2014.
- DIAS, T. M. R.; MOITA, G. F.; DIAS, P. M. Estudo e Análise Bibliométrica dos Pesquisadores Brasileiros que Atuam com Métodos Computacionais. In: IBERIAN LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING, 36., 2015. **Anais...** Rio de Janeiro, 2015.
- DIAS, T. M. R. et al. Uma Plataforma para Análise e Visualização de Redes de Colaboração Científica Institucional. In: COMPUTER ON THE BEACH, 2016. **Anais...** Florianópolis, 2016.
- SANTOS, L. R. F. et al. Caracterização dos Bolsistas de Produtividade em Pesquisa do CNPq a partir de Dados da Plataforma Lattes. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016. **Anais...** São Paulo: USP, 2016.

- OLIVEIRA, R. S. et al. Um processo para análise de redes científicas institucionais a partir de dados de publicações científicas. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016. **Anais...** São Paulo: USP, 2016.
- MOREIRA, L. R. F.; DIAS, T. M. R.; MOITA, G. F. Uma estratégia baseada em árvores genealógicas científicas para visualização da relação orientador-orientado. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016. **Anais...** São Paulo: USP, 2016.

Como consequência dos artigos publicados no V EBBC (V Encontro Brasileiro de Bibliometria e Cientometria 2016), os autores receberam convites para o envio de versões estendidas dos artigos para uma edição especial do periódico **Em Questão**. Tais artigos foram submetidos e estão em processo de revisão.

Além disso, outros trabalhos foram e estão em desenvolvimento explorando assuntos que objetivam analisar os dados dos currículos cadastrados na Plataforma Lattes com auxílio do arcabouço desenvolvido, sendo eles:

- **Caracterização do Programa de Bolsas de Produtividade em Pesquisa do CNPq a partir de Dados da Plataforma Lattes.** Dissertação de Mestrado defendida em 2015 pelo aluno Leandro Roberto Ferreira dos Santos no Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG.
- **Genealogia Científica, um Processo para Compreensão da Relação Orientador-Orientado.** Dissertação de Mestrado em andamento do aluno Tales Henrique José Moreira no Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG.
- **Análise de Redes de Palavras-chave como Mecanismo de Identificação de Tópicos de Pesquisa Emergentes.** Tese de Doutorado em andamento do aluno Jether Oliveira Gomes no Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG.
- **Uma Análise Cientométrica dos Pesquisadores Brasileiros da Área Interdisciplinar.** Tese de Doutorado da aluna Patrícia Mascarenhas Dias no Programa de Pós-Graduação em Modelagem Matemática e Computacional do CEFET-MG.

8.3 Extensões para Possíveis Trabalhos Futuros

Para explorar o potencial e a riqueza dos dados da Plataforma Lattes, algumas extensões para possíveis trabalhos futuros são sugeridas:

- **Análise da rede de colaboração dos programas de pós-graduação de excelência no Brasil:** esta análise visaria verificar como os programas de pós-graduação avaliados como de excelência pela CAPES (conceito 7) colaboram e

também como é composta a rede de colaboração científica de seus docentes. Para isso, os dados dos docentes de todos os programas de pós-graduação com o conceito máximo de avaliação da CAPES, independentemente de sua área do conhecimento, seriam extraídos, tendo como base para essa extração a lista de docentes presente nos cadernos de indicadores, disponibilizados pela CAPES. Com isso, seria possível analisar todos os programas de excelência e, a partir da produção científica desses pesquisadores, caracterizar as redes de colaboração científica adotando-se diversas técnicas baseadas em análise de redes sociais. Dessa forma, seria possível compreender como os docentes desses programas colaboram e qual o padrão de publicação do conjunto de docentes analisado em cada uma das áreas a que os programas estão vinculados.

- **Um estudo sobre os tópicos de pesquisa em evidência:** neste estudo, as palavras-chave de publicações seriam analisadas objetivando extrair conhecimento sobre os principais tópicos de interesse em cada área de pesquisa. A análise de palavras-chave tem recebido atenção devido à possibilidade de representar o conteúdo de um trabalho sem se preocupar com questões semânticas, o que possibilita representar, de forma significativa, o contexto de todo trabalho em análise. Diante disso, diversas métricas seriam aplicadas a todo o conjunto de publicações dos currículos da Plataforma Lattes, incluindo a modelagem de redes de palavras-chave, que possibilitam visualizar como palavras estão relacionadas, qual a frequência desses relacionamentos e quais seriam as palavras-chave mais influentes no conjunto analisado.
- **Análise comparativa sobre as diversas áreas de pesquisa:** neste estudo, seria realizada uma análise com o objetivo de traçar o perfil de cada uma das áreas de atuação dos pesquisadores brasileiros. Seriam analisados os perfis de cada conjunto de pesquisadores de uma determinada área considerando a sua proficiência, nível de formação, distribuição geográfica, padrão de publicação e orientações, dentre outros aspectos, que possibilitariam uma visão única e inédita das diversas áreas de pesquisa no Brasil e de como elas estariam evoluindo.
- **Análise da Mobilidade dos Pesquisadores no Cenário Científico Brasileiro:** neste estudo, o intuito seria analisar os dados da naturalidade de todos os indivíduos com currículos cadastrados na Plataforma Lattes e, a partir dos dados de formação acadêmica e vínculos profissionais, verificar como eles têm se deslocado em seu processo de formação e atuação. Tal estudo, possibilitaria obter uma visão sobre como ocorre o processo de formação dos indivíduos nos diversos níveis de capacitação acadêmica e profissional, identificando os maiores fluxos de deslocamento e formação, o que possibilitaria auxiliar as políticas de investimento na capacitação dos brasileiros.

Esses estudos teriam como objetivo mostrar o potencial que os dados da Plataforma Lattes oferecem para análise de diversas áreas de pesquisa. Todos os resultados a serem apresentados seriam de caráter inédito, tendo em vista a abrangência das análises que seriam realizadas. Por sua vez, isso corroboraria o potencial dos dados da Plataforma Lattes para a realização de diversos outros estudos de interesse da ciência brasileira.

Referências

ABBASI, A.; ALTMANN, J.; HOSSAIN, L. Identifying the effects os co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. **Informetrics**, v. 5, p. 594-607, 2011.

ADAMIC, L. The Small World Web. **European Research and Advanced Technology for Digital Libraries**. Springer Berlin Heidelberg, v. 1696, cap. 27. p. 443-452, 1999.

ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. **LattesMiner**: a multilingual DSL for information extraction from Lattes Platform. In: PROCEEDINGS OF THE COMPILATION OF THE CO-LOCATED WORKSHOPS ON DSM'11, TMC'11, AGERE!'11, AOPES'11, NEAT'11; VMIL'11. Portland, Oregon, USA: ACM, p. 85-92. 2011.

ARAÚJO, C. A. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, v. 12, n. 1, p. 11-32, jan./jun. 2006.

ARAÚJO, E. B. et al. Collaboration networks from a large cv database: dynamics, topology and bonus impact. **PLoS one**, v. 9, n. 3, p. e90537, 2014.

ARRUDA, D. et al. Brazilian computer science research: Gender and regional distributions. **Scientometrics**, v. 79, n. 3, p. 651-665, 2009.

BARABÁSI, A. L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, n. 5439, p. 509-512, 1999.

_____; _____; JEONG, H. Scale-free characteristics of random networks: the topology of the world-wide Web. **Physica A: Statistical Mechanics and its Applications**, v. 281, n. 1, p. 69-77, 2000.

_____. et al. Evolution of the social network of scientific collaborations. **Physica A: Statistical mechanics and its applications**, v. 311, n. 3, p. 590-614, 2002.

BARABÁSI, A. L.; OLTVAI, Z. N. Network biology: understanding the cell's functional organization. **Nature Reviews Genetics**, v. 5, n. 2, p. 101-113, 2004.

_____. **Linked (conectado)**: a nova ciência dos networks. São Paulo: Leopardo, 2009.

BENEVENUTO F.; LAENDER, A. H. F.; ALVES, B. L. How Connected are the ACM SIG Communities ?. **SIGMOD Record**, v. 44, n. 4, p. 57-63, 2015.

BHATTACHARYA, S.; BASU, P. Mapping a research area at the micro level using co-word analysis. **Scientometrics**, v. 43, n. 3, p. 359-372, 01 nov. 1998.

BOAVENTURA, M. et al. Caracterização Temporal das Redes de Colaboração Científica nas Universidades Brasileiras: Anos 2000-2013. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 3, 2014, **Anais ...** Brasília, 2014.

BRANDÃO, M. A. et al. Using link semantics to recommend collaborations in academic social networks. In: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON WORLD WIDE WEB COMPANION, 22, 2013, **Anais ...**, Rio de Janeiro, p. 833-840, 2013.

_____; DINIZ, A. M.; MORO, M. M. Using Topological Properties to Measure the Strength of Co-authorship Ties. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 5, 2016, **Anais ...** Porto Alegre, 2016.

BRITO, A. G. C.; QUONIAM, L.; MENA-CHALCO, J. P. Exploração da Plataforma Lattes por assunto: proposta de metodologia. **TransInformação**, v. 28, n. 1, p. 77-86, 2016.

CALLON, M.; COURTIAL, J. P.; LAVILLE, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. **Scientometrics**, v. 22, n. 1, p. 155-205, 01 nov. 1991.

CANDOTTI, E. **Ciência na Educação Popular**. Ciência e público: caminhos da divulgação científica no Brasil. Rio de Janeiro: Casa da Ciência-Centro Cultural de Ciência e Tecnologia da UFRJ. 2002.

CAÑIBANO, C.; BOZEMAN, B. Curriculum vitae method in science policy and research evaluation: the state-of-the-art. **Research Evaluation**, v. 18, n. 2, p. 86-94, 2009.

CARDOSO, O. N. P.; MACHADO, R. T. M. Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras. **Revista de Administração Pública**, v. 42, n. 3, p. 495-528, 2008.

CARRILHO-JUNIOR, J. R. **Desenvolvimento de uma Metodologia para Mineração de Textos**. 2007. 113 p. (Mestrado). Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro, 2007.

CHAIMOVICH, H. Brasil, ciência, tecnologia: alguns dilemas e desafios. **Estudos Avançados**, v. 14, n. 40, p. 134-143, 2000.

CHRISTEN, P. **Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection**. Springer, Berlin, 2012.

COLIZZA, V. et al. Detecting rich-club ordering in complex networks. **Nature physics**, v. 2, n. 2, p. 110-115, 2006.

CUNHA, M. D. V. et al. Redes de títulos de artigos científicos variáveis no tempo. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 2, **Anais...** Maceió, 2013.

DIAS, T. M. R. et al. Modelagem e Caracterização de Redes Científicas: Um Estudo Sobre a Plataforma Lattes. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 2, **Anais ...** Maceió, 2013.

_____; MOITA, G. F. A method for the identification of collaboration in large scientific databases. **Em Questão**, v. 21, p. 140-161, 2015.

DIGIAMPIETRI, L. A. et al. Minerando e caracterizando dados de currículos lattes. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 1, 2012, **Anais...** Curitiba, 2012.

_____; SANTIAGO, C. R. N.; ALVES, C. M. Predição de coautorias em redes sociais acadêmicas: um estudo exploratório em Ciência da Computação. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 2, 2013, **Anais...** Maceió, 2013.

_____; MUGNAINI, R.; ALVES, C. Analysis of Participation in Supervised Production of Advisors: A Case Study in Computer Science. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 2, 2013, **Anais ...** Maceió, 2013.

_____. et al. Análise da Rede dos Doutores que Atuam em Computação no Brasil. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 3, 2014, **Anais ...** Brasília, 2014a.

_____. et al. BraX-Ray: An X-Ray of the Brazilian Computer Science Graduate Programs. **PLoS ONE**, v. 9, n. 4, 2014b.

_____. **Análise da Rede Social Acadêmica Brasileira**. (Livre Docência). Escola de Artes Ciências e Humanidades da Universidade de São Paulo. São Paulo, 2015.

DING, Y.; CHOWDHURY, G. G.; FOO, S. Bibliometric cartography of information retrieval research by using co-word analysis. **Information Processing & Management**, v. 37, n. 6, p. 817-842, 2001.

DING, Y. Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. **Informetrics**, v. 5, n. 1, p. 187-203, 2011.

FADIGAS, I. et al. Análise de redes semânticas baseada em títulos de artigos de periódicos científicos: o caso dos periódicos de divulgação em educação matemática. *Educação Matemática Pesquisa*. **Revista do Programa de Estudos Pós-Graduados em Educação Matemática**, v. 11, 2010.

_____; PEREIRA, H. A network approach based on cliques. **Physica A: Statistical Mechanics and its Applications**, v. 392, n. 10, p. 2576-2587, 2013.

FALOUTSOS, M.; FALOUTSOS, P.; FALOUTSOS, C. On power-law relationships of the internet topology. **ACM SIGCOMM Computer Communication Review**, v. 29, n. 4, ACM, 1999.

FERNANDES, G. O.; SAMPAIO, J. O.; SOUZA, J. M. XMLattes - A Tool for Importing and Exporting Curricula Data. In: THE 2011 WORLD CONGRESS IN COMPUTER SCIENCE, COMPUTER ENGINEERING, AND APPLIED COMPUTING - WORLDCOMP, 2011, **Anais...** Las Vegas, Nevada, USA, 2011.

FERRAZ, R. R. N.; QUONIAM, L.; MACCARI, E. A. The use of ScriptLattes tool for extraction and on line availability of academic production from a departament of stricto sensu in management. In: INTERNATIONAL CONFERENCE ON INFORMATION SYSTEMS AND TECHNOLOGY MANAGEMENT, 11, São Paulo, 2014, **Anais...** São Paulo, p. 663-679, 2014.

FERREIRA, A. G. C. Bibliometria na avaliação de periódicos científicos. **Data Grama Zero**, v. 11, n. 3, p. 1-9, 2010.

FERREIRA, A. A.; GONÇALVES, M. A.; LAENDER, A. H. F. A Brief Survey of Automatic Methods for Author Name Disambiguation. **SIGMOD Record**, v. 41, n. 2, p. 15-26, 2012.

FURTADO, C. A. et al. A Spatiotemporal Analysis of Brazilian Science from the Perspective of Researchers' Career Trajectories. **PloS one**, v. 10, n. 10, p. e0141528, 2015.

GARFIELD, E. Citation indexes for science. A new dimension in documentation through association of ideas. **Science**, v. 122, p. 1123–1127, 1955.

_____. The use of JCR and JPI in Measuring Short and Long Term Journal Impact. In: **The Scientist**. Council of Scientific Editors Annual Meeting, May 9, 2000.

GLÄNZEL, W.; LETA, J.; THUIS, B. Science in Brazil. Part 1: A macro-level comparative study. **Scientometrics**, v. 67, n. 1, p. 67-86, 2006.

GOFFMAN, W.; NEWILL, V. A. Generalization of Epidemic Theory: An Application to the Transmission of Ideas. **Nature**, v. 204, n. 4955, p. 225-228, 1964.

_____; _____. Communication and Epidemic Processes. **Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences**, v. 298, n. 1454, p. 316-334, 1967.

GUEDES, V. L. S.; BORSCHIVER, S. Bibliometria: Uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 2005, **Anais...** Florianópolis, 2005.

