



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

UMA ANÁLISE DA PRODUÇÃO CIENTÍFICA BRASILEIRA EM PERIÓDICOS DE ACESSO ABERTO

PATRÍCIA MASCARENHAS DIAS

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

Belo Horizonte
Agosto de 2021

PATRÍCIA MASCARENHAS DIAS

**UMA ANÁLISE DA PRODUÇÃO CIENTÍFICA BRASILEIRA
EM PERIÓDICOS DE ACESSO ABERTO**

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

Belo Horizonte
Agosto de 2021

Dias, Patrícia Mascarenhas
D541a Uma análise da produção científica brasileira em periódicos de acesso aberto / Patrícia Mascarenhas Dias. – 2021.
134 f.

Tese de doutorado apresentada ao Programa de Pós-Graduação em Modelagem Matemática e Computacional.
Orientador: Gray Farias Moita.
Tese (doutorado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Publicações científicas – Teses. 2. Comunicação na ciência – Teses.
3. Bibliometria – Análise – Teses. 4. Curriculum vitae – Teses. I. Moita, Gray Farias. II. Centro Federal de Educação Tecnológica de Minas Gerais.
III. Título.

CDD 511.3



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

***“UMA ANÁLISE DA PRODUÇÃO CIENTÍFICA BRASILEIRA EM
PERIÓDICOS DE ACESSO ABERTO”.***

Tese de Doutorado apresentada por **Patrícia Mascarenhas Dias**, em 30 de agosto de 2021, 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. Adilson Luiz Pinto
Universidade Federal de Santa Catarina

Profª. Drª. Eloísa da Conceição Príncipe de Oliveira
Instituto Brasileiro de Informação em Ciência e Tecnologia

Profª. Drª. Elizabeth Fialho Wanner
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Arthur Rodrigo Bosco de Magalhães
Centro Federal de Educação Tecnológica de Minas Gerais

Visto e permitida à impressão,

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



Emitido em 22/11/2021

CÓPIA DE FOLHA DE ASSINATURAS Nº 1/2021 - DTI (11.47)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 22/11/2021 14:55)
ARTHUR RODRIGO BOSCO DE MAGALHAES
PROFESSOR DO MAGISTERIO SUPERIOR
DM (11.56.11)
Matrícula: 1524308

(Assinado digitalmente em 22/11/2021 16:21)
ELIZABETH FIALHO WANNER
PROFESSOR DO MAGISTERIO SUPERIOR
DECOM (11.56.03)
Matrícula: 1539394

(Assinado digitalmente em 22/11/2021 12:37)
GRAY FARIAS MOITA
PROFESSOR DO MAGISTERIO SUPERIOR
DTI (11.47)
Matrícula: 1243156

(Assinado digitalmente em 25/11/2021 18:11)
ELOÍSA DA CONCEIÇÃO PRÍNCIPE DE OLIVEIRA
ASSINANTE EXTERNO
CPF: 424.834.097-68

(Assinado digitalmente em 26/11/2021 14:21)
ADILSON LUIZ PINTO
ASSINANTE EXTERNO
CPF: 184.218.848-83

Para verificar a autenticidade deste documento entre em <https://sig.cefetmg.br/documentos/> informando seu número:
1, ano: **2021**, tipo: **CÓPIA DE FOLHA DE ASSINATURAS**, data de emissão: **22/11/2021** e o código de
verificação: **e35cd57bb7**

Dedico esta tese ao meu esposo Thiago, aos
meus filhos Guilherme e Eduarda, e aos
meus pais.

AGRADECIMENTOS

Para concretizar este trabalho, muitas pessoas tiveram um papel importante, que sem elas, seria ainda mais trabalhoso concluí-lo. O meu muito obrigado a todos que de forma direta e indiretamente contribuíram para o desenvolvimento desta tese.

De maneira especial, gostaria de agradecer ao Professor Gray Farias Moita, que aceitou orientar minha tese de doutorado. Os seus conselhos e sugestões foram determinantes para o resultado do meu trabalho. Agradeço imensamente pela paciência durante este momento tão difícil em que nos encontramos.

Agradeço a todos os professores do PPGMMC (Programa de Pós-Graduação em Modelagem Matemática e Computacional) que contribuíram significativamente durante o mestrado e doutorado.

Duas instituições foram também especiais para minha formação, que entenderam os momentos em que me ausentei, o meu agradecimento ao diretor e colegas da E.E. Joaquim Nabuco e também a todos colegas da UEMG (Universidade Estadual de Minas Gerais) unidade de Divinópolis.

Também é importante registrar meu agradecimento aos Professores Rodrigo Tomás Nogueira Cardoso e Adilson Luiz Pinto, 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, souberam entender os momentos que não pude estar presente, e mesmo assim incentivaram nesta importante caminhada.

“A mente que se abra a uma nova ideia,
jamais voltará ao seu tamanho original”.

(Albert Einstein)

RESUMO

No ambiente científico, um dos principais objetivos é a divulgação dos resultados alcançados pelos pesquisadores em suas pesquisas. Assim, a comunicação científica se caracteriza como um importante mecanismo para a divulgação de resultados de pesquisas nas diversas áreas do conhecimento. Nesse contexto, os artigos científicos se caracterizam como os principais elementos para a descrição dos estudos realizados e, conseqüentemente, dos resultados obtidos. A divulgação de resultados de pesquisas, principalmente em artigos científicos, possibilita a difusão das técnicas e métricas utilizadas, bem como dos resultados alcançados, a uma quantidade de interessados na temática investigada. Diante disso, identificar como os resultados de investigações científicas estão sendo publicados permite compreender como a comunicação científica tem sido utilizada para a divulgação dos estudos realizados. Nesse cenário, a publicação de artigos em periódicos de acesso aberto surge como um interessante mecanismo para a divulgação de pesquisas científicas, já que facilita e viabiliza o acesso a elas, tendo em vista que não existem barreiras, em especial financeiras, para acessar os conteúdos desse tipo de publicação. Considerando que a ampla divulgação e a facilidade de acesso a estudos científicos possibilitam e impulsionam a evolução das pesquisas, esta tese visa identificar e analisar o conjunto de pesquisadores que têm utilizado periódicos de acesso aberto para a divulgação dos resultados de seus estudos, permitindo ampla difusão e acesso a seus trabalhos. Além disso, é realizada uma análise ampla de suas publicações em periódicos de acesso aberto. Para tanto, são discutidos e analisados indicadores sobre a comunicação científica brasileira em periódicos de acesso aberto, tendo como fonte de dados os artigos registrados nos currículos cadastrados na Plataforma Lattes, especificamente as publicações em periódicos que estejam registrados no Directory of Open Access Journals (DOAJ). Descrevem-se os materiais e métodos utilizados, com ênfase na metodologia adotada para apresentar uma visão da comunidade científica que tem utilizado periódicos de acesso aberto. Configurou-se, assim, um retrato até então inédito da produção científica em periódicos de acesso aberto no Brasil, tendo em vista principalmente a quantidade de dados analisados. Os resultados revelaram uma comunidade composta por 370.388 indivíduos distribuídos em diversas áreas do conhecimento, além de demonstrar que a proporção de artigos publicados em periódicos de acesso aberto no Brasil tem aumentado de forma consistente ao longo dos últimos anos, apesar de este ainda não ser o meio preferido de publicação pelos pesquisadores. Revelou-se como algumas áreas tendem a ser mais representativas nesse meio de divulgação, influenciadas por pesquisadores altamente produtivos, e, ainda, como ocorre a colaboração científica entre os pesquisadores na elaboração dos artigos publicados. Por fim, foram aplicadas as principais leis bibliométricas no conjunto de dados analisados, revelando os principais periódicos em que ocorrem as publicações, a produtividade dos autores, considerando suas áreas de atuação, os principais tópicos de pesquisa das publicações, e uma análise qualitativa do conjunto de artigos, considerando suas citações.

PALAVRAS-CHAVE: Acesso Aberto; Produção Científica; Análise Bibliométrica; Plataforma Lattes.

ABSTRACT

In the scientific environment, one of the main objectives is to disseminate the results achieved by researchers in their research. Thus, scientific communication is characterized as an important mechanism for the dissemination of research results in different areas of knowledge. In this context, scientific articles are characterized as the main elements for the description of the studies carried out and, consequently, of the results obtained. The dissemination of research results, mainly in scientific articles, allows the dissemination of the techniques and metrics used, as well as the results achieved, to a number of people interested in the topic investigated. Therefore, identifying how the results of scientific investigations are being published allows us to understand how scientific communication has been used to disseminate the studies carried out. In this scenario, the publication of articles in open access journals emerges as an interesting mechanism for the dissemination of scientific research, as it facilitates access to them, considering that there are no barriers, especially financial, to access the contents of this type of publication. Considering that the wide dissemination and the ease of access to scientific studies enable and drive the evolution of research, this thesis aims to identify and analyze the set of researchers who have used open access journals to disseminate the results of their studies, allowing for wide dissemination and access to their work. In addition, an extensive analysis of their publications in open access journals is carried out. For this purpose, indicators on Brazilian scientific communication in open access journals are discussed and analyzed, having as data source the articles registered in the curricula in the Lattes Platform, specifically publications in journals that are registered in the Directory of Open Access Journals (DOAJ). The materials and methods used are described, with an emphasis on the methodology adopted to present a view of the scientific community that uses open access journals. Thus, it was configured a hitherto unprecedented view of scientific production in open access journals in Brazil, mainly in view of the amount of data analyzed. The results revealed a community composed of 370,388 individuals distributed in different areas of knowledge, in addition to demonstrating that the proportion of articles published in open access journals in Brazil has consistently increased over the last few years, although this is not yet the preferred means of publication by researchers. It was revealed how some areas tend to be more representative in this means of dissemination, influenced by highly productive researchers, and also how scientific collaboration between researchers occurs in the preparation of published articles. Finally, the main bibliometric laws were applied in the analyzed dataset, revealing the main journals in which publications occur, the authors' productivity, considering their areas of expertise, the main research topics of the publications, and a qualitative analysis of the set. of articles, considering their citations.

KEYWORDS: Open Acces; Scientifc Production; Bibliometric Analisys; Lattes Platform.

LISTA DE FIGURAS

Figura 1.1 - Evolução do investimento no Portal de Periódicos da CAPES (CAPES, 2020).....	19
Figura 2.1 - Diagrama com as iniciativas para o Acesso Aberto (Freire, 2011).....	27
Figura 2.2 - Fluxo com os benefícios da publicação em acesso aberto (Freire, 2011).....	30
Figura 4.1 - Arquitetura do LattesDataXplorer (Dias, 2016).	48
Figura 4.2 - LattesDataXplorer estendido.	49
Figura 4.3 - Método para identificação de publicações em periódicos de acesso aberto.....	51
Figura 4.4 - Quantitativo das publicações em periódicos registradas nos currículos.....	52
Figura 4.5 - Quantitativo das publicações em periódicos registradas.....	54
Figura 5.1 - Distribuição dos autores por níveis de formação.....	56
Figura 5.2 - Distribuição dos autores pelas suas grandes áreas de atuação.....	57
Figura 5.3 - Distribuição dos autores pelas suas áreas de atuação.....	58
Figura 5.4 - Distribuição geográfica dos autores.....	60
Figura 5.5 - Total de publicações por ano do conjunto selecionado.....	62
Figura 5.6 - Evolução temporal da produção de artigos em periódicos no Brasil.....	63
Figura 5.7 – Idioma de publicações dos artigos analisados.....	64
Figura 5.8 - Produção de artigos em periódicos de acesso aberto por grandes áreas.....	65
Figura 5.9 - Média de autores por artigo.....	71
Figura 5.10 - Rede de colaboração científica dos autores.....	72
Figura 5.11 - Componente gigante.....	73
Figura 6.1 – Divisão dos autores por área de conhecimento.....	82
Figura 6.2 – Dispersão dos valores observados e esperados.....	90
Figura 6.3 – Divisão das Palavras Identificadas nas Três Zonas Textuais.....	94
Figura 6.4 – Zona I: Informação Trivial ou Básica.....	95
Figura 6.5 – Zona II: Informação Interessante.....	95
Figura 6.6 – Zona III: Ruído.....	96
Figura 6.7 – Ajuste dos dados.....	97
Figura 6.8 – Distribuição da Mediana do Fator de Impacto ao Longo dos Anos.....	99
Figura 6.9 – Distribuição da Mediana do Fator de Impacto Excluídos os Outliers.....	100
Figura 6.10 – Média de Citações dos Artigos Analisados.....	102

LISTA DE QUADROS

Quadro 2.1: Modelos de negócio para acesso à informação científica	29
--	----

LISTA DE TABELAS

Tabela 2.1 - Princípios Bibliométricos e suas Aplicações (Guedes e Borschiver, 2005).....	33
Tabela 4.1 - Quantitativo dos dados extraídos.....	53
Tabela 5.1 - Produção per <i>capita</i> por área de atuação.....	66
Tabela 5.2 - Autores com maior quantidade de publicações em periódicos de acesso aberto.....	69
Tabela 5.3 – Métricas da Rede.....	74
Tabela 6.1- Zonas e seus respectivos periódicos.....	76
Tabela 6.2 – Primeiras três zonas identificadas.....	77
Tabela 6.3 – Principais autores com publicações no Núcleo.....	79
Tabela 6.4 – Distribuição dos autores que compoem o núcleo por grande Área de Atuação.....	81
Tabela 6.5 - Principais autores dos periódicos do núcleo e principais autores do conjunto selecionado	83
Tabela 6.6 – Produtividade dos Autores.....	84
Tabela 6.7 – Distribuição de frequência para o expoente de Lotka.....	85
Tabela 6.8 – Distribuição dos Autores Observados e Esperados.....	88
Tabela 6.9 – Resultado da aplicação do Teste de K-S.....	89
Tabela 6.10 – Etapas do Processo de Tratamento dos Dados.....	91
Tabela 6.11 – Frequência dos Principais Termos sem a Remoção de StopWords.....	92
Tabela 6.12 – Distribuição das palavras por posição (x) e suas frequências (y).....	93
Tabela 6.13 – Distribuição de $\tilde{x} = \log(x)$ e $\tilde{y} = \log(y)$	97
Tabela 6.14 – Divisão dos Periódicos em Conjuntos de Estratos.....	100
Tabela 6.15 – Quantidade de Publicações nos Estratos por Grandes Áreas.....	101
Tabela 6.16 – Mapa de Calor com as Médias, Medianas e Desvio Padrão de Citações.....	104
Tabela 6.17 – Percentual de Publicações sem Citações.....	105

LISTA DE ABREVIATURAS E SIGLAS

BOAI	<i>Budapest Open Access Initiative</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
C&T	Ciência e Tecnologia
DOAJ	<i>Directory of Open Access Journals</i>
HTML	<i>HyperText Markup Language</i>
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
eISSN	<i>Electronic International Standard Serial Number</i>
ISSN	<i>International Standard Serial Number</i>
MCTI	Ministério de Ciência, Tecnologia e Inovação
OA	<i>Open Access</i>
OAI – PMH	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
OSI	<i>Open Society Institute</i>
XML	<i>eXtensible Markup Language</i>

SUMÁRIO

1. INTRODUÇÃO.....	14
1.1. Descrição do Problema	17
1.2. Objetivos e Contribuições	20
1.3. Metodologia de Pesquisa	21
1.4. Estrutura do Trabalho	23
2. FUNDAMENTAÇÃO CONCEITUAL	25
2.1. Movimento de Acesso Aberto	25
2.2. Bibliometria	31
2.3. Plataforma Lattes	34
3. TRABALHOS CORRELATOS... ..	37
3.1. Trabalhos envolvendo publicações em acesso aberto.....	37
3.2. Trabalhos envolvendo Repositório de Dados da Plataforma Lattes	41
3.3. Considerações sobre Trabalhos Correlatos e Diferenciais Propostos.....	45
4. DESENVOLVIMENTO	47
4.1. Aquisição dos Dados	47
4.2. Filtragem e Tratamento dos Dados.....	49
4.3. Visão Geral dos Dados	52
5. CARACTERIZAÇÃO DOS DADOS	55
5.1. Caracterização dos autores.....	55
5.2. Publicações em Acesso Aberto no Brasil.....	61
5.3. Colaboração Científica	71
6. MÉTRICAS BIBLIOMÉTRICAS APLICADAS NOS ARTIGOS EM PERIÓDICOS DE ACESSO ABERTO	76
6.1. Relevância dos Periódicos – Lei de Bradford.....	76
6.2. Produtividade dos autores – Lei de Lotka	84
6.3. Principais Tópicos de Pesquisa – Lei de Zipf	90
6.4. Análise Qualitativa dos Periódicos em Acesso Aberto	98
7. CONSIDERAÇÕES FINAIS	107
7.1. Considerações	107
7.2. Publicações.....	111
7.3. Trabalhos Futuros.....	116
REFERÊNCIAS BIBLIOGRÁFICAS	117
ANEXOS	126

1. INTRODUÇÃO

Compreender a evolução do desenvolvimento científico e tecnológico de um país é de extrema importância, tendo em vista que isso possibilita identificar como o progresso das pesquisas nas diversas áreas do conhecimento tem evoluído historicamente. Além disso, tal compreensão permite identificar os principais tópicos de investigação, o perfil dos pesquisadores e as suas colaborações científicas, o que pode servir como base para diversas políticas de fomento à pesquisa científica.

Meadows (1999) afirma que a competição científica está estreitamente ligada ao grau de interação que os cientistas têm com seus pares nos mais variados canais de comunicação. A comunicação científica é tão vital para ciência quanto a própria pesquisa.

Segundo André (2005), com o nascimento da ciência ocidental, na antiga Grécia, veio a comunicação científica e a troca de conhecimento, ainda que de maneira informal. No século XVII, nascem as revistas científicas, consideradas a base do sistema moderno de comunicação científica. Ainda segundo o autor, verifica-se que existe um entendimento de que somente o compartilhamento do conhecimento científico pode diminuir as desigualdades sofridas de forma crônica no planeta, devido à importância desse conhecimento para o processo de desenvolvimento.

O tradicional formato impresso de comunicação da ciência vem aos poucos dando espaço para os novos formatos eletrônicos, devido à ascensão da tecnologia de informação e comunicação. No contexto das pesquisas e estudos científicos, a comunicação científica surge nos dias atuais como um elemento central em diversos níveis de discussão, com ênfase na divulgação de artigos científicos em periódicos, atualmente um dos principais meios de comunicação para esse fim.

Mueller (1999) afirma que o periódico científico desempenha pelo menos quatro funções essenciais: certificação da ciência com o aval da comunidade científica; canal de comunicação entre os cientistas e de divulgação mais ampla da ciência; arquivo ou memória científica e registro da autoria da descoberta científica.

Já Mueller (2006) destaca que o periódico científico se caracteriza como o canal mais desejado e eficiente para a divulgação de resultados de pesquisas, além de ser aquele que dá maior visibilidade ao pesquisador que divulga seus resultados, tendo em vista que tal meio é aceito e validado pela comunidade científica. Para a autora, embora o periódico científico não seja, mesmo nos dias atuais, o veículo preponderante em todas as áreas do conhecimento, as revistas científicas indexadas e arbitradas pelos pares tornaram-se o centro do sistema tradicional de comunicação científica. Seu *status* de canal preferencial foi concedido pela comunidade científica, que lhe delega a atribuição de confirmar a autoria da descoberta científica e conferir prestígio aos cientistas e aos periódicos científicos por meio de um sistema de avaliação baseado em indicadores, como quantidade de publicações, índices de citação e visibilidade internacional.

De acordo com vários estudos, os periódicos — principalmente os disponíveis em formato eletrônico — estão em crescimento desde a última década. Pode-se falar que os periódicos, em todas as áreas do conhecimento, têm o papel de ser um filtro para o reconhecimento dos trabalhos aceitos. Para Rodrigues e Oliveira (2012), a publicação em uma revista reconhecida pela área é a forma mais aceita para registrar a originalidade do trabalho e para confirmar que os trabalhos foram confiáveis o suficiente para superar o ceticismo da comunidade científica.

Para Freire (2011, p. 20), a publicação eletrônica realizada de forma colaborativa trouxe mudanças significativas nos aspectos formais da comunicação científica. O periódico científico, principalmente nas ciências naturais, é considerado o principal suporte de comunicação da ciência e foi criado com o objetivo de acelerar o processo de publicação e a disseminação de resultados de pesquisas, cuja qualidade é verificada pela avaliação por pares. Além de vislumbrar melhorias na qualidade dos conteúdos por meio da expansão do processo de avaliação, os periódicos científicos eletrônicos incorporam um conjunto de vantagens em relação aos impressos, porque:

- Reduzem o espaço de tempo entre a submissão de originais de pesquisa e sua publicação nos canais competentes. Essa característica favorece grande parte das áreas de pesquisa que dependem da velocidade das informações correntes para o avanço do conhecimento;
- Indexam seus conteúdos em bases de dados eletrônicas internacionais, permitindo a inclusão de serviços de contagens de citações, de alerta, de personalização e de cruzamento de referências bibliográficas;
- Ampliam a visibilidade e proporcionam a recuperação imediata dos conteúdos, com menor esforço por parte dos potenciais leitores;
- Ampliam e diversificam o processo de avaliação pré e pós-publicação;
- Reduzem os custos do processo de publicação e disseminação dos conteúdos;
- Permitem maior interatividade entre leitores, autores e editores;
- Possibilitam a hipertextualidade, com o auxílio de um conjunto de links internos e externos ao texto;
- Incorporam recursos de hipermídia, com o emprego de áudio e vídeo dentro da estrutura dos documentos.

Nesse contexto, nos primeiros anos do século XXI o Movimento de Acesso Aberto, cuja definição é tornar disponível o texto integral das publicações, permitindo a qualquer usuário ler, fazer download, copiar, distribuir, imprimir, pesquisar ou referenciar o texto integral desses artigos, recolhe-los para indexação, introduzi-los como dados em software, ou usá-los para outro qualquer fim legal, sem barreiras financeiras, legais ou técnicas que não sejam inseparáveis ao próprio acesso a uma conexão à Internet (BOAI, 2012).

Apesar de alguns outros estudos apresentarem definições distintas para “acesso aberto” e “acesso livre”, no contexto deste trabalho não se faz distinção entre esses termos.

Ainda que existam inúmeros benefícios que os periódicos de acesso aberto (*Open Access*) proporcionam, existe também a necessidade de um esforço conjunto para que o principal elemento de todo o processo, a informação científica, esteja acessível a todos os interessados. Para isso, algumas iniciativas já vêm sendo empreendidas. Entre elas estão a criação de

repositórios digitais para armazenar e organizar a literatura científica de acordo com padrões internacionais de interoperabilidade e a busca da conscientização dos principais atores envolvidos no processo de produção, editoração e avaliação da informação científica, para que disponibilizem tais conteúdos em ambientes digitais de acesso aberto ao público em geral.

Neubert, Rodrigues e Goulart (2012) afirmam que o acesso aberto assume um importante papel em todo o contexto da atividade científica, pois permite ao pesquisador ter acesso aos resultados de outros estudos sem as barreiras de custo e as dificuldades de acesso, além de promover a visibilidade e a divulgação dos resultados das atividades científicas de cada pesquisador e de cada universidade.

A publicação científica em acesso aberto faz parte de um cenário mais amplo em prol da abertura do conhecimento em geral (acesso aberto, dados abertos, recursos educacionais abertos, software livre, licenças abertas) e constitui essencialmente um movimento em direção à concepção da informação e do conhecimento como bens públicos (FURNIVAL e SILVA-JEREZ, 2017).

Governantes de vários países reconhecem que o acesso aberto aos dados, à informação e ao conhecimento contribui decisivamente para os avanços da pesquisa científica e inovação, além de maximizar o valor derivado de investimentos públicos, trazer benefícios para a economia e a sociedade e inserir os países em desenvolvimento no sistema global de ciência, contribuindo para seu desenvolvimento econômico e social (OCDE, 2004).

Segundo Freire (2011), limitar a reprodução e a disseminação de um material na sociedade traz consequências negativas, uma vez que se não houver contato com a obra, o autor não é reconhecido pela sociedade e esta não compartilha dos potenciais benefícios advindos das novas descobertas.

Nesse sentido, Lawrence (2001) apresenta números que justificam e motivam a publicação em periódicos de acesso aberto. Para o autor, os artigos de acesso livre apresentam 336% mais citações que os artigos que são disponibilizados de forma impressa, com mediana igual a 158%.

Concomitantemente, vale destacar que geralmente há uma quantidade limitada de recursos para fomentar a pesquisa e um grande número de pesquisadores ou instituições interessadas nesses recursos. Logo, quanto mais amplo e preciso for esse entendimento sobre a produção científica, maior será a possibilidade de se determinar os recursos de maneira correta. Contudo, esse tipo de avaliação é uma tarefa extremamente complexa, pois envolve a análise de diferentes características, tanto quantitativas como qualitativas. Além disso, não há um consenso sobre quais medidas ou características devem ser consideradas para a avaliação da produtividade científica (DIGIAMPIETRI, 2015).

Tendo em vista que grande parte das pesquisas científicas no país é financiada com recursos públicos, geralmente em instituições de ensino ou centros de pesquisa públicos, é de se esperar que os resultados de tais estudos sejam divulgados sem nenhum tipo de barreira, principalmente financeira. Nesse contexto, aliados às vantagens que as publicações em acesso aberto apresentam, como disponibilidade, visibilidade e acessibilidade, diversos esforços estão

sendo empregados para que cada vez mais artigos científicos sejam publicados em periódicos de acesso aberto.

Diante disso, compreender como as publicações de um determinado conjunto de pesquisadores vêm sendo realizadas em periódicos de acesso aberto possibilita identificar um panorama do estágio atual desse tipo de comunicação no Brasil. Esse esforço permite, ainda, verificar se em determinadas áreas do conhecimento, esse tipo de publicação tende a ser mais frequente.

Esse tipo de estudo se caracteriza como um importante mecanismo para avaliar a evolução das publicações em periódicos de acesso aberto pelos pesquisadores brasileiros, permitindo verificar se as políticas de incentivo para a publicação de pesquisas nesse formato de comunicação têm alcançado resultados satisfatórios.

1.1 Descrição do Problema

A ciência apresentou um crescimento impressionante nas últimas décadas quanto ao número de registros de publicações de artigos científicos (RIDKER e RIFAI, 2013). Tal crescimento é verificado por meio de resultados de esforços coletivos de pesquisadores de todas as áreas de conhecimento. Embora essa sistematização social tenha existido em épocas anteriores, o novo paradigma da tecnologia da informação, ou seja, a internet e seus serviços, fornecem a base fundamental para sua ampliação (CASTELLS, 2010).

Para Ferreira (2010), a comunidade científica tem a necessidade e o compromisso de tornar público todo o conhecimento produzido durante o desenvolvimento de suas pesquisas. A divulgação desse conhecimento tem ajudado a compreender melhor o mundo, e, com isso, permite o desenvolvimento da sociedade e de padrões para a vida humana.

Ao ser comunicado formalmente, o saber científico se torna público e incorpora-se a um conjunto de publicações resultantes da comunicação científica chamado Literatura Científica, que constitui a base de acumulação do conhecimento e referência para novas pesquisas.

O acesso à informação científica tem sido um grande desafio para países em desenvolvimento como o Brasil. Com a crise dos periódicos, surgida em função dos altos custos na manutenção das assinaturas das revistas científicas, o acesso a esse tipo de informação ficou bastante limitado. Embora essa crise tenha começado em meados dos anos 80, ainda não existe nenhuma solução definitiva. Com as tecnologias da informação e da comunicação, surge a iniciativa de arquivos abertos, a qual define um modelo de interoperabilidade entre bibliotecas e repositórios digitais, possibilitando alternativas para a comunicação científica (KURAMOTO, 2006).

O acesso aberto para a literatura científica surge com a convergência de dois fatores. Um deles é que cientistas e investigadores têm a boa vontade de publicar suas pesquisas em revistas sem remuneração, ajudando, assim, a investigação e a propagação do conhecimento. O outro fator

é o alto crescimento do acesso à internet para grande parcela da sociedade. Inúmeros estudos comprovam que revistas de acesso aberto recebem consideravelmente mais citações quando comparadas com as do sistema tradicional por assinatura (LAWRENCE, 2001; KURTZ, 2004; BRODY *et. al.* 2004; ANTELMAN, 2004; EYSENBACH, 2006).

Furnival e Silva-Jerez (2017) destacam que apenas cerca de 30% do total dos 2,5 milhões de artigos científicos publicados no mundo anualmente está disponível em canais de acesso aberto. Os autores ressaltam que existem poucas pesquisas no Brasil que exploram as opiniões e percepções da comunidade científica brasileira sobre o acesso aberto. Dentre as motivações para a baixa quantidade de material disponível em relação ao acesso aberto, são destacadas:

- Pouco conhecimento da questão e das implicações do acesso aberto;
- Desinteresse pelo acesso aberto, como resultado da falsa concepção de que o acesso à literatura científica não é um problema para a maioria dos cientistas no mundo;
- Desconhecimento sobre as revistas prestigiadas em acesso aberto que já existem, devido ao hábito de publicar e ler exclusivamente as revistas já conhecidas (a maioria dependente de assinatura, não de acesso aberto), vistas como mais seguras para a progressão na carreira profissional;
- Desconhecimento dos repositórios de acesso aberto como forma de disseminar seus trabalhos e como forma complementar à publicação tradicional em revistas fechadas e pagas;
- Percepção do procedimento de arquivamento como tecnicamente complexo e demandante do tempo precioso do pesquisador;
- Força do sistema de premiação na comunidade científica, pelo qual o prestígio pode contar muito mais do que a disseminação rápida e ampla. Além disso, novos títulos de revistas entrando em cena ficam com uma relativa desvantagem – portanto, com menor reputação e prestígio – diante dos títulos mais consolidados.

Nesse contexto, é possível perceber, ainda, que na maioria dos casos os recursos públicos se fazem presentes nas etapas de formação de um pesquisador. Quando estudante, em geral o pesquisador é apoiado por escolas e universidades públicas ou recebe subsídios públicos para auxiliar sua formação. É o caso das bolsas para incentivo a pesquisa. Já como pesquisador, recebe financiamento público para desenvolver suas pesquisas e apoio à participação em eventos científicos nacionais e internacionais para apresentar os resultados das pesquisas. Por fim, uma vez publicados, os resultados da pesquisa dependem da utilização de recursos públicos para a aquisição dos periódicos em que se encontram essas pesquisas. Por se tratar de revistas científicas que reúnem vários artigos em um único volume, é necessário adquirir o conjunto de artigos publicados naquelas revistas, o que nem sempre reflete as necessidades do pesquisador ou das instituições de pesquisa que prestam o financiamento (FREIRE, 2011).

No Brasil, o Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES¹) é uma das principais fontes de acesso a dados de publicações científicas. Nele é possível acessar um acervo contendo mais de 45 mil títulos com texto completo, 130 bases

¹ <http://www.periodicos.capes.gov.br/>

referenciais e 12 bases dedicadas exclusivamente a patentes, além de livros, enciclopédias e obras de referência, normas técnicas, estatísticas e conteúdo audiovisual (CAPES, 2018).

Porém, o Portal de Periódicos da CAPES não se caracteriza como uma iniciativa de acesso livre, já que o órgão governamental paga pelas assinaturas dos periódicos disponibilizados gratuitamente para as instituições associadas. Somente no ano de 2020 o custo foi de aproximadamente R\$ 480 milhões de reais (CAPES, 2021). Na Figura 1.1 é possível observar a evolução do custo do Portal de Periódicos da CAPES em milhões de reais entre os anos de 2004 a 2020.