HAYASHI, M. C. P. I. Sociologia da Ciência, Bibliometria e Cientometria: Contribuições para a Análise da Produção Científica. In: SEMINÁRIO DE EPISTEMOLOGIA E TEORIAS DA EDUCAÇÃO, 2012, **Anais...** São Paulo, 2012.

HECKNER, M.; MÜHLBACHER, S.; WOLFF, C. Tagging tagging: analysing user keywords in scientific bibliography management systems. **Journal of Digital Information (JODI)**, v. 9, n. 2, 2008.

HIRSCH, J. E. An index to quantify an individual's scientific research output. **Proceedings of the National academy of Sciences of the United States of America**, v. 102, n. 46, p. 16569-16572, 2005.

HOFFNAGEL, J. C. A prática de citação em trabalhos acadêmicos. **Cadernos de Linguagem e Sociedade**, v. 10, n. 1, p. 71, 2009.

HOU, H.; KRETSCHMER, H.; LIU, Z. The structure of scientific collaboration networks in Scientometrics. **Scientometrics**, v. 75, n. 2, p. 189-202, 2008.

HUANG, J. et al. Collaboration over time: characterizing and modeling network evolution. In: PROCEEDINGS OF THE 2008 INTERNATIONAL CONFERENCE ON WEB SEARCH AND DATA MINING, ACM, p. 107-116, 2008.

INOMATA, D. O. et al. Mapeamento de conhecimento: Identificação de palavras através de coocorrência. **RDBCi: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 13, n. 2, p. 279-297, 2015.

JORIO, A. et al. Brazilian science towards a phase transition. **Nature materials**, v. 9, n. 7, p. 528-531, 2010.

KAUER, A.; MOREIRA, V. Evolução dos Temas de Interesse do SBBB ao Longo dos Anos. In: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 2013, **Anais...**, Recife, 2013.

- KE, H. R.; CHEN, Y. N. Structure and pattern of social tags for keyword selection behaviors. **Scientometrics**, v. 92, n. 1, p. 43-62, 2012.
- KUMAR, R.; NOVAK, J.; TOMKINS, A. Structure and Evolution of Online Social Networks. In: YU, P. S.; HAN, J.; FOLOUTSOS, C. (Ed.). **Link Mining: Models, Algorithms, and Applications**. New York: Springer. Cap. 13, p.337-357, 2010.
- KUMAR, S. Co-authorship networks: a review of the literature. **Aslib Journal of Information Management**, v. 67, n. 1, p. 55-73, 2015.
- LAENDER, A. H. F. et al. Assessing the research and education quality of the top Brazilian Computer Science graduate programs. **SIGCSE Bulletin**, v. 40, n. 2, p. 135-145, 2008.
- _____; et al. CiênciaBrasil – The Brazilian Portal of Science and Technology. In: SEMINÁRIO INTEGRADO DE SOFTWARE E HARDWARE, 38, 2011, **Anais...** Natal, 2011.
- LANE, J. Let's make science metrics more scientific. **Nature**, v. 464, n. 7288, p. 488-489, 2010.
- LÉDA, D. B.; MANCEBO, D. REUNI: heteronomia e precarização da universidade e do trabalho docente. **Educação & Realidade**, v. 34, n. 1, p. 49-59, 2009.
- LEE, P. C.; SU, H. N.; CHAN, T. Y. Assessment of ontology-based knowledge network formation by Vector-Space Model. **Scientometrics**, v. 85, n. 3, p. 689-703, 2010.
- LEMIEUX, V.; OUIMET, M. **Análise estrutural das redes sociais**. Lisboa: Instituto Piaget. 128 p, 2008.
- LETA, J.; MEIS, L. A profile of science in Brazil. **Scientometrics**, v. 35, n.1, p.33-44, 1996.
- _____; GLÄNZEL, W.; THIJIS, B. Science in Brazil. Part 2: Sectoral and institutional research profiles. **Scientometrics**, v. 67, n. 1, p. 87-105, 2006.
- _____; THIJIS, B.; GLÄNZEL, W. A macro-level study of science in Brazil: seven years later. **Encontros Bibli**, v. 18, n. 36, p. 51-66, 2013.
- LIBEN-NOWELL, D.; KLEINBERG, J. The link-prediction problem for social networks. **Journal of the American Society for Information Science and Technology**, v. 58, n. 7, p. 1019-1031, 2007.
- LIMA, H. et al. Aggregating productivity indices for ranking researchers across multiple areas. In: PROCEEDINGS OF THE 13TH ACM/IEEE-CS JOINT CONFERENCE ON DIGITAL LIBRARIES, ACM, p. 97-106, 2013.
- _____. et al. Assessing the profile of top Brazilian computer science researchers. **Scientometrics**, v. 103, n. 3, p. 879-896, 2015.

Liu, B. **Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data**. Second Edition, Springer, Berlin, 2011.

LOPES, G. R.; et al. Ranking Strategy for Graduate Programs Evaluation. In: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY AND APPLICATIONS - IEEE, 7, **Anais...** Austrália, p. 59-64, 2011.

_____. **Avaliação e Recomendação de Colaborações em Redes Sociais Acadêmicas**. 2012. 128p. (Doutorado). Instituto de Informática UFRGS, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

_____; et al. Scientific Collaboration in Research Networks: A Quantification Method by Using Gini Coefficient. **International Journal of Computer Science and Applications**, v. 9, n. 2, p. 15-31, 2012.

MALTRÁS, B. B. **Los indicadores bibliométricos: fundamentos y aplicación al análisis de la ciencia, Documentación y gestión de la información**, Trea, 2003.

MEADOWS, A. J.; LEMOS, A. A. B. **A comunicação científica**. Brasília: Briquet de Lemos/livros, 1999.

MENA-CHALCO, J. P.; CESAR-JUNIOR, R. M. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31-39, 2009.

_____; DIGIAMPIETRI, L. A.; CESAR-JUNIOR, R. M. Caracterizando as redes de coautoria de currículos Lattes. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 1, 2012, **Anais...** Curitiba, 2012.

_____; DIGIAMPIETRI, L. A.; OLIVEIRA, L. B. Perfil de produção acadêmica dos programas brasileiros de pós- graduação em Ciência da Computação nos triênios 2004-2006 e 2007-2009. **Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS**, Porto Alegre, v. 18, 2012.

_____; CESAR-JUNIOR, R. M. Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. In: HAYASHI, M. C. P. I.; LETA, H. E. J. (Orgs.). **Bibliometria e Cientometria: reflexões teóricas e interfaces**. São Carlos: Pedro & João, p. 109-128, 2013.

_____. et al. Brazilian bibliometric coauthorship networks. **Journal of the Association for Information Science and Technology**, v. 65, n. 7, p. 1424-1445, 2014.

MENEGHINI, R. The key role of collaborative work in the growth of Brazilian science in the last ten years. **Scientometrics**, v. 35, n. 3, p. 367-373, 1996.

- MENEZES, V. S. D. A. **Análise de Redes Sociais Científicas**. 2012. (Tese de Doutorado). Programa de Engenharia de Sistemas e Computação, Universidade Federal do Rio de Janeiro, UFRJ/COPPE, 2012.
- MILGRAM, S. The small world problem. **Psychology today**, v. 2, n. 1, p. 60-67, 1967.
- MISLOVE, A. et al. Measurement and analysis of online social networks. In: PROCEEDINGS OF THE 7TH ACM SIGCOMM CONFERENCE ON INTERNET MEASUREMENT, 7, San Diego, 2007, **Anais...** San Diego, California, USA, ACM. p. 29-42, 2007.
- MONTOLIO, S. L.; DOMINGUEZ, D.; LARRIBA-PEY, J. L. Research endogamy as an indicator of conference quality. **SIGMOD Record**, v. 42, n. 2, p. 11-16, 2013.
- MOREL, C. M.; MOREL, R. L. M. Estudo sobre a produção científica brasileira, segundo os dados do ISI. **Ciência da Informação**, v. 6, n. 2, p. 99-109, 1978.
- MUGNAINI, R.; JANNUZZI, P.; QUONIAM, L. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. **Ciência da Informação**, v. 33, n. 2, p. 123-131, 2004.
- _____; LEITE, P.; LETA, J. Fontes de informação para análise de internacionalização da produção científica brasileira. **PontodeAcesso**, v. 5, n. 3, p. 87-102, 2011.
- _____. 40 anos de bibliometria no Brasil: da bibliografia estatística à avaliação da produção científica nacional. In: HAYASHI, M. C. P. I.; LETA, J. (Eds.). **Bibliometria e cientometria: reflexões teóricas e interfaces**. (Editorial Científico-Técnica ed., pp. 37-58). São Carlos: Pedro e João Editores, 2013.
- _____. et al. Comunicação científica no Brasil (1998-2012): indexação, crescimento, fluxo e dispersão. **Transinformação**, v. 26, n. 3, p. 239-252, 2014.
- NAVARRO, G. A guided tour to approximate string matching. **ACM Comput. Surv.**, v. 33, n. 1, p. 31-88, 2001.
- NEWMAN, M. E. J. The structure of scientific collaboration networks. **Proceedings of the National Academy of Sciences**, v. 98, n. 2, p. 404-409, 2001a.
- _____. Scientific collaboration networks.I. Network construction and fundamental results. **Physical Review E**, v. 64, n. 1, p. 016131_1-016131_8, 2001b.
- _____. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. **Physical Review E**, v. 64, n. 1, p. 016132_1-016132_7, 2001c.
- _____. The Structure and Function of Complex Networks. **SIAM Review**, v. 45, n. 2, p. 167-256, 2003.

NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration. **Proceedings of the National Academy of Sciences**, v. 101, suppl 1, p. 5200-5205, 2004.

_____. **Networks: an introduction**. Oxford University Press, 2010.

NISHIMURA, S. P. A precarização do trabalho docente como necessidade do capital: um estudo sobre o REUNI na UFRGS. **Universidade e Sociedade**, n. 53, p. 48-59, 2014.

OTTE, E.; ROUSSEAU, R. Social network analysis: a powerful strategy, also for the information sciences. **Journal of Information Science**, n. 28, 2002.

PACHECO, R. C. et al. Uma análise da pesquisa em engenharia e ciências mecânicas no Brasil a partir dos dados da Plataforma Lattes. Rio de Janeiro, **ABCM Engenharia**, v. 12, n. 1, p. 16-22, 2007.

PEREIRA, H. B. B. et al. Semantic networks based on titles of scientific papers. **Physica A: Statistical Mechanics and its Applications**, v. 390, n. 6, p. 1192-1197, 15 mar. 2011.

PEREZ-CERVANTES, E. et al. Using link prediction to estimate the collaborative influence of researchers. In: IEEE 9TH INTERNATIONAL CONFERENCE ON ESCIENCE (ESCIENCE), 9. **Anais...** China, Beijin, p. 293-300, 2013.

PETERSEN, A. M. et al. Persistence and uncertainty in the academic career. **Proceedings of the National Academy of Sciences**, v. 109, n. 14, p. 5213-5218, 2012.

PODLUBNY, I. Comparison of Brazilian impact expressed by the numbers of citations in different fields of science. **Scientometrics**, v. 64, n. 1, p. 95-99, 2005.

PRITCHARD, A. Statistical bibliography or bibliometrics? **Journal of Documentation**, v. 4, n. 25, p. 348-349, 1969.

REVOREDO, K. et al. Mining scientific literature for analysis of collaboration in research communities. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 1, Curitiba, **Anais...** Curitiba, 2012.

ROOS, D. H. et al. Brazilian scientific production in areas of biological sciences: a comparative study on the modalities of full doctorate in Brazil or abroad. **Scientometrics**. v. 98, n. 1, p. 415-427, 2014.

SIDONE, O. J. G.; HADDAD, E. A.; MENA-CHALCO, J. P. A Ciência nas Regiões Brasileiras: Evolução da Produção e das Redes de Colaboração Científica. **Transinformação**. v. 28, n. 1, p. 15-31, 2016.

SILVA, M. R.; HAYASHI, C. R. M.; HAYASHI, M. C. P. I. Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. **InCID: Revista de Ciência da Informação e Documentação**. v. 2, n. 1, p. 110-129, 2011.

- SILVA, T. H. P. et al. The Impact of Academic Mobility on the Quality of Graduate Programs. **D-Lib Magazine**. v. 22, n. 9/10, 2016.
- SONNENWALD, D. H. Scientific collaboration. **Annual Review of Information Science and Technology**, v. 41, n. 1, p. 643-681, 2007.
- SOUZA, J. et al. Análise de redes de palavras baseada em títulos extraídos de um sistema de atendimento. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 3, Brasília, **Anais ...**, Brasília, 2014.
- SCOTT, J. **Social network analysis: a handbook**. 2. ed. London: SAGE, 2009.
- STRÖELE, V.; ZIMBRÃO, G.; SOUZA, J. M. Análise de redes sociais científicas: modelagem multi-relacional. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 1, Curitiba, **Anais ...** Curitiba, 2012.
- SZWARCFITER, J. L. **Grafos e algoritmos computacionais**. 2. ed. Rio de Janeiro: Campus, 1986.
- THOMAZ, P. G.; ASSAD, R. S.; MOREIRA, L. F. P. Uso do Fator de impacto e do índice H para avaliar pesquisadores e publicações. **Arq. Bras. Cardiol.**, São Paulo, v. 96, n. 2, p. 90-93, Feb. 2011.
- TUESTA, E. F. et al. Analysis of an Advisor-Advisee Relationship: Na Exploratory Study of the Area of Exact and Earth Sciences in Brazil. **Plos ONE**, v. 10, n. 5, 2015.
- VANTI, N. A. P. Da Bibliometria à Webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, v. 31, n. 2, p. 152-162, 2002.
- WANDERLEY, A. J. et al. Identificando correlações entre métricas de Análise de Redes Sociais e o *h-index* de pesquisadores de Ciência da Computação. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 3, Brasília, **Anais...** Brasília, 2014.
- WASSERMAN, S.; FAUST, K. **Social network analysis: methods and applications**. 19. ed. [S.l.]: Social network analysis: methods and applications, 2009.
- ZANOTTO, E. D. Scientific and technological development in Brazil. The widening gap. **Scientometrics**, v. 55, n. 3, p. 411-419, 2002.
- ZHU, D. et al. Small-word phenomenon of keywords network based on complex network. **Scientometrics**, v. 97, n. 2, p. 435-442, 2013.

Apêndices

Neste apêndice são apresentados os gráficos com a produção científica em cada uma das áreas do conhecimento, bem como o percentual de crescimento dessa produção nos últimos anos. E ainda, as médias de citações dos artigos em periódico para cada grande área.

Ciências Agrárias - Artigos em Anais de Congresso

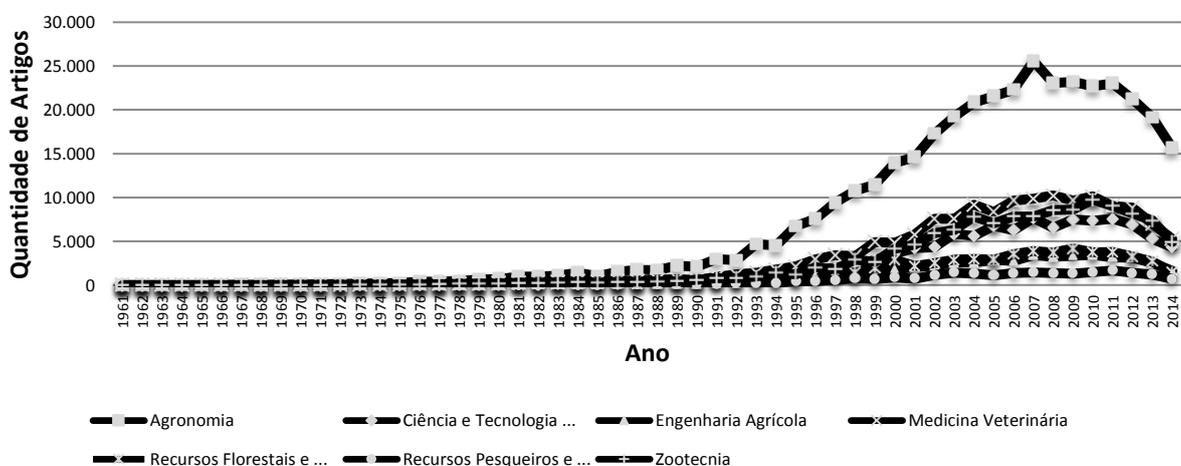


Figura A.1: Quantidade de artigos em anais de congresso – Ciências Agrárias.

Ciências Agrárias - Artigos em Periódico

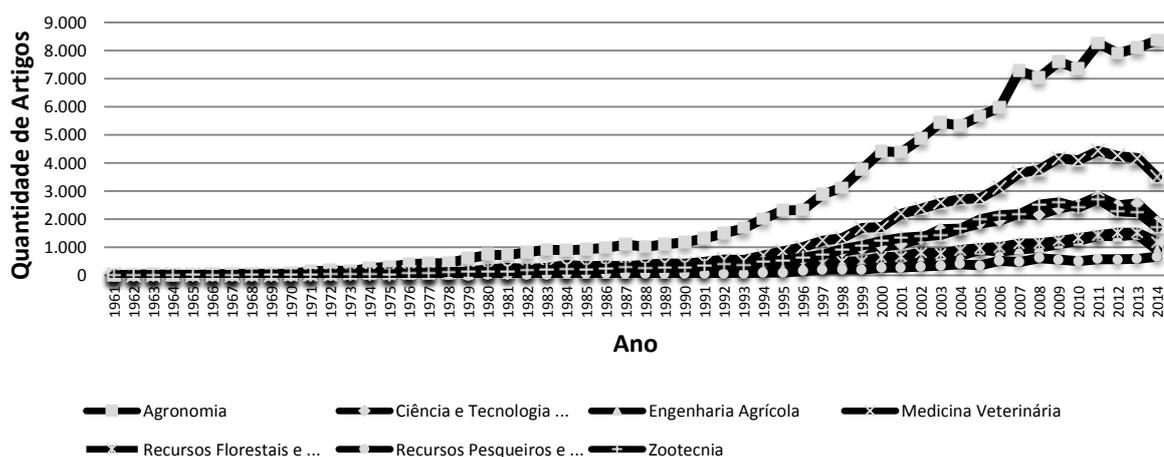


Figura A.2: Quantidade de artigos em periódico – Ciências Agrárias.

Ciências Biológicas - Artigos em Anais de Congresso

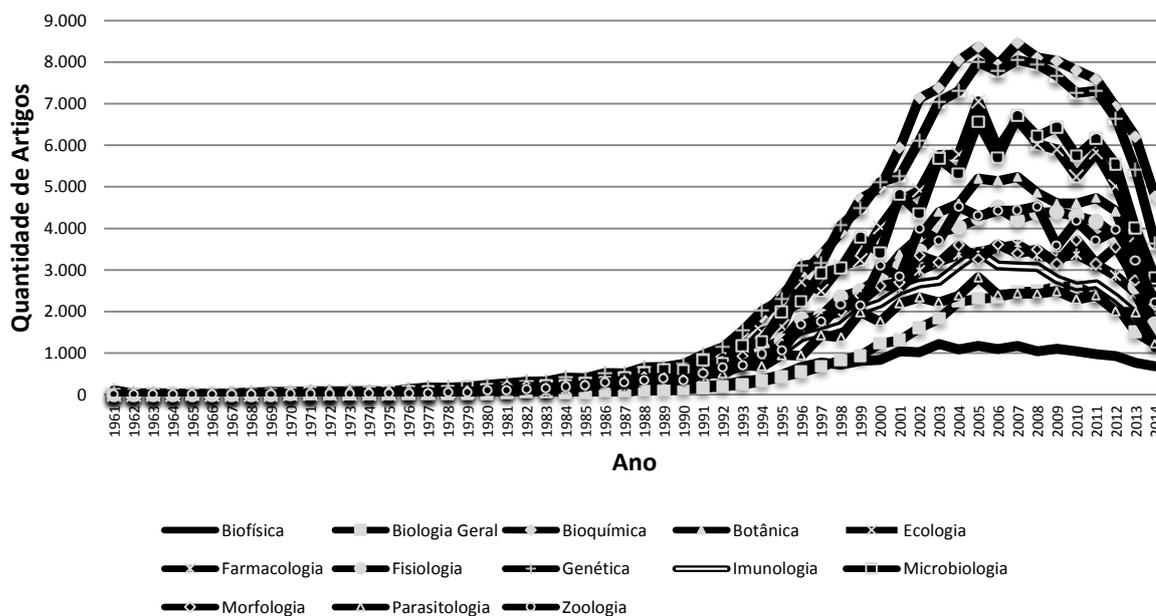


Figura A.3: Quantidade de artigos em anais de congresso – Ciências Biológicas.

Ciências Biológicas - Artigos em Periódico

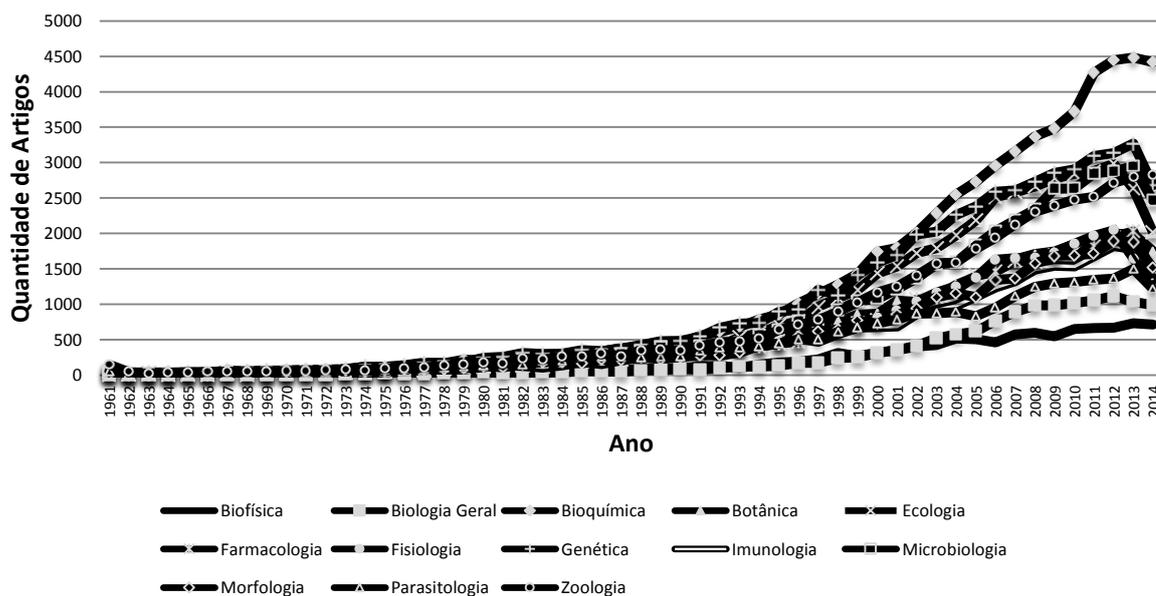


Figura A.4: Quantidade de artigos em periódico – Ciências Biológicas.

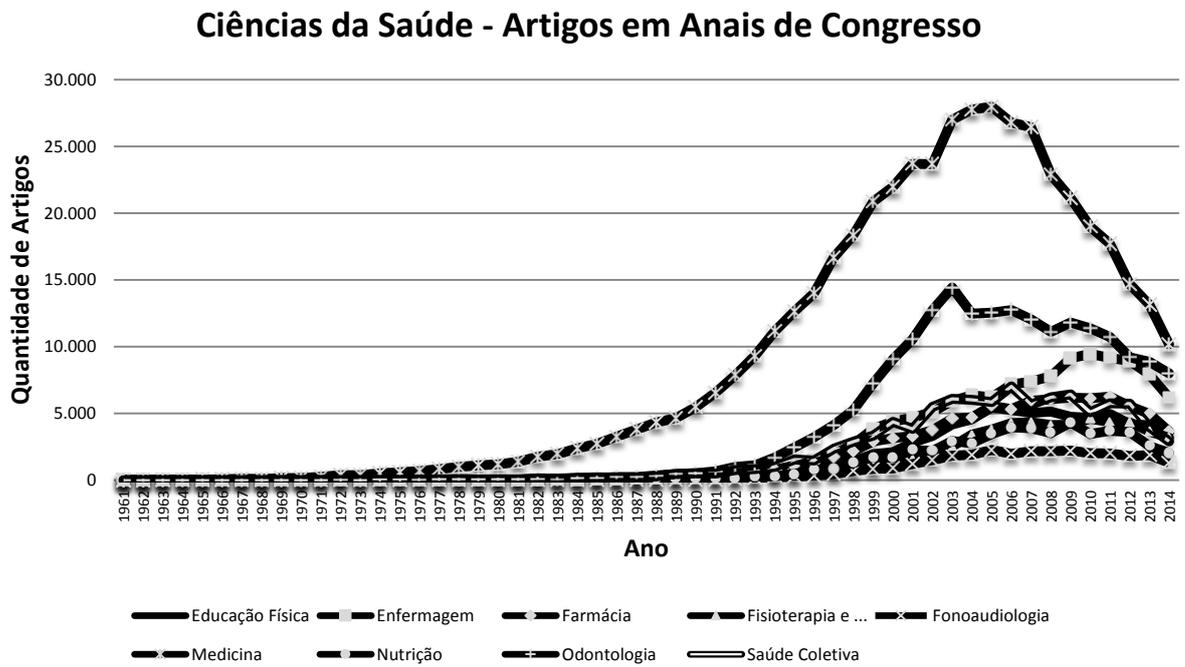


Figura A.5: Quantidade de artigos em anais de congresso – Ciências da Saúde.

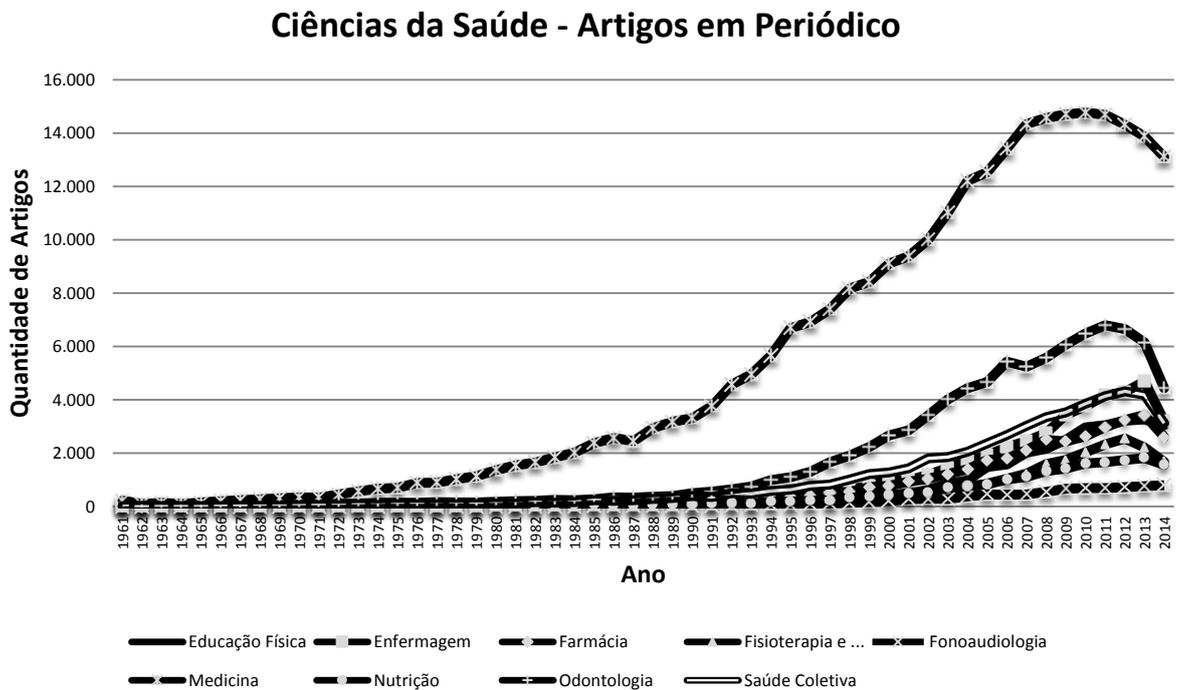


Figura A.6: Quantidade de artigos em periódico – Ciências da Saúde.

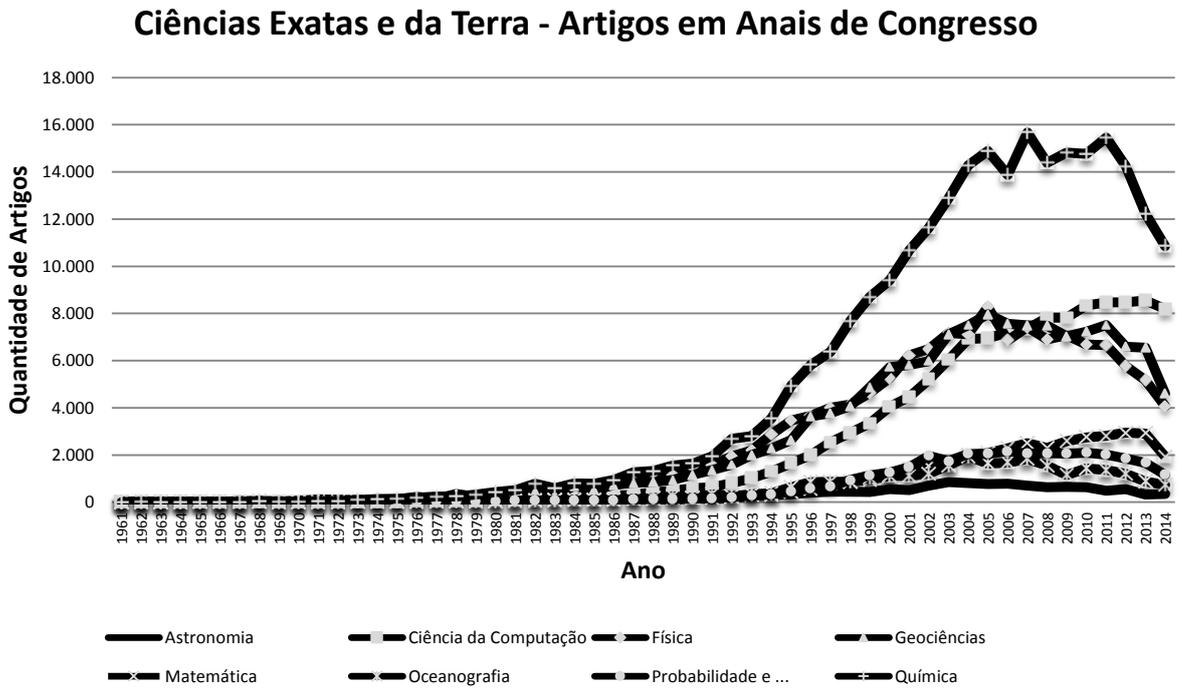


Figura A.7: Quantidade de artigos em anais de congresso – Ciências Exatas e da Terra.