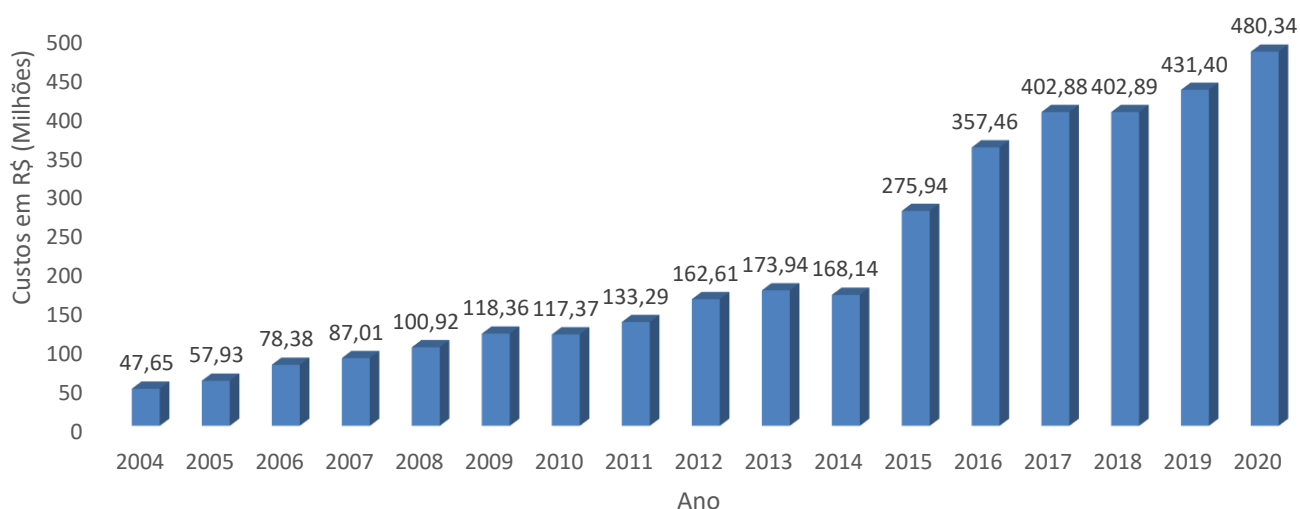


Figura 1.1 - Evolução do investimento no Portal de Periódicos da CAPES (CAPES, 2021).

É possível observar como o investimento no Portal de Periódicos da CAPES tem sido cada vez maior. Somente no ano de 2014, é que houve uma queda de aproximadamente 3,3% referente ao ano anterior. Percebe-se, ainda, que durante os anos de 2017 e 2018 o investimento foi praticamente o mesmo, e que, de acordo com a CAPES², a previsão para se investir no ano de 2021 no portal é de aproximadamente 377 milhões de reais. Ao considerar o período compreendido, entre 2004 e 2020, o investimento já realizado pela CAPES para manter o portal chegou a aproximadamente 3,5 bilhões de reais.

A preocupação com a acessibilidade dos resultados de pesquisas, principalmente as produzidas com financiamento público, está presente em diversos documentos que recomendam governos e agências financiadoras a formular políticas mandatórias para arquivamento da produção científica dos pesquisadores.

No entanto, observa-se que uma grande parte da literatura científica publicada em periódicos é administrada por editoras. Como um dos seus intuitos é obter lucro, elas limitam o acesso dos resultados de pesquisas científicas para o público que tem que pagar para obter esse acesso.

² <https://www.gov.br/capes/pt-br/acesso-a-informacao/institucional/orcamento-evolucao-em-reais>

Logo, uma atividade como a produção científica, que geralmente é financiada pelos cofres públicos, por vezes tem acesso limitado a seu conteúdo, tendo em vista que é necessário pagamento para se ter acesso ao conteúdo completo. Além disso, em algumas situações, tal pagamento também é realizado com recursos públicos. Devido a esse estado, vários segmentos vinculados ao processo de produção e propagação da informação estão trabalhando para que a literatura seja acessível e sem custos para qualquer interessado.

Sendo assim, o problema principal abordado nesta tese é analisar a produção científica brasileira publicada em periódicos de acesso aberto, compreendendo como tem evoluído a divulgação científica nacional nesse meio de publicação, bem como verificar quais são os autores que têm optado por divulgar suas pesquisas em acesso aberto. Para tratar desse assunto, o presente trabalho extrai, processa e analisa todos os artigos científicos publicados em periódicos de todos os indivíduos com currículos cadastrados na Plataforma Lattes, para em seguida avaliar quais dessas publicações foram realizadas em periódicos de acesso aberto. Assim, é possível analisar a evolução temporal desse meio de publicação, o perfil dos autores e as áreas do conhecimento mais representativas, além de realizar a avaliação dessa produção com adoção de métricas bibliométricas, dentre outras análises. Os resultados dessas investigações podem servir de subsídio para políticas e estratégias de incentivo à publicação em periódicos de acesso aberto.

1.2 Objetivos e Contribuições

Buscando maior entendimento sobre como tem evoluído a quantidade de publicações em periódicos de acesso aberto pelos pesquisadores brasileiros ao longo dos anos, este estudo tem como objetivo geral analisar a produção científica brasileira em periódicos de acesso aberto, tendo como principal fonte de dados todos os currículos cadastrados na Plataforma Lattes.

Especificamente, são explorados os seguintes desafios:

- Compreensão do histórico do movimento de acesso aberto;
- Identificação dos repositórios de dados a serem utilizados;
- Extração e integração de dados;
- Identificação dos pesquisadores que publicam em periódicos de acesso aberto;
- Identificação do perfil dos pesquisadores;
- Análise temporal sobre a produção científica em periódicos de acesso aberto;
- Modelagem, caracterização e análises de redes de colaboração científica do conjunto analisado;
- Análises bibliométricas para a determinação de indicadores sobre a pesquisa científica em periódicos de acesso aberto.

Com isso, pode ser possível responder algumas questões, como:

- 1) Qual o perfil dos autores brasileiros que têm publicado artigos em periódicos de acesso aberto?

- 2) Como a produção de artigos em periódicos de acesso aberto no Brasil tem evoluído ao longo dos anos?
- 3) Quais as áreas mais representativas na publicação de artigos em acesso aberto?
- 4) Como têm colaborado os autores na produção de artigos em periódicos de acesso aberto?
- 5) Quais são os principais periódicos de acesso aberto utilizado pelo conjunto analisado?
- 6) Quem são os principais autores de artigos em periódicos de acesso aberto no país?
- 7) Quais são os principais tópicos de pesquisa nesse meio de comunicação?
- 8) Como os artigos publicados em periódicos de acesso aberto têm sido citados?

O propósito é explorar dados da produção científica realizada por indivíduos com currículos cadastrados na Plataforma Lattes e verificar quais dessas publicações foram realizadas em periódicos de acesso aberto, contribuindo para compreender como as publicações nesse meio de divulgação têm evoluído ao longo dos anos e traçar o perfil dos seus autores. Além disso, pretende-se verificar quais são os pesquisadores e áreas do conhecimento mais representativos nesse meio de divulgação, e avaliar com métricas bibliométricas a produção científica dos pesquisadores.

Ademais, abordagens considerando a quantidade de indivíduos e publicações em análise tornam possível avaliar seu impacto, considerando critérios qualitativos, como o número de citações dos artigos. Pretende-se, ainda, explorar diversas métricas bibliométricas para melhor compreender o conjunto de dados analisados.

1.3 Metodologia de Pesquisa

Para este trabalho, inicialmente definiu-se como tema a análise das publicações em periódicos de acesso aberto dos pesquisadores brasileiros. Tendo em vista a grande quantidade de indivíduos e registros de publicações e, ainda, o reconhecimento mundial da comunidade científica, foram utilizados os dados curriculares cadastrados na Plataforma Lattes, a saber, 6.548.210 currículos (coletados em janeiro/2020).

A escolha da Plataforma Lattes para a extração está relacionada ao fato de que ela possui uma vasta quantidade de dados, pois trata da integração de dados científicos, profissionais e acadêmicos e permite que a atualização dos dados individuais seja realizada pelos próprios pesquisadores. Esses currículos possuem informações sobre formação acadêmica, áreas de pesquisa, atuação profissional e orientações acadêmicas, além de produções técnicas e científicas.

Os currículos que fazem parte da Plataforma Lattes se tornaram um padrão nacional utilizado para a avaliação individual das atividades científicas e acadêmicas, já que eles agregam dados de pesquisadores de todas as áreas do conhecimento, tornando a Plataforma Lattes uma fonte extremamente extensa e valiosa para analisar e compreender o comportamento de grupos de pesquisa (DIGIAMPIETRI *et al.*, 2012).

Apesar de os dados dos currículos da Plataforma Lattes serem disponibilizados livremente, eles são visualizados por interface de consulta disponibilizada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que apresenta os currículos individualmente, sem possibilidade de agrupamentos, análises e comparações com outros currículos. Diante disso, técnicas e ferramentas para a extração dos dados curriculares se fazem necessárias (DIAS,2016).

Logo, para análise de todo o conjunto de dados e para se contemplar os objetivos propostos neste trabalho, diversos métodos e técnicas são propostos e implementados. Diante disso, a seguir, é apresentada uma visão geral dos procedimentos metodológicos explorados nesta tese.

- Inicialmente, para compreensão do histórico do movimento de acesso aberto, e para subsidiar a fundamentação conceitual desta tese, são analisados e compilados os principais trabalhos que apresentam estudos sobre a produção científica em acesso aberto.
- Posteriormente, para a construção do repositório de dados a ser analisado, são identificados todos os periódicos de acesso aberto, para, dessa maneira, identificar todos os pesquisadores com currículos cadastrados na Plataforma Lattes que possuem publicações científicas nesses periódicos. Pela grande quantidade de dados registrados nos currículos, se faz necessário o uso de técnicas de extração e análise de dados para a compreensão do conjunto a ser analisado.
- Em seguida, o conjunto de indivíduos que possuem publicações em periódicos de acesso aberto é analisado e, a partir disso, são identificados os perfis dos pesquisadores que têm apresentado os resultados de suas pesquisas nesse meio de divulgação. Depois, são comparados seus perfis de formação acadêmica, áreas de atuação, distribuição geográfica, entre outros, o que pode proporcionar a obtenção de conhecimento até então inédito sobre esse conjunto de dados. Porém, a identificação e a análise dos indivíduos que têm publicado em periódicos de acesso aberto não são de fácil realização, e atualmente não existem ferramentas automatizadas ou repositórios específicos para essa finalidade, considerando um contexto global de publicações científicas. Logo, se faz necessária a implementação de métodos que possam analisar os pesquisadores de forma coletiva, utilizando suas publicações como critério de seleção.
- Com todo o conjunto de dados dos indivíduos selecionado, é possível verificar, baseando-se em seus históricos de publicação, como tem evoluído a publicação de artigos em periódicos de acesso aberto, bem como analisar quais são as áreas do conhecimento mais representativas nesse meio de publicação. As análises levam em consideração os dados registrados em cada um dos currículos dos pesquisadores que compõem o conjunto.
- Finalmente, foram aplicadas algumas das principais métricas bibliométricas: Bradford — produtividade dos periódicos; Lotka — produtividade dos autores; Zipf

— frequência das palavras no conjunto de dados selecionado. Em seguida, é apresentada uma análise qualitativa dos artigos em periódicos em acesso aberto, considerando as citações recebidas e o fator de impacto dos periódicos.

Para avaliar os principais resultados obtidos, foram realizadas análises quantitativas, resultando em informações que possibilitam identificar uma visão geral dos pesquisadores que têm divulgado suas pesquisas em periódicos de acesso aberto, bem como, uma evolução temporal desse tipo de publicação. Com base nessas análises, são apresentados resultados que apresentam um retrato da produção científica brasileira em periódicos de acesso aberto, e ainda, informações sobre como têm colaborado os pesquisadores brasileiros que divulgam o resultado de suas pesquisas nesse meio de divulgação. A adoção das principais métricas bibliométricas também viabilizou uma análise abrangente do conjunto selecionado, até então inédita, tendo em vista a quantidade de dados analisados da produção de artigos neste meio de divulgação da comunidade científica brasileira.

1.4 Estrutura do Trabalho

O restante deste trabalho está organizado da seguinte maneira:

- No Capítulo 2, é apresentada uma revisão bibliográfica que aborda os assuntos permeados por esta pesquisa. Esse capítulo inclui a apresentação de conceitos relacionados ao movimento de acesso aberto e a seu histórico, bem como a estudos e análises bibliométricas.
- No Capítulo 3, são apresentados, discutidos e comparados alguns trabalhos relacionados aos assuntos abordados neste estudo. São incluídos trabalhos que focam a produção científica em acesso aberto de diversas áreas do conhecimento e análises que utilizam os currículos da Plataforma Lattes como fonte de dados.
- No Capítulo 4, são abordados em detalhe os materiais e métodos utilizados nesta tese para a extração dos dados e seleção do conjunto de indivíduos a serem analisados. Esse capítulo descreve o mecanismo de extração dos currículos cadastrados na Plataforma Lattes e a forma pela qual as publicações em periódicos de acesso aberto foram identificadas.
- No Capítulo 5, tem-se uma caracterização geral dos pesquisadores, bem como a distribuição destes em suas áreas e grandes área de atuação, além de sua distribuição geográfica. Também é focada a evolução das publicações de artigos em periódicos de acesso aberto de modo geral, e a evolução desses artigos por grandes áreas. Além disso, abordam-se redes de colaboração que mostram como os

pesquisadores selecionados têm trabalhado em conjunto nas publicações selecionadas.

- No Capítulo 6 é realizado um estudo bibliométrico dos dados extraídos. Para tanto, algumas métricas bibliométricas foram selecionadas, a saber: Bradford, Lotka e Zipf. Com isso, foi possível avaliar, no conjunto de dados analisados, os principais periódicos, os autores mais produtivos e os tópicos de pesquisa mais frequentes. Ademais, é apresentada uma análise qualitativa de artigos dos periódicos de acesso aberto com base em suas citações e fator de impacto dos periódicos.
- Por fim, no Capítulo 7 estão as considerações finais, destacando as contribuições e os resultados obtidos. Esse capítulo inclui a apresentação das publicações resultantes dos trabalhos realizados no decorrer do estudo e as possibilidades de trabalhos futuros.

2. FUNDAMENTAÇÃO CONCEITUAL

Este capítulo apresenta um estudo bibliográfico relativo aos assuntos permeados por esta tese. Na próxima seção, serão apresentados definições e conceitos sobre a temática do movimento de acesso aberto (Seção 2.1). Em seguida, serão apresentados conceitos sobre bibliometria, bem como sobre as principais métricas adotadas para análises bibliométricas (Seção 2.2). Já na Seção 2.3 são apresentados conceitos relacionados aos currículos cadastrados na Plataforma Lattes.

2.1 Movimento de Acesso Aberto

O termo “acesso aberto” é uma tradução do inglês *Open Access* (OA), que está relacionado especificamente ao acesso gratuito à informação científica na internet, principalmente de artigos científicos que passaram pela revisão por pares e foram aceitos ou publicados em uma revista científica especializada. Sua filosofia está fundamentada na premissa de que a pesquisa científica é, em sua maioria, financiada com recursos públicos, portanto os seus resultados deveriam estar disponíveis e acessíveis publicamente, sem nenhum custo para os usuários. Destaca-se que os pesquisadores não escrevem por dinheiro, mas para maximizar a visibilidade, o uso e o impacto dos resultados de suas pesquisas. Embora o processo de editoração e divulgação de um artigo envolva gastos, eles não devem ser repassados aos leitores (FREIRE, 2011).

Kuramoto (2006) afirma que no final da década de 90 surgiram diversas manifestações em favor do acesso aberto à informação científica, como consequência das dificuldades encontradas em face da crise dos periódicos científicos. Buscaram-se alternativas de solução para manter o acesso a essas revistas por meio da formação de consórcios, criando-se portais de acesso às revistas eletrônicas, mas as negociações com os editores foram e ainda são difíceis. Com o estabelecimento de ferramentas de *software* para a construção de repositórios digitais e publicações periódicas eletrônicas em conformidade com o formato de acesso aberto, o modelo de acesso aberto ganha consistência. Logo, diversas declarações em favor desse modelo foram (e ainda estão) sendo publicadas, intensificando, assim, a implantação de publicações periódicas eletrônicas e repositórios institucionais e temáticos de acesso livre.

Uma das principais mudanças ocorridas na comunicação científica teve seu ápice com a crise dos periódicos na década de 1990, quando o abusivo aumento no custo das assinaturas tornou inviável a manutenção de coleções por instituições ou pesquisadores. Uma das principais alternativas encontradas foi a mudança de paradigma do movimento de acesso livre à informação científica, marco para a comunicação científica. Esse movimento teve seu primeiro documento, a Declaração de Budapeste, lançado em reunião promovida pelo *Open Society Institute* (OSI), marco inicial em que foram definidos princípios e estratégias para concretizar e assegurar o acesso livre à informação científica (CHALHUB e PINHEIRO, 2011).

Para Suber (2007), o acesso aberto impulsiona o retorno dos esforços realizados em pesquisas com investimento público, tornando os resultados destas acessíveis com maior facilidade. Independentemente dos significados que o termo encerra, o acesso aberto deve ser discutido com base em diferentes aspectos, dentre os quais estão aqueles concernentes ao âmbito a que se refere: à literatura ou ao conhecimento nela registrado. É importante ressaltar que o acesso aberto ao conhecimento científico se refere tanto aos aspectos formais quanto aos informais do processo de comunicação científica (LEITE, 2006).

Mueller (2006) destaca que os canais mais importantes em acesso aberto são os periódicos científicos eletrônicos, os repositórios temáticos, os repositórios institucionais e o autoarquivamento em páginas pessoais. Ressalta-se ainda que, nesse âmbito, os periódicos científicos eletrônicos já estão consolidados, com uma grande quantidade de periódicos em acesso aberto sendo, inclusive, reconhecidos por diversas áreas do conhecimento.

De acordo com a Declaração de Berlin (2003), o acesso aberto à informação científica deve atender a duas condições: 1) o(s) autor(es) e o(s) detentor(es) dos direitos autorais concede(m) a todos os usuários o direito de acesso gratuito, irrevogável e irrestrito e uma licença para copiar, usar, distribuir, transmitir e exibir o trabalho publicamente, bem como para distribuir obras derivadas, em qualquer suporte digital, para qualquer propósito responsável, sujeito à correta atribuição da autoria. Além disso, todos os usuários devem ter o direito de fazer um pequeno número de cópias impressas para seu uso pessoal; 2) uma versão completa da obra e todos os materiais suplementares, incluindo uma cópia da licença como acima definida, são depositados, num formato eletrônico normalizado e apropriado, em pelo menos um repositório que utilize normas técnicas adequadas (como as definidas pelo Open Archive) e seja mantido por uma instituição acadêmica, sociedade científica, organismo governamental ou outra organização estabelecida que pretenda promover o acesso aberto, a distribuição irrestrita, a interoperabilidade e o armazenamento a longo prazo (FREIRE, 2011).

Na Figura 2.1, é mostrado como os movimentos para o acesso aberto surgiram no início dos anos 90, motivados pela ascensão da tecnologia da comunicação e da informação. A iniciativa para os arquivos abertos deu-se com a construção de um ferramental para arquivamento dos arquivos digitais. Já os movimentos para o acesso aberto tiveram como objetivo a conscientização dos cientistas e a consequente propagação do conhecimento científico para o avanço da ciência.

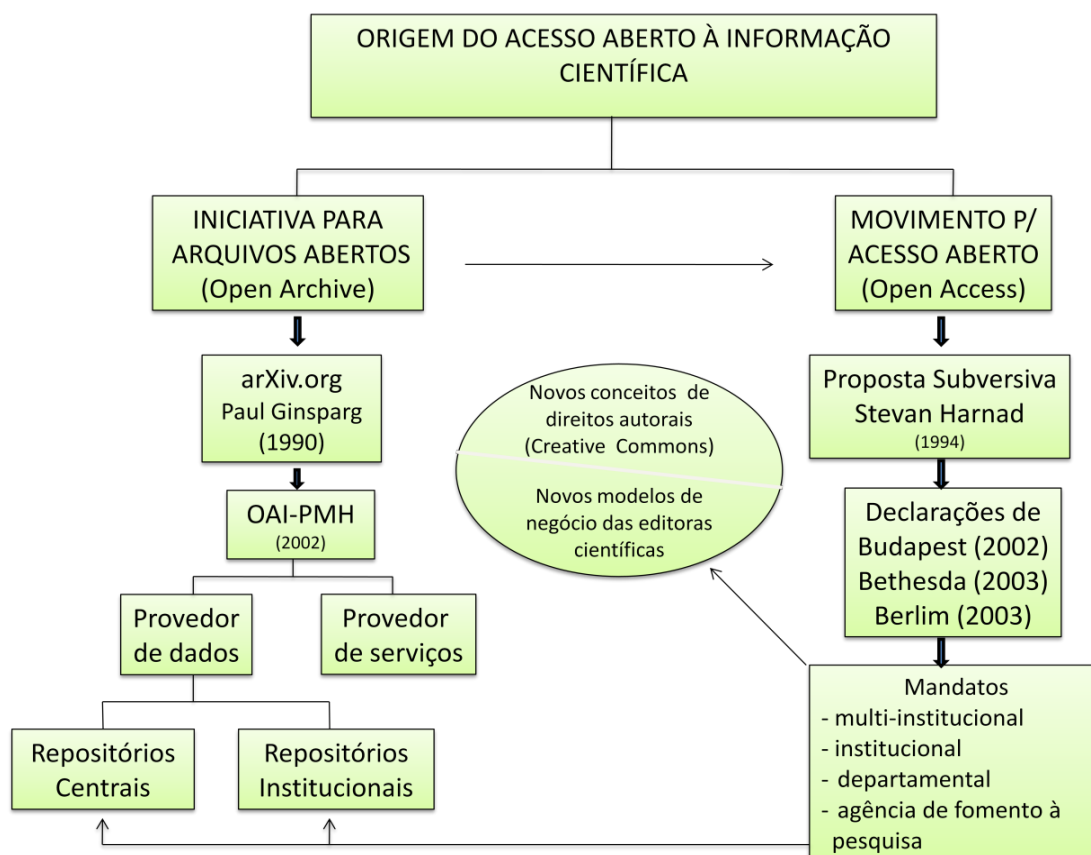


Figura 2.1 - Diagrama com as iniciativas para o Acesso Aberto (Freire, 2011).

Com a ideia de que os pesquisadores tinham interesse em disseminar seus conhecimentos para o maior número possível de pares e ampliar a visibilidade de seus artigos, o professor da ciência cognitiva Stevan Harnad iniciou o movimento de acesso aberto com a chamada “proposta subversiva”: postar toda a produção científica proveniente de resultados de pesquisas na internet para acesso livre e irrestrito (OAKERSON e O'DONNELL, 1995).

Lançada em 2002, a Declaração de Budapeste, publicada pela Budapest Open Access Initiative (BOAI), propõe pela primeira vez o termo *open access* a todas as disciplinas e países, com o intuito de acelerar internacionalmente o processo de disponibilização na internet dos artigos de todos os segmentos acadêmicos. Ainda em 2002, na Convenção de Santa Fé, Novo México, EUA, foi discutido o modelo arXiv, no qual foram introduzidos dispositivos que têm a capacidade de se comunicar com os diferentes tipos de arquivos pelo chamado *Protocol for Metadata Harvesting* (OAI – PMH).

Surgida posteriormente, de acordo com Silva e Alcará (2008), a *Bethesda on Open Access Publishing*³ é resultado da reunião que ocorreu no *Haward Hughes Medicinal Institute* em abril de 2003, com o objetivo de estimular a discussão sobre o acesso aberto na comunidade de pesquisadores da área de Biomédica, bem como de delinear princípios para obter apoio formal das agências de financiamento, sociedades científicas, editores, bibliotecários e pesquisadores para a publicação dos resultados das pesquisas científicas.

³ <https://legacy.earlham.edu/~peters/fos/bethesda.htm>

A Declaração de Berlim, de 2003, foi escrita por pesquisadores de universidades europeias, e trata do acesso livre ao conhecimento com foco nas ciências e humanidades, caracterizado como universal e parte do alicerce científico. São apresentadas diretrizes para as contribuições de acesso livre, tais como a concessão completa de todos os direitos dos autores, com imposição de uso responsável, e, ainda, para os formatos das obras, que devem ser completos e disponíveis em repositórios adequados às normas técnicas, de modo a garantir o acesso irrestrito (ANDRADE e MURIEL-TORRADO, 2017).

Nesse contexto, em termos práticos, surgem os provedores de dados. As instituições de ensino e pesquisa são exemplos de provedores de dados que coordenam e mantêm o padrão OAI-PMH. O Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), responsável por um importante repositório de teses e dissertações mantido pelas instituições de ensino superior brasileiras, é considerado um exemplo de provedor de acesso que usa mecanismos de coleta de metadados apresentados pelos provedores de dados.

Destaca-se que os repositórios digitais possuem duas características-chave na maioria das instituições de pesquisa. São elas: 1) servir como elemento catalisador no sistema de comunicação científica, por expandir o acesso à pesquisa, afirmar o controle sobre a comunidade acadêmica e destacar a relevância das instituições e bibliotecas que os apoiam; 2) potencial para servir como indicadores tangíveis da qualidade de uma instituição e demonstrar a relevância científica, social e econômica de suas atividades de pesquisa, assim como aumentar a visibilidade, o *status* e o valor público da instituição (CROW, 2002).

Os repositórios digitais de acesso aberto são arquivos eletrônicos destinados ao armazenamento e à preservação da produção científica gerada por um pesquisador, ou por um departamento ou instituição de ensino e pesquisa. Eles possuem mecanismos de interoperabilidade que permitem a integração de vários repositórios na formação de uma única base global de informação científica. Nesse sentido, quando um pesquisador realiza o *upload* (autoarquivamento) de seu artigo para um dos repositórios listados na Figura 2.1, suas pesquisas tornam-se mundialmente disponíveis na internet (FREIRE, 2011).

Um repositório digital é composto por variados tipos de documentos, em suas diversas tipologias documentais, de várias disciplinas. Entre essa diversidade de conteúdos, pode-se citar: *e-books*, artigos científicos, relatórios técnicos, teses e dissertações, *pré* e *post-prints*, peças publicitárias e artigos de anais de congressos.

Segundo Leite (2009), os repositórios digitais de acesso aberto, embora apresentem algumas semelhanças com as bibliotecas digitais, possuem atributos específicos que os distinguem destas. Os repositórios digitais de acesso aberto, além de abrigarem exclusivamente materiais científicos referentes à produção intelectual de uma instituição de ensino e pesquisa, desenvolvem-se num contexto diferente das bibliotecas digitais. Para esse autor, os *softwares* dos repositórios digitais de acesso aberto foram desenhados para atender particularidades do processo de comunicação científica, principalmente no que tange à gestão da informação.

Outro aspecto importante a se destacar com relação ao acesso aberto à publicação científica está relacionado ao modelo de negócio adotado pelos periódicos. De acordo com Cockerill

(2006), a característica de um modelo de negócio para acesso aberto universal ao conhecimento científico depende da dissociação, em relação ao consumidor final, dos custos inerentes ao processo de publicação, isto é, o leitor, ao acessar essas informações, deve estar livre de encargos técnicos, legais ou financeiros. Freire (2011) apresenta uma síntese dos modelos de negócio das editoras científicas (Quadro 2.1).

Modelos de negócio	Origem dos recursos para publicação	Acesso à informação		Observações
		Vantagens	Desvantagens	
Assinatura individual ou institucional (versão impressa)	Editoras	Acesso à informação certificada pelos pares	Acesso restrito aos assinantes	O orçamento das bibliotecas não permite a aquisição de todas as revistas de interesse, devido aos altos custos e ao progressivo aumento de novos fascículos.
Big Deal (assinatura institucional – versão eletrônica)	Consórcio entre bibliotecas	Maior poder de negociação junto às editoras	Acesso restrito aos membros das instituições consorciadas. A aquisição em forma de “pacotes de revistas” oferece pouca flexibilidade para a escolha e cancelamento de títulos individuais.	A aquisição das revistas por pacotes nem sempre contempla as necessidades de informação de uma instituição ou de um país.
Pay-per-view (pagar para ver)	Editoras	Permite a aquisição do artigo no lugar da revista	Nem todos os leitores contam com recursos para taxa de <i>download</i> .	Poucas revistas disponibilizam essa opção de acesso.
Author-pay (autor paga) ou híbrido	Autor ou suas instituições filiadas	Acesso livre na internet. Os leitores são livres de encargos para acessar o artigo.	Nem todos os pesquisadores ou possíveis patrocinadores possuem recursos para bancar os custos de publicação.	Poucas revistas ainda adotam esse modelo.
Delayed Open Access (acesso postergado)	Editoras	Acesso livre após um período de embargo da editora a partir da data de publicação do artigo. Atende às necessidades de algumas áreas do conhecimento que não dependem da velocidade de informações correntes.	O atraso ao acesso à informação inviabiliza o avanço do conhecimento para áreas que dependem da velocidade de informações correntes.	O período de embargo varia extraordinariamente entre editoras e áreas do conhecimento.
Full Open Access (acesso aberto total)	Universidades e instituições de financiamento à pesquisa (públicas e privadas)	Acesso aberto universal na internet. Maior visibilidade e fator de impacto		Poucas revistas adotam esse modelo. As editoras comerciais temem que esse modelo possa comprometer a sustentabilidade do sistema por assinatura.
Self-archiving (autoarquivamento)	Editora	Arquivamento de uma cópia do artigo, publicado em revistas por assinatura, em um repositório de acesso aberto.	Nem sempre as editoras permitem esse comportamento por parte dos autores. Algumas editoras cobram ou exigem um período de embargo.	Poucos autores têm adotado esse comportamento. Falta de consciência e incentivo são as causas mais prováveis.

Quadro 2.1: Modelos de negócio para acesso à informação científica (Freire, 2011).

Considerando as modalidades de acesso aberto, observa-se que o modelo de negócio de acesso aberto opcional ao autor ou híbrido (*Author-pay*) é um modelo em que as revistas tradicionais por assinatura deixam como opção para o autor pagar o custo da publicação, tornando esse artigo, em particular, em acesso aberto, mesmo que os outros artigos sejam de acesso restrito.

Já o modelo de acesso aberto postergado (*Delayed Open Access*) disponibiliza os artigos em canal de acesso aberto após o tempo de sua publicação tradicional por assinatura. Para determinadas áreas do conhecimento, em que a velocidade de informações não é impactante, é uma opção considerada.

O modelo de negócio para o acesso aberto total (*Full Open Access*) adere aos princípios existentes nos movimentos de acesso aberto, em que é permitido ao leitor o uso e o reuso do artigo sem nenhuma restrição, desde que com fins não comerciais. Esse modelo se caracteriza como uma concorrência significativa para as editoras comerciais, tendo em vista que podem comprometer a sustentabilidade do sistema por assinatura.

Por fim, há o autoarquivamento (*Self-archiving*), que consiste em armazenar, em repositórios de acesso aberto, uma cópia dos trabalhos publicados nas revistas. Tal modelo tem tido pouca adesão por parte dos autores, tendo como principais limitadores a falta de consciência e incentivo e as restrições impostas por algumas editoras, o que o torna pouco difundido.

Tendo em vista os diversos modelos de publicação em acesso aberto, cada qual com suas vantagens e limitações, é possível identificar diversos benefícios ao se publicar os artigos em periódicos de acesso aberto. Tais benefícios formam um fluxo que deve ser considerado por todos os autores no momento de realizar a publicação de seus artigos (Figura 2.2).



Figura 2.2 - Fluxo com os benefícios da publicação em acesso aberto (FREIRE, 2011).

Sabe-se que as editoras ligadas ao modelo de acesso aberto não têm fins lucrativos, pois as fontes que o financiam são públicas. Quando os trabalhos são inseridos na internet, há visibilidade e aumento no impacto, formando, assim, um ciclo vicioso que impacta a reputação e o prestígio da revista e, principalmente, do autor.

No Brasil, diversos esforços têm sido realizados com o objetivo de dar maior visibilidade à produção científica local. Um deles é o IBICT (Instituto Brasileiro de Informação em Ciência e Tecnologia), referência em acesso livre à informação científica, que ao longo dos anos vem propondo ações de incentivo para fornecer maior visibilidade à produção científica brasileira. Dentre as iniciativas, uma que se destaca é o lançamento do Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica (Anexo A), endereçado às universidades, aos institutos de pesquisa, à comunidade científica, às agências de fomento e às editoras comerciais de publicação, objetivando:

- Promover o registro da produção científica brasileira, em consonância com o paradigma do acesso livre à informação;
- Disseminar a produção científica brasileira, em consonância com o paradigma do acesso livre à informação;
- Estabelecer uma política nacional de acesso livre à informação científica;
- Buscar apoio da comunidade científica em prol do acesso livre à informação científica.

Outras iniciativas que se destacam são as propostas de políticas que visam ao avanço do acesso aberto, como a Portaria nº 13 da CAPES, de 15 de fevereiro de 2006 (Anexo B), que institui a obrigatoriedade de realização do depósito digital de dissertações e teses provenientes dos programas de pós-graduação *stricto sensu* reconhecidos no país. Ainda, há o Projeto de Lei nº 1120/2017 (Anexo C), que dispõe sobre o processo de disseminação da produção técnico-científica pelas instituições de ensino superior no Brasil. Além disso, recentemente, a Portaria CTA nº 01/2019 da FAPESP (Anexo D) instituiu a política para o acesso aberto às publicações resultantes de auxílios e bolsas da FAPESP.

2.2 Bibliometria

Conforme a produção científica mundial cresce de forma expressiva, surgem também alguns problemas referentes a esse crescimento, aos quais respondem mecanismos para controle e monitoramento. Nesse contexto, a bibliometria surge como uma ferramenta estatística que possibilita estruturar e gerar indicadores distintos de gestão de informação e conhecimento.

A bibliometria tem como objetivo desenvolver padrões e modelar matematicamente os processos para as medições, possibilitando, a partir dos resultados, traçar previsões e tomar as possíveis decisões.

Por meio de suas técnicas, a bibliometria procura estudar os aspectos quantitativos da ciência e da produção científica como uma atividade que envolve características sociais, econômicas e políticas. Ela fornece um instrumental para estudos que visam mapear o conhecimento

científico e extrair informações, bem como explicita a compreensão de como a produção científica tem sido realizada (HAYASHI, 2012).

Pritchard (1969) define a bibliometria como um campo de estudos cujo objetivo é 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, grupos de pesquisa, instituições, regiões geográficas, organizações ou eventos.

A bibliometria também pode ser definida como o estudo dos aspectos quantitativos da produção, disseminação e uso da informação. Nela são desenvolvidos métodos matemáticos focando a construção de processos, cujos resultados podem ser utilizados para embasar tomadas de decisão (Tague-Sutcliffe, 1992).

Para Vanti (2002), vale ressaltar algumas vantagens de se utilizar as técnicas 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 as revistas mais relevantes de uma determinada área;
- Mensurar a cobertura das revistas secundárias;
- Identificar os usuários de uma disciplina;
- Prever as tendências de publicação;
- 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 cocitação;
- Determinar o desempenho dos sistemas de recuperação da informação;
- Avaliar os aspectos estatísticos da linguagem, das palavras e das frases;
- Avaliar a circulação e o uso de documentos em um centro de documentação;
- Medir o crescimento de determinadas áreas e o surgimento de novos tópicos.

Nesse contexto, diversas técnicas podem ser aplicadas à literatura científica com o intuito de se avaliar indicadores bibliométricos. Guedes e Borschiver (2005) destacam três leis importantes na bibliometria: 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. Estudando a distribuição dos artigos em termos de variáveis de proximidade ou de afastamento, com o objetivo de descobrir a extensão na qual artigos de um assunto científico específico apareciam em periódicos destinados a outros assuntos, as observações de Bradford culminaram, em 1934, na formulação da lei da dispersão. Verifica-se, na Lei de Bradford, a qual 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).

Já a Lei de Lotka, proposta em 1926, também conhecida como Lei do Quadrado Inverso, propõe que um número restrito de pesquisadores produz muito em determinada área do conhecimento, enquanto um grande volume de pesquisadores produz pouco. Lotka estudou os autores presentes no *Chemical Abstracts* entre 1909 e 1916, e identificou que grande parte da produção científica é realizada por poucos autores. A produção desse número reduzido de autores se iguala em quantidade ao desempenho de muitos autores que possuem baixo volume de publicação. A representação desse princípio pode ser expressa matematicamente da seguinte maneira: o número de autores que publica n artigos (n é igual à quantidade de artigos) é igual a $1/n^2$ dos autores que publicam somente um artigo (MACHADO-JUNIOR, *et al.*, 2014).

A Lei de Zipf está relacionada à frequência de ocorrência de palavras em um dado texto. Essa lei desenvolveu e estendeu uma lei empírica observada por Estoup, em 1916, a qual estabelece uma relação entre a posição de uma palavra e a frequência de seu aparecimento em um texto longo. A Lei de Zipf é assim formulada: $r \cdot f = c$, sendo que “ r ” é a posição da palavra, “ f ” é a frequência e “ c ” é uma constante. Zipf extraiu sua lei do princípio geral do “esforço mínimo”, segundo o qual uma palavra cujo custo de utilização seja pequeno ou cuja transmissão demande um esforço mínimo é frequentemente usada em um texto grande (KLEINUBING, 2010).

Na Tabela 2.1, Guedes e Borschiver (2005) resumizam um conjunto de princípios bibliométricos, relacionados a seus objetos de estudo e a suas aplicações para a melhor compreensão da comunicação científica.

Tabela 2.1 - Princípios Bibliométricos e suas Aplicações (Guedes e Borschiver, 2005).

Bibliometria		
Leis e Princípios	Focos de Estudo	Principais Aplicações
Lei de Bradford	Periódicos	Estimar o grau de relevância de periódicos em dada área do conhecimento
Lei de Lotka	Autores	Estimar o grau de relevância de autores em dada área do conhecimento
Lei de Zipf	Palavras	Indexar artigos científicos e tecnológicos automaticamente

Além dessas 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, em que as ideias de um trabalho são o material infectante, e os leitores as pessoas suscetíveis à doença.

A característica multidisciplinar da bibliometria permite a aplicação de suas métricas em áreas distintas de pesquisa, possibilitando, desse modo, uma análise quantitativa para a medição da produção científica e dos mecanismos de disseminação do conhecimento. Dessa maneira, as

análises bibliométricas viabilizam experimentos quantitativos e estatísticos utilizando técnicas para essas análises complexas sobre a produção científica.

No contexto deste projeto, as análises bibliométricas irão quantificar diversos indicadores sobre a produção científica brasileira em acesso aberto.

2.3 Plataforma Lattes

Nesta tese, a base curricular da Plataforma Lattes é a principal fonte de dados utilizada. A Plataforma Lattes, que tem contínuo crescimento na quantidade de currículos nela cadastrados, possui relevância e reconhecimento internacional (Dias, 2016). Grande parte dos programas de financiamento à 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 a análise da produção científica brasileira.

Para Lane (2010), existem diversos esforços empregados na construção de repositórios confiáveis de dados científicos que poderiam permitir análises com o objetivo de explorar e compreender como a ciência tem evoluído. No entanto, são destacados vários problemas que comprometem o sucesso de tais iniciativas. Exemplos disso são informações dependentes de *softwares* específicos, repositórios dependentes de financiamentos transitórios (e esforços para poder combinar tais repositórios) 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 deve ser incentivada por órgãos federais, instituições e órgãos financiadores. A autora ainda destaca que a Plataforma Lattes é uma das fontes de dados de pesquisadores mais confiáveis atualmente.

A Plataforma Lattes representa a experiência do CNPq na integração de bases de dados de currículos, de grupos de pesquisa e de instituições em um único sistema de informações. Sua dimensão atual se estende não só às ações de planejamento, gestão e operacionalização do fomento do CNPq, mas também aos feitos de outras agências de fomento federais e estaduais, das fundações estaduais de apoio à ciência e tecnologia, das instituições de ensino superior e dos institutos de pesquisa. Além disso, ela se tornou estratégica não só para as atividades de planejamento e gestão, mas também para a formulação das políticas do Ministério da Ciência, Tecnologia, Inovações e Comunicações (CNPq, 2021).

A Plataforma Lattes é composta por três diretórios: 1) um de currículos, responsável por armazenar informações sobre a vida pregressa e atual de pesquisadores e estudantes do país; 2) um de grupos de pesquisa do Brasil, que contém informações sobre os seus integrantes, linhas de pesquisa e setores de atividades envolvidos; 3) um de instituições, que registra todas as organizações ou entidades que estabeleçam algum tipo de relacionamento com o CNPq. No presente estudo, somente o diretório de currículos foi utilizado.

Atualmente, o diretório curricular da Plataforma Lattes é composto por mais de 7 milhões de currículos cadastrados (março/2021). Cada indivíduo pode possuir somente um único currículo, e nele podem ser registradas informações gerais do indivíduo, seu histórico de formação acadêmica, sua produção científica, técnica, cultural e artística, suas participações em eventos e orientações, entre outros. Cada currículo é disponibilizado em uma página HTML (*Hypertext Markup Language*) para visualização, também sendo possível obter uma versão em formato XML (*Extensible Markup Language*) contendo todos os dados registrados em cada currículo.

A seguir, são apresentadas as informações mais relevantes de cada um dos nove módulos que compõem os currículos:

- **Dados Gerais:** nesse módulo é possível cadastrar informações pessoais, formação acadêmica e complementar, atuação profissional, idiomas e outras informações relevantes.
- **Projetos:** módulo que dá acesso aos itens “Projetos de pesquisa”, “Projetos de extensão”, “Projetos de ensino” e “Projetos de desenvolvimento”. Nesse item é possível editar e cadastrar com detalhamento os projetos realizados.
- **Produção Bibliográfica:** agrupa informações a respeito da produção bibliográfica (publicações) realizada pelo indivíduo, como artigos em anais de eventos ou em periódicos, livros e capítulos de livros, entre outros.
- **Produção Técnica:** esse módulo traz informações a respeito da produção técnica realizada, como patentes, *softwares*, mapas, produtos, processos técnicos, cartas, maquetes e editorações, entre outras.
- **Orientações:** apresenta informações a respeito de orientações ou supervisões realizadas pelo indivíduo nos diversos níveis de formação, tanto as já concluídas como as em andamento.
- **Produção Cultural:** informações a respeito da produção artística ou cultural do indivíduo, como apresentação de obra artística, arranjo musical, composição musical, sonoplastia e outros.
- **Eventos:** agrupa informações a respeito de eventos dos quais o usuário tenha participado (organização, entre outros).
- **Bancas:** traz informações a respeito das bancas nas quais o usuário esteve presente, como bancas de trabalhos de conclusão ou de comissões julgadoras.
- **Citações:** possibilita cadastrar os índices em três entre as bases mais relevantes, a saber: *Web of Science*, Scielo e Scopus.

É importante ressaltar que o acesso aos currículos cadastrados na Plataforma Lattes é realizado com o auxílio de uma interface de consulta que apresenta cada currículo individualmente, não ofertando nenhuma ferramenta de análise dos dados curriculares de forma abrangente. Diante disso, para uma análise mais detalhada de conjuntos de indivíduos, se fazem necessários a adoção de técnicas e o desenvolvimento de ferramentas para coleta, tratamento e análises desses dados.

Diante disso, apesar dos desafios envolvidos em seu processo de tratamento, os currículos cadastrados na Plataforma Lattes se caracterizam como uma interessante fonte de dados para

estudos que visam verificar a produção científica brasileira em periódicos de acesso aberto, utilizando métricas bibliométricas como método de análise.

3. TRABALHOS CORRELATOS

Este capítulo apresenta o resumo do conjunto de trabalhos relacionados a esta tese. Primeiramente, nas seções (3.1 e 3.2), são apresentados trabalhos envolvendo o movimento de acesso aberto, além de estudos que utilizaram os currículos da Plataforma Lattes como principal fonte de dados para suas análises. Por fim, na Seção 3.3, são apresentadas as considerações sobre os trabalhos correlatos e os diferenciais propostos.

3.1 Trabalhos envolvendo publicações em acesso aberto

Nesta seção, são apresentados trabalhos que de alguma forma analisaram publicações em periódicos de acesso aberto, com o objetivo de compreender como tem evoluído a divulgação de estudos científicos nesse meio de comunicação. Para isso, destacam-se as principais iniciativas e movimentos para o aumento das publicações nesse formato.

No trabalho de Silva e Alcará (2008), são analisadas as políticas de acesso aberto à informação científica e as propostas de ação, com ênfase nas iniciativas governamentais em diferentes países. Foi identificado que o movimento de acesso livre à informação científica já era preocupação oficialmente registrada em vários países, embora com diferentes graus de desenvolvimento. Entre tais diferenças estão as próprias determinações das políticas, já que algumas obrigam instituições públicas e pesquisadores a disponibilizar em acesso aberto os resultados de suas pesquisas, enquanto outras apenas sugerem o envolvimento e a participação de pesquisadores e instituições no movimento.

Oliveira e Chalhub (2009) analisaram as revistas científicas ibero-americanas que aderiram ao movimento de acesso aberto, integrantes do *Directory of Open Access Journal* (DOAJ). Especificamente, esse estudo visou identificar as revistas da região ibero-americana incluídas no repositório, suas instituições editoras, periodicidade e áreas de cobertura, além de verificar sua inserção no movimento de acesso livre, por meio da análise de sua participação em outros espaços virtuais. Os resultados apontaram crescente adesão das revistas científicas da região. Em âmbito internacional, o Brasil e a Espanha ocupam a segunda e a quarta posições, respectivamente. As unidades de ensino e pesquisa representam a maioria das instituições editoras. Destaca-se, ainda, a diversidade das áreas das revistas, que vão desde a Engenharia até a Linguística, com participação expressiva da Medicina. Conclui-se que dados específicos de um diretório parecem fortalecer a legitimação do movimento de acesso livre por parte dos atores envolvidos no sistema de comunicação científica.

Em Chalhub e Pinheiro (2010), as autoras objetivaram identificar o tipo de adesão das universidades federais ao acesso aberto, a distribuição geográfica das instituições e os tipos de repositórios, além de discutir a funcionalidade, a usabilidade e a interatividade dos portais institucionais. No trabalho, foi utilizado o Portal do Ministério da Educação para identificar as universidades públicas e seus portais. Os dados foram coletados eletronicamente nos sites das universidades, adotando método de análise documental. Os resultados apresentam um quadro

diversificado de implantação das políticas de acesso aberto. As características gerais das universidades pesquisadas denotam concentração nas Regiões Sudeste (19), Nordeste (13) e Sul (11), havendo, porém, uma mudança quantitativa nas Regiões Norte (8) e Centro-Oeste (6), com a criação de universidades. Algumas universidades estão em estágio avançado de implantação dessas políticas, apresentando repositório institucional com a produção científica de seus pesquisadores. Os resultados apontam que há, mesmo que em estágio inicial na maioria das instituições, esforço de atendimento às políticas públicas de acesso aberto à informação científica. O enfoque maior é o acesso a repositórios eletrônicos e repositórios de teses e dissertações, enquanto os repositórios institucionais representam número residual nas instituições públicas federais brasileiras.

O trabalho de Freire (2011) propôs um sistema de gestão da produção científica no CNPq por meio de repositório digital de acesso aberto. Para tanto, recorre-se à literatura científica pertinente para a compreensão do processo de comunicação científica e sua formalização em canais competentes. Utilizam-se também os documentos institucionais do CNPq e da CAPES para identificar elementos de reconhecimento dessas agências quanto à disseminação e ao acesso à informação. Além disso, são usados os bancos de dados da *Thomson Reuters* (ISI), do DOAJ e do Qualis da CAPES⁴, visando comparar a qualidade das revistas de acesso aberto, considerando o fator de impacto da base ISI e dos estratos do Qualis da CAPES como critérios para essa análise. Finalmente, aplica-se um questionário aos pesquisadores-bolsistas de produtividade em pesquisa do CNPq, visando conhecer suas razões para publicar artigos científicos, o nível de conhecimento sobre acesso aberto à informação científica e sua disposição em adotar os canais abertos para disseminação dos resultados de suas pesquisas.

Já em Chalhub e Pinheiro (2011), são identificados os principais canais de comunicação científica de acesso aberto utilizados por pesquisadores, e analisados os fatores intervenientes na adesão ao autoarquivamento de sua produção científica. O trabalho tinha como objetivo identificar os principais canais de comunicação científica em acesso aberto utilizados por pesquisadores de universidades públicas do estado do Rio de Janeiro. Foi utilizada a listagem de 47 Comitês de Assessoramento do CNPq para Bolsas de Produtividade em Pesquisa e efetuada uma amostragem probabilística estratificada por área de conhecimento, seguindo a divisão por Comitê de Assessoramento (Ciências Agrárias, Ciências Biológicas, Ciências Exatas e da Terra, Ciências da Saúde, Ciências Humanas, Ciências Sociais Aplicadas, Engenharias, e Linguística Letras e Artes). A partir da seleção dos pesquisadores contemplados pelo programa de Bolsa de Produtividade em Pesquisa do CNPq no ano de 2010, cuja relação está disponível no site desse órgão federal, foram identificados aqueles vinculados a universidades públicas do estado do Rio de Janeiro com cursos de pós-graduação *stricto sensu*. Após a identificação dos endereços eletrônicos dos selecionados, foi enviada correspondência contendo em anexo um formulário com questões fechadas e abertas sobre estas categorias: comportamento informacional, publicação de acesso aberto e adesão a repositório institucional.

De maneira geral, os resultados da pesquisa apontam para uma mudança na postura desses pesquisadores com relação à publicação de resultados de pesquisa em canais de acesso aberto. Algumas áreas apresentam publicações em canais formais de comunicação científica, como em

⁴<https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/veiculoPublicacaoQualis/listaConsultaGeralPeriodicos.jsf>

periódicos eletrônicos, e autoarquivamento em repositórios institucionais ou temáticos. Outras se inserem mais em iniciativas individuais ou de grupos de pesquisa, muitas vezes antecipando as políticas institucionais. Os pesquisadores foram unânimes com relação às vantagens da publicação em acesso aberto, e a democratização do conhecimento foi apontada pela maioria como a principal vantagem dessa adesão. Além desse aspecto, também aparece nas falas dos pesquisadores o benefício da comunicação entre pares — “trocas”, “parcerias” e “diálogos” — no processo de produção do conhecimento. Ainda, é sinalizada a importância da utilização desse canal aberto de comunicação em dois momentos distintos: para que o pesquisador acesse a informação para suas pesquisas e para que disponibilize seus resultados, possibilitando-lhes maior visibilidade e impacto.

Já no trabalho de Taga (2016), são analisados artigos de pesquisa sobre acesso aberto indexados na base de dados Scopus, publicados entre os anos de 2001 e 2015, visando propor um esquema de classificação dos subtemas e artigos de pesquisa sobre o acesso aberto, além de verificar a evolução dos subtemas identificados e organizar uma bibliografia sobre esse tema. Foram analisados 347 artigos de pesquisa e, como resultado, foi proposto um esquema de classificação sobre acesso aberto, sendo estabelecidas oito categorias de classificação. O estudo revelou um contínuo e crescente interesse da comunidade de pesquisadores em realizar estudos de caso dedicados à análise do desenvolvimento ou evolução do acesso aberto em relação a certos grupos, instituições, regiões e períodos.

No trabalho de Pavan e Barbosa (2017), foi analisada a existência de uma política nacional de financiamento público para a publicação de artigos em acesso aberto, examinando 29 agências de fomento à pesquisa. Para isso, com auxílio de questionários e análises dos sites institucionais, foi possível identificar que só uma minoria das agências possui uma política para financiar custos com as taxas de publicação dos artigos. Isso sugere a necessidade de uma reavaliação do sistema de fomento brasileiro e de criação de políticas estratégicas.

Percebe-se que diversos outros trabalhos também utilizaram questionários junto a conjuntos de pesquisadores para coleta de dados. É o caso do trabalho de Furnival e Silva-Jerez (2017), que se propôs a explorar de que modo várias dimensões do acesso aberto à literatura científica são percebidas por pesquisadores brasileiros, identificando também seus hábitos de publicação e de uso e citação de fontes em acesso aberto. Os autores levantaram opiniões de uma amostra de pesquisadores da comunidade científica brasileira, buscando identificar os fatores que afetam sua aceitação ou resistência à publicação de suas pesquisas nas revistas de acesso aberto ou ao depósito de cópias dos seus trabalhos em repositórios de acesso aberto. Foi aplicado um questionário a 643 doutores vinculados a universidades brasileiras, oriundos de todas as áreas do conhecimento. O questionário era composto por 29 questões de múltipla escolha. Foram coletadas informações de pesquisadores de 31 institutos de ensino superior e institutos de pesquisa do Brasil. A estratificação por áreas do conhecimento mostrou uma distribuição relativamente equilibrada entre elas, sendo as Ciências da Saúde (21%), seguidas pelas Ciências Exatas e da Terra (20,1%), as áreas mais representativas, e a menor porcentagem, de 1,1%, relativa à área de Linguística, Letras e Artes. Como resultado, foi observado que a maioria dos pesquisadores que responderam ao questionário detém conhecimento sobre o acesso aberto e apoiam seus princípios e algumas de suas ações, principalmente em relação à publicação de revistas nesse formato, o que se reflete também nos seus hábitos de uso e citação

dessas fontes. Porém, existe certa indefinição por parte deles com relação ao status do copyright, o que tem limitado a publicação de seus trabalhos em repositórios de acesso aberto como os repositórios institucionais. Por isso, é destacada a necessidade de maiores iniciativas informativas para esclarecer esses aspectos.

Diversos outros trabalhos têm analisado a visão dos pesquisadores sobre as publicações em acesso aberto, como os de Park e Qin (2007); Fry *et al.* (2011); Moore (2011); Bongiovani, Gómez e Miguel (2012); Suber (2012) e Santos e Monteiro (2013).

Visando explorar o cenário nacional e internacional e, dessa forma, apresentar uma investigação que busca uma solução tecnológica para efetivar o acesso aberto a dados de pesquisa, Pavão *et al.* (2018) propõem uma metodologia dividida em cinco etapas. São elas: a) identificação de práticas de acesso aberto a dados de pesquisa em instituições brasileiras; b) mapeamento de seus usuários e suas necessidades; c) proposta de um portal *web* para reunir a comunidade nacional; d) levantamento dos serviços e soluções tecnológicas existentes no cenário internacional para o compartilhamento de dados de pesquisa; e) proposição de recomendações para o apoio à criação de repositórios de dados de pesquisa em instituições nacionais e à sua agregação a uma rede de pesquisa em acesso aberto a dados de pesquisa. Como resultado, são propostas iniciativas e estratégias internacionais para a criação de repositório de dados de pesquisa e para a criação de comunidades de práticas em torno do assunto.

Para Minniti, Santoro e Belli (2018), a disseminação do movimento de acesso aberto nos países da América Latina e Caribe, impulsionada pelo crescimento de iniciativas regionais e nacionais, como a criação de bibliotecas de revistas digitais em acesso aberto, e o estabelecimento de políticas governamentais de apoio, tem fornecido evidências do significativo papel do acesso aberto para a participação desses países na produção científica global. No trabalho, são mapeadas as publicações em acesso aberto dos países da América Latina e Caribe, por meio de uma análise bibliométrica das publicações indexadas pela Web of Science (WoS) e pelo SciELO, durante o período de 2005 a 2017. É constatado que as publicações têm se intensificado significativamente no período examinado, e que, embora haja um crescimento na quantidade de publicações nesse formato, em alguns países a sua magnitude não se traduz em um peso relativo do acesso aberto na quantidade total de publicações.

No trabalho de Pavan e Barbosa (2018), as autoras analisam documentos publicados em acesso aberto entre os anos de 2012 e 2016 por autores que tenham afiliação brasileira e identificam o perfil dessas publicações. Para isso, dados de 930 periódicos e 63.847 documentos foram coletados da WoS. Destaca-se ainda que a produção científica brasileira em acesso aberto é caracterizada por um perfil endógeno, e que ainda se fazem necessárias políticas para estimular a publicação de artigos em acesso aberto, principalmente em periódicos internacionais.

Costa, Weitzel e Leta (2020) apresentam um estudo do nível de adesão da elite brasileira na publicação em periódicos de Acesso Aberto para divulgação dos resultados de suas pesquisas. Para tal estudo, foram coletados, na Plataforma Lattes, dados pessoais e acadêmicos e dados da produção científica referente ao período de 2000 a 2015, divididos em quadriênios, do grupo de 1.205 pesquisadores considerados como elite. Apesar da redução da adesão dos

pesquisadores classificados na faixa “muito baixa”, foi detectado no grupo da elite que somente 10% dos pesquisadores foram classificados como “alta” e “muito alta” adesão. Portanto, apesar de o grupo de pesquisadores brasileiros ser considerado elite, por mais que esteja em ascensão, a adesão dos periódicos de acesso aberto ainda é muito baixa.

Além desses trabalhos, diversos outros estudos têm analisado a produção científica em canais de acesso aberto, como Swan (2008); Rodrigues e Oliveira (2012); Leta *et al.* (2017); Marchiori *et al.* (2018) e Zhang e Watson (2018).

3.2 Trabalhos envolvendo Repositório de Dados da Plataforma Lattes

Para Lane (2010), em publicação na revista Nature, a Plataforma Lattes é uma das fontes de dados de pesquisadores mais confiáveis entre as existentes. Por isso, tendo em vista que este projeto visa analisar a produção científica em acesso aberto dos pesquisadores brasileiros, nesta seção são apresentados alguns trabalhos relacionados ao presente estudo, que utilizam o repositório curricular da Plataforma Lattes como principal fonte de dados.

Diversos trabalhos que utilizam o repositório de dados curriculares da Plataforma Lattes têm como objetivo a utilização ou o desenvolvimento de ferramentas para tratamento e análise dos dados. No trabalho de Cardoso e Machado (2008), é aplicada e analisada uma ferramenta de mineração de dados para extrair conhecimento referente à produção científica de pesquisadores de uma instituição de ensino e pesquisa do país, tendo como fonte de dados os currículos cadastrados na Plataforma Lattes desses pesquisadores. São destacadas algumas limitações na análise dos resultados, indicando que ainda seria preciso padronizar o modo de preenchimento dos currículos cadastrados na Plataforma Lattes para, dessa forma, refinar as análises e obter indicadores. Como resultado, foi possível gerar um banco de dados estruturado para auxiliar na elaboração de novas políticas de gestão científica e tecnológica, tendo em vista o aperfeiçoamento do sistema de ensino superior brasileiro.

Mena-Chalco e Cesar-Junior (2009) desenvolveram uma ferramenta de código fonte aberta, denominada ScriptLattes, para a extração de dados de um conjunto de currículos cadastrados na Plataforma Lattes. Sua primeira versão, lançada em 2005, foi desenvolvida com a finalidade de auxiliar na elaboração de relatórios sobre a produção bibliográfica dos docentes do Departamento de Ciência da Computação do IME/USP. Inicialmente, os relatórios eram baseados apenas nas informações cadastradas nos currículos. Posteriormente, foram sendo disponibilizadas novas funcionalidades, como mapas de geolocalização baseados nos endereços profissionais de cada currículo e grafos de colaborações. Os 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.

No trabalho de Mena-Chalco, Digiampietri e Oliveira (2012), os autores analisam os programas de pós-graduação em Ciência da Computação brasileiros, identificando seu 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. Essa pesquisa 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. Os resultados obtidos pela análise mostram que a área de Computação brasileira se caracteriza por publicar preferencialmente trabalhos completos em anais de congressos (54%), ao que se segue a publicação de artigos completos em periódicos (14%). Ainda nesse trabalho, foi mostrado que no segundo triênio (2007-2009) houve um aumento na publicação de artigos em periódicos.

Em Digiampietri *et al.* (2014) é apresentada uma estratégia para caracterizar os programas de pós-graduação em Ciência da Computação brasileiros e a relação entre eles. No trabalho, são feitas análises dos programas sob diversas perspectivas. São descritas: a classificação sob cada uma dessas perspectivas, a correlação entre as métricas de avaliação utilizadas, a maneira 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. Utilizando os relatórios trienais da CAPES, foram identificados 732 professores nos programas avaliados. De posse dos nomes dos professores, foi possível obter o identificador para o currículo de cada um deles, com o objetivo de extraí-los, posteriormente, 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 que contêm dados sobre conferências e periódicos de publicações científicas. Essas informações descrevem, entre outros, o nível de cada um dos periódicos ou conferências e a quantidade de citações de um determinado trabalho publicado. 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, dessa forma, caracterizar uma colaboração.

No trabalho de Trucolo e Digiampietri (2014) os autores identificam tendências por meio de mineração de texto e de análise histórica das produções científicas de doutores da área de Ciência da Informação. Inicialmente, obtiveram-se os dados dos currículos cadastrados na Plataforma Lattes do conjunto analisado. Posteriormente, foi realizada a extração automática dos termos mais importantes inseridos nos títulos das publicações e, finalmente, foram feitas regressões lineares e não lineares dos índices de importância dos termos extraídos, baseados em frequência. Como resultado, foi possível obter informações gerais sobre as tendências identificadas para a área de Ciência da Informação em curto, médio e longo prazo. O trabalho apresenta e aplica uma metodologia de identificação de tendências que ainda pode ser considerada um primeiro passo diante do potencial da análise de tendências para a produção científica nacional. Além disso, informações gerais sobre as tendências identificadas e os comportamentos dessas tendências ao longo do tempo são discutidas.

No estudo realizado por Boaventura *et al.* (2014) é caracterizada 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, e 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 endogamia 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 pelos autores são estas: (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 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.

Ferraz *et al.* (2015) utilizam o ScriptLattes para identificar e disponibilizar a produção científica brasileira relacionada à dengue. Inicialmente, foram geradas listas contendo os identificadores Lattes dos pesquisadores em dengue no país, para possibilitar o acesso direto ao currículo de cada um deles. Essa listagem foi a entrada para que a ferramenta ScriptLattes extraísse e gerasse relatórios com informações desses currículos. Foram compilados indicadores de desempenho dos pesquisadores e de seus grupos de pesquisa, analisando o andamento de suas produções e, dessa forma, contribuindo para a prospecção dos resultados das pesquisas. Como resultado, foi possível apresentar uma análise apurada, no âmbito da Saúde Coletiva, do andamento das pesquisas em dengue no país, possibilitando a criação ou reformulação de estratégias voltadas ao planejamento de pesquisas relacionadas ao diagnóstico, tratamento e prevenção de importantes condições clínicas.

Em Dias (2016) é realizada uma análise da produção científica brasileira, tendo como principal fonte de dados os currículos cadastrados na Plataforma Lattes. Diferentemente de outros trabalhos, o autor realizou um estudo abrangente que inicialmente analisou todos os mais de 5 milhões de currículos cadastrados na época do estudo. São apresentados métodos e técnicas que compõem um *framework* para a extração, tratamento e análise dos dados, denominado *LattesDataXplorer*. Com os dados coletados e armazenados localmente, métricas bibliométricas foram aplicadas e, a partir disso, foram analisadas as redes de colaboração dos doutores brasileiros, bem como a produção científica dos docentes em programas de pós-graduação no Brasil. Além disso, foi feito um abrangente estudo dos bolsistas de produtividade em pesquisa do CNPq. Nesse trabalho, que analisou todo o repositório de dados curriculares da Plataforma Lattes, foi possível traçar um retrato até então inédito sobre a evolução da colaboração e da produção científica brasileira ao longo do tempo.