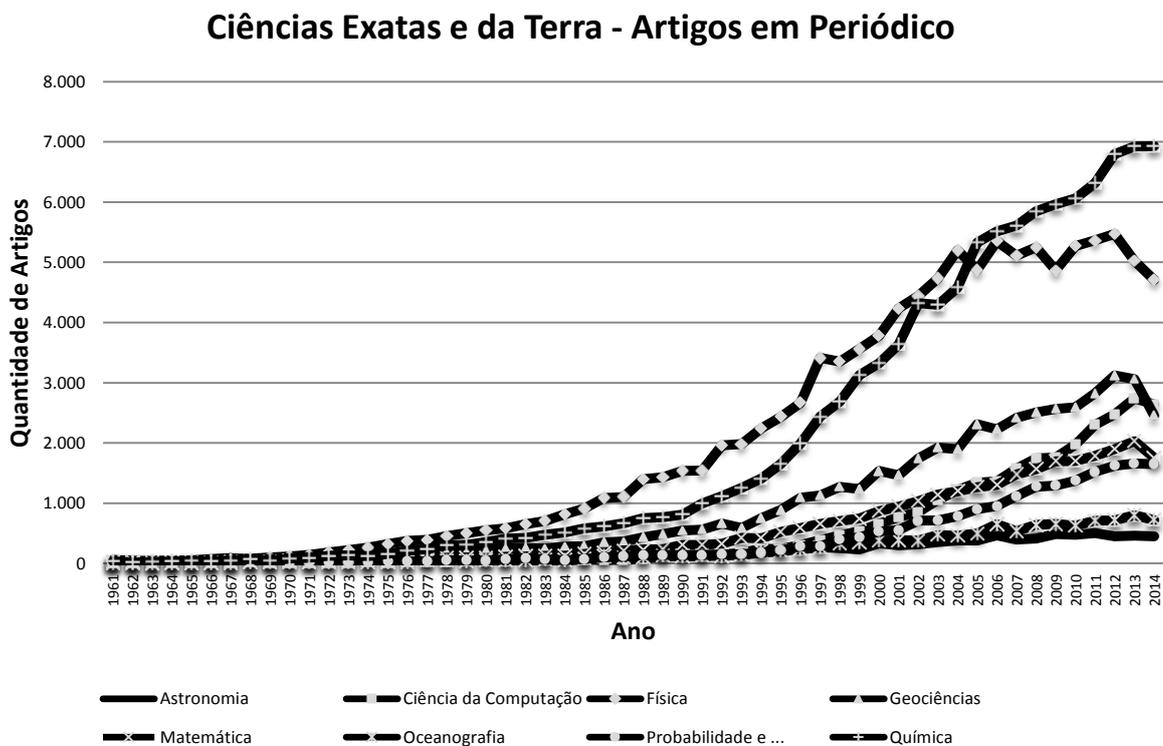


Figura A.8: Quantidade de artigos em periódico – Ciências Exatas e da Terra.

Ciências Humanas - Artigos em Anais de Congresso

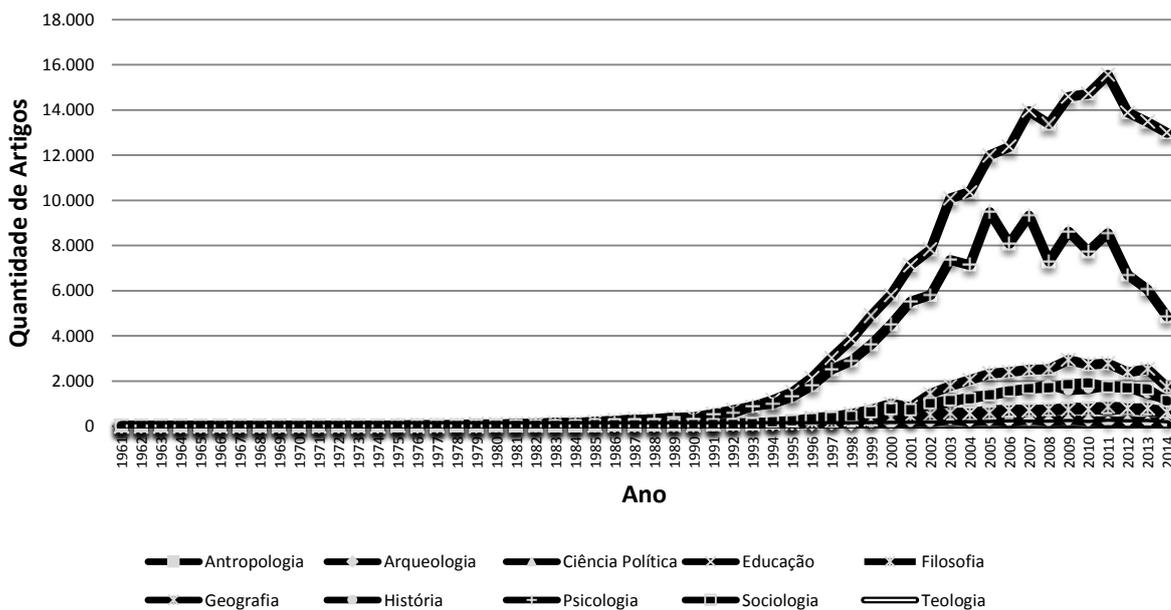


Figura A.9: Quantidade de artigos em anais de congresso – Ciências Humanas.

Ciências Humanas - Artigos em Periódico

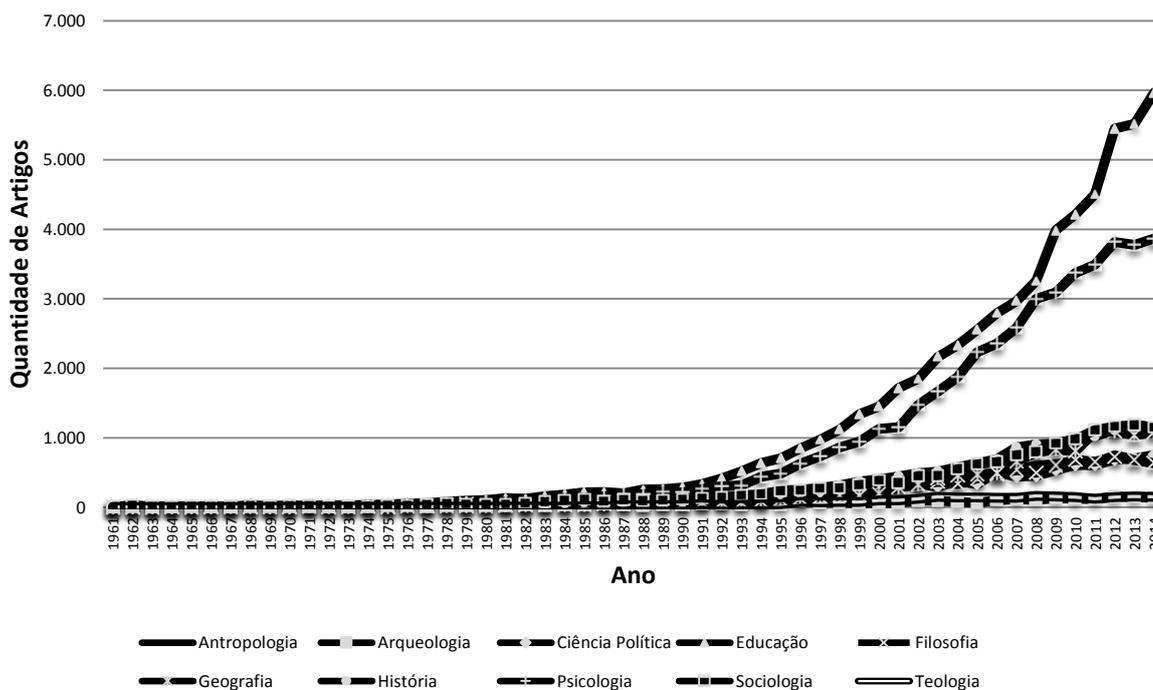


Figura A.10: Quantidade de artigos em periódico – Ciências Humanas.

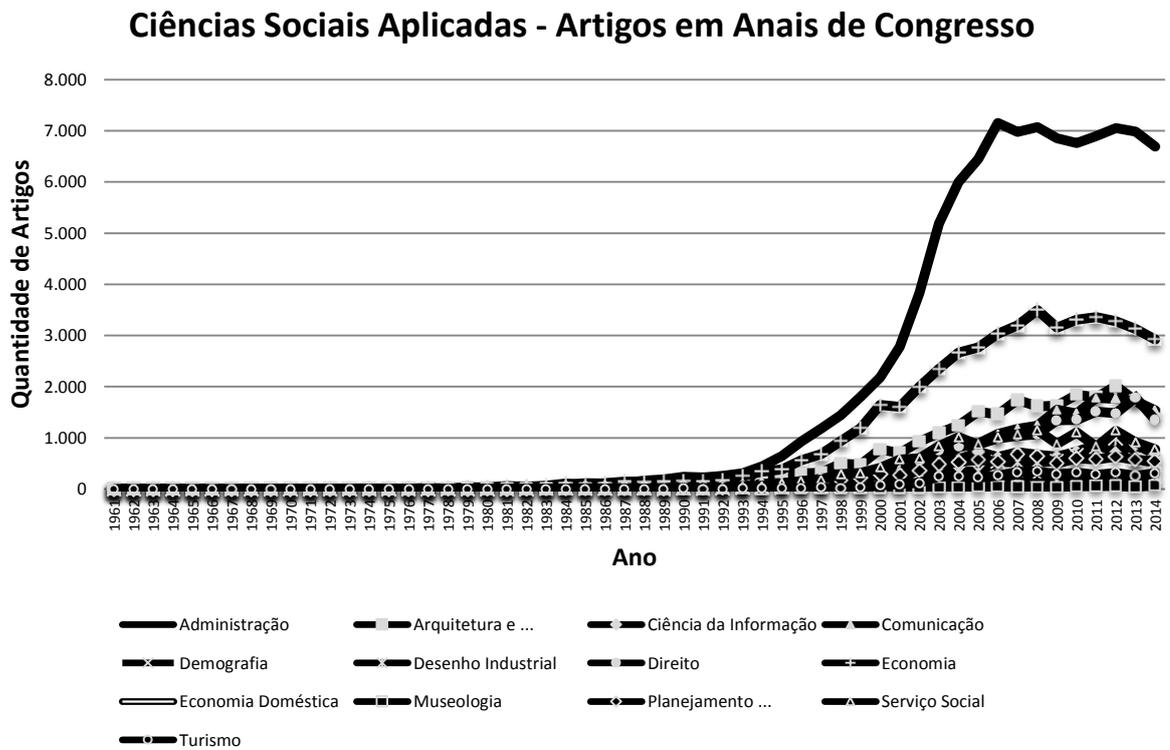


Figura A.11: Quantidade de artigos em anais de congresso – Ciências Sociais Aplicadas.

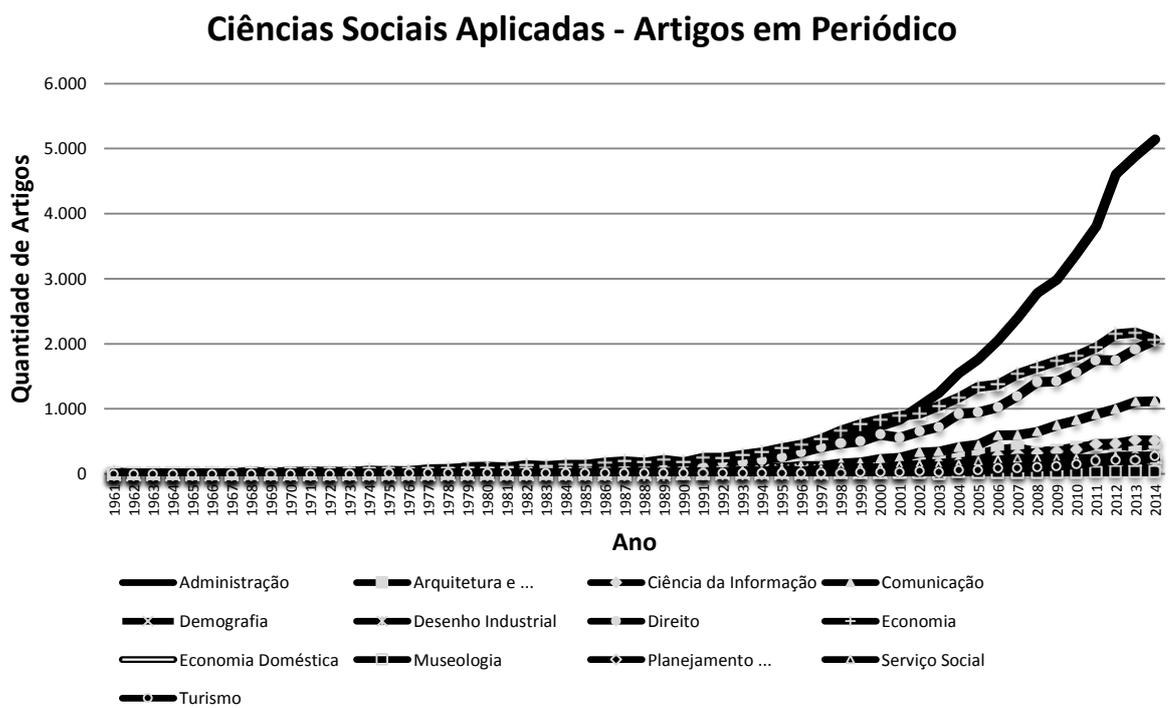


Figura A.12: Quantidade de artigos em periódico – Ciências Sociais Aplicadas.

Engenharias - Artigos em Anais de Congresso

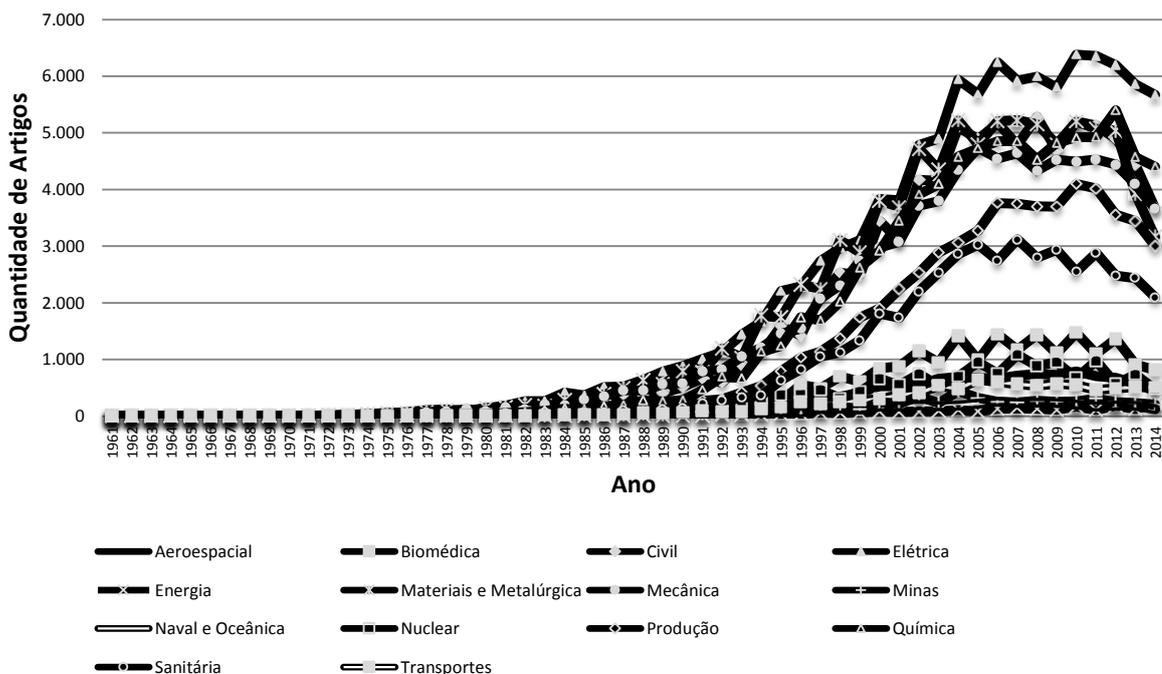


Figura A.13: Quantidade de artigos em anais de congresso – Engenharias.

Engenharias - Artigos em Periódico

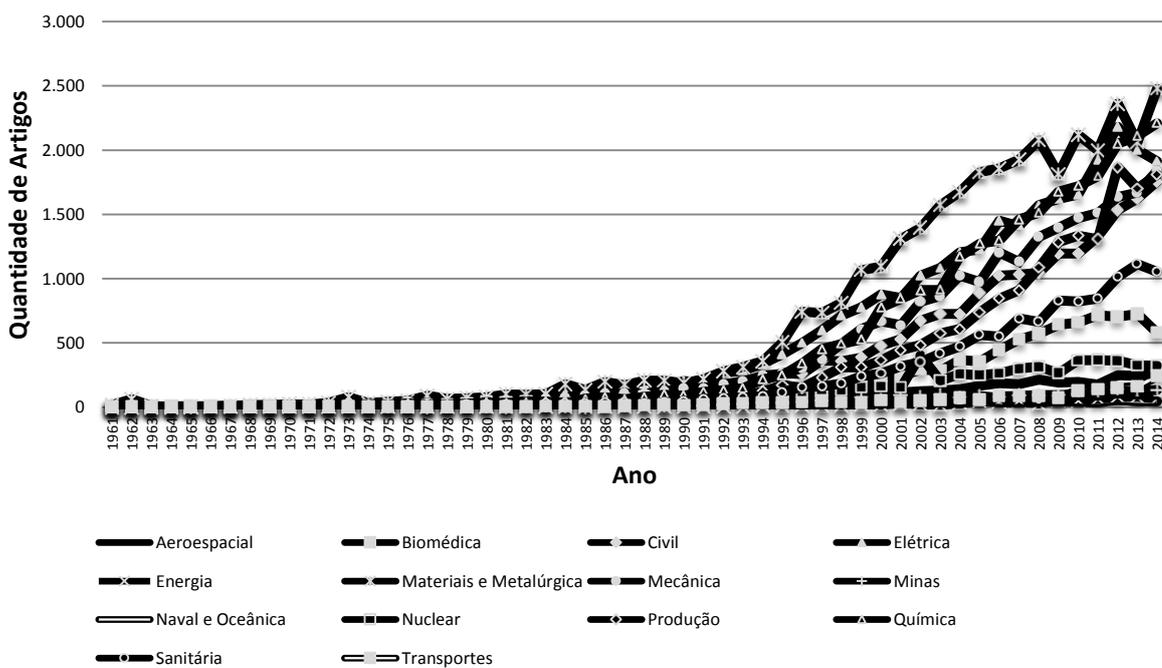


Figura A.14: Quantidade de artigos em periódico – Engenharias.

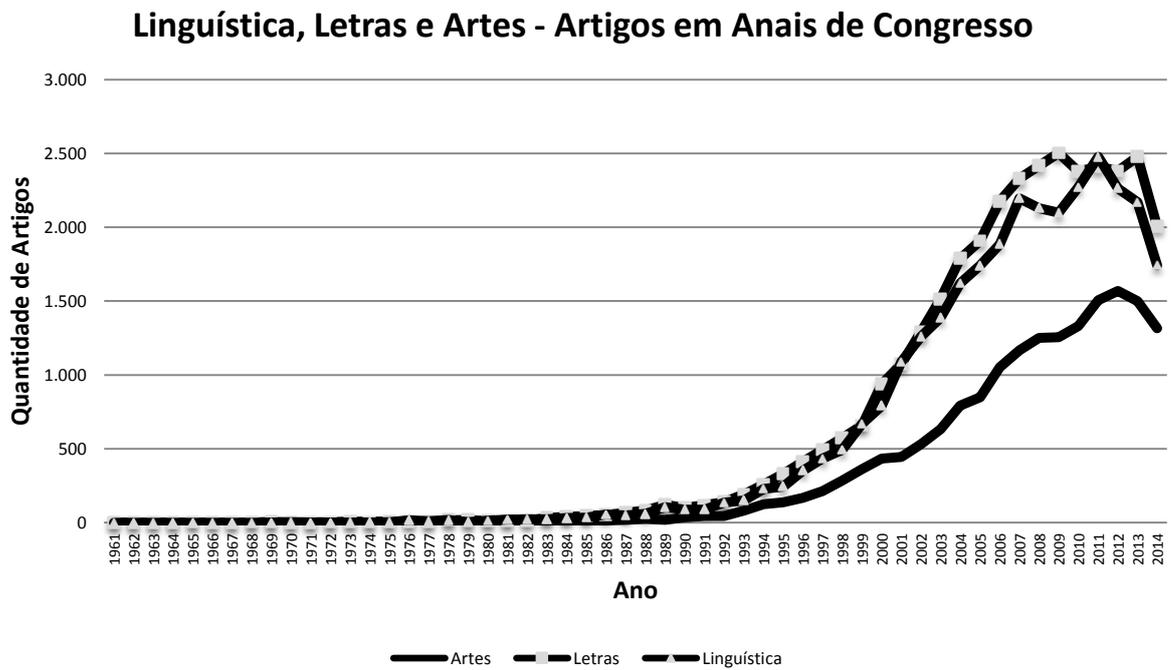


Figura A.15: Quantidade de artigos em anais de congresso – Linguística, Letras e Artes.

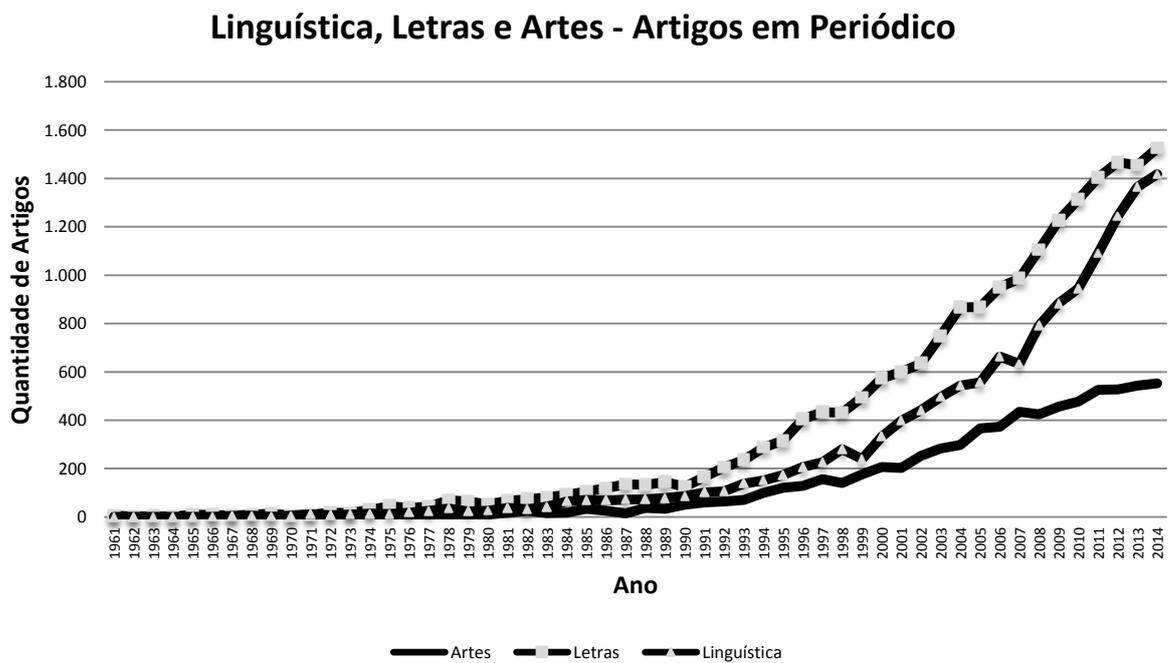


Figura A.16: Quantidade de artigos em periódico – Linguística, Letras e Artes.

Tabela A.1: Percentual de crescimento na produção de artigos em anais de congresso.

Áreas	Anos										Total
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	
Ciências Agrárias											
Agronomia	3,49%	3,43%	14,70%	-9,84%	0,66%	-2,27%	1,49%	-7,81%	-9,79%	-18,26%	-24,18%
Ciência e Tecnologia ...	21,17%	-7,72%	22,40%	-14,29%	14,26%	-1,80%	2,80%	-8,77%	-23,52%	-17,80%	-13,27%
Engenharia Agrícola	12,33%	-3,68%	23,77%	-5,97%	4,09%	2,53%	-7,73%	-1,07%	-19,09%	-36,49%	-31,31%
Medicina Veterinária	-9,09%	15,73%	2,17%	3,53%	-5,49%	4,72%	-10,61%	-2,28%	-20,30%	-25,20%	-46,81%
Recursos Florestais e ...	-1,35%	20,44%	10,84%	-4,99%	12,40%	-10,05%	0,59%	-13,57%	-17,44%	-44,82%	-47,96%
Recursos Pesqueiros e ...	-15,61%	22,62%	7,03%	-5,20%	-4,76%	13,11%	12,66%	-15,93%	-14,99%	-41,01%	-42,09%
Zootecnia	-9,68%	10,97%	0,01%	9,49%	0,64%	13,09%	-7,30%	-9,63%	-9,81%	-33,39%	-35,60%
Ciências Biológicas											
Biofísica	7,91%	-6,73%	7,40%	-11,49%	6,44%	-5,60%	-6,60%	-4,92%	-17,24%	-10,81%	-41,64%
Biologia Geral	3,04%	1,69%	4,99%	1,10%	4,54%	0,23%	-9,40%	-8,68%	-30,23%	-19,53%	-52,25%
Bioquímica	4,11%	-5,08%	6,44%	-4,02%	-0,96%	-2,89%	-2,73%	-8,90%	-10,24%	-23,23%	-47,51%
Botânica	13,29%	-1,17%	1,87%	-7,35%	-5,20%	0,00%	2,83%	-6,68%	-28,40%	-27,67%	-58,49%
Ecologia	22,23%	-16,10%	11,83%	-8,96%	-2,09%	-11,15%	11,01%	-14,01%	-23,78%	-33,56%	-64,59%
Farmacologia	1,90%	2,68%	2,38%	-7,40%	-1,35%	2,31%	-7,15%	-8,72%	-14,35%	-39,42%	-69,11%
Fisiologia	5,16%	6,08%	-6,38%	3,94%	0,32%	-2,16%	-2,98%	-8,40%	-31,98%	-35,19%	-71,59%
Genética	9,52%	-2,68%	3,43%	-1,27%	-3,47%	-5,44%	0,77%	-9,17%	-18,49%	-32,78%	-59,57%
Imunologia	10,63%	-9,40%	-0,84%	-0,52%	-9,67%	-6,06%	1,92%	-12,47%	-17,17%	-40,42%	-83,99%
Microbiologia	23,67%	-13,27%	17,80%	-7,35%	3,57%	-10,80%	7,20%	-10,00%	-27,70%	-29,80%	-46,70%
Morfologia	-9,84%	11,43%	-5,97%	2,79%	-9,99%	18,28%	-15,40%	12,93%	-22,62%	-21,23%	-39,61%
Parasitologia	18,61%	-14,95%	1,42%	-0,45%	2,81%	-8,07%	3,97%	-15,79%	-1,10%	-38,31%	-51,85%
Zoologia	-5,08%	2,91%	0,32%	2,10%	-20,90%	16,64%	-11,08%	6,68%	-18,85%	-31,32%	-58,59%
Ciências da Saúde											
Educação Física	19,53%	-3,49%	-4,17%	2,10%	-7,60%	-2,08%	6,75%	-15,38%	-2,05%	-12,87%	-19,25%
Enfermagem	-2,22%	15,40%	2,89%	5,54%	17,20%	3,54%	-3,16%	-2,43%	-11,34%	-21,87%	3,54%
Farmácia	18,10%	-4,31%	11,77%	2,10%	1,61%	-0,60%	1,95%	-11,45%	-10,30%	-26,43%	-17,56%
Fisioterapia e ...	11,72%	12,41%	1,61%	-4,86%	0,73%	10,73%	-5,53%	-0,90%	-15,33%	-12,97%	-2,38%
Fonoaudiologia	19,01%	-12,36%	9,26%	-0,05%	1,98%	-8,27%	-0,79%	-9,64%	3,02%	-30,63%	-28,45%
Medicina	0,81%	-4,33%	-1,20%	-13,49%	-7,67%	-10,15%	-6,74%	-17,07%	-10,48%	-22,54%	-92,86%
Nutrição	25,92%	14,14%	-1,56%	-9,17%	21,64%	-19,65%	6,96%	-2,97%	-28,79%	-20,04%	-13,52%
Odontologia	0,47%	1,63%	-5,56%	-7,72%	6,27%	-3,63%	-5,78%	-14,10%	-3,70%	-9,72%	-41,82%
Saúde Coletiva	-2,92%	20,04%	-19,17%	9,42%	3,45%	-20,83%	13,36%	-0,43%	-30,53%	-27,01%	-54,63%
Ciências Exatas e da Terra											
Astronomia	-5,15%	0,52%	-9,65%	-9,69%	1,26%	-1,25%	-24,61%	16,53%	-42,73%	-0,31%	-75,08%
Ciência da Computação	0,52%	3,44%	2,13%	6,46%	0,01%	6,61%	1,70%	0,01%	0,93%	0,32%	22,14%
Física	16,23%	-17,01%	8,07%	-7,06%	2,55%	-5,43%	-0,40%	-13,38%	-10,38%	-21,32%	-48,13%
Geociências	6,30%	-5,07%	-0,94%	-0,15%	-6,04%	2,74%	3,94%	-12,34%	-0,56%	-29,68%	-41,80%
Matemática	4,61%	9,30%	9,91%	-8,22%	12,13%	6,63%	2,22%	4,73%	-1,12%	-33,79%	6,39%
Oceanografia	-15,85%	2,50%	8,69%	-15,27%	-25,52%	23,85%	-4,83%	-11,26%	-26,78%	-17,21%	-81,69%
Probabilidade e ...	0,20%	5,47%	-5,05%	1,37%	-1,01%	1,65%	-3,30%	-8,31%	-10,04%	-30,05%	-49,08%
Química	4,29%	-6,85%	13,08%	-8,21%	3,00%	-0,43%	4,68%	-7,88%	-14,09%	-11,21%	-23,63%