Sidone, Haddade e Mena-Chalco (2017) identificam padrões espaciais de colaboração no Brasil e medem o papel da proximidade geográfica na determinação da interação entre pesquisadores. Utilizando um banco de dados composto por mais de um milhão de indivíduos e sete milhões de publicações, todos extraídos dos currículos cadastrados na Plataforma Lattes, são consolidadas informações sobre a colaboração inter-regional de pesquisa em termos de redes de coautoria científica entre 4.615 municípios, durante o período de 1992 a 2009, o que possibilitou analisar uma quantidade significativa de dados. Os efeitos da distância geográfica na colaboração foram medidos para diferentes áreas, estimando modelos de interação espacial. Os principais resultados fornecem fortes evidências de desconcentração geográfica da colaboração nos últimos anos. A distância continua sendo um fator significativo na determinação da intensidade do fluxo de conhecimento em redes de colaboração no Brasil, pois

um aumento de 100 km entre dois pesquisadores reduz a probabilidade de colaboração em uma média de 16%, embora existam variações entre as redes de diferentes áreas.

Já em Gomes, Dias e Moita (2018) são analisados os tópicos de pesquisa publicados por pesquisadores doutores ao longo dos últimos 55 anos de trajetória da ciência brasileira, no intuito de mapear o conhecimento científico e identificar temas em destaque. Para isso, são realizadas a caracterização geral e uma análise dos dados referentes às palavras-chave dos artigos publicados em anais de congressos e em periódicos pelos doutores que têm currículos cadastrados na Plataforma Lattes. Os resultados mostraram uma visão geral sobre as palavras-chave utilizadas pelos doutores e, assim, possibilitaram a identificação e a análise dos principais tópicos de pesquisa desenvolvidos por eles. Como principais contribuições, destacam-se a estratégia de análise de palavras-chave de publicações científicas dos artigos analisados e a constatação da validade dos resultados, do ponto de vista quantitativo, para o auxílio ao entendimento do desenvolvimento e evolução da ciência brasileira.

No trabalho de Dias, Moita e Dias (2019), os autores apresentam uma caracterização geral do conjunto de doutores brasileiros com currículos cadastrados na Plataforma Lattes e, posteriormente, realizam uma análise temporal da rede de colaboração científica desses doutores. Para isso, todas as publicações científicas do conjunto foram utilizadas para a identificação e caracterização das redes de colaboração. Como resultado, é apresentado de que maneira a colaboração científica do conjunto analisado foi sendo intensificada ao longo dos anos, com destaque para as grandes áreas de atuação de cada doutor.

Em Pessoa-Junior *et al.* (2020), os autores classificam um conjunto de 263.264 indivíduos com currículos cadastrados na Plataforma Lattes e doutorado concluído nas suas áreas de atuação. Após isso, utilizando as publicações científicas registradas nos currículos de cada indivíduo, foram caracterizadas redes de colaboração científica. O objetivo era realizar uma análise sobre as colaborações interdisciplinares do conjunto. Como resultado, foi possível verificar que aproximadamente 35% das colaborações são interdisciplinares e pelo menos 57% dos pesquisadores possuem pelo menos uma colaboração interdisciplinar.

Santiago, Affonso e Dias (2020) analisaram a participação de mulheres na ciência brasileira, utilizando como base de dados o conjunto de doutoras que possuem seus currículos cadastrados na Plataforma Lattes. Os dados foram coletados e selecionados, obtendo-se um conjunto de 125.515 currículos com registro de gênero feminino e a formação acadêmica doutorado concluída. Os dados das doutoras foram agrupados de acordo com as grandes áreas de atuação e formação acadêmica/titulação e, por meio deles, foi possível analisar a evolução acadêmica e a produção científica e tecnológica do conjunto de maneira temporal.

Logo, 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 e sobre diversas perspectivas de análise: Pacheco *et al.* (2007); Arruda *et al.* (2009); Alves; Yanasse; Soma (2011); Roos *et al.* (2014); Lima *et al.* (2015); Brito; Quoniam; Mena-Chalco (2016); Silva *et al.* (2016); Hilário; Grácio (2017); Caparelli; Digiampietri (2018); Dias; Moreira; Dias (2018); Tuesta *et al.* (2019).

3.3 Considerações sobre Trabalhos Correlatos e Diferenciais Propostos

Na literatura revisada, verificou-se um grande esforço para compreender a opinião dos pesquisadores sobre a divulgação do resultado de suas pesquisas em formato que possa ser acessado de forma fácil, livre e sem restrições. Dentre os principais trabalhos, destacam-se algumas iniciativas que avaliam, com o auxílio de questionários, o sentimento de pesquisadores com reconhecida relevância em suas áreas de atuação sobre a divulgação de suas pesquisas em formato de acesso aberto.

No entanto, apesar de ser possível identificar certo interesse por parte dos pesquisadores que responderam a tais questionários, percebe-se que, entre eles, os interessados em realizar publicações em acesso aberto ainda são pouco representativos, e que vários deles ainda possuem dúvidas sobre questões legais relativas à divulgação nesse formato.

Outro elemento a ser destacado diz respeito à quantidade de indivíduos que foram avaliados. Esses estudos trabalharam com conjuntos específicos e, conseqüentemente, com uma pequena parcela dos autores de publicações em acesso aberto, que não representam de forma significativa a produção científica brasileira nesse formato. Tendo em vista que, em geral, os avaliados são de um mesmo nível de formação ou de uma mesma região geográfica, as conclusões não podem representar de forma efetiva o perfil nacional dos autores.

Para além disso, alguns outros trabalhos objetivam analisar como instituições de ensino ou pesquisa têm mantido seus repositórios institucionais de acesso aberto, bem como avaliar quais esforços são necessários para que seus pesquisadores utilizem esses repositórios para arquivamento dos resultados de suas pesquisas.

Considerando os trabalhos que analisam publicações em periódicos de acesso aberto, percebe-se que a maioria analisa pequenos conjuntos de indivíduos, além de utilizar, em geral, repositórios de dados internacionais, negligenciando, dessa forma, publicações de algumas áreas do conhecimento e não representando de forma significativa a produção brasileira em acesso aberto como um todo.

A maioria dos trabalhos que avaliam o movimento de acesso aberto não tem como objeto principal de estudo as publicações nesse tipo de formato, e sim os repositórios ou os periódicos em acesso aberto. Logo, apesar de os trabalhos apresentados nesse capítulo serem importantes para compreender as iniciativas existentes e as opiniões dos pesquisadores brasileiros, bem como os principais repositórios em acesso aberto no Brasil, um estudo abrangente sobre os pesquisadores brasileiros que têm publicado artigos em periódicos de Acesso Aberto de forma ampla se faz necessário.

Diante disso, o presente estudo apresenta-se como a primeira análise abrangente sobre o perfil dos pesquisadores brasileiros que publicam em periódicos de Acesso Aberto, constituindo-se como uma análise até então inédita sobre a evolução das publicações nesse formato no Brasil.

Também são realizadas análises com adoção das principais leis bibliométricas (Lei de Bradford, Lei de Lotka e Lei de Zipf) no conjunto de dados selecionados. Como resultado, são caracterizados os periódicos de maior relevância do conjunto (Lei de Bradford), os autores mais representativos (Lei de Lotka) e os principais tópicos de investigação dos artigos considerados (Lei de Zipf).

Além disso, é também a primeira análise sobre a produção científica em periódicos de acesso aberto, tendo como fonte de dados todo o conjunto de publicações registradas em todos os currículos cadastrados na Plataforma Lattes. Considerando a relevância dos dados analisados, objeto de diversos estudos, os resultados a serem aqui obtidos são de extrema relevância para a compreensão da produção científica brasileira.

O conteúdo desta tese está inserido no contexto do processamento, caracterização, modelagem e análise de grandes volumes de dados sobre o perfil de pesquisadores e suas produções científicas, com o auxílio de métodos computacionais, métricas bibliométricas e visualização de dados. Para isso, no próximo capítulo serão apresentados os métodos e componentes utilizados na identificação e tratamento dos dados, que posteriormente serão analisados.

4. DESENVOLVIMENTO

Neste capítulo é descrito o arcabouço de componentes que suporta o desenvolvimento desta tese. Inicialmente, na Seção 4.1, é apresentado o processo utilizado para a aquisição do conjunto de dados a serem analisados. Posteriormente, na Seção 4.2, é apresentado o processo de filtragem e tratamento dos dados para o propósito deste trabalho. Por fim, é apresentada uma visão geral dos dados que serão analisados (Seção 4.3).

4.1 Aquisição dos Dados

No processo de extração dos dados para as análises realizadas neste trabalho, foram utilizados os dados curriculares da Plataforma Lattes do CNPq. Grande parte dos editais de financiamento de projetos de pesquisa, realizados por diversos órgãos de fomento, utiliza dados cadastrados nos currículos dos proponentes como uma das formas de avaliação das propostas. Logo, há um grande incentivo para que os pesquisadores mantenham as informações de seus currículos atualizadas, o que torna esses currículos uma excelente fonte de dados para análises. Por essa mesma razão, vários trabalhos têm utilizado a Plataforma Lattes como fonte de dados para diversos estudos, como redes de colaborações, análise de produção científica, genealogia acadêmica, entre outros.

Como já mencionado, atualmente a Plataforma Lattes conta com mais de 7 milhões de currículos cadastrados (segundo dados de março de 2021) e o aumento no número de usuários tem sido constante, impulsionado, principalmente, pelos órgãos governamentais e por agências de fomento que incentivam o cadastro e atualizações dos currículos.

Para as análises da produção científica brasileira em periódicos de acesso aberto, considerou-se o conjunto de todos os indivíduos com currículos cadastrados na Plataforma Lattes. A seleção desse conjunto de dados para as análises tem como motivação o fato de que a maioria dos pesquisadores brasileiros possui currículos cadastrados, e mesmo pesquisadores em início de carreira são incentivados a se registrar e manter seus currículos atualizados. Assim, o conjunto de indivíduos analisado neste trabalho compreende grande parte dos pesquisadores em atuação no Brasil.

Considerando que a maioria dos trabalhos correlatos analisou apenas grupos específicos de indivíduos, e tendo em vista que a manipulação de grandes quantidades de currículos da Plataforma Lattes não é uma tarefa trivial, já que existem problemas que envolvem recuperação de informação e algoritmos eficientes para manipulação de grandes volumes de dados, o *LattesDataXplorer* (DIAS, 2016), um *framework* para extração e tratamento dos dados, foi utilizado (Figura 4.1).

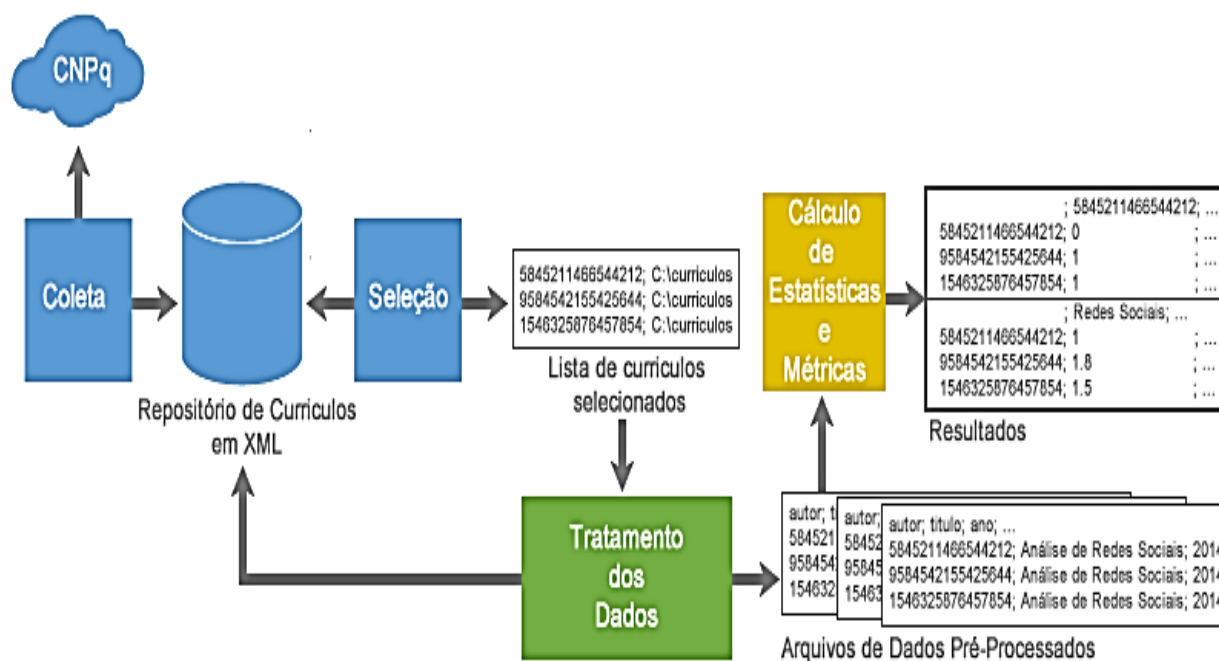


Figura 4.1 - Arquitetura do *LattesDataExplorer* (Dias, 2016).

O *LattesDataExplorer* é responsável por englobar todo um conjunto de técnicas e métodos de coleta e tratamento dos dados utilizados neste trabalho. Ele é composto por um conjunto de componentes que viabiliza a coleta e o tratamento dos currículos. O processo de extração de todos os dados curriculares da Plataforma Lattes é dividido em componentes que objetivam minimizar o custo computacional necessário para essa tarefa. 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 coletados possuem. 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 utilizado um componente denominado “Seleção”, que compõe subgrupos de currículos com base em informações presentes em seus registros (DIAS, 2016).

Logo, utilizando o *LattesDataExplorer* é possível agrupar um conjunto de currículos com base em parâmetros desejados. Na busca pelos parâmetros em cada um dos currículos, independentemente ou não da seção em que ele seja encontrado, esses currículos são selecionados e formam um grupo para análise. Seus dados são organizados em uma lista de currículos selecionados.

Posteriormente, tendo como entrada uma determinada listagem de currículos gerada pelo componente de Seleção, ou mesmo utilizando uma listagem global, com todo o repositório local de currículos, o componente “Tratamento dos dados” processa cada um dos currículos que compõem o conjunto a ser analisado. Tal processamento visa extrair dos currículos informações de interesse e agrupá-las em arquivos de dados pré-processados. O objetivo é gerar conjuntos de dados específicos para a aplicação de métricas de análise, não sendo mais necessário acessar o conjunto de currículos toda vez que uma análise precisa ser realizada. Com isso, os currículos,

que possuem uma grande quantidade de dados a serem processados, são acessados uma única vez. Como exemplos de arquivos de dados pré-processados, é possível destacar os que agrupam informações sobre orientações, formações acadêmicas, colaboração e produção científica, bem como os que tratam de projetos de pesquisa e produção técnica.

Finalmente, o componente denominado “Cálculo de estatísticas e métricas” tem como entrada um arquivo de dados pré-processados específico para a métrica a ser aplicada. Dessa maneira, para, por exemplo, analisar a produção científica de um determinado conjunto de pesquisadores, somente o arquivo que contém dados de produção científica é utilizado. Isso otimiza as análises, já que os arquivos sumarizam todos os dados de análise e não se faz mais necessário acessar os currículos e nem processar dados que não serão utilizados.

4.2 Filtragem e Tratamento dos Dados

Conforme já explanado, um currículo cadastrado na Plataforma Lattes pode conter diversas informações capazes de auxiliar na compreensão da evolução da ciência brasileira sob diversas perspectivas. No entanto, para atender aos propósitos desta tese, somente dados de publicações de artigos em periódicos de acesso aberto, bem como de seus autores, foram levados em conta. Diante disso, foi proposta uma extensão do *LattesDataXplorer*, com a inclusão de componentes *a priori* inexistentes, a qual avaliasse, para cada artigo publicado em periódico (a saber, 7.841.860), de cada um dos indivíduos (a saber, 6.548.210)⁵, se o periódico no qual aquele artigo havia sido publicado era de acesso aberto (Figura 4.2). Logo, com a proposta dessa extensão, somente os autores e as publicações de artigos em periódicos de acesso aberto foram analisados.

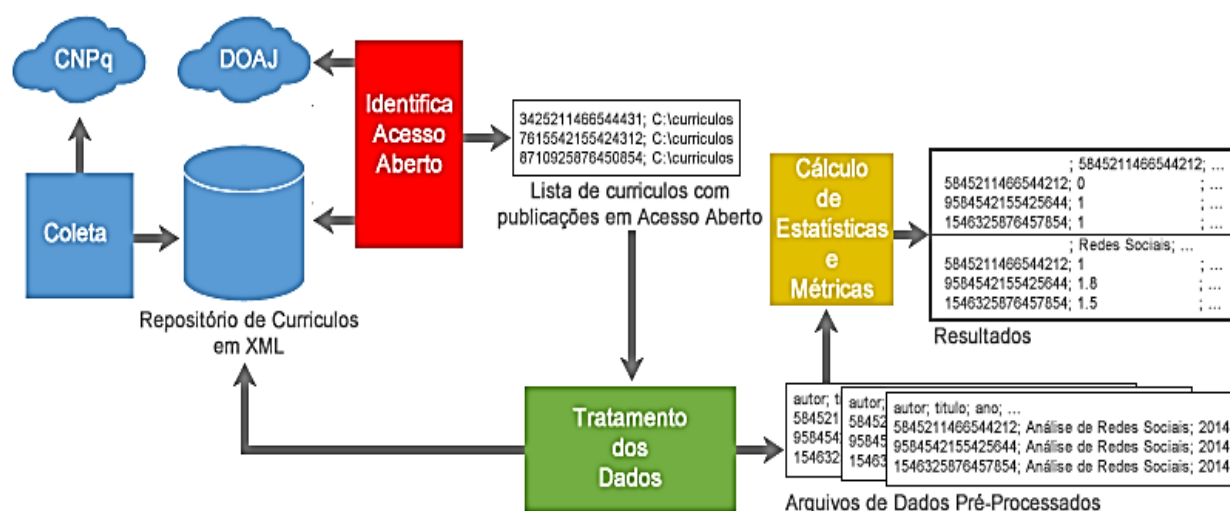


Figura 4.2 - LattesDataXplorer estendido.

⁵ Dados coletados em janeiro de 2020.

Inicialmente, utilizando o *LattesDataXplorer*, coletaram-se todos os currículos cadastrados na Plataforma Lattes em janeiro de 2020, os quais foram armazenados no repositório local (Repositório de Currículos em XML). Depois, o componente desenvolvido e denominado “Identifica Acesso Aberto” foi utilizado para recuperar todos os periódicos de acesso aberto registrados no portal do *Directory of Open Access Journals* (DOAJ), diretório *on-line* que indexa e fornece acesso a periódicos com acesso aberto. Em janeiro de 2020, o DOAJ indexava 14.170 periódicos e 4.543.939 artigos. O DOAJ tem sido fonte de dados e referencial sobre periódicos de acesso aberto para diversos estudos, como em Lone, Rather e Shah (2008); Swan (2008); Oiveira e Chalhub (2009); Gumpenberger, Ovalle-Perandones e Gorraiz (2013); Bjork e Salomon (2015); Morrison *et al.* (2015); Fukuzawa (2017); Costa (2018); Martín-Martin *et al.* (2018); Sanchez-Gil, Gorraiz e Melero-Fuentes (2018); Sarvo (2018).

Com todo o conjunto de currículos cadastrados na Plataforma Lattes armazenado no “Repositório de currículos em XML”, foram aplicados os métodos propostos e implementados no componente “Identifica Acesso Aberto”, para identificar os pesquisadores que têm publicado em periódicos de acesso aberto no Brasil.

Visando otimizar ao máximo possível o processamento dos currículos, toda vez que uma publicação analisada em 1 currículo cujo ISSN ou o eISSN do periódico estivesse contido na relação de periódicos em acesso aberto extraída do DOAJ, imediatamente o identificador do currículo em análise era inserido na lista de currículos selecionados, e o próximo currículo era analisado.

Após a análise de todos os currículos que compõem o repositório local, uma lista contendo todos os currículos com publicações em acesso aberto foi gerada, e ela se tornou a base para o componente “Tratamento de Dados”, que passou a incorporar os métodos propostos neste trabalho (Figura 4.3).

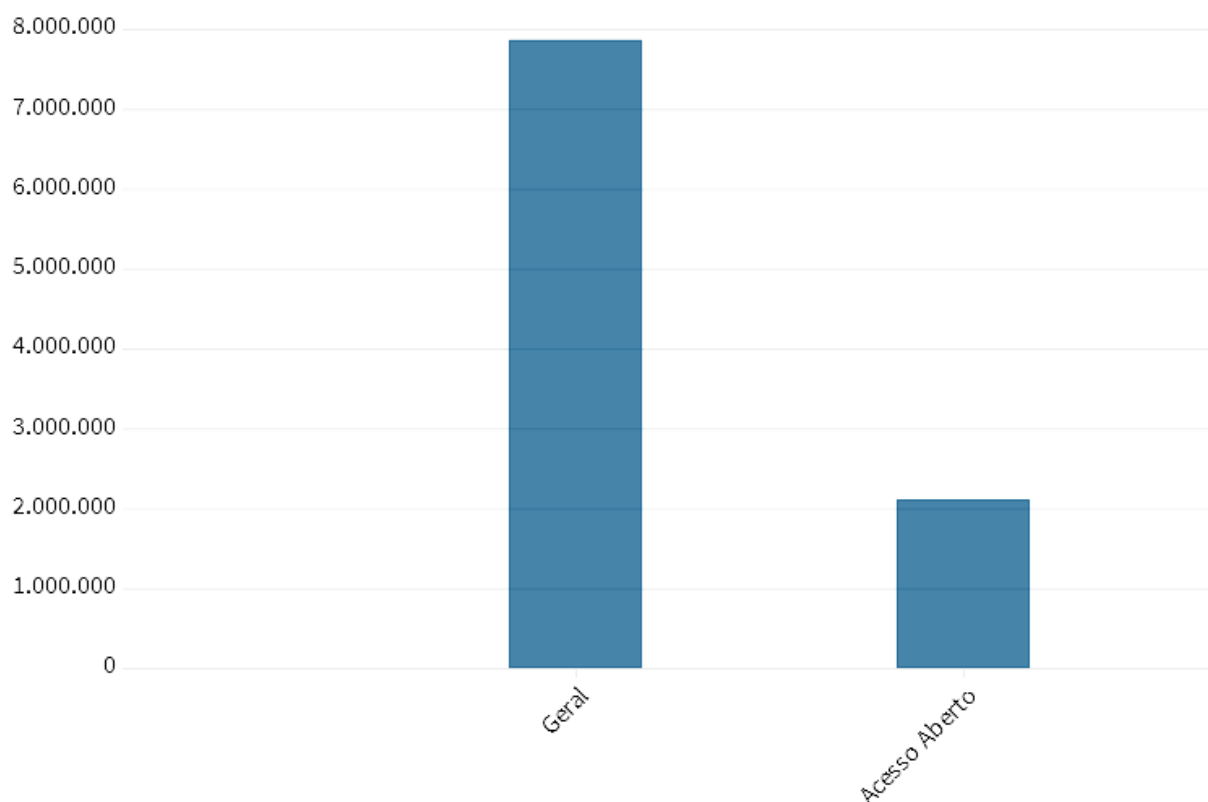


Figura 4.4 - Quantitativo das publicações em periódicos registradas nos currículos.

Como pôde ser observado, do conjunto total de artigos publicados em periódicos, considerando todo o histórico de publicações registrado em todos os currículos cadastrados na Plataforma Lattes (7.841.860), um percentual de 26,65% (2.090.015) foi publicado em periódicos de acesso aberto, levando em conta a relação dos periódicos recuperados do DOAJ. Esse percentual de publicações em periódicos de acesso aberto é relevante, acima de tudo, por considerar todo o histórico de publicações de cada pesquisador. Percebe-se que as publicações em periódicos de acesso aberto vêm recebendo atenção e adesão por parte dos pesquisadores, apresentando-se como uma tendência na disseminação e na comunicação científica, sobretudo nos últimos anos.

4.3 Visão Geral dos Dados

Além dos artigos publicados em periódicos, o que constitui o principal objeto de estudo deste trabalho, os currículos analisados também possuem diversos outros dados que podem ser utilizados em métricas para melhor compreender o perfil dos autores selecionados. É importante destacar a diversidade dos dados que estão registrados nesse conjunto de currículos, como os relativos a artigos em anais de congressos, artigos em periódicos, livros e trabalhos técnicos.

Tendo em vista o formato do repositório utilizado, principalmente com relação as análises quantitativas da produção científicas, ressalta-se que um mesmo trabalho pode estar inserido

em currículos distintos, já que pode ter sido realizado em colaboração. Consequentemente, nos currículos cadastrados na Plataforma Lattes, um trabalho pode aparecer várias vezes, pois pode ser registrado por cada um de seus autores.

Na Tabela 4.1 são apresentados os quantitativos dos dados de publicações de todos os indivíduos com currículos cadastrados na Plataforma Lattes (na coluna “Todo Repositório”), bem como do conjunto de autores que já publicaram algum artigo em periódico de acesso aberto (na coluna “Conjunto Selecionado”). A coluna “Global” apresenta o somatório das publicações encontradas nos currículos, enquanto na coluna “Único” estão os dados sem duplicatas, ou seja, um mesmo trabalho encontrado em mais de um currículo foi contabilizado uma única vez nessa coluna. Logo, pela redução na quantidade apresentada na coluna “Único”, é possível observar que alguns pesquisadores têm trabalhado de forma colaborativa mais intensamente.

Tabela 4.1 - Quantitativo dos dados extraídos.

	Todo Repositório		Conjunto Selecionado	
	Global	Únicos	Global	Únicos
Artigos em anais de congresso	17.081.698	8.942.688	11.224.397	6.566.691
Artigos em periódicos	7.841.860	3.289.080	6.125.160	2.690.208
Capítulos de Livros	1.916.816	929.834	1.329.067	661.322
Livros	773.901	393.928	429.209	228.491
Textos em jornais e revistas	1.912.254	518.611	889.089	234.227
Trabalhos Técnicos	2.570.462	744.458	1.611.903	379.564

No contexto desta pesquisa, considerando a produção científica, serão analisados somente os artigos publicados em periódicos de acesso aberto. Logo, o principal elemento de análise é o conjunto de artigos publicados em periódicos. Esse conjunto é composto por quase oito milhões de artigos (7.841.860) e, considerando somente os artigos únicos, representa 41,94% do conjunto global.

Utilizando a extensão proposta neste trabalho para o *LattesDataXplorer*, foram identificados todos os autores que publicaram pelo menos um artigo em periódico de Acesso Aberto (370.388). Esses autores — apesar de serem uma pequena quantidade de indivíduos, em relação a todo o conjunto cadastrado na Plataforma Lattes (5,65%) — possuem uma grande representatividade no contexto das publicações em periódicos. Aproximadamente 78% dos artigos globais e 81% dos artigos únicos publicados em periódicos foram realizados pelo conjunto de indivíduos que possuem publicações em periódicos de acesso aberto (Figura 4.5).

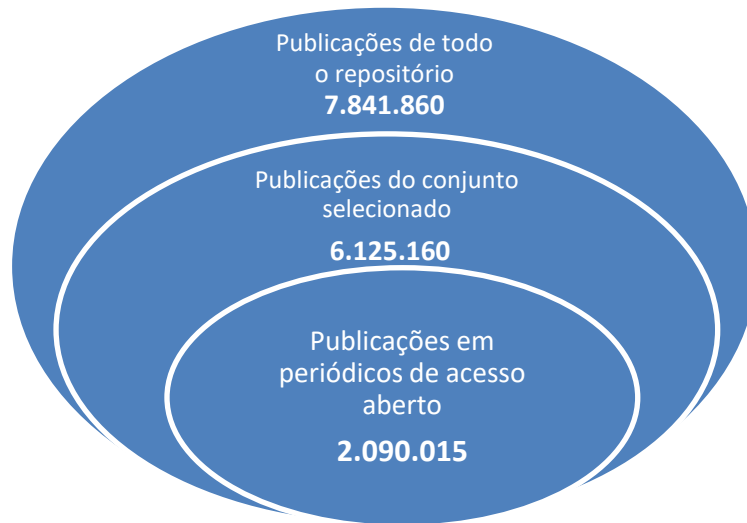


Figura 4.5 - Quantitativo das publicações em periódicos registradas.

Diante disso, é possível notar a representatividade do conjunto a ser analisado nesta pesquisa. Tendo em vista que ele engloba uma parcela considerável dos autores que têm publicado artigos em periódicos no Brasil, os resultados apresentados podem proporcionar uma visão inédita sobre a evolução dos artigos em acesso aberto, bem como servir como base para diversos outros trabalhos.

Levando em conta o que foi exposto até aqui, no próximo capítulo serão apresentadas diversas análises, para inicialmente descrever os indivíduos que compõem o conjunto, suas produções de artigos em acesso aberto e suas colaborações.

5. CARACTERIZAÇÃO DOS DADOS

Este capítulo apresenta uma caracterização geral dos dados analisados. Inicialmente é apresentada uma visão do perfil dos pesquisadores que têm divulgado suas pesquisas em periódicos de acesso aberto (Seção 5.1). Em seguida (Seção 5.2), é analisada a produção científica em periódicos de acesso aberto no Brasil, com ênfase na evolução temporal e na publicação por áreas do conhecimento. Por fim, é analisada a colaboração científica do conjunto selecionado (Seção 5.3).

5.1 Caracterização dos autores

Realizar uma caracterização dos autores analisados nesta tese possibilita compreender quem são os pesquisadores que têm publicado artigos em periódicos de acesso aberto, possibilitando, dessa forma, obter um panorama da comunidade científica que tem publicado artigos nesse meio de comunicação. O conjunto de artigos publicados em periódicos de acesso aberto analisados neste trabalho tem como autores um total de 370.388 indivíduos, o que representa aproximadamente 6% do conjunto total de indivíduos com currículos cadastrados na Plataforma Lattes, de acordo com coleta em 01/2020. Ressalta-se que esse percentual de autores é bem inferior ao quantitativo de artigos em periódicos de acesso aberto, que representam aproximadamente 27% do total geral de publicações de artigos em periódicos de todos os indivíduos. Esse percentual de publicação está muito próximo do apresentado por Furnival e Silva-Jerez (2017), que destacam que apenas aproximadamente 30% do total de artigos científicos publicados no mundo anualmente está disponível em canais de acesso aberto. Logo, identifica-se aqui que o percentual de publicações em periódicos de acesso aberto no Brasil está ligeiramente menor do que a média mundial de publicações nesse formato⁶.

Uma informação importante para compreender o perfil dos autores que já publicaram em periódicos de acesso aberto diz respeito a seu nível de formação acadêmica. Tal informação é importante pois possibilita compreender o perfil acadêmico dos autores que têm publicado em periódicos desse formato no país. Em um currículo cadastrado na Plataforma Lattes é possível registrar todo o histórico de formação acadêmica concluída e em andamento de um indivíduo, do ensino fundamental até o doutorado. Tendo isso em vista, na Figura 5.1 é apresentada a distribuição do conjunto de autores de acordo com seu maior nível de formação acadêmica concluído.

⁶ O conjunto de dados utilizado para as análises dos autores selecionados está disponibilizado como um dataset público em: <https://doi.org/10.5281/zenodo.5176811>.