Ciências Humanas											
Antropologia	16,75%	6,89%	-2,01%	0,00%	13,15%	-0,97%	5,50%	-3,71%	-13,84%	-23,32%	-1,56%
Arqueologia	27,49%	-22,68%	38,94%	-25,26%	27,78%	-13,77%	-19,75%	17,80%	-23,11%	-16,18%	-8,74%
Ciência Política	3,33%	9,85%	10,34%	0,47%	20,68%	-4,25%	-1,21%	2,45%	-10,51%	-17,53%	13,62%
Educação	1,63%	3,11%	12,85%	-4,11%	9,01%	0,91%	5,67%	-10,70%	-3,00%	-3,57%	11,80%
Filosofia	5,33%	18,38%	0,85%	5,23%	-1,61%	2,86%	13,13%	-7,39%	1,14%	-21,78%	16,15%
Geografia	13,79%	3,05%	4,05%	1,32%	15,86%	-7,07%	1,91%	-14,24%	6,26%	-31,21%	-6,27%
História	4,83%	11,27%	8,49%	10,73%	-13,09%	3,97%	8,73%	-4,18%	-16,52%	-17,07%	-2,84%
Psicologia	32,83%	-14,72%	15,36%	-21,88%	18,24%	-10,21%	10,46%	-21,71%	-9,37%	-20,01%	-21,02%
Sociologia	15,37%	12,81%	6,06%	2,17%	9,25%	2,91%	-9,53%	-2,32%	-2,02%	-32,85%	1,85%
Teologia	11,00%	15,91%	-10,75%	0,00%	12,89%	-30,39%	73,24%	-19,51%	9,09%	-49,07%	12,40%
Ciências Sociais Aplicadas											
Administração	7,40%	10,95%	-2,46%	1,33%	-3,03%	-1,42%	2,03%	2,26%	-0,96%	-4,17%	11,94%
Arquitetura e ...	11,70%	-3,70%	19,48%	-7,41%	0,99%	13,20%	-2,87%	12,40%	-14,21%	-15,40%	14,18%
Ciência da Informação	12,67%	-14,13%	15,34%	-6,93%	-5,95%	16,46%	16,30%	-15,19%	18,60%	-20,09%	17,07%
Comunicação	5,00%	6,31%	9,57%	5,26%	4,74%	-4,78%	3,76%	-0,96%	-2,44%	-10,41%	16,05%
Demografia	10,81%	15,85%	-46,90%	50,73%	-43,64%	54,84%	-21,25%	-1,06%	-8,02%	1,16%	12,53%
Desenho Industrial	3,59%	4,62%	-2,99%	-1,16%	8,19%	8,38%	24,81%	20,00%	-41,57%	-9,57%	14,32%
Direito	3,68%	3,27%	5,05%	2,79%	5,54%	0,67%	11,84%	-2,12%	10,89%	-24,50%	17,12%
Economia	3,56%	9,87%	5,20%	9,42%	-9,84%	4,92%	1,51%	-2,29%	-4,60%	-6,55%	11,20%
Economia Doméstica	33,08%	-10,37%	28,93%	-32,69%	58,10%	-18,67%	11,85%	-27,15%	-26,36%	-27,16%	-10,46%
Museologia	20,84%	0,00%	-18,75%	-23,08%	6,67%	21,25%	23,71%	7,02%	-31,15%	9,52%	16,04%
Planejamento ...	8,57%	-5,79%	26,26%	-6,05%	-21,04%	19,88%	-3,98%	8,98%	-9,51%	-4,38%	12,95%
Serviço Social	-14,41%	17,53%	4,91%	7,67%	-23,46%	25,88%	-25,25%	37,39%	-19,84%	-15,01%	-4,58%
Turismo	-10,08%	17,49%	27,10%	4,20%	-17,87%	12,28%	-12,81%	13,62%	-18,93%	0,78%	15,78%
Engenharias											
Aeroespacial	-2,59%	-1,77%	7,52%	4,34%	-2,28%	8,64%	-10,86%	-3,97%	-15,34%	-22,30%	-38,61%
Biomédica	-29,88%	44,92%	-18,65%	22,42%	-22,98%	33,18%	-25,39%	24,11%	-33,14%	-10,96%	-16,37%
Civil	-4,21%	5,91%	-7,50%	10,36%	-11,30%	11,05%	-1,40%	-4,64%	-9,40%	-17,19%	-28,31%
Elétrica	-4,38%	9,96%	-5,19%	1,20%	-3,20%	9,98%	-0,36%	-2,53%	-5,47%	-3,21%	-3,20%
Energia	-24,30%	39,26%	6,20%	-5,84%	-12,40%	31,33%	-35,67%	40,00%	-34,76%	9,02%	12,83%
Materiais e Metalúrgica	-8,08%	8,59%	0,29%	-1,42%	-6,81%	8,31%	-2,46%	-0,67%	-22,10%	-19,34%	-43,70%
Mecânica	9,07%	-3,99%	2,00%	-6,58%	4,41%	-0,73%	0,89%	-2,01%	-7,79%	-10,58%	-15,30%
Minas	-27,21%	-21,50%	-3,57%	11,93%	-5,51%	0,78%	20,46%	-20,51%	-6,05%	-13,73%	-64,91%
Naval e Oceânica	16,93%	15,38%	-8,24%	2,14%	-2,09%	-11,97%	-16,99%	9,36%	-4,28%	-30,73%	-30,48%
Nuclear	36,19%	-20,90%	43,96%	-18,54%	7,36%	-28,90%	44,51%	-41,38%	28,72%	-35,10%	15,92%
Produção	6,78%	14,84%	-0,32%	-1,23%	-0,11%	10,82%	-1,95%	-11,42%	-3,29%	-12,78%	1,34%
Química	3,12%	2,69%	0,02%	-6,66%	6,51%	1,97%	0,16%	9,56%	-15,32%	-3,65%	-1,60%
Sanitária	5,91%	-9,25%	13,61%	-10,44%	5,08%	-13,21%	13,02%	-13,84%	-1,89%	-13,92%	-24,93%
Transportes	33,26%	-5,46%	-6,44%	-5,11%	6,69%	-4,18%	-16,18%	6,29%	4,49%	-0,98%	12,39%
Linguística, Letras e Artes											
Artes	2,93%	8,91%	7,93%	4,28%	0,24%	3,14%	8,06%	4,25%	-6,52%	-16,21%	17,01%
Letras	6,71%	14,05%	7,03%	3,78%	3,52%	-5,00%	1,43%	-1,24%	3,99%	-18,98%	15,28%
Linguística	7,40%	8,27%	16,55%	-3,10%	-1,41%	8,19%	9,07%	-8,64%	-4,15%	-19,64%	12,55%

Tabela A.2: Percentual de crescimento na produção de artigos em periódico.

Áreas	Anos										Total
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	
Ciências Agrárias											
Agronomia	6,67%	5,23%	21,73%	-3,19%	7,94%	-3,29%	12,41%	-4,33%	2,52%	3,27%	48,97%
Ciência e Tecnologia ...	10,96%	5,24%	12,88%	-2,18%	10,64%	5,25%	12,45%	-11,50%	2,85%	-25,46%	21,13%
Engenharia Agrícola	30,00%	-7,03%	7,92%	7,89%	13,20%	10,40%	13,73%	-0,50%	4,52%	-37,50%	42,63%
Medicina Veterinária	1,07%	14,18%	16,44%	2,91%	10,90%	-1,63%	8,23%	-4,17%	-1,77%	-16,16%	29,98%
Recursos Florestais e ...	7,56%	3,51%	12,57%	1,77%	6,62%	8,33%	7,69%	6,09%	0,26%	-21,79%	32,63%
Recursos Pesqueiros e ...	-10,58%	46,48%	-8,46%	28,57%	-8,01%	-10,12%	14,03%	-0,87%	3,50%	12,67%	67,21%
Zootecnia	20,18%	7,27%	1,68%	15,49%	3,58%	-7,03%	11,82%	-15,59%	-1,44%	-23,77%	12,18%
Ciências Biológicas											
Biofísica	-1,75%	-8,70%	24,46%	4,17%	-8,85%	18,86%	2,31%	1,05%	9,09%	-2,46%	38,20%
Biologia Geral	7,49%	24,31%	15,91%	10,57%	0,20%	3,25%	4,03%	5,58%	-7,25%	-4,73%	59,36%
Bioquímica	6,94%	8,43%	6,80%	6,62%	3,45%	6,83%	14,81%	4,10%	0,81%	-1,38%	57,40%
Botânica	9,11%	11,38%	5,62%	4,59%	2,17%	5,96%	4,32%	3,26%	0,10%	-12,28%	34,23%
Ecologia	10,80%	15,06%	2,55%	3,22%	4,28%	2,38%	1,44%	3,47%	-11,74%	-23,33%	8,13%
Farmacologia	16,76%	4,50%	9,84%	-0,12%	5,23%	3,20%	4,53%	6,86%	4,37%	-30,36%	24,81%
Fisiologia	9,73%	18,39%	1,29%	0,91%	3,84%	7,35%	5,87%	4,07%	-19,51%	-21,87%	10,06%
Genética	5,26%	8,78%	0,73%	4,56%	4,69%	1,65%	6,58%	1,36%	4,21%	-17,99%	19,82%
Imunologia	12,76%	11,86%	7,44%	12,27%	3,42%	-0,39%	9,24%	8,76%	-2,47%	-24,49%	38,40%
Microbiologia	14,68%	11,27%	7,52%	6,86%	12,62%	0,04%	8,01%	1,16%	2,61%	-16,50%	48,27%
Morfologia	-4,59%	22,16%	1,19%	15,58%	6,87%	0,65%	1,42%	10,20%	-0,58%	-19,57%	33,32%
Parasitologia	-7,06%	15,08%	17,09%	11,46%	4,10%	1,23%	2,59%	1,26%	9,83%	-18,70%	36,87%
Zoologia	12,28%	8,86%	9,63%	8,36%	3,77%	3,43%	1,66%	7,95%	2,87%	1,15%	59,96%
Ciências da Saúde											
Educação Física	52,47%	6,63%	27,65%	7,50%	15,20%	18,85%	2,70%	5,65%	1,34%	-6,57%	131,43%
Enfermagem	12,47%	15,30%	10,34%	12,27%	24,93%	9,63%	10,83%	3,12%	9,24%	-31,67%	76,47%
Farmácia	24,91%	3,16%	17,49%	18,49%	-2,68%	8,10%	13,01%	9,16%	5,67%	-24,13%	73,18%
Fisioterapia e ...	22,65%	20,44%	18,03%	30,07%	9,23%	16,91%	14,31%	9,51%	-13,01%	-24,44%	103,71%
Fonoaudiologia	18,37%	-2,88%	3,42%	18,98%	23,93%	2,25%	0,73%	4,36%	6,27%	3,80%	79,24%
Medicina	2,69%	6,99%	6,93%	1,57%	0,91%	0,44%	-0,52%	-2,52%	-3,10%	-5,52%	7,88%
Nutrição	8,52%	21,62%	10,03%	19,58%	6,94%	13,41%	2,86%	5,02%	4,49%	-14,17%	78,31%
Odontologia	5,12%	16,45%	-3,43%	6,39%	8,71%	6,88%	4,91%	-2,22%	-7,54%	-27,80%	7,47%
Saúde Coletiva	15,10%	13,33%	13,11%	10,60%	5,31%	9,07%	7,43%	4,38%	-3,06%	-25,44%	49,83%
Ciências Exatas e da Terra											
Astronomia	-0,26%	22,75%	-15,09%	3,81%	18,09%	-1,86%	5,70%	-10,98%	2,91%	-2,18%	22,89%
Ciência da Computação	9,76%	1,57%	15,89%	10,60%	1,09%	11,87%	16,80%	7,61%	10,66%	-3,76%	82,10%
Física	-6,35%	9,88%	-4,43%	2,66%	-7,30%	8,30%	1,78%	2,01%	-8,22%	-6,25%	-7,91%
Geociências	22,01%	-3,72%	8,54%	3,85%	2,31%	0,90%	9,15%	10,40%	-1,99%	-19,78%	31,66%
Matemática	5,76%	3,08%	13,10%	5,96%	8,44%	0,12%	5,12%	6,27%	6,91%	-13,66%	41,11%
Oceanografia	8,06%	30,65%	-17,44%	19,44%	2,50%	-4,58%	13,12%	1,27%	11,87%	-9,99%	54,91%
Probabilidade e ...	14,80%	5,78%	17,12%	13,99%	1,57%	6,35%	10,63%	6,71%	2,28%	-0,54%	78,70%
Química	16,32%	3,45%	1,72%	4,24%	1,98%	1,63%	4,21%	7,65%	1,87%	0,06%	43,13%