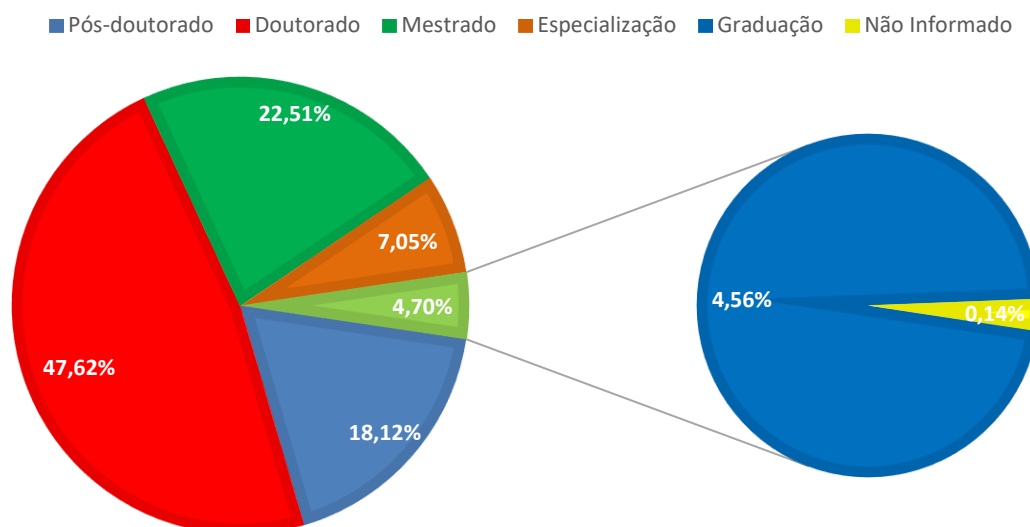


Figura 5.1 - Distribuição dos autores por níveis de formação.

Como pode ser observado, apesar de o pós-doutorado não ser um nível de formação, os autores que o fizeram foram separados do conjunto de doutores para fins de estatística e melhor visualização. Percebe-se que a maioria dos autores possuem doutorado concluído (47,62%). Em segundo lugar, estão os autores com mestrado (22,51%), seguidos dos que realizaram o pós-doutorado (18,12%), de modo que somente esses três grupos totalizam 88,25% do conjunto analisado. Além disso, observa-se que apenas 0,14% dos autores não informaram níveis de formação concluídos em seus currículos, e que 4,56% possuem apenas graduação concluída. A ausência e baixa representatividade de autores nos níveis mais inferiores de formação acadêmica pode ser justificada pelo fato de que as publicações científicas, principalmente aquelas realizadas em periódicos, são, em geral, resultado de pesquisas realizadas na pós-graduação.

Diante disso, fica evidente que a maioria do conjunto de autores identificados por já terem publicado trabalhos em periódicos de acesso aberto é composta por indivíduos com alto nível de formação acadêmica. Isso justifica sua significativa representatividade no conjunto total de publicações de artigos em periódicos, mesmo que seja um pequeno percentual do conjunto total de indivíduos com currículos cadastrados na Plataforma Lattes.

Ao analisar os autores das publicações realizadas em periódicos de acesso aberto pelas suas grandes áreas de atuação, é possível verificar sua distribuição por essas áreas, bem como analisar quais delas têm maior representatividade, considerando a quantidade de indivíduos que já publicaram pelo menos um trabalho nesse meio de publicação (Figura 5.2).

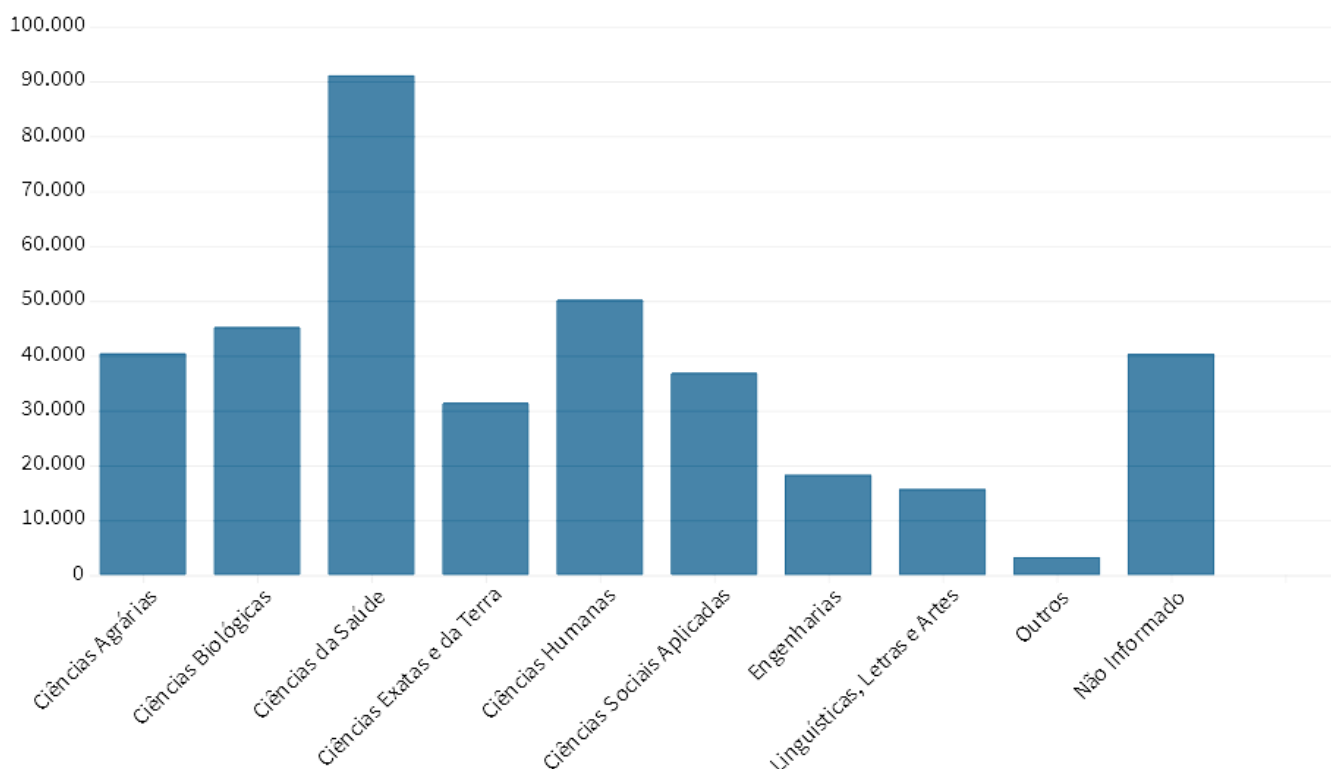


Figura 5.2 - Distribuição dos autores pelas suas grandes áreas de atuação.

Ressalta-se que o primeiro registro identificado em cada um dos currículos foi utilizado para determinar a principal grande área de atuação dos autores, tendo em vista que nos currículos da Plataforma Lattes é possível realizar o registro de até três grandes áreas de atuação, segundo a distribuição de áreas do CNPq. Como pode ser observado, a grande área de Ciências da Saúde se destaca com a maior quantidade de indivíduos (24,53%), seguida pelas grandes áreas de Ciências Humanas (13,49%), Ciências Biológicas (12,14%), Ciências Agrárias (10,85%) e Ciências Sociais Aplicadas (9,89%). Já a menor quantidade de autores pertence à grande área de Linguística, Letras e Artes (4,17%), com quantidade próxima à de Engenharias (4,87%).

Destaca-se, ainda, uma significativa quantidade de indivíduos (10,81%) que não informaram grande área de atuação em seus currículos, o que foi categorizado como “Não Informado”. Na análise desse conjunto de indivíduos que não informaram grande área, nota-se que a maioria é composta por autores que ainda estão em processo de formação e que possivelmente ainda não definiram suas áreas de atuação. Além disso, também foi possível identificar um pequeno conjunto de indivíduos (1,03%) que informaram a grande área “Outros”.

Além da grande área de atuação, também é possível informar nos currículos as áreas, subáreas e especialidades em que um determinado indivíduo atua. Ao analisar as áreas de atuação do conjunto de indivíduos, é possível notar grande diversidade na distribuição destes em cada grande área, bem como uma distribuição irregular na quantidade de áreas que cada grande área possui. Diante disso, é importante realizar uma análise baseada nas áreas de atuação dos indivíduos (Figura 5.3).

Como pode ser observado, e considerando a distribuição de áreas do CNPq, não existe uma distribuição uniforme da quantidade de áreas em cada grande área. No recorte de Linguística, Letras e Artes há apenas três áreas, enquanto as grandes áreas de Ciências Biológicas e Engenharias possuem 14 áreas cada uma. No momento do registro das grandes áreas de atuação, o indivíduo pode não informar o campo “área”. Nesses casos, os indivíduos também foram categorizados como “Não Informado”. Devido à pequena quantidade de indivíduos (1,03%) que informaram “Outros” como grande área, a análise de suas áreas não foi considerada.

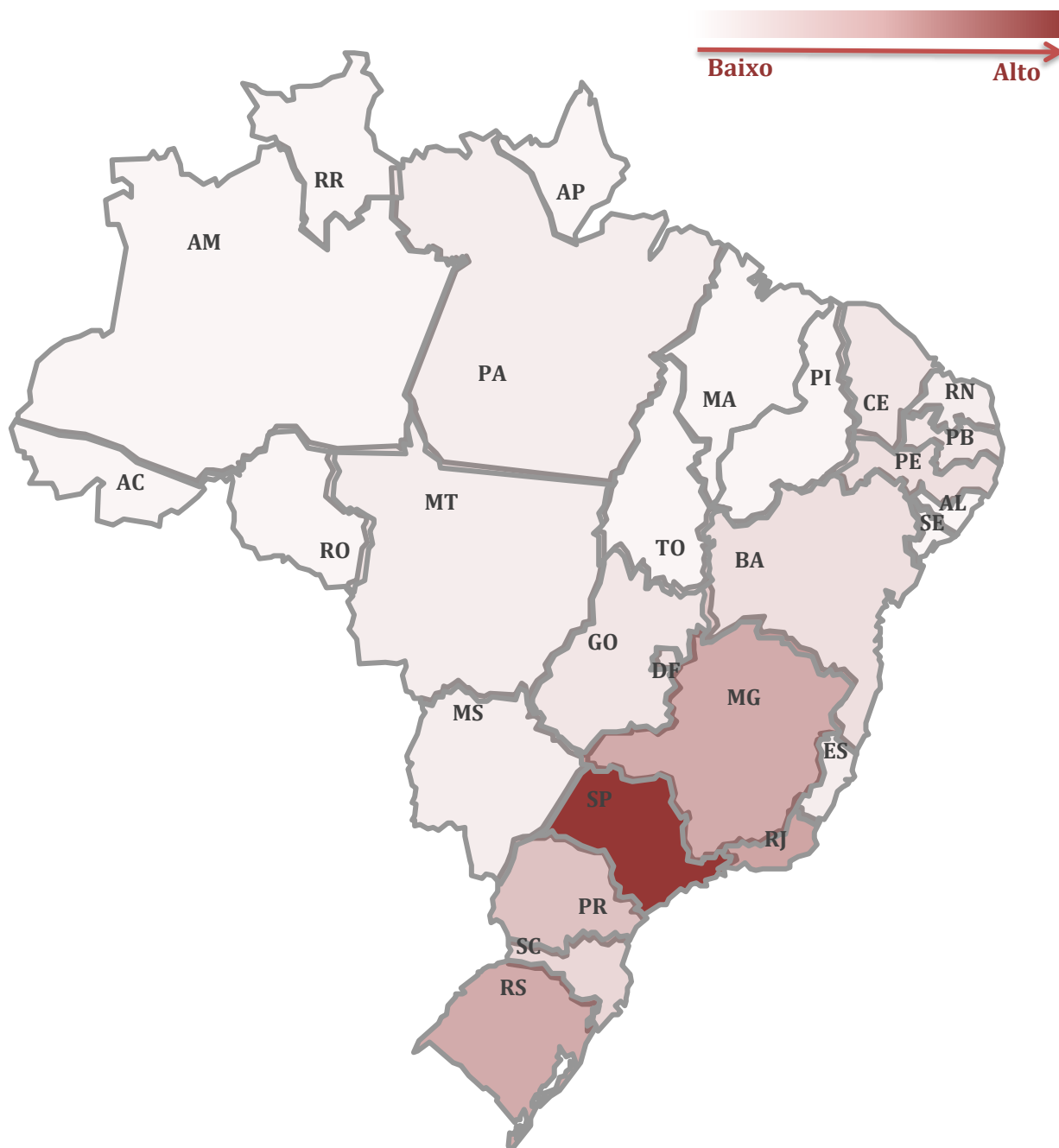
Com o intuito de verificar as áreas do conhecimento mais representativas do conjunto analisado, destaca-se Medicina (32.966), compondo a grande área de Ciências da Saúde, como a mais representativa, conforme já mostrado. A representatividade da área de Medicina é tão considerável que somente ela possui praticamente a mesma quantidade de indivíduos que a soma total das grandes áreas de Linguística, Letras e Artes e Engenharias juntas. Na sequência, destacam-se as áreas de Educação (16.840), Agronomia (15.719), Enfermagem (12.710), Administração (11.717), Química (10.501) e Psicologia (10.483). Somente essas sete áreas são responsáveis por abrigarem aproximadamente 30% do conjunto total de indivíduos analisados.

Destaca-se, ainda, a grande representatividade de Letras, que comporta 51,91% dos indivíduos de sua grande área, mesmo que essa seja a grande área com a menor quantidade de indivíduos. Vale, ainda, ressaltar a distribuição uniforme nas grandes áreas de Ciências Biológicas e Engenharias, cujas áreas mais representativas comportam apenas 12,75% (Ecologia) e 16,24% (Engenharia Civil) dos pesquisadores, respectivamente.

Considerando as áreas de atuação com a menor quantidade de indivíduos, destacam-se as áreas de Engenharia Naval e Oceânica, com apenas 58 indivíduos, Economia Doméstica, com 62, Museologia, com 146, e Engenharia de Energia, com 153, que se caracterizam como as menos representativas no conjunto analisado.

Por outro lado, evidencia-se que algumas grandes áreas do conhecimento se destacam por possuírem áreas em que uma grande quantidade de indivíduos está atuando, como é o caso da grande área de Ciências da Saúde, inflacionada por Medicina, Enfermagem e Odontologia, bem como da grande área de Ciências Agrárias, inflacionada principalmente por Agronomia, Medicina Veterinária e Zootecnia.

Considerando, ainda, informações sobre os autores dos artigos em periódicos de acesso aberto com currículos cadastrados na Plataforma Lattes, é possível verificar como esses autores estão distribuídos geograficamente no país. Na Figura 5.4 é apresentada a distribuição geográfica dos autores, utilizando como fonte de dados os seus endereços profissionais.



UF	AC	AL	AM	AP	BA	CE	DF	ES	GO
Quantidade	559	2.006	2.672	607	9.129	7.303	7.378	3.605	5.731
UF	MA	MG	MS	MT	PA	PB	PE	PI	PR
Quantidade	2.463	25.225	3.342	3.345	4.821	5.406	7.857	2.376	17.572
UF	RJ	RN	RO	RR	RS	SC	SE	SP	TO
Quantidade	25.522	4.532	886	522	23.356	10.580	2.324	59.389	1.406

Figura 5.4 - Distribuição geográfica dos autores.

Considerando somente os currículos que possuem endereço profissional informado, é possível observar que a maioria dos indivíduos está concentrada nas regiões Sudeste e Sul do país, com destaque para o estado de São Paulo, que contém aproximadamente 16% do conjunto de indivíduos, percentual bem superior à soma de indivíduos de diversos outros estados. Em seguida, destacam-se os estados do Rio de Janeiro, Minas Gerais e Rio Grande do Sul. O somatório do quantitativo desses quatro estados é maior do que o somatório de todos os outros estados juntos. Esse cenário, não por acaso, também coincide com índices nacionais que apresentam desigualdades semelhantes, como em níveis de escolaridade e de distribuição de renda. Além disso, uma hipótese para explicar tal distribuição pode ser a grande concentração de universidades e centros de pesquisas nesses estados. Além desses autores, foi possível identificar que 52 indivíduos do conjunto analisado declararam em seus currículos que estão atuando em instituições estrangeiras, sendo Colômbia (7 autores) e Portugal (5 autores) os países mais representativos.

Como se viu, nesta seção foi possível verificar o perfil dos autores brasileiros que têm publicado artigos em periódicos de acesso aberto. Essa análise é importante pois possibilita conhecer os níveis de formação acadêmica e as áreas de atuação desses autores, bem como a maneira como estão distribuídos geograficamente. Essa análise se justifica pois possibilita compreender de forma ampla, pela primeira vez, o retrato dos autores brasileiros que têm publicado em periódicos de acesso aberto.

5.2 Publicações em Acesso Aberto no Brasil

De acordo com Dias (2016), há uma tendência de os pesquisadores com níveis mais elevados de formação serem mais produtivos. Isso deve-se à necessidade de divulgação de suas pesquisas, que geralmente são realizadas em seu processo de formação na pós-graduação ou em colaborações originadas de orientações também realizadas na pós-graduação. Considerando que a quantidade de publicações vem aumentando ano após ano, impulsionada principalmente pela facilidade de divulgação e acesso às informações pela *web*, e pelo elevado grau de formação acadêmica do conjunto de autores analisado neste trabalho, a Figura 5.5 apresenta a evolução dos principais meios de publicações do conjunto de autores que têm publicado artigos em periódicos de acesso aberto. Para as análises apresentadas, foi compreendido o espaço temporal entre os anos de 1961, data da primeira publicação identificada, e 2019, ano definido como final, tendo em vista que a coleta foi realizada em janeiro de 2020.

Os dados apresentados correspondem a trabalhos únicos registrados, independentemente de sua inserção em mais de um currículo. Com relação aos principais meios de publicação, a quantidade de artigos publicados em periódicos teve crescimento significativo e constante a partir de 2001, apresentando queda a partir de 2017, diferentemente do apresentado em Dias (2016), que afirma que a partir de 2015 a produção de artigos em periódicos no Brasil, considerando todos os indivíduos com currículos cadastrados na Plataforma Lattes, sofreu redução. Logo, percebe-se que o conjunto analisado neste projeto, diferentemente do conjunto

total de autores com currículos cadastrados na Plataforma Lattes, tem publicado cada vez mais artigos em periódicos.

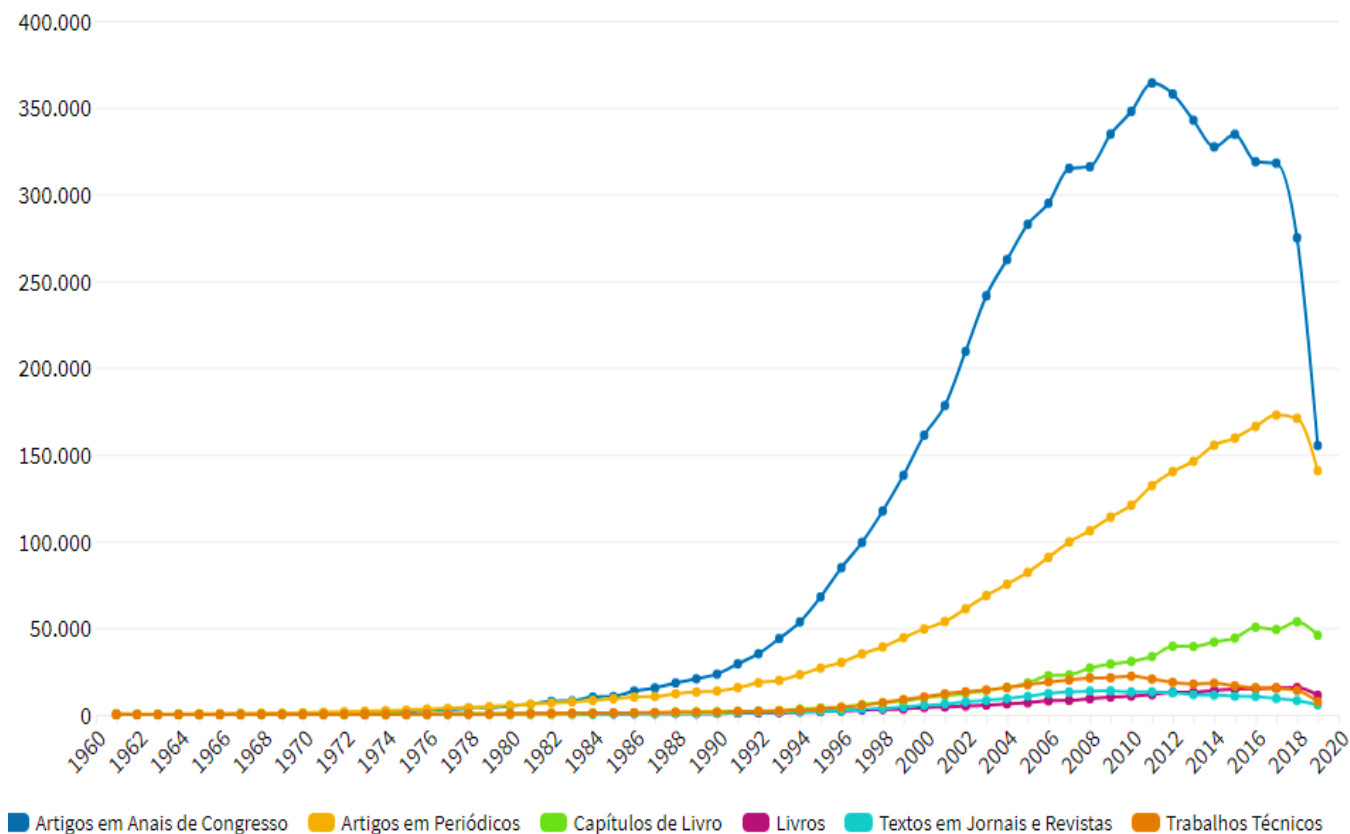


Figura 5.5 - Total de publicações por ano do conjunto selecionado

Situação diferente ocorre com a publicação de artigos em anais de congresso, principal meio de publicação de artigos no país. Essa via teve crescimento constante principalmente do início dos anos 1990 até 2011, quando atingiu seu ápice de publicações (363.792). Após esse período, houve queda muito representativa, principalmente após o ano de 2015. Vários trabalhos têm identificado essa queda na produção de artigos em anais de congressos (DIAS, 2016), (GOMES *et al.*).

Uma hipótese de explicação para a queda na quantidade de publicações, principalmente nos últimos anos, pode ser a falta de atualização dos currículos, já que trabalhos publicados recentemente podem não ter sido registrados pelos autores. No entanto, como a coleta dos dados foi realizada em janeiro de 2020 e somente foram consideradas as publicações até 2019, espera-se que esse efeito seja o menor possível.

Diante do exposto, é evidente a importância de análises sobre os artigos publicados em periódicos, tendo em vista serem um dos principais meios de divulgação de resultados de pesquisas. Além disso, os artigos publicados em periódicos também são frequentemente utilizados como indicadores para a avaliação de pesquisadores, instituições de pesquisa e até mesmo programas de pós-graduação. Logo, no contexto desta pesquisa, que visa analisar as

publicações em periódicos de acesso aberto no Brasil, realizar uma análise temporal da evolução desse meio de divulgação se faz necessário (Figura 5.6).

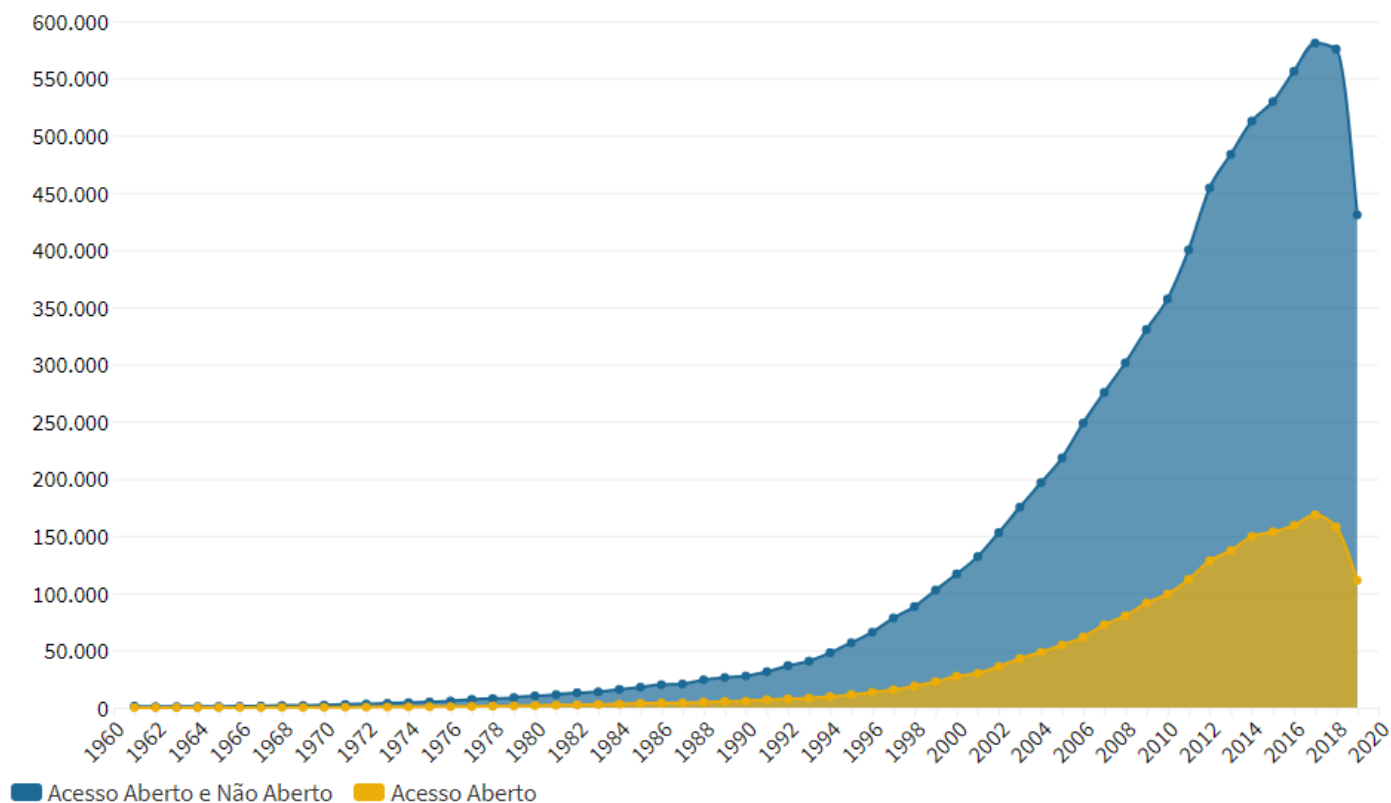


Figura 5.6 - Evolução temporal da produção de artigos em periódicos no Brasil.

O crescimento mais acentuado na publicação de artigos em periódicos de acesso aberto se deu a partir de 2000, quando esse tipo de publicação representava aproximadamente 24% do conjunto total de artigos publicados em periódicos. Após esse período, com o crescimento na quantidade de publicações ao longo dos anos, acentuado em 2011, a representatividade dos artigos em periódicos de acesso aberto se tornou mais significativa, sendo 29,02% do conjunto total no ano de 2017. Com a queda acentuada na quantidade geral de artigos no último ano analisado, os artigos em periódicos de acesso aberto passaram a ter 25,83% do conjunto total. Isso corrobora a tendência e o crescente interesse da comunidade científica em publicar os resultados de suas pesquisas de forma aberta.

No intuito de melhor compreender o perfil das publicações em periódicos de acesso aberto realizadas pela comunidade científica brasileira, um importante elemento de análise é referente ao idioma em que as publicações analisadas foram publicadas (Figura 5.7).

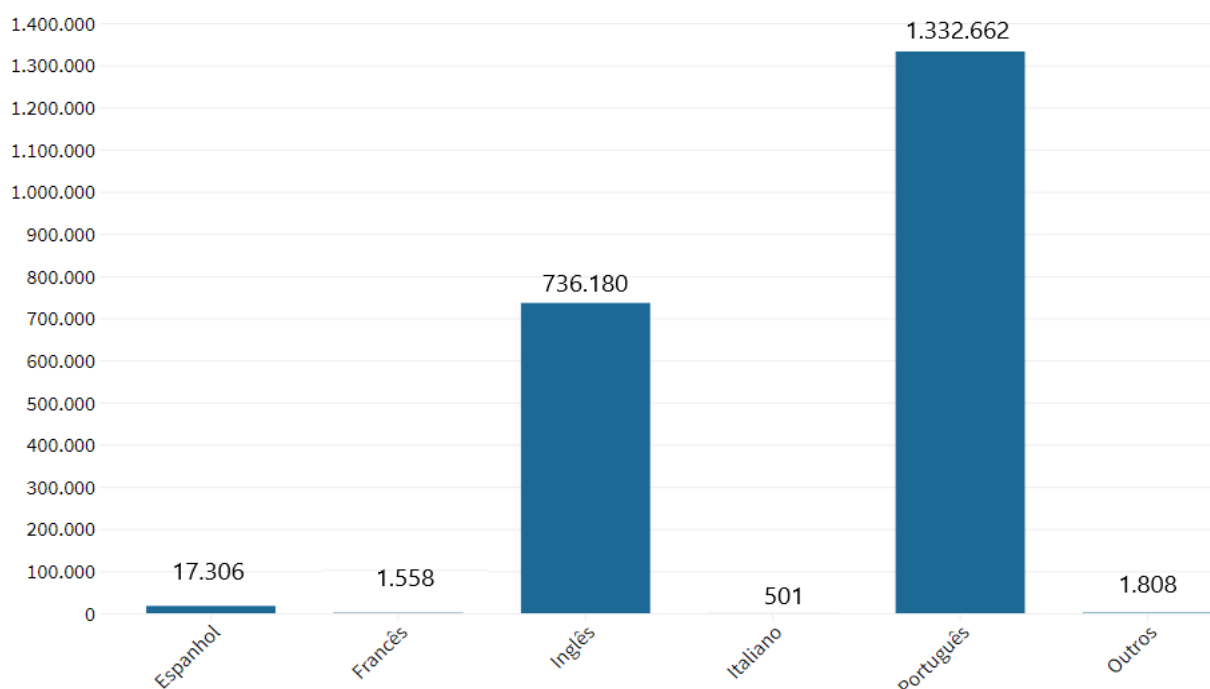


Figura 5.7 - Idioma de publicações dos artigos analisados.

A informação referente ao idioma de uma determinada publicação é inserida pelo indivíduo no momento de registro do artigo, selecionando um idioma dentre uma lista previamente cadastrada. Logo, erros ortográficos no registro desta informação não existem. Porém, por não ser um registro obrigatório, não são raros cadastros em que o autor não informa esse campo, deixando o artigo sem a informação. No entanto, do conjunto total de artigos analisado nesta pesquisa, somente 85 artigos não possuem informação sobre o idioma da publicação.

Consequentemente, como era de se esperar, o idioma de maior representatividade é o Português (63,76%) seguido em menor quantidade do Inglês (35,22%). Em geral, as publicações em Português são de artigos publicados em periódicos nacionais, sendo a maioria, e justificando dessa forma a sua representatividade. Já os artigos escritos em Inglês em sua maioria são de publicações em periódicos internacionais, sendo este o idioma aceito pela maioria dos periódicos do exterior.

Com uma quantidade inferior, surgem os idiomas Espanhol, Francês e Italiano. Ressalta-se, ainda, um total de 1.808 publicações que possuem outros idiomas informados, como Alemão, Chinês e Japonês. Tais publicações sofrem, em parte, influência de colaborações de pesquisadores fluentes nesses idiomas. Logo, devido à predominância de publicações em Português, se faz importante destacar as limitações que tais publicações podem sofrer, em relação a serem vistas pela comunidade científica internacional.