Ciências Humanas											
Antropologia	11,03%	-3,17%	22,49%	5,88%	5,05%	0,48%	0,96%	17,85%	-7,64%	-5,66%	47,27%
Arqueologia	-18,42%	62,90%	14,85%	-8,62%	24,53%	-7,58%	-4,10%	31,62%	-5,84%	-6,21%	83,14%
Ciência Política	-10,88%	49,40%	-12,15%	2,49%	17,70%	14,29%	-0,33%	13,20%	2,48%	7,25%	83,46%
Educação	9,57%	9,40%	6,26%	9,60%	22,50%	5,62%	7,01%	20,89%	1,29%	8,18%	100,30%
Filosofia	24,18%	12,98%	13,62%	-6,93%	21,53%	13,08%	-3,95%	10,67%	-6,61%	-6,49%	72,07%
Geografia	4,21%	4,04%	34,76%	19,88%	4,94%	-2,53%	34,78%	0,00%	-4,87%	3,96%	99,17%
História	9,33%	9,48%	24,39%	4,99%	1,44%	6,43%	4,91%	8,78%	7,17%	-4,69%	72,23%
Psicologia	19,04%	5,65%	9,54%	16,07%	3,04%	9,16%	3,50%	9,43%	-1,10%	2,36%	76,67%
Sociologia	12,73%	5,16%	15,49%	6,51%	14,71%	6,41%	13,28%	4,87%	1,63%	-2,71%	78,09%
Teologia	0,00%	-1,44%	0,73%	19,57%	-6,06%	-7,10%	-13,89%	18,55%	5,44%	-2,58%	13,22%
Ciências Sociais Aplicadas											
Administração	13,92%	16,54%	16,63%	16,39%	7,40%	13,32%	12,40%	21,01%	6,08%	5,22%	128,92%
Arquitetura e ...	11,03%	21,52%	7,63%	-16,71%	10,03%	11,88%	7,65%	3,44%	-8,65%	14,08%	61,91%
Ciência da Informação	22,16%	28,84%	7,58%	8,39%	6,50%	8,14%	23,39%	1,09%	10,99%	-0,97%	116,10%
Comunicação	9,83%	31,99%	0,68%	8,92%	15,46%	10,04%	11,56%	8,83%	11,02%	0,54%	108,87%
Demografia	54,10%	-2,13%	-10,87%	23,17%	5,94%	-0,93%	15,09%	4,92%	9,38%	9,29%	107,95%
Desenho Industrial	51,92%	-21,52%	20,97%	33,33%	29,00%	27,13%	13,41%	20,97%	-1,78%	16,29%	189,73%
Direito	2,61%	8,36%	16,70%	18,08%	0,64%	10,07%	11,84%	-0,29%	9,93%	7,10%	85,02%
Economia	14,19%	2,77%	11,87%	6,64%	5,92%	4,44%	7,45%	10,32%	1,07%	-4,97%	59,70%
Economia Doméstica	0,00%	205,26%	-29,31%	-14,63%	-14,29%	-3,33%	24,14%	38,89%	-44,00%	-25,00%	137,73%
Museologia	-12,50%	-9,52%	-36,84%	25,00%	20,00%	-5,56%	35,29%	47,83%	-5,88%	6,25%	64,07%
Planejamento ...	27,68%	6,99%	11,76%	3,51%	18,08%	-5,74%	2,03%	29,35%	1,92%	1,13%	96,72%
Serviço Social	45,32%	1,49%	13,66%	-3,43%	3,11%	-2,16%	1,32%	11,30%	-7,81%	-3,81%	58,99%
Turismo	-10,34%	76,92%	1,09%	17,20%	12,84%	19,51%	36,05%	4,00%	2,88%	22,43%	182,59%
Engenharias											
Aeroespacial	20,14%	12,57%	-4,79%	19,55%	-10,28%	2,60%	-10,15%	40,11%	-1,21%	11,02%	79,58%
Biomédica	-3,27%	25,63%	18,16%	8,35%	12,43%	2,80%	8,48%	-1,82%	3,13%	-20,00%	53,91%
Civil	22,48%	15,20%	0,78%	1,45%	14,24%	0,08%	10,95%	15,60%	5,74%	8,26%	94,80%
Elétrica	1,83%	18,34%	-2,89%	11,49%	3,12%	2,10%	16,31%	13,30%	-8,21%	-4,40%	50,98%
Energia	64,52%	-11,76%	6,67%	-37,50%	103,33%	-14,75%	25,00%	20,00%	5,13%	-17,07%	143,55%
Materiais e Metalúrgica	9,19%	1,59%	3,71%	7,84%	-12,75%	16,88%	-5,71%	18,27%	-12,78%	20,23%	46,47%
Mecânica	-4,79%	23,51%	-6,07%	17,26%	5,51%	5,29%	2,58%	8,34%	2,14%	9,99%	63,77%
Minas	16,00%	-16,09%	24,66%	3,30%	1,06%	5,26%	16,00%	12,07%	-3,85%	7,20%	65,61%
Naval e Oceânica	41,18%	-22,92%	2,70%	5,26%	12,50%	-11,11%	-7,50%	48,65%	-20,00%	4,55%	53,31%
Nuclear	-4,23%	5,62%	14,45%	4,65%	-15,87%	37,74%	0,27%	-0,82%	-10,47%	-0,62%	30,72%
Produção	21,35%	14,21%	7,82%	19,23%	18,34%	3,89%	-1,95%	42,81%	-8,89%	6,29%	123,11%
Química	8,53%	1,89%	12,72%	3,49%	10,91%	2,56%	4,42%	14,08%	2,98%	4,69%	66,25%
Sanitária	19,03%	-2,49%	25,68%	-3,77%	25,15%	-0,84%	2,67%	19,74%	9,97%	-5,30%	89,85%
Transportes	-24,00%	50,88%	1,16%	2,30%	-15,73%	66,67%	11,20%	6,47%	8,11%	51,88%	158,93%
Linguística, Letras e Artes											
Artes	22,82%	1,91%	16,62%	-2,30%	7,29%	4,61%	10,27%	0,19%	3,04%	1,66%	66,11%
Letras	0,12%	9,45%	3,79%	11,97%	11,05%	7,10%	6,93%	4,42%	-0,75%	4,81%	58,87%
Linguística	2,39%	19,42%	-4,67%	25,12%	11,62%	6,90%	15,45%	14,02%	9,73%	3,81%	103,79%

Citações - Ciências Agrárias

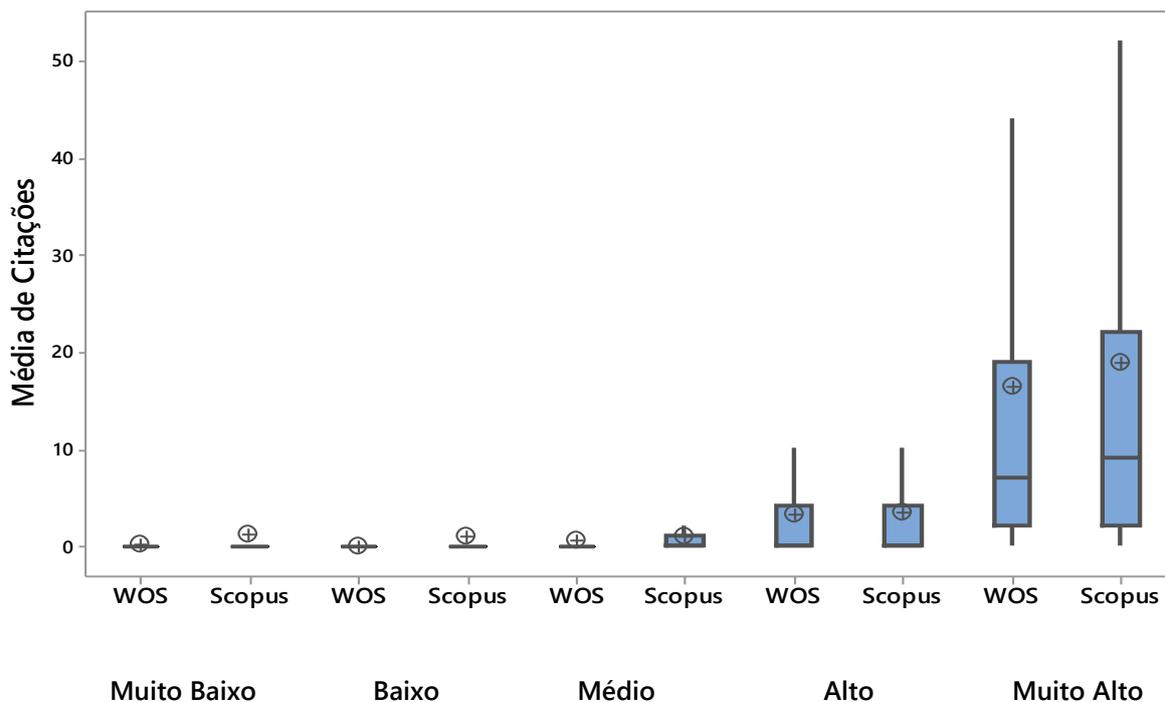


Figura A.17: Média de citações em periódico – Ciências Agrárias.

Citações - Ciências Biológicas

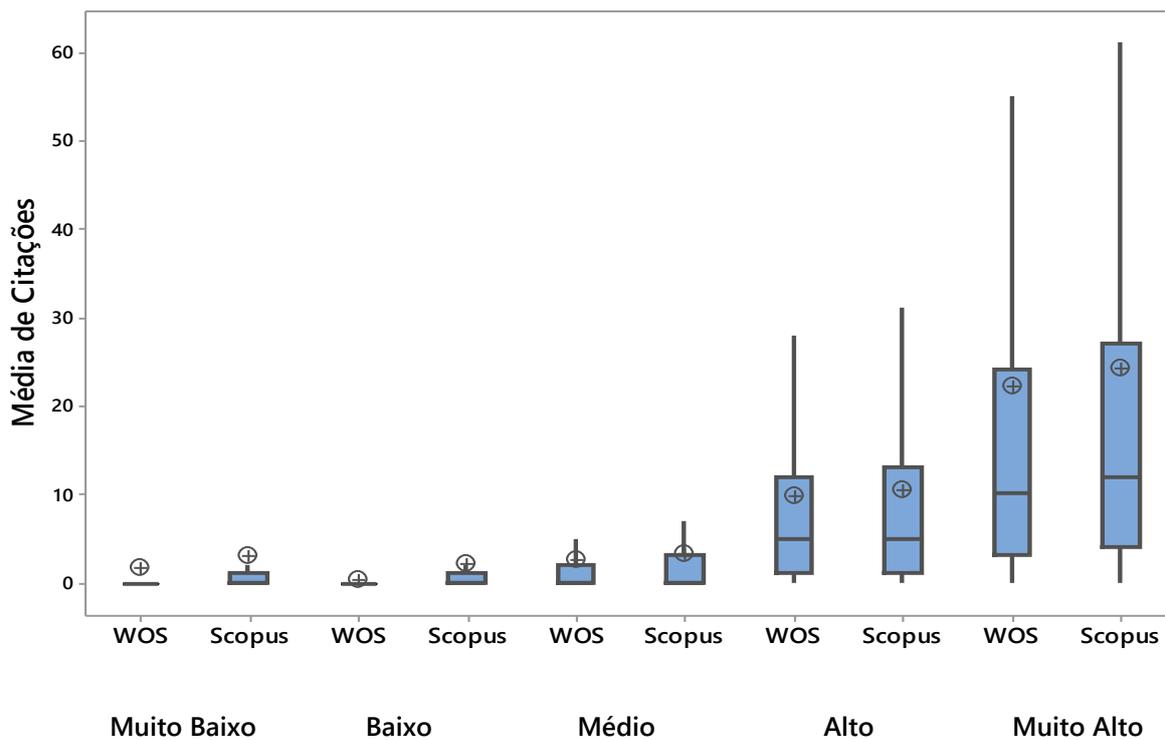


Figura A.18: Média de citações em periódico – Ciências Biológicas.

Citações - Ciências da Saúde

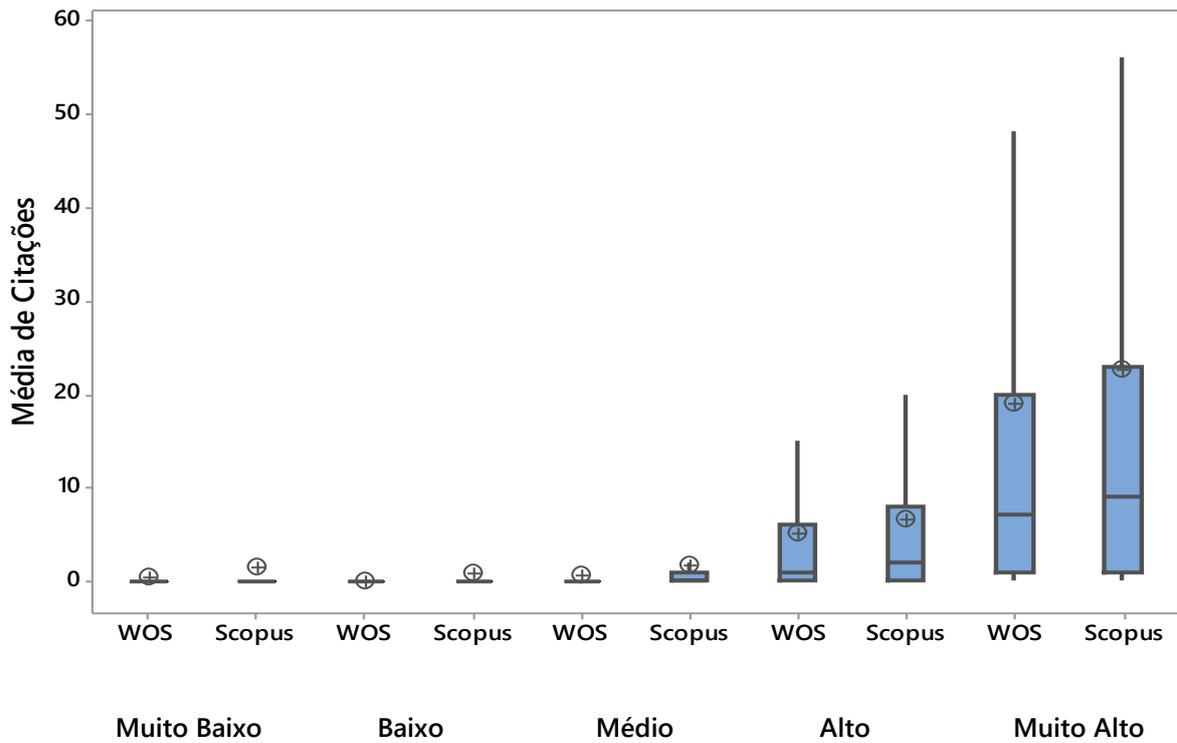


Figura A.19: Média de citações em periódico – Ciências da Saúde.

Citações - Ciências Exatas e da Terra

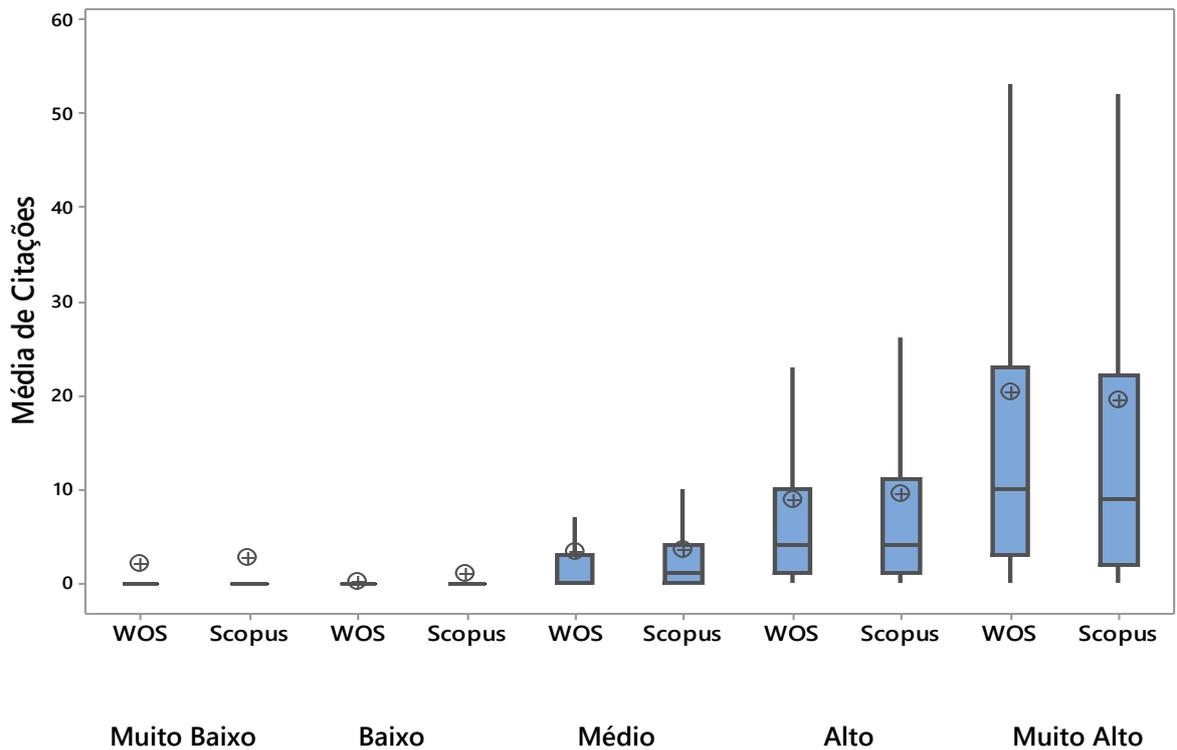


Figura A.20: Média de citações em periódico – Ciências Exatas e da Terra.