Ao considerar temporalmente as publicações realizadas em periódicos de acesso aberto, é possível verificar que autores de determinadas áreas do conhecimento tendem a ser mais produtivos, tornando tais áreas muito representativas em relação a outras. A evolução temporal das publicações, considerando as grandes áreas de atuação de seus autores, pode ser visualizada na Figura 5.8. Para essa análise são consideradas publicações globais, já que o mesmo trabalho pode ter sido realizado em colaboração com autores de áreas distintas.

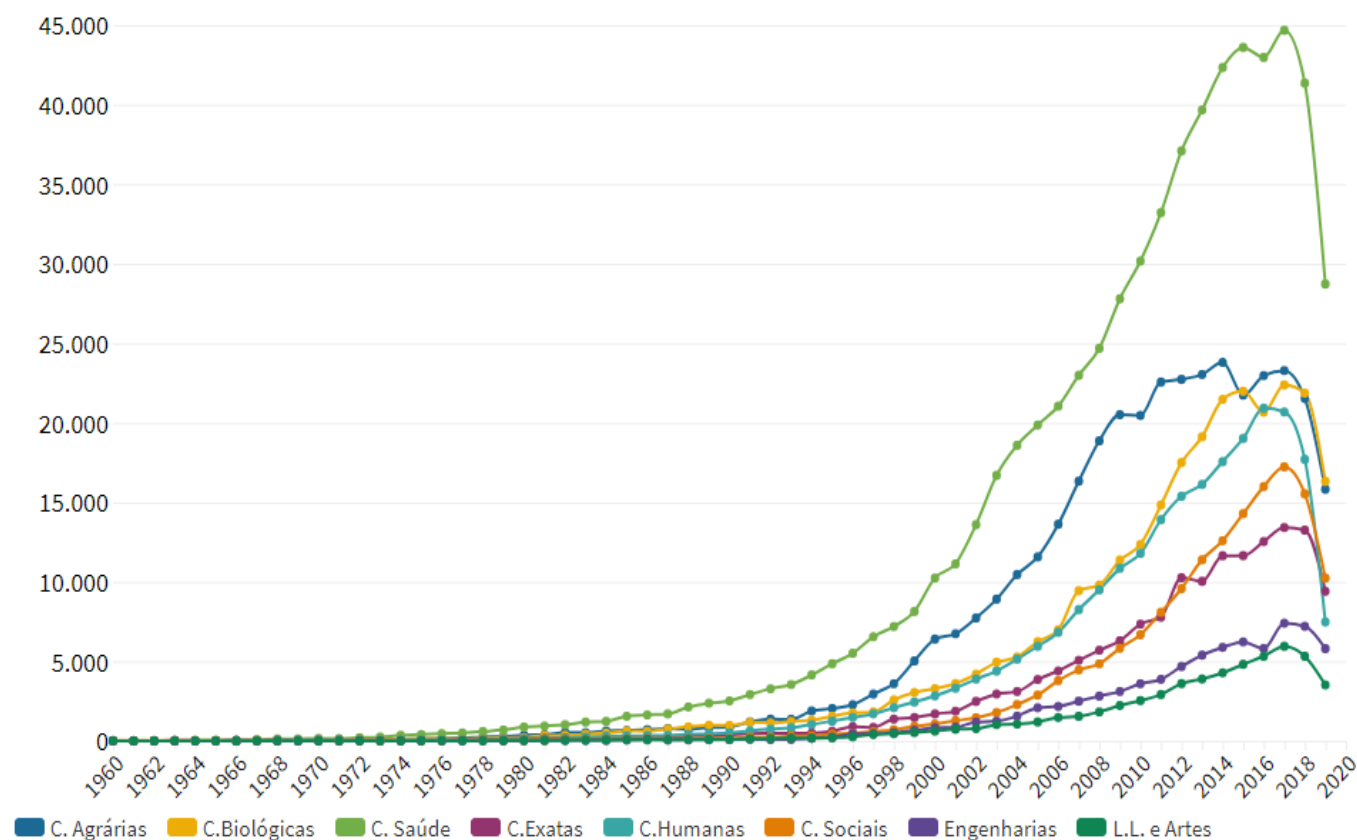


Figura 5.8 - Produção de artigos em periódicos de acesso aberto por grandes áreas.

Observa-se que a grande área de Ciências da Saúde se caracteriza como a mais representativa, tendo em vista o crescimento constante na quantidade de publicações, principalmente a partir de 1999. De forma semelhante, mas com menor intensidade, destacam-se também as grandes áreas de Ciências Agrárias e Ciências Biológicas, com produção elevada de artigos. No entanto, destaca-se que, apesar de não ter sido identificada queda na quantidade de publicações de artigos em periódicos no conjunto global de autores analisado, nessas duas últimas grandes áreas houve redução a partir de 2013 e 2014, respectivamente.

Considerando até o ano de 2019, último ano avaliado na análise, somente a grande área de Ciências da Saúde é detentora de aproximadamente 31% dos artigos publicados em periódicos de acesso aberto. De forma mais modesta, com as menores quantidades de publicações, surgem as grandes áreas de Linguística, Letras e Artes e Engenharias, com quantidade de publicações bem inferior à das outras áreas.

A análise das grandes áreas mais representativas pode ser influenciada pela quantidade de autores que cada uma delas possui, já que, conforme apresentado anteriormente, algumas áreas possuem uma quantidade muito grande de autores, diferentemente de outras, com quantidades muito pouco significativas. Por isso, na Tabela 5.1 é apresentada uma análise *per capita* de cada grande área de atuação. Como produção *per capita* entende-se a razão da quantidade de publicações pelo total de autores em cada área de atuação.

Tabela 5.1 - Produção *per capita* por área de atuação.

Ciências Agrárias			
Área	Autores	Artigos	Per Capita
Agronomia	15.719	147.622	9,39
Ciência e Tecnologia de Alimentos	4.187	28.544	6,82
Engenharia Agrícola	1.690	18.018	10,66
Medicina Veterinária	8.819	75.386	8,55
Recursos Florestal e Engenharia Florestal	3.185	26.075	8,19
Recursos Pesqueiros e Engenharia de Pesca	992	7.526	7,59
Zootecnia	5.490	65.662	11,96
Não Informado	92	918	9,98
	40.174	369.751	9,20

Ciências Biológicas			
Área	Autores	Artigos	Per Capita
Biofísica	582	3.325	5,71
Biologia geral	3.618	12.746	3,52
Bioquímica	5.443	29.979	5,51
Biotecnologia	447	2.090	4,68
Botânica	3.760	25.201	6,70
Ecologia	5.754	30.244	5,26
Farmacologia	1.867	12.646	6,77
Fisiologia	2.565	17.535	6,84
Genética	5.230	36.434	6,97
Imunologia	2.421	17.449	7,21
Microbiologia	4.549	30.254	6,65
Morfologia	2.183	13.452	6,16
Parasitologia	2.056	19.531	9,50
Zoologia	4.275	25.877	6,05
Não Informado	228	933	4,09
	44.978	277.696	6,17

Ciências da Saúde			
Área	Autores	Artigos	Per Capita
Educação Física	6.472	44.057	6,81
Enfermagem	12.710	84.448	6,64
Farmácia	6.295	30.069	4,78
Fisioterapia e Terapia Ocupacional	7.558	32.233	4,26
Fonoaudiologia	2.268	13.871	6,12
Medicina	32.966	268.877	8,16
Nutrição	4.701	22.422	4,77
Odontologia	10.003	68.440	6,84
Saúde Coletiva	7.227	71.772	9,93
Não Informado	289	2.366	8,19
	90.489	638.555	7,06

Ciências Exatas e da Terra			
Área	Autores	Artigos	Per Capita
Astronomia	206	722	3,50
Ciência da Computação	4.905	14.008	2,86
Física	4.975	28.697	5,77
Geociências	5.365	30.733	5,73
Matemática	2.781	8.102	2,91
Oceanografia	897	4.766	5,31
Probabilidade e Estatística	1.483	14.338	9,67
Química	10.501	53.684	5,11
Não Informado	53	189	3,57
	31.166	155.239	4,98

Ciência Humanas			
Área	Autores	Artigos	Per Capita
Antropologia	2.131	11.089	5,20
Arqueologia	419	1.525	3,64
Ciência Política	2.585	11.904	4,61
Educação	16.840	84.274	5,00
Filosofia	2.902	12.165	4,19
Geografia	4.295	18.514	4,31
História	5.846	24.369	4,17
Psicologia	10.483	52.113	4,97
Sociologia	3.491	18.088	5,18
Teologia	580	3.330	5,74
Não Informado	119	524	4,40
	49.691	237.895	4,79

Ciências Sociais Aplicadas			
Área	Autores	Artigos	Per Capita
Administração	11.684	62.400	5,34
Arquitetura e Urbanismo	1.988	5.730	2,88
Ciência da Informação	1.836	9.245	5,04
Comunicação	4.395	18.398	4,19
Demografia	193	1.595	8,26
Desenho Industrial	702	2.180	3,11
Direito	8.465	24.847	2,94
Economia	3.999	20.496	5,13
Economia Doméstica	62	193	3,11
Museologia	146	562	3,85
Planejamento Urbano e Regional	623	3.003	4,82
Serviço social	1.518	4.912	3,24
Turismo	906	3.636	4,01
Não Informado	62	259	4,18
	36.579	157.456	4,30

Engenharias			
Área	Autores	Artigos	Per Capita
Eng. Aeroespacial	255	1.074	4,21
Eng. Biomédica	563	3.488	6,20
Eng. Civil	2.920	11.158	3,82
Eng. de Energia	153	588	3,84
Eng. de Materiais e Metalurgia	2.391	13.911	5,82
Eng. de Minas	206	719	3,49
Eng. de Produção	2.374	12.499	5,26
Eng. de Transportes	271	1.201	4,43
Eng. Elétrica	2.343	6.649	2,84
Eng. Mecânica	1.884	6.324	3,36
Eng. Naval e Oceânica	58	124	2,14
Eng. Nuclear	210	998	4,75
Eng. Química	2.480	10.212	4,12
Eng. Sanitária	1.849	8.988	4,86
Não Informado	33	87	2,64
	17.990	78.020	4,34

Linguística, Letras e Artes			
Área	Autores	Artigos	Per Capita
Artes	3.134	8.580	2,74
Letras	8.052	28.183	3,50
Linguística	4.204	21.039	5,00
Vazia	40	138	3,45
	15.430	57.940	3,76

O maior valor de produção *per capita* é da área de Zootecnia, a saber, 11,96. Essa área, apesar de possuir apenas 5.490 autores (1,48% do conjunto total), produziu 65.662 artigos, o que representa 3,14% do total de publicações. Esse perfil é bem distinto do das outras áreas que também compõem a grande área de Ciências Agrárias, que possuem produção *per capita* inferior à de Zootecnia, mesmo algumas áreas possuindo maior quantidade de autores. Já a segunda área com o maior valor de produção *per capita* (9,93) é a de Saúde Coletiva, que possui uma quantidade relativamente baixa de autores (7.227), mas com significativa produção de artigos em periódicos de acesso aberto (71.722).

Já as menores produções *per capita* são as das áreas de Engenharia Naval e Oceânica (2,14), Engenharia Elétrica (2,84), Ciência da Computação (2,86), Arquitetura e Urbanismo (2,88) e Matemática (2,91). Importante ressaltar que a grande área com menor produção *per capita* é a grande área de Linguística, Letras e Artes, com valor de 3,76, influenciada principalmente pela área de Artes, com produção *per capita* de apenas 2,74, bem inferior à produção das outras duas componentes da grande área.

As áreas de Ciência da Computação e Matemática estão vinculadas à grande área de Ciências Exatas e da Terra, a qual, apesar de possuir algumas áreas com valores significativos de produção *per capita*, como as de Probabilidade e Estatística (9,67) e Física (5,77), acaba tendo seu quantitativo geral influenciado pela baixa produtividade *per capita* das duas primeiras.

A área de Medicina se caracteriza como a área com a maior quantidade de autores (32.966), e também com a maior quantidade de publicações (268.877). No entanto, a sua produção *per capita* é de apenas 8,16, valor inferior ao de diversas outras áreas de atuação, inclusive ao da área de Saúde Coletiva (9,93), que também está vinculada à grande área de Ciências da Saúde. Logo, apesar de possuir um grande conjunto de autores e grande quantidade de publicações, a produção *per capita* de artigos dos autores da área de Medicina é inferior ao de algumas outras áreas.

Nesse momento se faz importante destacar as características distintas das diversas áreas do conhecimento quanto ao meio de publicação dos resultados de suas pesquisas. Algumas delas tendem a direcionar a publicação dos seus resultados em outros meios, como por exemplo em artigos publicados em anais de congressos ou até mesmo em livros ou capítulos de livros.

Diante disso, ressalta-se que a grande área de Ciências da Saúde, apesar de se caracterizar como a grande área com a maior quantidade de autores (90.489) e publicações em periódicos de acesso aberto (638.555), tem produção *per capita* igual a 7,06, valor bem inferior ao da grande área de Ciências Agrárias (9,20) e próximo ao de outras grandes áreas, como a de Ciências Biológicas (6,17).

Nesse contexto, observa-se ainda que, mesmo considerando a produção *per capita* de cada área de conhecimento, é possível identificar que determinadas áreas sofrem a influência de autores extremamente produtivos, em detrimento de uma grande maioria, que produz muito pouco. A análise dos autores mais produtivos será feita de maneira mais detalhada na Seção 6.2.

Para demonstrar o impacto da produção de alguns autores altamente produtivos, a Tabela 5.2 apresenta os 24 autores com a maior quantidade de publicações em periódicos de acesso aberto, bem como informações adicionais sobre suas áreas de atuação e produção em periódicos não abertos. Essa quantidade de autores foi determinada calculando a raiz quadrada do total de indivíduos do conjunto selecionado por duas vezes consecutivas. Ou seja, do conjunto inicial contendo 370.388 indivíduos, resultou-se inicialmente em 608 e consequentemente, em 24 autores.

Tabela 5.2 - Autores com maior quantidade de publicações em periódicos de acesso aberto.

Posição	Titulação	Instituição	Grande Área	Área	Artigos em Periódicos de Acesso Aberto	Artigos em Periódicos não Abertos	Tempo de Carreira	Bolsista de Produtividade em Pesquisa do CNPq
1	Doutorado	Universidade Federal do Paraná	Ciências da Saúde	Medicina	544 (70,3%)	230 (29,7%)	35	Não
2	Doutorado	Universidade Federal do Rio de Janeiro	Ciências da Saúde	Medicina	481 (47,4%)	534 (52,6%)	28	1A
3	Doutorado	Universidade Federal de Viçosa	Ciências Agrárias	Zootecnia	401 (62,3%)	243 (37,7%)	36	1A
4	Doutorado	Universidade Federal de Viçosa	Ciências Exatas e da Terra	Probabilidade e Estatística	399 (62,3%)	241 (37,7%)	28	1C
5	Doutorado	Universidade Federal de Viçosa	Ciências Biológicas	Genética	379 (59,8%)	255 (40,2%)	30	1A
6	Doutorado	Universidade Federal de Minas Gerais	Ciências da Saúde	Saúde Coletiva	374 (79,4%)	97 (20,6%)	19	2
7	Doutorado	Universidade Estadual Paulista	Não Informado	Não Informado	354 (45,5%)	423 (54,5%)	36	Não
8	Doutorado	Universidade Federal de São Paulo	Ciências Biológicas	Farmacologia	347 (28,0%)	892 (72,0%)	42	1A
9	Doutorado	Universidade Federal de Lavras	Ciências Agrárias	Agronomia	335 (64,9%)	181 (35,1%)	35	1A
10	Doutorado	Universidade de São Paulo	Ciências da Saúde	Medicina	332 (40,7%)	484 (59,3%)	30	1B
11	Doutorado	Universidade Estadual de Maringá	Ciências da Saúde	Enfermagem	329 (67,3%)	160 (32,7%)	22	1C
12	Doutorado	Faculdade Jesuíta de Filosofia e Teologia	Ciências Humanas	Teologia	328 (69,3%)	145 (30,7%)	46	Não
13	Doutorado	Faculdade de Medicina de São José do Rio Preto	Ciências da Saúde	Medicina	325 (72,5%)	123 (27,5%)	30	2
14	Doutorado	Universidade Federal de Santa Catarina	Ciências da Saúde	Enfermagem	321 (72,5%)	122 (27,5%)	25	1A
15	Doutorado	Universidade Federal de Viçosa	Ciências Agrárias	Agronomia	316 (31,9%)	675 (68,1%)	39	1A
16	Doutorado	Universidade de São Paulo	Ciências Agrárias	Medicina Veterinária	315 (47,5%)	348 (52,5%)	35	1A
17	Doutorado	Faculdade de Medicina do ABC	Ciências da Saúde	Fisioterapia e Terapia Ocupacional	314 (58,3%)	225 (41,7%)	17	1B
18	Doutorado	Universidade de São Paulo	Ciências da Saúde	Medicina	308 (58,9%)	215 (41,1%)	29	Não
19	Doutorado	Universidade de São Paulo	Ciências Biológicas	Zoologia	306 (77,9%)	87 (22,1%)	45	Não
20	Doutorado	Pontifícia Universidade Católica do Rio de Janeiro	Ciências Exatas e da Terra	Física	305 (57,9%)	222 (42,1%)	21	1C
21	Doutorado	Universidade Federal de Goiás	Ciências Agrárias	Zootecnia	300 (88,0%)	41 (12,0%)	37	1D
22	Doutorado	Universidade Federal do Rio de Janeiro	Ciências Exatas e da Terra	Física	299 (46,6%)	342 (53,4%)	24	1C
23	Doutorado	Universidade de São Paulo	Ciências da Saúde	Medicina	298 (44,2%)	376 (55,8%)	38	Não
24	Doutorado	Universidade Federal do Rio de Janeiro	Ciências Exatas e da Terra	Física	296 (47,6%)	326 (52,4%)	31	1B

Os dados apresentados mostram que existe realmente uma parcela de indivíduos com produtividade muito acima da média. Além disso, há especificação das grandes áreas de atuação dos 24 autores mais produtivos na publicação de artigos em periódicos de acesso aberto, de modo que é possível observar que existem, entre eles, representantes de cinco grandes áreas, além de um indivíduo que não informou área e nem grande área de atuação.

Destaca-se, nesse sentido, a grande área de Ciências da Saúde, com dez autores entre os 24 mais produtivos. Ao avaliar suas áreas, a área de Medicina sobressai, com seis autores a ela vinculados. Todos os autores mais produtivos do conjunto selecionado estão vinculados a universidades de diversos estados do país, sendo as mais representativas a Universidade de São Paulo (5) e a Universidade de Viçosa (4). Além disso, observa-se que todos eles possuem doutorado concluídos ou realizaram pós-doutorado.

Avaliando cada um dos currículos desse conjunto de indivíduos, identificou-se que, em geral, eles pertencem a pesquisadores com reconhecida relevância em suas áreas de atuação e com significativa experiência em pesquisas. Dentre os 24 indivíduos com maior produção científica em periódicos de acesso aberto identificados, 18 deles são Bolsistas de Produtividade em Pesquisa do CNPq, sendo 8 na Modalidade 1A. Isso corrobora a significatividade de sua contribuição em suas áreas de atuação e áreas de pesquisa.

O autor com a maior quantidade de publicações (544), atuante na área de Medicina, possui mais publicações do que o somatório de todos os autores de algumas áreas, como as de Engenharia Naval e Oceânica (124) e Economia Doméstica (193). Aproximadamente 70% de toda a produção de artigos em periódicos desse autor está publicada em periódicos de acesso aberto.

Dentre os dez autores mais produtivos, cinco são Pesquisadores de Produtividade em Pesquisa do CNPq na Modalidade 1A, contando, ainda, além dos artigos em periódicos, com uma significativa quantidade de orientações de mestrado e doutorado, sendo que alguns deles possuem centenas de orientações concluídas nesses níveis de formação acadêmica, o que pode contribuir de forma direta em sua produção científica.

Em um contexto geral, todos os autores mais produtivos possuem publicações tanto em periódicos de acesso aberto quanto no formato não aberto, sendo que alguns deles possuem mais de 1.000 artigos em periódicos em todo seu histórico de publicações, como os autores das posições 2 e 8 da Tabela 5.2. Destes, um autor é da área de Medicina e o outro, da Farmacologia. Além das publicações em periódicos, observou-se, ainda, para todos os autores, elevada produção científica em outros meios de publicação, principalmente de artigos publicados em anais de congresso.

Diante do exposto, é possível identificar que, apesar da grande quantidade de autores que têm publicado artigos em periódicos de acesso aberto, somente uma pequena parcela deles é altamente produtiva. A maioria deles, ao contrário, possui uma produção tímida: vários dos autores analisados, que juntos representam aproximadamente 63% do conjunto de indivíduos, possuem apenas uma (130.793), duas (64.638) ou três publicações (38.598).

Logo, destaca-se a importância de análises bibliométricas mais específicas para que se possa identificar os autores mais produtivos e que representam de forma efetiva a elite nacional dos autores adeptos de publicações em periódicos de acesso aberto. Para tanto, algumas métricas, como a da Lei de Lotka (Seção 6.2), podem fornecer subsídios.

Ainda, na análise da produção científica do conjunto de indivíduos considerado neste trabalho, é possível verificar que a colaboração entre os autores tem aumentado ao longo dos anos. Isso justifica uma análise da rede de colaboração desses autores, avaliando seus perfis de colaboração institucional do conjunto, e, ainda, as áreas do conhecimento que têm trabalhado de forma mais colaborativa, conforme apresentado na próxima seção.

5.3 Colaboração Científica

Diante do crescimento da forma colaborativa de se realizar as publicações analisadas neste trabalho, um estudo sobre a rede de colaboração científica surge como uma alternativa para melhor compreender como colaboram os indivíduos analisados. Considerando a colaboração científica dos autores de artigos em periódicos de acesso aberto, percebe-se que a partir do ano de 2000 os trabalhos passaram a ser realizados de forma mais colaborativa. Atualmente, os trabalhos publicados em periódicos de acesso aberto possuem uma média de 5,3 autores por publicação (Figura 5.9).

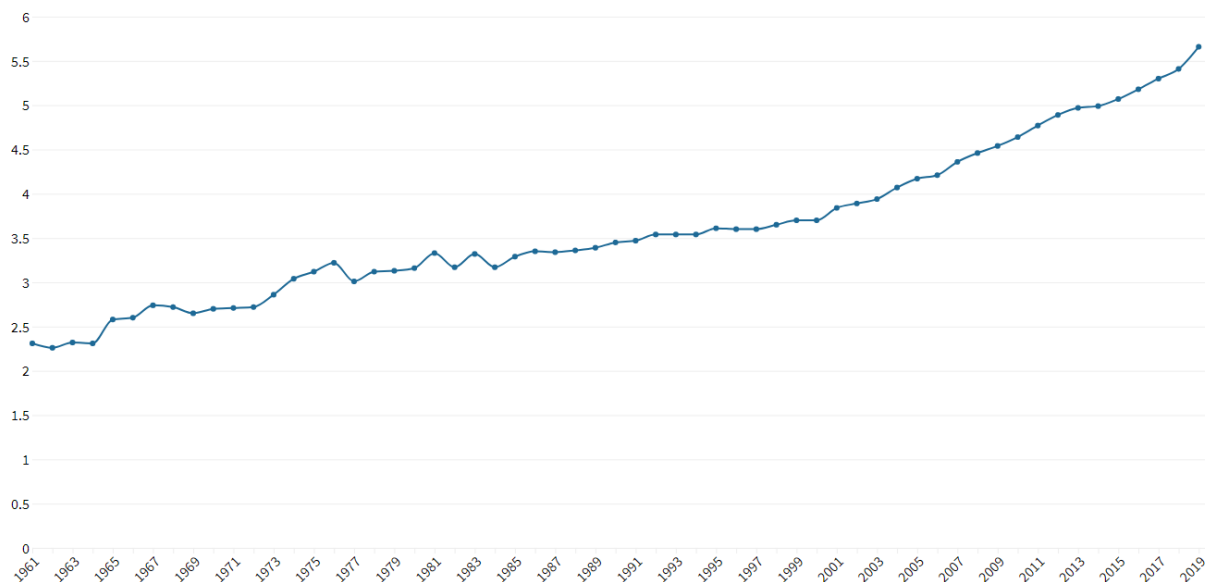


Figura 5.9 - Média de autores por artigo.

É possível observar que a média de coautores vem crescendo constantemente, desde 1961 — primeiro ano considerado nesta análise — até 2019, com exceção dos anos de 1977, 1982 e 1984, que apresentaram ligeira queda. Foi possível observar, no conjunto de publicações analisadas (2.090.015), que aproximadamente 29% delas foram elaboradas em autoria única.

Essas publicações não foram consideradas para as análises aqui apresentadas, já que não caracterizam colaborações.

Para todas as publicações em periódicos de acesso aberto, foi aplicado o método *ISCooll*, proposto por Dias e Moita (2015), para a identificação e caracterização das colaborações científicas. Com a adoção do método, todas as publicações identificadas como sendo realizadas em colaboração geravam *cliques*⁷, e eram, posteriormente, inseridas por justaposição na rede de colaboração. Com isso, foi possível gerar a rede contendo todas as colaborações realizadas entre todos os indivíduos analisados (Figura 5.10).

Para fins de visualização, foram aplicados métodos que visam concentrar os nós mais conectados da rede e expulsar para as extremidades os nós com poucas conexões ou que não possuem nenhuma conexão (nós isolados). Além disso, as cores dos nós foram aplicadas considerando suas grandes áreas de atuação. Logo, ao observar a rede de colaboração, é possível identificar como são gerados grupos mais conectados, cujas ligações tornam-se mais intensas, principalmente por influência das grandes áreas de atuação dos indivíduos, mas também por sua atuação nas mesmas instituições ou em regiões geográficas próximas. Do total de indivíduos analisados (370.388), 38.173, ou seja, aproximadamente 1% deles, estão isolados na rede.

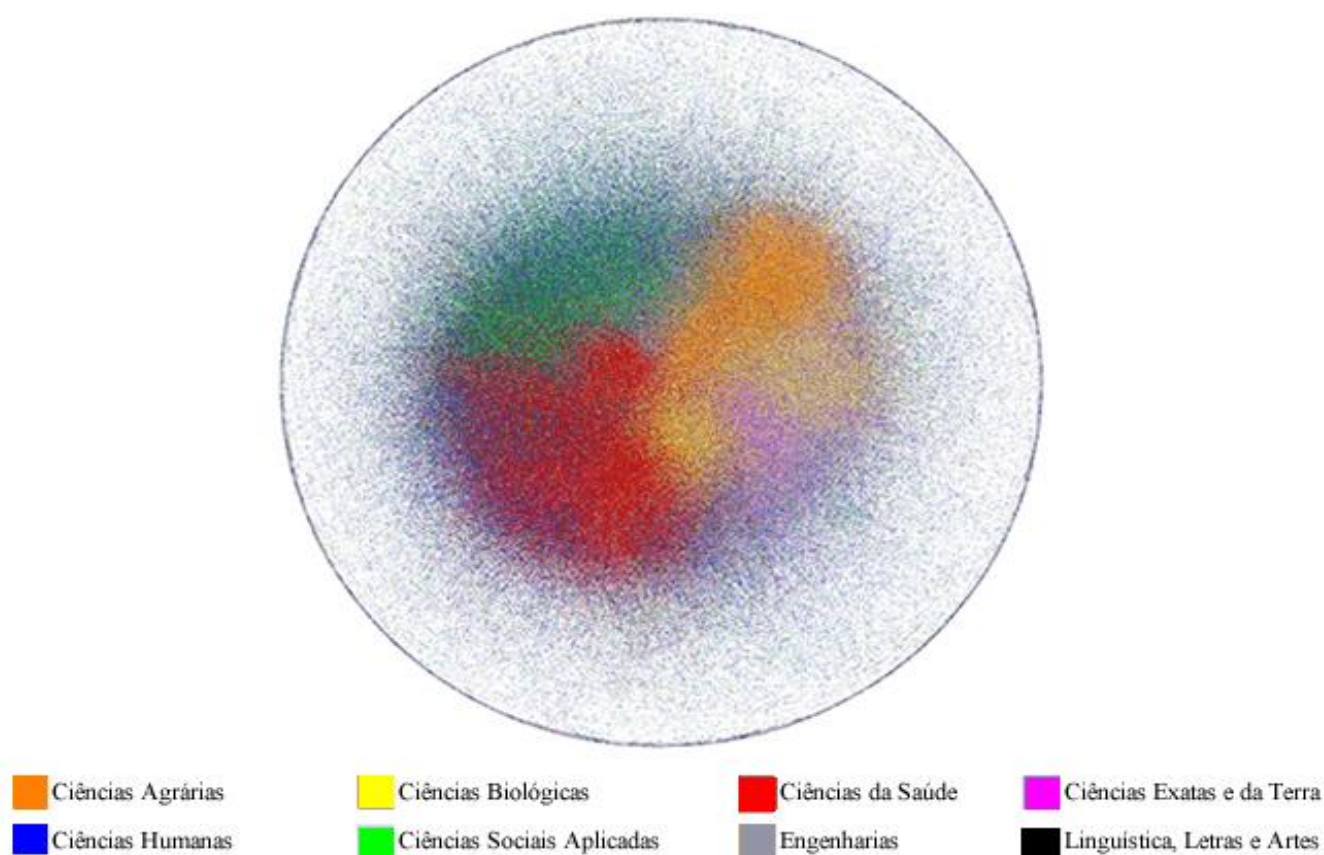


Figura 5.10 - Rede de colaboração científica dos autores.

⁷ Uma clique pode ser compreendida como um conjunto de vértices no qual todos os vértices estão diretamente vinculados uns aos outros por um conjunto de arestas.

É possível observar, então, que existe uma certa quantidade de nós isolados, que se concentram nas extremidades da rede e que se caracterizam por representarem indivíduos que publicaram de forma isolada em periódicos de acesso aberto ou com colaboradores que não estão cadastrados na Plataforma Lattes. Diante disso, algumas das métricas a serem utilizadas visam analisar a maior componente conexa da rede, ou seja, a que possui a maior quantidade de nós conectados (Figura 5.11).

Quando se visualiza somente a componente gigante da rede, responsável por agrupar 87,27% dos nós e 99,81% das conexões, fica claro que a grande maioria das colaborações está interligada na mesma componente. Destaca-se, ainda, que alguns grupos tendem a colaborar de forma mais intensa, por exemplo, os das grandes áreas de Ciências da Saúde e Ciências Agrárias, que formam um cluster bem agrupado, caracterizando-os, assim, como os grupos com o maior índice de colaborações entre as grandes áreas.

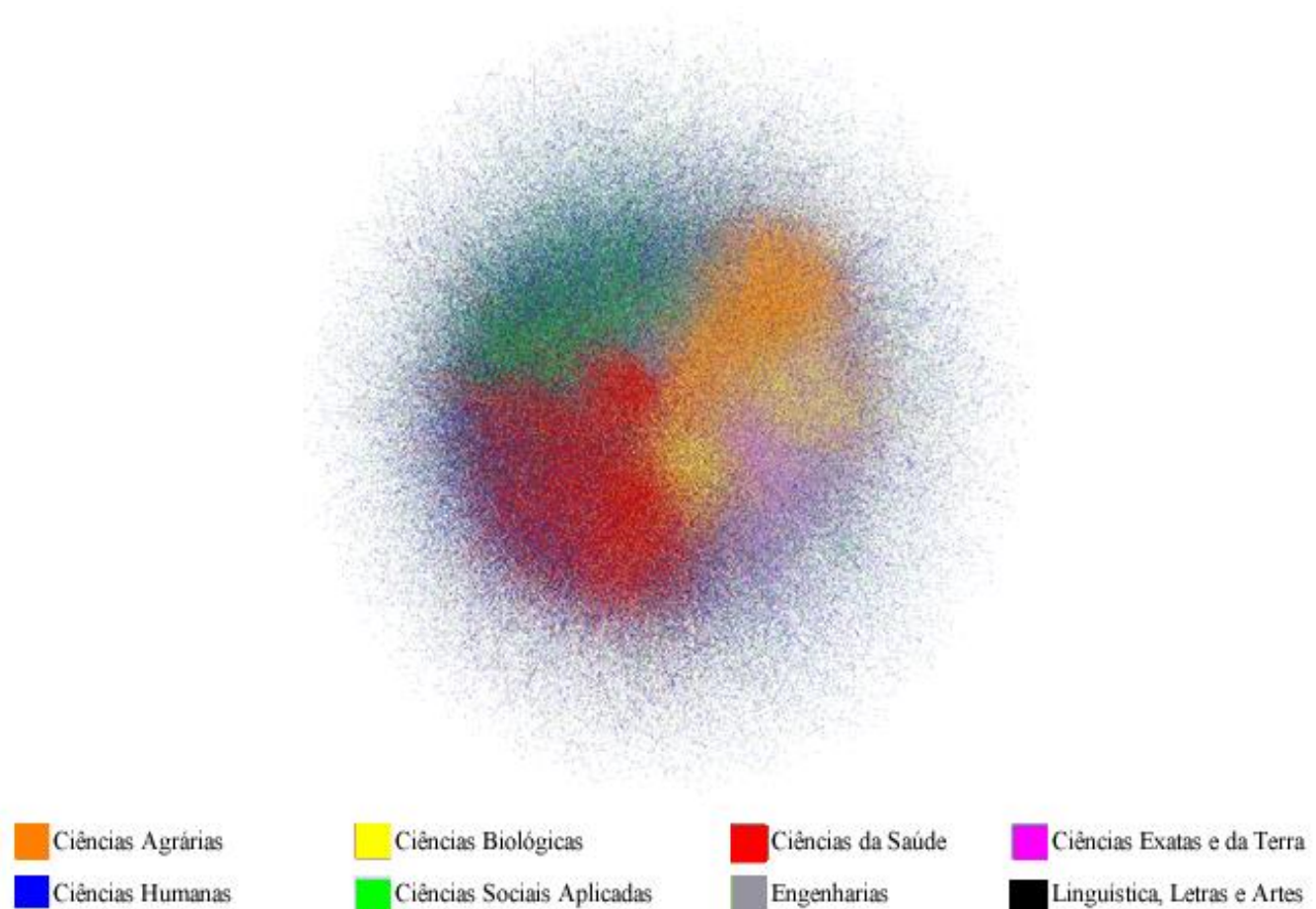


Figura 5.11 - Componente gigante.

Percebe-se, ainda, que algumas grandes áreas tendem a colaborar entre si de forma mais intensa, como é o caso das grandes áreas de Ciências Sociais Aplicadas e Ciências Humanas e das grandes áreas de Ciências da Saúde e Ciências Biológicas. Por outro lado, não é possível observar agrupamentos mais intensos de indivíduos das grandes áreas de Engenharias e

Linguística, Letras e Artes com baixa colaboração, principalmente devido a sua baixa representatividade no conjunto analisado. Para melhor compreensão da rede de colaboração caracterizada, a Tabela 5.3 apresenta uma sumarização das métricas tradicionais de análise de redes geralmente utilizadas.

Tabela 5.3 – Métricas da Rede.

Métrica	Valor
Quantidade de Nós	370.388
Quantidade de Colaborações	3.040.188
Grau Médio	12,7
Nós no Componente Gigante	323.254
% de Nós no Componente Gigante	87,27%
Colaborações na Componente Gigante	3.034.365
Densidade da Rede	0,00004432
Diâmetro da Rede	11
Caminho Mínimo Médio	4,7
Quantidade de Nós Isolados	38.173

Pelas métricas avaliadas, fica evidente que a componente conexa agrega a maioria dos nós (87,27%) e, consequentemente, a maioria das colaborações (99,81%). Além disso, percebe-se que a densidade da rede, que representa a razão/proporção entre o número existente de colaborações e o número de colaborações possíveis, é baixa, ou seja, existe uma quantidade relativamente pequena de colaborações diante do total possível de colaborações na rede.

As colaborações também podem ser medidas pelo grau médio dos nós que compõem a rede, o qual se refere ao número médio de colaborações que cada um dos nós possui. Neste caso, o grau médio da rede analisada é igual a 12,7, que indica a quantidade média de coautores que cada um dos indivíduos possui, entre os outros que estão na rede. Essa medida também sofre influência de autores com uma quantidade muito elevada de colaboradores. Foi possível identificar que entre os indivíduos do conjunto analisado existem autores com centenas de colaboradores, impulsionando o grau médio da rede, sendo esses em sua maioria, das grandes áreas de Ciências da Saúde e Ciências Biológicas.

Já o diâmetro da rede, que se refere ao tamanho da maior distância geodésica entre qualquer par de nós da componente gigante, é igual a 11. Isso significa que existe um caminho mínimo entre dois nós quaisquer da rede com pelo menos 10 outros nós entre eles. Tal fato está relacionado ao caminho mínimo médio (4,7), que corresponde a quantidade média de colaborações entre dois indivíduos em todos os caminhos mínimos existentes na rede.

Ao verificar os indivíduos com grau igual a zero, ou seja, sem colaboração na rede, foi possível observar que, em geral, seus trabalhos possuem coautores, mas estes são geralmente estrangeiros que não possuem currículos cadastrados na Plataforma Lattes e que, portanto, não foram considerados no conjunto em análise desta tese.

No contexto da análise de redes de colaboração, as métricas aqui utilizadas possibilitam compreender a estrutura e a influência de determinados nós na rede. Diversas outras métricas

podem ser aplicadas com o intuito de melhor compreender a estrutura da colaboração científica de artigos em periódicos de acesso aberto, e até mesmo como ela tem evoluído. No entanto, no contexto deste trabalho, as métricas apresentadas servem como um primeiro esforço para avaliar como o conjunto analisado tem colaborado no processo de publicação de artigos.

6. MÉTRICAS BIBLIOMÉTRICAS APLICADAS NOS ARTIGOS EM PERIÓDICOS DE ACESSO ABERTO

Este capítulo apresenta uma análise com aplicação de métricas bibliométricas nos artigos publicados em periódicos de acesso aberto. Inicialmente avalia-se a relevância dos periódicos, utilizando para isso a Lei de Bradford (Seção 6.1). Após isso, é analisada a produtividade dos autores com auxílio da Lei de Lotka (Seção 6.2). Posteriormente, um estudo sobre a frequência dos principais temas utilizados no conjunto de publicações é realizado com auxílio da Lei de Zipf (Seção 6.3). Por fim, na Seção 6.4, uma análise de atributos de qualidade com citações e Fator de Impacto (FI) é realizada.

6.1 Relevância dos Periódicos – Lei de Bradford

Adotou-se a mesma estratégia utilizada no Capítulo 4, em que cada uma das publicações foi analisada comparando se o seu ISSN ou E-ISSN estava no conjunto de periódico do DOAJ. Isso se deu no intuito de identificar os periódicos em acesso aberto registrados nos currículos da Plataforma Lattes, e dessa forma foram recuperados 9.202 periódicos. Aqui é importante ressaltar que, apesar do conjunto de periódicos utilizado como validador do DOAJ possuir 14.170 periódicos, ao analisar os 2.090.015 de artigos em periódicos de todo o conjunto selecionado, foram recuperados 9.202 (65%) periódicos que estavam no DOAJ. Logo, percebe-se que, apesar de significativa representatividade, ainda existe um conjunto de periódicos que não receberam contribuições da comunidade científica brasileira.

Para a aplicação da Lei de Bradford, foram considerados todos os periódicos em acesso aberto e, respectivamente, o total de artigos desses periódicos cadastrados em todos os currículos da Plataforma Lattes, com o objetivo de se utilizar esse quantitativo de artigos e consequentemente dividi-los em zonas.

Segundo a Lei de Bradford, inicialmente deve-se dividir o conjunto de periódicos em três zonas, sendo a primeira chamada de Núcleo, em que se agrupam os periódicos mais relevantes do conjunto, ou seja, aqueles que possuem a maior quantidade de publicações. A Tabela 6.1 apresenta as zonas identificadas do conjunto, conforme os cálculos da Lei de Bradford.

Tabela 6.1- Zonas e seus respectivos periódicos.

ZONAS	Quantidades de periódicos
Núcleo	81
Zona 1	366
Zona 2	8.755

A partir da proposta da Lei de Bradford, o Núcleo contém uma quantidade representativa de publicações, porém com poucos periódicos. A Zona 1 tem um número considerável tanto de periódicos quanto de artigos e a Zona 2 tem um número bem representativo de periódicos, mas com menor quantidade de artigos. No entanto, a lei é aplicável só se existir a proporcionalidade, $1:n:n^2:n^3$. Logo, visto que não foi possível obter a mesma constante da divisão Zona 2 pela Zona 1 e da Zona 1 pelo núcleo, foram feitos novos cálculos para se obter zonas que resultassem em uma mesma constante. Logo, foi possível identificar pelos novos cálculos um total de 15 zonas, seguindo uma divisão que atendeu ao critério da proporcionalidade.

A Tabela 6.2 mostra a divisão das três primeiras zonas identificadas dentre as 15, com seus respectivos periódicos e quantidades de artigos. Foi adotado Pareto para escolha das 3 zonas apresentadas.

Tabela 6.2 – Primeiras três zonas identificadas.

ZONAS	Periódicos	Quantidade de artigos	
NÚCLEO	1932-6203	35.508	134.946
	0066-782X	25.907	
	0103-8478	22.688	
	0102-0935	17.241	
	0100-204X	16.997	
	0100-4042	16.605	
ZONA 1	0004-282X	15.983	135.722
	1516-3598	15.752	
	0102-311X	14.996	
	0103-5053	14.287	
	0037-8682	14.146	
	0074-0276	13.760	
	0100-879X	12.865	
	1413-8123	12.186	
	0100-0683	10.942	
	0100-736X	10.805	
ZONA 2	1679-0359	10.560	139.097
	0100-2945	10.440	
	1413-7054	10.220	
	0102-0536	9.861	
	0100-7203	9.724	
	2045-2322	9.723	
	0034-8910	9.026	
	0104-4230	8.756	
	0004-2749	8.663	
	0001-3765	7.859	
	0100-6991	7.752	
	1807-5932	7.430	
	0021-7557	7.392	
	1517-7491	7.352	
	1415-4757	7.220	
	0034-7167	7.119	

Como pode ser observado, o núcleo, que é composto por seis periódicos em acesso aberto, possui um total de 134.946 artigos cadastrados nos currículos cadastrados na Plataforma Lattes. Essa quantidade é próxima da quantidade de artigos dos periódicos que compõem as zonas 1 e 2, no entanto, com uma quantidade de periódicos bem inferior.

O periódico com maior quantidade de artigos que compoem o Núcleo é a PLoS One, com 35.508 artigos. A PLoS One se caracteriza por ser um periódico interdisciplinar que aceita somente artigos escritos na língua inglesa, publicados após a aceitação. Destaca-se por possuir o maior Fator de Impacto (FI) entre os 6 periódicos existentes no Núcleo, sendo este de 2,776.

Posteriormente, destaca-se o periódico Arquivos Brasileiros de Cardiologia, da Área da Saúde. Este aceita publicações bilíngue, sendo submetidos os manuscritos em Português e/ou Inglês, com FI igual a 1,679.

O periódico Ciência Rural ocupa a terceira posição no Núcleo, com FI igual a 0,553 e se caracteriza por ser da Área das Ciências Agrárias. Tem periodicidade mensal, com publicações em Português e Inglês. O quantitativo de seus artigos no conjunto analisado representa aproximadamente 64% do total de artigos do periódico PLoS One (1ª posição).

Já o periódico Arquivo Brasileiro de Medicina Veterinária e Zootecnia (4ª posição) tem o menor FI dentre os periódicos do Núcleo, correspondente a 0,286. Ressalta-se que, apesar de não ter um valor de FI tão significativo, ele tem uma grande quantidade de artigos atribuídos.

O periódico Pesquisa Agropecuária Brasileira, que ocupa a quinta posição do Núcleo, possui periodicidade mensal, com FI igual a 0,546, e aceita trabalhos somente manuscritos em inglês. Possui qualis B1 nas área da Ciências Agrárias I e Zootecnia/Recursos Pesqueiros, de acordo com o último qualis disponível (2013-2016). Complementando o Núcleo, na sexta posição está o periódico Química Nova, com periodicidade mensal, permitindo submissões de artigos em português, espanhol e inglês e com um valor de FI igual 0,646. Destaca-se, também, por possuir qualis A2 nas Áreas da Ciências Ambiental e Educação.

Já a Zona 1, composta por 10 periódicos, possui em sua maioria periódicos da área de Ciências da Saúde, com exatamente seis periódicos (60%), variando seus valores do FI entre 2,368 e 0,757. O periódico que se encontra na Zona 1 com maior valor de FI é Memórias do Instituto Oswaldo Cruz, que publica 12 edições por ano e aceita somente artigos em Inglês.

Posteriormente, a Zona 2 composta por 16 periódicos, se caracteriza por possuir periódicos em sua maioria das grandes áreas de Ciências da Saúde e Ciências Agrárias. A área de Ciências da Saúde se destaca por possuir aproximadamente 56% dos periódicos nesta zona, e a de Ciências Agrárias possui 25%.

Ao considerar os artigos dos periódicos em acesso aberto que compoem o Núcleo, é possível identificar e analisar os autores desses artigos. Logo, uma análise considerando suas áreas de atuação foi realizada (Tabela 6.3). Nota-se que a área de atuação que mais se destaca é a área de Medicina, que representa aproximadamente 63% dos autores que se encontram com publicações no Núcleo.

Tabela 6.3 – Principais autores com publicações no Núcleo.

Posição	Titulação	Instituição	Grande Área	Área	Artigos em Periódicos no Núcleo	Artigos em Periódicos fora do Núcleo	Tempo de Carreira	Bolsista de Produtividade em Pesquisa do CNPq
1	Livre Docência	Universidade do Estado do Rio de Janeiro	Ciências da Saúde	Medicina	240 (99,2%)	2 (00,8%)	31	Não
2	Doutorado	Universidade de São Paulo	Ciências da Saúde	Medicina	214 (72,1%)	83 27,9%)	34	Não
3	Doutorado	Universidade de São Paulo	Ciências da Saúde	Medicina	193 (33,2%)	389 (66,8%)	37	Não
4	Doutorado	Instituto Dante Pazzanese de Cardiologia	Ciências da Saúde	Medicina	178 (33,8%)	349 (66,2%)	54	Não
5	Doutorado	Universidade Federal de São Paulo	Ciências da Saúde	Medicina	167 (34,6%)	315 (65,4%)	37	Não
6	Doutorado	Não Informado	Ciências da Saúde	Medicina	158 (97,5%)	4 (2,5%)	50	Não
7	Doutorado	Universidade de São Paulo	Ciências da Saúde	Medicina	155 (41,0%)	223 (59,0%)	40	Não
8	Doutorado	Universidade Federal de Minas Gerais	Ciências Agrárias	Zootecnia	139 (47,2%)	155 (52,8%)	47	Não
9	Doutorado	Universidade de São Paulo	Ciências da Saúde	Medicina	138 (54,5%)	115 (45,5%)	39	Não
10	Doutorado	Fundação Universitária de Cardiologia	Ciências da Saúde	Medicina	132 (97,0%)	4 (3,0%)	24	Não
11	Doutorado	Fundação Universitária de Cardiologia	Ciências da Saúde	Medicina	128 (53,5%)	111 (46,5%)	33	2
12	Doutorado	Universidade Federal Fluminense	Ciências da Saúde	Medicina	116 (54,5%)	97 (45,5%)	25	Não
13	Doutorado	Universidade Tecnológica Federal do Paraná	Ciências Agrárias	Agronomia	101 (45,9%)	119 (54,1%)	31	Não
14	Doutorado	Instituto Dante Pazzanese de Cardiologia	Ciências da Saúde	Medicina	100 (29,0%)	244 (71,0%)	25	2
15	Doutorado	Universidade Federal do Paraná	Ciências da Saúde	Medicina	99 (89,2%)	12 (10,8%)	24	Não
16	Doutorado	Universidade Federal de Santa Maria	Ciências Exatas e da Terra	Probabilidade e Estatística	98 (30,2%)	227 (69,8%)	18	1A
17	Doutorado	Universidade Federal de São Paulo	Ciências da Saúde	Medicina	96 (30,6%)	218 (69,4%)	47	Não
18	Doutorado	Universidade Federal de São Paulo	Ciências da Saúde	Medicina	95 (85,6%)	16 (14,4%)	31	Não
19	Doutorado	Universidade Federal de Minas Gerais	Ciências Agrárias	Medicina Veterinária	93 (75,0%)	31 (25,0%)	23	Não

20	Doutorado	Universidade Federal de Minas Gerais	Ciências Agrárias	Zootecnia	90 (38,3%)	145 (61,7%)	33	Não
21	Doutorado	Universidade Federal de Viçosa	Ciências Biológicas	Genética	88 (13,9%)	546 (86,1%)	30	1A
22	Doutorado	Universidade Federal do Rio Grande do Sul	Ciências da Saúde	Medicina	87 (96,7%)	3 (3,3%)	41	Não
23	Doutorado	Universidade de São Paulo	Ciências da Saúde	Medicina	86 (38,7%)	136 (61,3%)	25	Não
24	Doutorado	Universidade Federal de Santa Maria	Ciências Agrárias	Ciência e Tec. de Alimentos	85 (25,4%)	250 (74,6%)	30	1D

Observa-se, ainda, na Tabela 6.3, que 8 dos 24 autores mais produtivos possuem mais de 70% de suas publicações em acesso aberto nos periódicos que compõem o Núcleo. Tais autores atuam principalmente nas áreas de medicina (7) e medicina veterinária (1), demonstrando uma forte tendência desses indivíduos em publicar os resultados de suas pesquisas em acesso aberto. Além disso, destacam-se algumas áreas que possuem uma quantidade significativa de indivíduos com publicações em periódicos que se encontram no Núcleo, tais como as áreas de Zootecnia, Agronomia, Probabilidade e Estatística, Genética e Ciência e Tecnologia de Alimentos.

Ressalta-se, ainda, que dos 24 autores com as maiores quantidades de publicações no Núcleo, apenas 5 são bolsistas de produtividade em pesquisa do CNPq, distribuídos nos níveis 1A (2), 1D (1) e 2 (2). Já na Tabela 6.4, apresenta-se a divisão de todos os autores com publicações nos periódicos que compõem o Núcleo, distribuídos em suas grandes áreas de atuação.

Tabela 6.4 – Distribuição dos autores que compoem o núcleo por grande Área de Atuação.

Grande Área	Quantidades de autores
Ciências Agrárias	13.936
Ciências Biológicas	12.391
Ciências da Saúde	10.855
Ciências Exatas e da Terra	6.037
Ciências Humanas	592
Ciências Sociais Aplicadas	328
Engenharias	1.262
Linguística, Letras e Artes	30
Outros	231
Não Informado	3.248
Total	48.910

Como pode ser observado, a grande área de Ciências Agrárias possui a maior quantidade de autores no Núcleo, que equivale a aproximadamente 28,5% do total de autores que compõem o Núcleo. Consequentemente, dentre as grandes áreas mais representativas destaca-se a grande área de Ciências Biológicas, com aproximadamente 25% do conjunto, a de Ciências da Saúde, com aproximadamente 22% do conjunto e, ainda, Ciências Exatas e da Terra, com 12,3%. Ao considerar as quatro grandes áreas mais representativas, elas contemplam mais de 80% do conjunto dos autores que publicam em periódicos do Núcleo. Tal fato está diretamente relacionado ao perfil dos periódicos do Núcleo que, em geral, contempla essas áreas.

É possível ainda verificar que a menor quantidade de autores no Núcleo é da grande área de Linguísticas, Letras e Artes, representando somente 0,061% do total de autores do Núcleo. Tendo em vista especificamente as grandes áreas de atuação dos autores das publicações dos periódicos que compõem o Núcleo, é possível identificar a distribuição e representatividade desse conjunto de autores em cada um dos periódicos (Figura 6.1).

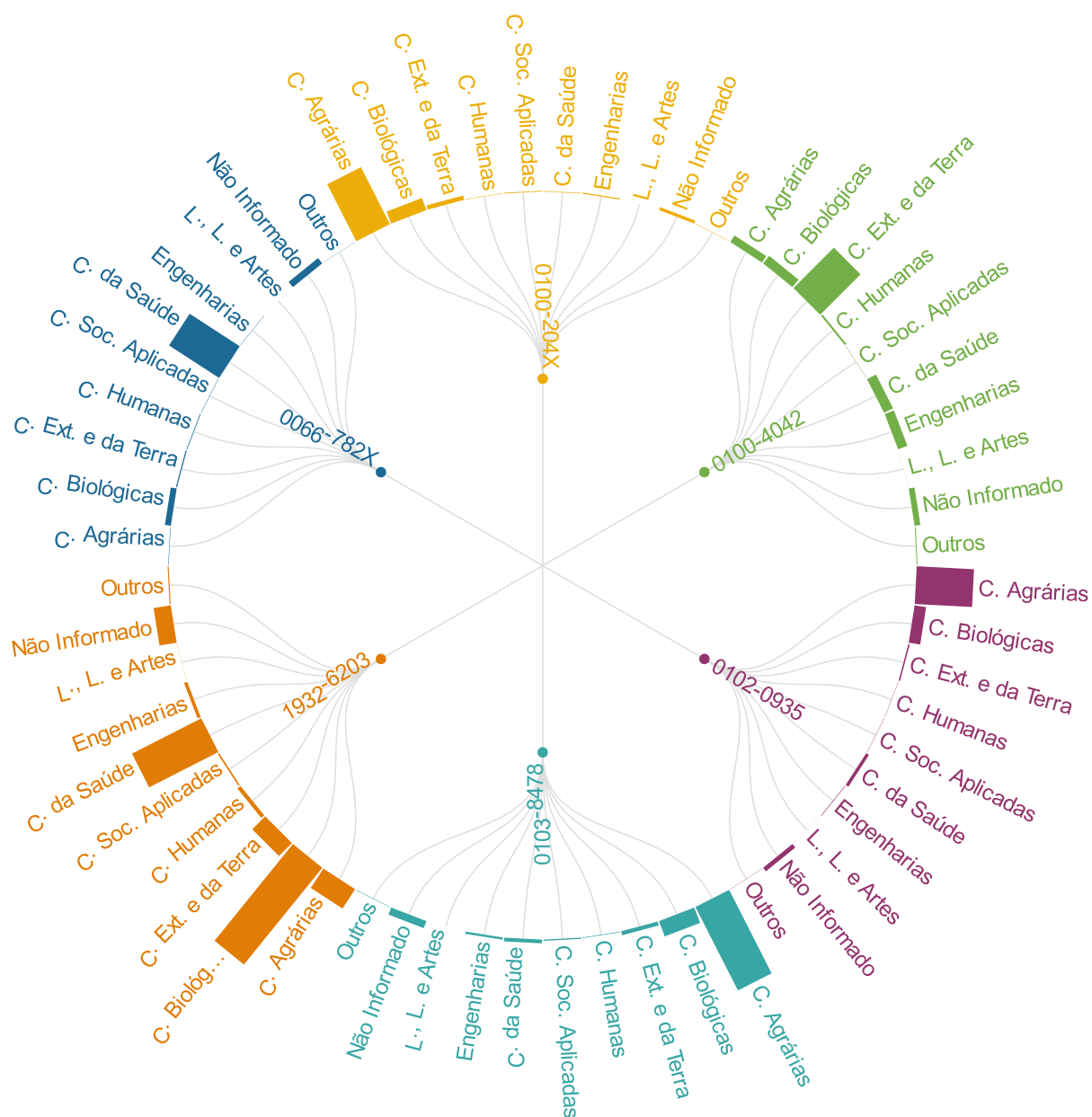


Figura 6.1 – Divisão dos autores por área de conhecimento.

Observa-se, conforme já apresentado, que a grande área de Ciências Agrárias se destaca como a Grande Área com a maior quantidade de autores com publicações no Núcleo. É importante destacar a influência dessa grande área que, consequentemente, possui a maior representatividade em 3 dos 6 periódicos do Núcleo. Destaca-se, na sequência, a Grande Área de Ciências da Saúde, seguida de Ciências Biológicas e Ciências Exatas e da Terra.

A representatividade de cada Grande Área do Conhecimento sofre influência significativa de alguns autores que são altamente produtivos. A Tabela 6.5 apresenta os dados dos autores mais produtivos dos periódicos que compõem o Núcleo e do conjunto global com publicações em periódicos de acesso aberto.

Tabela 6.5 – Principais autores dos periódicos do núcleo e principais autores do conjunto selecionado.

Posição	Instituição	Grande Área	Posição	Instituição	Grande Área
1	Universidade do Estado do Rio de Janeiro	C. da Saúde	1	Universidade Federal do Paraná	C. Saúde
2	Universidade de São Paulo	C. da Saúde	2	Universidade Federal do Rio de Janeiro	C. Saúde
3	Universidade de São Paulo	C. da Saúde	3	Universidade Federal de Viçosa	C. Agrárias
4	Instituto Dante Pazzanese de Cardiologia	C. da Saúde	4	Universidade Federal de Viçosa	C Ex. Terra
5	Universidade Federal de São Paulo	C. da Saúde	5	Universidade Federal de Viçosa	C. Biológicas
6	Não Informado	C. da Saúde	6	Universidade Federal de Minas Gerais	C. da Saúde
7	Universidade de São Paulo	C. da Saúde	7	Universidade Estadual Paulista	Não Informado
8	Universidade Federal de Minas Gerais	C. Agrárias	8	Universidade Federal de São Paulo	C. Biológicas
9	Universidade de São Paulo	C. da Saúde	9	Universidade Federal de Lavras	C. Agrárias
10	Fundação Universitária de Cardiologia	C. da Saúde	10	Universidade de São Paulo	C. da Saúde
11	Fundação Universitária de Cardiologia	C. da Saúde	11	Universidade Estadual de Maringá	C. da Saúde
12	Universidade Federal Fluminense	C. da Saúde	12	Faculdade Jesuíta de Filosofia e Teologia	C. Humanas
13	Universidade Tecnológica Federal do Paraná	C. Agrárias	13	Faculdade de Medicina de São José do Rio Preto	C. da Saúde
14	Instituto Dante Pazzanese de Cardiologia	C. da Saúde	14	Universidade Federal de Santa Catarina	C. da Saúde
15	Universidade Federal do Paraná	C. da Saúde	15	Universidade Federal de Viçosa	C. Agrárias
16	Universidade Federal de Santa Maria	C. Ex. Terra	16	Universidade de São Paulo	C. Agrárias
17	Universidade Federal de São Paulo	C. da Saúde	17	Faculdade de Medicina do ABC	C. da Saúde
18	Universidade Federal de São Paulo	C. da Saúde	18	Universidade de São Paulo	C. da Saúde
19	Universidade Federal de Minas Gerais	C. Agrárias	19	Universidade de São Paulo	C. Biológicas
20	Universidade Federal de Minas Gerais	C. Agrárias	20	Pontifícia Universidade Católica do Rio de Janeiro	C Ex. Terra
21	Universidade Federal de Viçosa	C. Biológicas	21	Universidade Federal de Goiás	C. Agrárias
22	Universidade Federal do Rio Grande do Sul	C. da Saúde	22	Universidade Federal do Rio de Janeiro	C. Ex. Terra
23	Universidade de São Paulo	C. da Saúde	23	Universidade de São Paulo	C. Saúde
24	Universidade Federal de Santa Maria	C. Agrárias	24	Universidade Federal do Rio de Janeiro	C Ex. Terra

Ao se realizar uma comparação entre os autores mais produtivos que têm publicado nos periódicos pertencentes ao Núcleo e os mais produtivos em todo o conjunto selecionado, ou seja, os autores que têm os maiores quantitativos de publicações em periódicos do Núcleo e em todo o conjunto de periódicos de acesso aberto, é possível verificar a baixa similaridade entre os conjuntos. Observa-se que somente um único autor está presente nos dois conjuntos, sendo esse autor da grande área de Ciências Biológicas. Mesmo tendo um número expressivo de autores da grande área de Ciências da Saúde, estes não estavam presentes nos dois conjuntos. Logo, percebe-se que os autores mais produtivos, considerando todos os artigos em todos os periódicos, não possuem grande representatividade nos principais periódicos.

Diante disso, ressalta-se que mesmo identificando autores muitos produtivos na publicação de artigos em periódicos de acesso aberto, estes, em geral, não são os autores com maior produção de artigos nos principais periódicos. Isso refere-se àqueles que compõem o núcleo, identificados pela adoção da Lei de Bradford. Tal fato justifica um estudo sobre a produtividade dos autores para uma melhor compreensão sobre o conjunto em análise. Para tanto, a Lei de Lotka pode servir como parâmetro.

6.2 Produtividade dos autores – Lei de Lotka

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 expressivas os melhores candidatos à aquisição de reconhecimento e acesso a recursos para melhorar ainda mais as suas pesquisas. Logo, a Lei do Quadrado Inverso foi um modelo estatístico utilizado para verificar se a produtividade com número de autores que totalizam n contribuições, num determinado campo, se aproxima de $1/n^2$, que possui somente uma publicação.

De acordo com Cruz *et al.* (2010), a Lei de Lotka possui uma premissa básica, segundo a qual alguns autores, teoricamente com maior importância em determinada área de conhecimento, publicam muito, e vários autores com menor importância publicam pouco. Nesse contexto, a Tabela 6.6 apresenta a produtividade por autor em periódicos de acesso aberto de todo o conjunto analisado neste trabalho.

Tabela 6.6 – Produtividade dos Autores.

Nº de contribuições por autor	Nº de autores	Frequência Relativa
1	130.793	35,5544%
2	64.638	17,5710%
3	38.425	10,4454%
4	25.400	6,9047%
5	18.086	4,9165%
6	13.433	3,6516%
7	10.467	2,8453%
8	8.262	2,2459%
9	6.695	1,8200%
10	5.536	1,5049%
11	4.660	1,2668%
12	3.978	1,0814%
13	3.470	0,9433%
14	2.861	0,7777%
15	2.595	0,7054%
⋮	⋮	⋮
379	1	0,0003%
399	1	0,0003%
401	1	0,0003%
481	1	0,0003%
544	1	0,0003%
Total	370.388	

Como pode ser observado, existem poucos autores com um quantitativo muito elevado de publicação, tendo o autor mais produtivo 544 artigos em periódicos de acesso aberto. Somente os 5 autores mais produtivos são responsáveis por 2.204 artigos (0,60%). Como era de se esperar, a maioria dos autores possui poucas contribuições. Um conjunto composto por 233.856 autores (63,13%) possui uma, duas ou três contribuições apenas.

A Lei de Lotka estabelece que o número de autores que contribuem com um artigo constituirá o maior grupo, cerca de 60% do total de autores (GUPTA e KARISIDDIPPA, 1999). Ao analisar o conjunto de dados sobre a produtividade dos autores considerados neste trabalho, percebe-se que o percentual dos autores que publicaram apenas um artigo (35,55%) está bem abaixo do que o padrão da Lei de Lotka sugere, de aproximadamente 60%.

Aplicando o modelo conhecido como Poder Inverso Generalizado para calcular a produtividade dos autores, como proposto no trabalho de Alvarado (2006), cuja distribuição teórica é calculada pela Equação (1), temos:

$$Y_x = cx^{-n} \quad (1)$$

Para $x = 1, 2, 3, \dots, x_{max}$

Onde:

Y_x é a probabilidade de que um autor publique x artigos.

c e n são os dois parâmetros a serem calculados a partir dos dados coletados.