Citações - Ciências Humanas

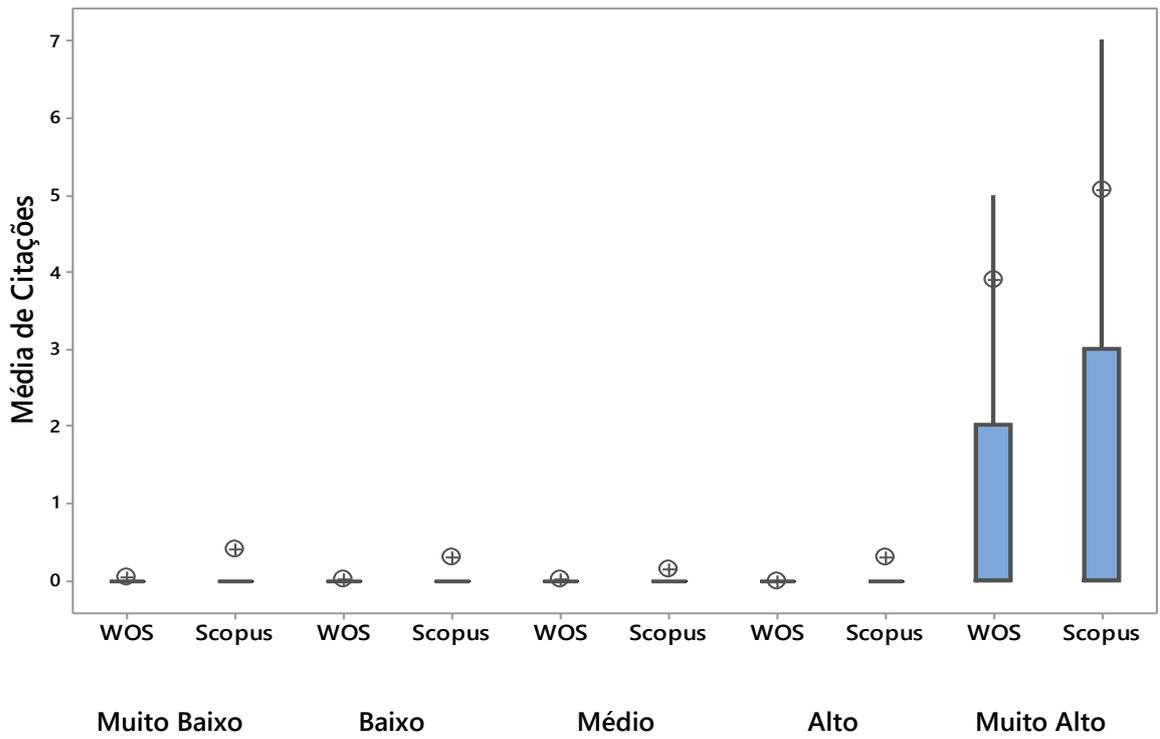


Figura A.21: Média de citações em periódico – Ciências Humanas.

Citações - Ciências Sociais Aplicadas

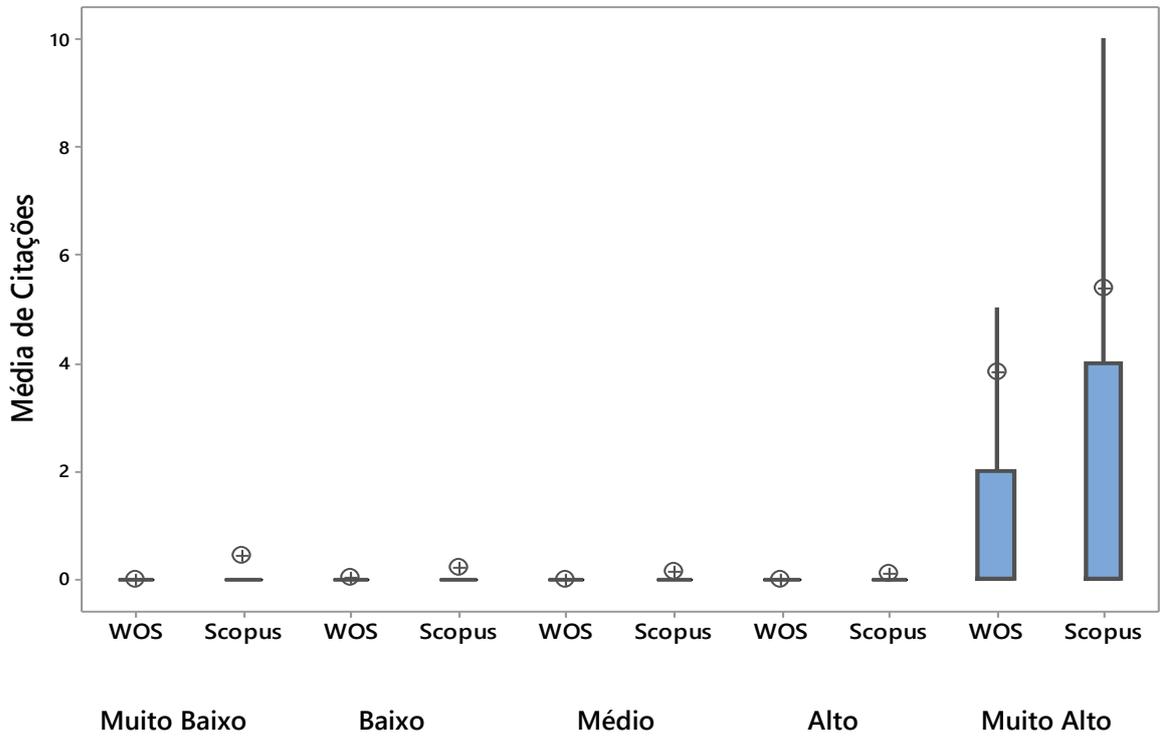


Figura A.22: Média de citações em periódico – Ciências Sociais Aplicadas.

Citações - Engenharias

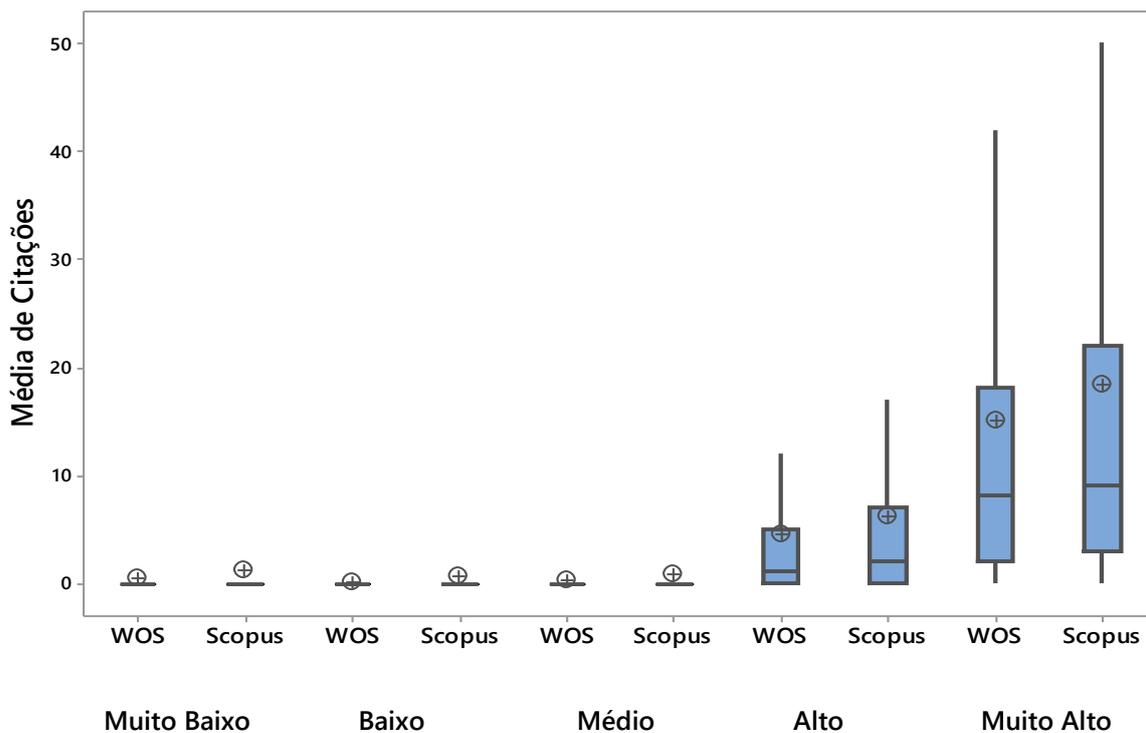


Figura A.23: Média de citações em periódico – Engenharias.

Citações - Linguística, Letras e Artes

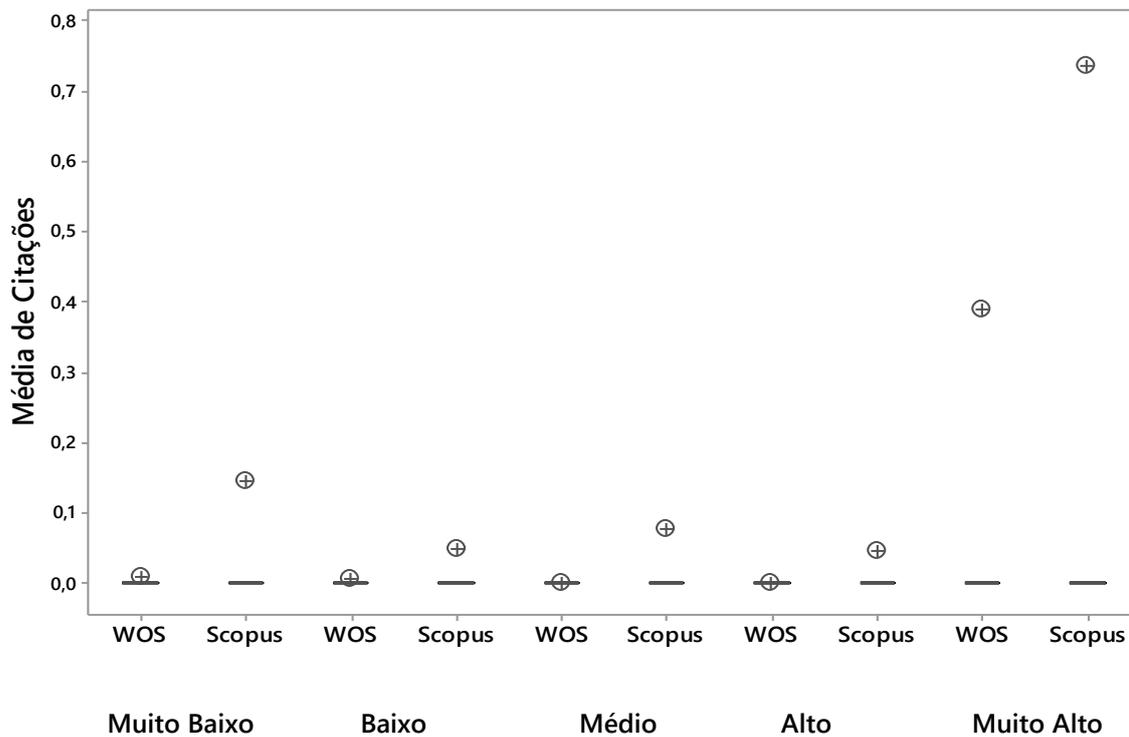


Figura A.24: Média de citações em periódico – Linguística, Letras e Artes.

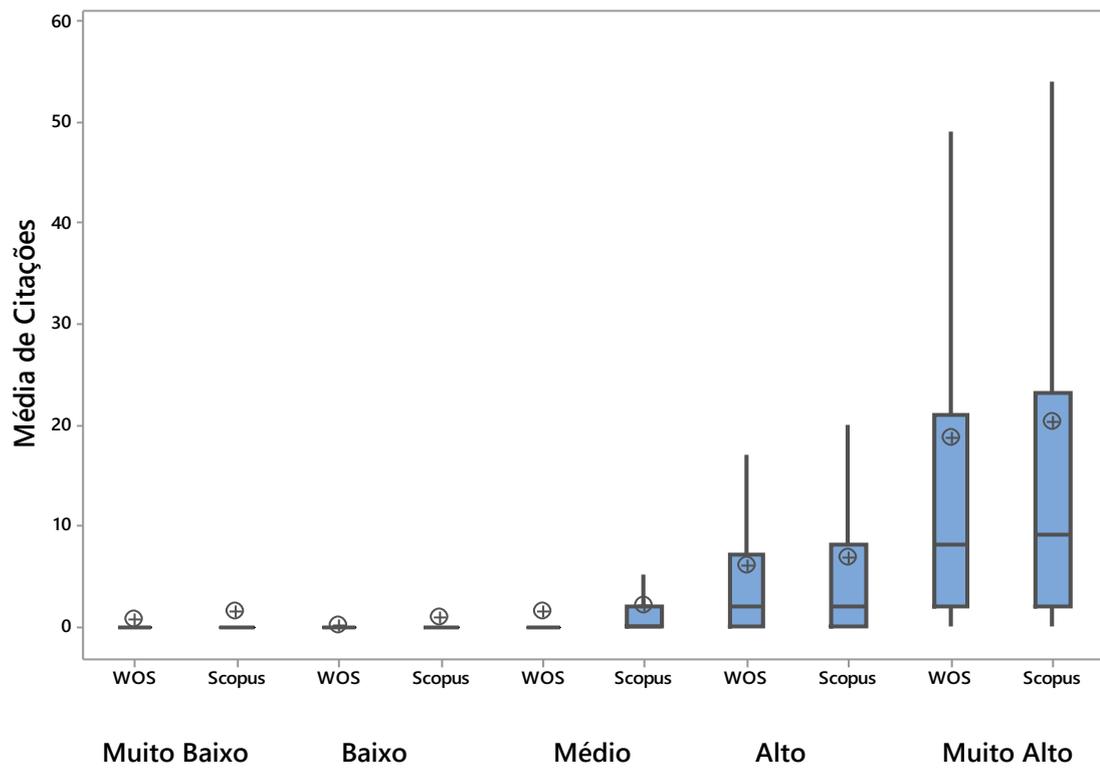


Figura A.25: Média de citações – bolsistas.