Logo, na Tabela 6.7, são apresentados os cálculos pertinentes para se calcular o expoente (n) de Lotka, considerando os dados analisados.

Tabela 6.7 – Distribuição de frequência para o expoente de Lotka.

x	y	$X = \log(x)$	$Y = \log(y)$	XY	X^2
1	130.793	0	5,116585	0	0
2	64.638	0,30103	4,810488	1,448101	0,090619
3	38.425	0,477121	4,584614	2,187417	0,227645
4	25.400	0,60206	4,404834	2,651974	0,362476
5	18.086	0,69897	4,257343	2,975755	0,488559
6	13.433	0,778151	4,128173	3,212343	0,605519
7	10.467	0,845098	4,019822	3,397144	0,714191
8	8.262	0,90309	3,917085	3,53748	0,815572
9	6.695	0,954243	3,825751	3,650694	0,910579
10	5.536	1	3,743196	3,743196	1
11	4.660	1,041393	3,668386	3,82023	1,084499
12	3.978	1,079181	3,599665	3,884691	1,164632
13	3.470	1,113943	3,540329	3,943726	1,24087
14	2.861	1,146128	3,456518	3,961612	1,313609
15	2.595	1,176091	3,414137	4,015337	1,383191
⋮	⋮	⋮	⋮	⋮	⋮
379	1	2,578639	0	0	6,64938

399	1	2,600973	0	0	6,76506
401	1	2,603144	0	0	6,776361
481	1	2,682145	0	0	7,193902
544	1	2,735599	0	0	7,483501
Total					
Geral	370.388	499,9983	315,5898	504,9284	1.044,156

Como pode ser observado, na coluna 1 e coluna 2 são apresentados os totais de contribuições por autor e total de autores, respectivamente. Percebe-se que o primeiro valor de x (uma contribuição) contém o maior número de autores (y). Na terceira coluna apresentam-se os valores do logaritmo (na base 10) da coluna 1. Na quarta coluna estão os valores de logaritmo (base 10) dos números de autores (coluna 2). Em sequência, na coluna 5 multiplicaram-se os valores encontrados das colunas 3 e 4. Na última coluna (6), elevou-se ao quadrado os valores da coluna 3 e, após todos os cálculos, fez-se o somatório de cada coluna para aplicar os valores na Equação (2).

Como primeiro passo no teste da Lei de Lotka, determina-se o valor do parâmetro n , obtido pelo método do Mínimo Quadrado Linear, desenvolvido por Legendre e aperfeiçoado por Galton e Pearson (BRUNI, 2013).

$$n = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2} \quad (2)$$

Onde:

N = número de pares observados

X = logaritmo de x

Y = logaritmos de y

Aplicou-se o resultado do parâmetro n (-2,32708), calculado com a Equação (2), para se obter o valor do parâmetro c , que é o número teórico de autores com um único artigo. Ele é estimado pela Função Zeta Inversa de Riemann e ajustada por Pao (1985), conforme a Equação (3), possibilitando gerar a distribuição teórica de Lotka.

$$c = \frac{1}{\sum_{x=1}^{P-1} \frac{1}{x^n} + \frac{1}{(n-1)P^{n-1}} + \frac{1}{P^n} + \frac{n}{24(P-1)^{n+1}}} \quad (3)$$

Onde:

x = número de contribuição por autor.

n = valor do parâmetro calculado pelo método do Mínimo Quadrado Linear.

P = número de pares observados.

Tendo em vista a análise sobre a produtividade dos autores que publicaram artigos em periódicos de acesso aberto, obtida utilizando a Lei de Lotka, testou-se a qualidade do ajuste dos dados coletados com o teste Kolmogorov-Smirnov(K-S) para aceitar ou rejeitar a hipótese de que a maioria dos artigos são publicados por poucos autores. O teste de K-S é útil em julgar o grau de proximidade das frequências, por ser um método simples para dados não

paramétricos. No trabalho de Coile (1977), o autor sugere o teste como um poderoso método estatístico.

A Tabela 6.8 apresenta os valores observados e esperados da distribuição dos autores. Tais valores serão utilizados para se aplicar o teste de K-S, que permitirá testar a existência de diferenças entre a frequência acumulativa dos valores observados e esperados.

Tabela 6.8 – Distribuição dos Autores Observados e Esperados.

x	Y	% de Autores	% Acumulativa de Autores	% de Autores c. x^{-n}	% Acumulativa de Autores c. x^{-n}	Dmax
1	130.793	0,3555442592	0,355544259	0,7050595070	0,705059507	0,349515248
2	64.638	0,1757102431	0,531254502	0,1405092419	0,845568749	0,314314247
3	38.425	0,1044535117	0,635708014	0,0546923323	0,900261081	0,264553067
4	25.400	0,0690466935	0,704754708	0,0280016749	0,928262756	0,223508049
5	18.086	0,0491645078	0,753919215	0,0166596800	0,944922436	0,191003221
6	13.433	0,0365159147	0,79043513	0,0108994745	0,955821911	0,165386781
7	10.467	0,0284532181	0,81888348	0,0076140378	0,963435948	0,1445476
8	8.262	0,0224592040	0,841347552	0,0055803717	0,96901632	0,127668768
9	6.695	0,0181995123	0,859547065	0,0042425514	0,973258871	0,113711807
10	5.536	0,0150489171	0,874595982	0,0033200588	0,97657893	0,101982949
11	4.660	0,0126676217	0,887263603	0,0026596328	0,979238563	0,09197496
12	3.978	0,0108136908	0,898077294	0,0021721243	0,981410687	0,083333393
13	3.470	0,0094327569	0,907510051	0,0018029782	0,983213666	0,075703614
14	2.861	0,0077772673	0,915287319	0,0015173793	0,984731045	0,069443726
15	2.595	0,0070541799	0,922341498	0,0012923118	0,986023357	0,063681858
	⋮	⋮	⋮	⋮	⋮	⋮
379	1	0,0000027184	0,999989127	0,0000007039	0,999606614	0,000382512
399	1	0,0000027184	0,999991845	0,0000006245	0,999607239	0,000384606
401	1	0,0000027184	0,999994563	0,0000006173	0,999607856	0,000386707
481	1	0,0000027184	0,999997282	0,0000004043	0,99960826	0,000389021
544	1	0,0000027184	1	0,0000003036	0,999608564	0,000391436
Total Geral	370.388					

Aplicando os valores n e c (Equação (2) e Equação (3)), foram calculados os valores esperados para distribuição dos autores. Para isso, foi necessária a elaboração da Tabela 6.8. Os valores x e y estão apresentados nas colunas 1 e 2, respectivamente. A coluna 3 contém a porcentagem dos autores. Em seguida, a coluna 4 possui a porcentagem acumulativa dos dados da coluna 3. Já a coluna 5 contém os valores teóricos relativos calculados com a Equação (1). Por sua vez, a coluna 7 possui as diferenças entre a coluna 4 e a coluna 6.

Em seguida identificou-se a desviação máxima (D_{max}) na coluna da diferença, ou seja, o máximo valor absoluto encontrado entre a diferença das colunas 4 e 6, sem considerar os sinais negativos da desviação máxima calculada.

Para encontrar a região de rejeição, utilizou-se os valores críticos com nível de significância (risco de erro) de 1%, 5%, 10%, 15% e 20% encontrados na tabela de teste de K-S (Anexo D). Como o valor de N é maior que 50, utilizou-se a fórmula constante na tabela de teste de K-S.

Os valores críticos ($D_{crítico}$) do teste de K-S estão apresentados na Tabela 6.9. Logo, foi possível realizar a comparação dos valores críticos encontrados com o valor da coluna D_{max} . Sendo assim, existem apenas duas interpretações possíveis: se a distribuição se ajusta ou não à Lei de Lotka. Se D_{max} for maior que o valor de $D_{crítico}$, rejeita-se a hipótese de que a distribuição se ajusta à Lei de Lotka. Se D_{max} for menor que o valor de $D_{crítico}$ aceita-se a hipótese de que a distribuição se ajusta à Lei de Lotka.

Tabela 6.9 – Resultado da aplicação do Teste de K-S.

Nível de Significância	1%	5%	10%	15%	20%
$D_{crítico}$	0,0027	0,0022	0,0020	0,0019	0,0018
$D_{máximo} = 0,349515$					
Resultado do teste H_0	REJEIÇÃO	REJEIÇÃO	REJEIÇÃO	REJEIÇÃO	REJEIÇÃO

Percebe-se que a comparação dos valores $D_{crítico}$ para os todos os níveis de significância foi feita com o $D_{max} = 0,349515$, e rejeitou-se todos os valores críticos encontrados com cada nível de significância, pois $D_{max} > D_{crítico}$. Diante disso, comprova-se que a distribuição de produtividade das publicações em periódicos de acesso aberto do conjunto analisado neste trabalho não se aplica à Lei de Lotka.

Conforme apresentado na dispersão dos dados, há um afastamento entre os valores observados e esperados da distribuição de produtividade dos autores (Figura 6.2).

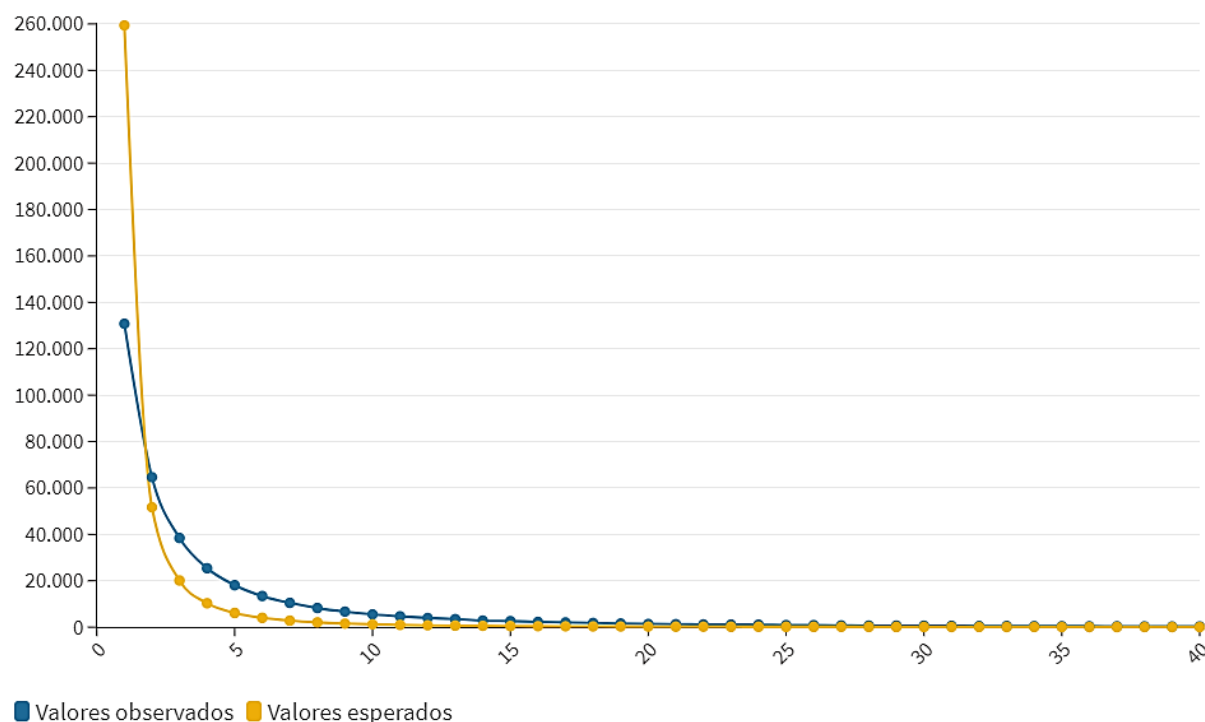


Figura 6.2 – Dispersão dos valores observados e esperados.

Dessa forma, os maiores afastamentos ocorrem nas primeiras frequências da distribuição dos autores. O modelo propõe 128.575 a mais do que os 130.793 autores que publicam um único artigo em periódico de acesso aberto. É válido salientar que, para a construção do gráfico, admitiu-se somente até 40 autores, de modo a facilitar a visualização dos afastamentos, já que estes ocorrem com maior intensidade no início das frequências da distribuição.

Diante disso, concluiu-se que a distribuição dos autores que publicaram em periódicos de acesso aberto não está dentro da normalidade dos dados propostos pela Lei de Lotka.

6.3 Principais Tópicos de Pesquisa – Lei de Zipf

Para a definição do conjunto de dados a serem analisados nesta seção, optou-se por extrair as palavras dos títulos dos artigos publicados em periódicos de acesso aberto (2.090.015 palavras). A escolha da extração das palavras dos títulos dos artigos em detrimento das palavras-chave vinculadas aos artigos se deve ao fato de que, aproximadamente, apenas 17% dos artigos analisados possuíam palavras-chave vinculadas a eles. Além disso, diversos trabalhos têm utilizados as palavras dos títulos das publicações como objeto de análise (CUNHA *et al.* 2013, VINKERS *et al.* 2015, MRYGLOD *et al.* 2016 e RONDA-PUPO 2016).

Adicionalmente, tendo em vista que o cadastramento das palavras-chave de um artigo científico em seus currículos é de inteira responsabilidade dos respectivos pesquisadores, e que isso é feito livremente por eles, conclui-se que podem ser inseridos quaisquer conjuntos de caracteres como uma palavra-chave. A partir disto, geralmente tem-se uma coleção muito grande de palavras-chave, sem nenhum padrão (GOMES, 2018).

Logo, para as análises aqui realizadas, os títulos de todas as publicações do conjunto identificado foram considerados. Tais títulos passaram por um processo de tratamento de dados que visou identificar as palavras que posteriormente serão objeto de análise. Todas as etapas do processo de tratamento podem ser visualizadas na Tabela 6.10.

Tabela 6.10 – Etapas do Processo de Tratamento dos Dados.

ETAPA DO ALGORITMO	RESULTADO
Recebimento do Título	UMA ESTRATÉGIA PARA IDENTIFICAÇÃO DE ARTIGOS EM PERIÓDICOS DE ACESSO ABERTO NA PLATAFORMA LATTES
LowerCase	uma estratégia para identificação de artigos em periódicos de acesso aberto na plataforma lattes
StopWords_PT	estratégia identificação artigos periódicos acesso aberto plataforma lattes
StopWords_EN	estratégia identificação artigos periódicos acesso aberto plataforma lattes
Identificação de Termos	estratégia
	identificação
	artigos
	periódicos
	acesso
	aberto
	plataforma lattes

Como pode ser observado, o título de cada artigo é recuperado e, dessa forma, o processo de tratamento dos dados é inicializado. Na etapa de *LowerCase*, todas as palavras são convertidas para minúsculas, com a proposta de padronizar o conjunto, bem como evitar que palavras sejam mapeadas em tópicos distintos, por algumas possuírem letras em maiúsculas e outras não. Já no processo de remoção de *StopWords* (*StopWords_PT* e *StopWords_EN*), são removidos todos os termos que não possuem valores semânticos significativos para caracterizar um tópico de pesquisa e, com isso, diminuir o volume de palavras a serem processadas e analisadas. Foram removidos os *StopWords* inicialmente em português e posteriormente em inglês, tendo em vista que esses são os idiomas mais utilizados, conforme já apresentado.

O objeto inicial de análise é o título dos artigos, nos quais, se existe uma preocupação com a descrição geral do estudo a ser apresentado, a quantidade de *stopWords* é significativa. Isso se dá diferentemente das palavras-chave, justificando a remoção para as análises a serem

realizadas. Após isso, na última etapa *Identificação de Termos*, as palavras são separadas em tópicos que irão compor um dicionário de termos para a contagem das frequências.

Inicialmente foram identificados 28.636.958 termos, considerando todas as palavras dos títulos dos artigos. Após a remoção das duplicatas, o conjunto foi reduzido a 423.364 palavras únicas. Posteriormente, com a remoção das *StopWords*, o conjunto passou a ter um total de 393.896 palavras que se tornaram objeto de análise. A Tabela 6.11 apresenta os termos com a maior frequência antes da remoção das *StopWords*.

Tabela 6.11 – Frequência dos Principais Termos sem a Remoção de *StopWords*.

Posição	Frequência	Palavras
1	1.741.809	de
2	762.560	of
3	716.058	e
4	529.940	a
5	513.477	do
6	507.828	da
7	501.705	in
8	463.628	and
9	452.657	em
10	373.515	the
11	234.604	no
12	230.966	na
13	217.452	o
14	156.366	para
15	149.550	com

Como pode ser observado, e tendo em vista a característica dos títulos das publicações que em geral necessitam da utilização de *StopWords* na sua composição, todos os primeiros 15 termos identificados são *StopWords* em português ou inglês. Tendo em vista a frequência muito significativa destes termos, justificou-se a sua exclusão.

Conforme citado no capítulo anterior, aproximadamente 64% das publicações em periódicos de acesso aberto são em português, logo se justifica uma quantidade considerável de termos neste idioma. A Tabela 6.12 apresenta o resultado da extração e ordenação pela frequência das palavras dos títulos de cada artigo analisado, após todo o tratamento dos dados.

Tabela 6.12 – Distribuição das palavras por posição (x) e suas frequências (y).

Posição (x)	Frequência(y)	Palavras
1	88.485	brazil
2	85.470	brasil
3	72.100	estudo
4	71.618	avaliação
5	69.314	análise
6	58.977	saúde
7	46.863	educação
8	44.131	study
9	43.411	brazilian
10	42.365	rio
11	41.411	patients
12	38.754	diferentes
13	37.788	produção
14	37.586	ensino
15	37.395	estado
⋮	⋮	⋮
393.894	1	zzaa
393.895	1	zzgam
393.896	1	zzgamma

Como pode ser observado, mesmo após a remoção das *stopWords* é possível verificar que dentre as palavras mais frequentes a maioria está em português, com algumas dessas palavras em sua versão em inglês, como as duas palavras mais frequentes. Já nas últimas posições encontram-se palavras com uma frequência muito baixa. Percebe-se que tais palavras não possuem conteúdo semântico, sendo uma hipótese para a existência de tais palavras os erros de digitação no momento de cadastro do título da publicação em um determinado currículo. Percebe-se, ainda, que dentre as palavras mais frequentes, se encontram tópicos que geralmente fazem parte dos títulos das publicações, já que são importantes para indicar métodos, técnicas, objetos ou localidades.

No intuito de avaliar o conjunto de palavras que estão vinculadas às publicações de artigos em periódicos de acesso aberto, utilizou-se a Lei de Zipf, conforme apresentada na Seção 2.2. No trabalho de Quoniam (1992), o autor descreve a curva de Zipf como dividida em três zonas de distribuição:

- Zona I - Informação trivial ou básica: define os temas centrais da análise bibliométrica;
- Zona II - Informação interessante: localiza-se entre as Zonas I e III e mostra os temas periféricos, uma informação potencialmente inovadora. É aí que as transferências de tecnologia relacionadas aos novos temas devem ser consideradas;

- Zona III - Ruído: tem como característica possuir conceitos ainda não emergentes, sendo impossível afirmar se eles serão emergentes ou se são apenas ruído estatístico.

Logo, conforme apresentado no trabalho de Cassettari *et al.* (2015) e aplicando os métodos descritos no estudo dos autores, o conjunto de palavras identificadas nesta tese, após todo o tratamento de dados já apresentado, foi dividido em três zonas textuais de distribuição (Figura 6.3).

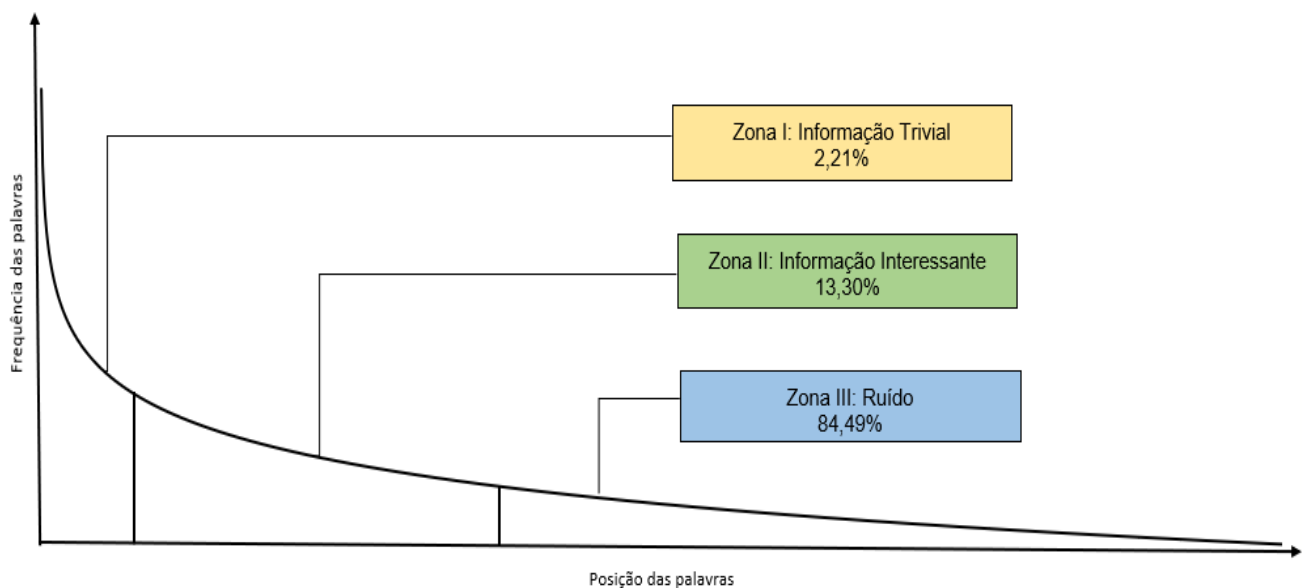


Figura 6.3 – Divisão das Palavras Identificadas nas Três Zonas Textuais.

A primeira zona identificada (Zona I) possui 2,21% das palavras analisadas. Tais palavras, que são as mais frequentes, descrevem quais são os temas centrais do conjunto analisado. Apesar de contemplar um baixo percentual de palavras, a frequência delas corresponde a 47,93% de todo o conjunto, comprovando a sua representatividade. Já a Zona II, que possui 13,3% das palavras, engloba um conjunto de tópicos que ocorrem em menor frequência que os da Zona I. Por não serem palavras utilizadas com tanta frequência, são caracterizadas como temas emergentes, já que se classificam como informação potencialmente inovadora. Por fim, a Zona III, que possui a maioria das palavras (84,49%), se caracteriza por agregar tópicos com baixa frequência, considerados ruídos. Aqui, cabe destacar como já apontado, os problemas originados do livre cadastramento dos dados das publicações por parte dos indivíduos em seus currículos, em que é real a inserção de dados incorretos, seja por erros de digitação ou até mesmo de codificação no momento de copiar um texto de documentos em diversos formatos. Um total de 149.891 palavras possuem apenas uma ocorrência.

Existe, portanto, uma grande disparidade entre os conjuntos de palavras que compõem cada uma das Zonas identificadas. No intuito de melhor compreender cada uma dessas Zonas, diversas técnicas de análise e visualização de dados podem ser aplicadas. A Figura 6.4 apresenta uma nuvem de palavras da Zona 1.

Por fim, na Zona III, que possui a maior quantidade de palavras, aquelas com as maiores frequências (2.426 palavras) possuem 16 ocorrências, e se caracterizam por possuir, em geral, tópicos pouco utilizados, caracterizados como ruídos (Figura 6.6).



Figura 6.6 – Zona III: Ruído.

Tendo em vista a grande quantidade de palavras com a mesma frequência, na nuvem de palavras estão destacadas aquelas que iniciam com a letra “a”, devido à ordem alfabética considerada pelos algoritmos de geração das nuvens. Ressalta-se uma quantidade significativa de palavras para todas as frequências até a frequência mínima igual a um (149.891 palavras). Em geral, as palavras com apenas uma ocorrência são resultado de termos novos propostos, novos acrônimos, fórmulas, ou termos técnicos, como os nomes de proteínas inseridos nos títulos. No entanto, em sua maioria são ruídos originados de erros ortográficos, provavelmente de digitação, como “wonen”, “brsil”, “www”, “qualitativa” e “brasileiro”.

No intuito de melhor compreender o conjunto das palavras segundo a Lei de Zipf, é possível estimar o comportamento da distribuição com o logaritmo dos valores existentes nas colunas 1 e 2 da Tabela 6.10 (ver Tabela 6.13).

Tabela 6.13 – Distribuição de $\tilde{x} = \log(x)$ e $\tilde{y} = \log(y)$.

$\tilde{x} = \log(x)$	$\tilde{y} = \log(y)$	Palavras
0	4,94687	brazil
0,30103	4,931814	brasil
0,477121	4,857935	estudo
0,60206	4,855022	avaliação
0,69897	4,840821	análise
0,778151	4,770683	saúde
0,845098	4,67083	educação
0,90309	4,644744	study
0,954243	4,6376	brazilian
1	4,627007	rio
1,041393	4,617116	patients
1,079181	4,588317	diferentes
1,113943	4,577354	produção
1,146128	4,575026	ensino
1,176091	4,572814	estado
⋮	⋮	⋮
5,595379	0,00000	zxaa
5,59538	0,00000	zzgam
5,595382	0,00000	zzgamma

Observa-se na Figura 6.7 o ajuste dos dados que apresenta a reta obtida pelo método dos mínimos quadrados (reta em vermelho), ajustando-se com os dados obtidos pelo logaritmo da posição versus o logaritmo da frequência das palavras dos títulos dos artigos.

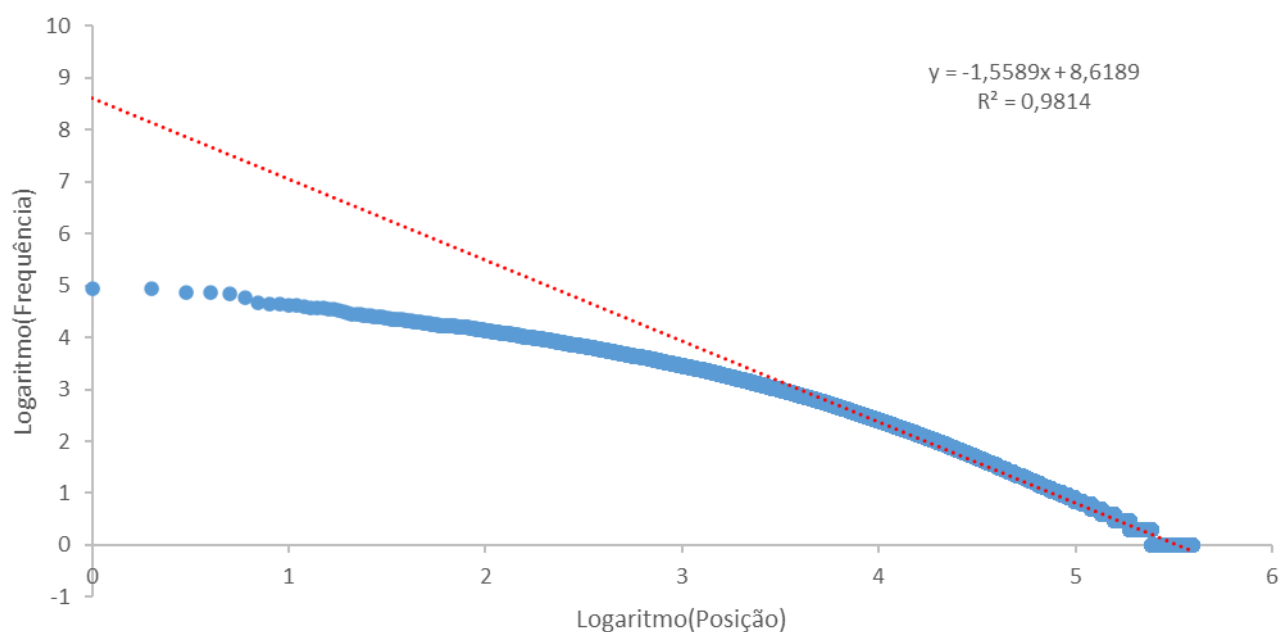


Figura 6.7 – Ajuste dos dados.

Como pode ser observado, os dados apresentados são os valores da Tabela 6.13, que parecem se alinhar no intervalo de logaritmo da posição ($\tilde{x}$) maior que 3,5. Com o Método dos Mínimos Quadrados é possível obter a equação da reta (Equação (4)), a fim de encontrar a função que melhor se ajuste aos dados para minimizar o erro. Com o uso desse método, encontra-se a equação:

$$\tilde{y} = -1,5589\tilde{x} + 8,6189 \quad (4)$$

Substituindo $\tilde{y} = \log y$ e $\tilde{x} = \log x$, obteve-se a Equação (5):

$$y = 415.814.854,6x^{-1,5589} \quad (5)$$

A Equação 5 é uma forma aproximada da relação entre frequência e posição de palavras, medida por uma Lei da Potência empírica conhecida atualmente por Lei de Zipf. Com isso, é possível perceber que mesmo que a reta obtida pelo Método dos Mínimos Quadrados não tenha se ajustado para os valores iniciais de $\tilde{x}$, ela se ajustou para um intervalo significativo dos dados a partir de $\tilde{x}$ maior que 3,5.

Logo, com a adoção do Método dos Mínimos Quadrados não foi possível acomodar satisfatoriamente os dados para os valores iniciais de $\tilde{x}$, conforme apresentado. Porém, diversos trabalhos têm sugerido adaptações na Lei de Zipf para se obter um modelo que melhor se adapte. No entanto, outros estudos descrevem que a Lei de Zipf é válida para $\tilde{x}$ em apenas um determinado intervalo (CLAUSET, SHALIZI E NEWMAN, 2009).

Ressalta-se que no estudo aqui apresentado foi realizada a remoção das *StopWord*, tendo em vista a sua alta frequência de utilização nos títulos dos artigos, podendo acarretar algum impacto na análise. Além disso, também é importante destacar que não foi possível realizar a unificação das palavras no singular e plural, bem como a utilização de palavras compostas, tendo em vista que seria necessária a adoção de técnicas como as de Processamento de Linguagem Natural, radicalização e n-gramas, as quais vão além do escopo desta tese.

6.4 Análise Qualitativa dos Periódicos em Acesso Aberto

Esta seção visa apresentar características qualitativas das publicações dos autores com artigos em periódicos de acesso aberto cadastrados na Plataforma Lattes com auxílio de algumas métricas. Essa caracterização qualitativa possibilita compreender, com adoção de indicadores qualitativos, a evolução das publicações em periódicos de acesso aberto ao longo dos últimos anos, nas várias áreas de conhecimento. Ressalta-se que essa análise não objetiva comparar ou avaliar as métricas qualitativas utilizadas, apenas adotar as métricas mais adotadas para compreender o conjunto de dados analisados.

O Fator de Impacto (FI) proposto por Garfield (1955) é considerado como informação muito valiosa para as análises qualitativas. Atualmente, o FI é divulgado anualmente pelo *Journal of Citation Report – JCR*, conforme o cálculo realizado pela Clarivate. 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 esse valor pelo número de artigos publicados pelo periódico no mesmo período. Vale mencionar que vários pesquisadores utilizam esse índice como critério de seleção para publicação de seus artigos, para que seu trabalho tenha maior visibilidade. Daí a importância do FI para qualificar o periódico.

Avaliando o FI dos periódicos em acesso aberto analisados nesta pesquisa, a Figura 6.8 apresenta as medianas em que se publicaram os artigos ao longo dos anos. Como pode ser observado, o ápice das medianas se deu no ano 2000, com valor igual a 0,5340. Após o ano 2000, por três anos consecutivos foi apresentada uma queda e, posteriormente, um crescimento por dois anos. Após isso, a partir de 2006, houve várias quedas até a chegada na mediana igual 0, no ano de 2010, que permanece até o último ano analisado nesta pesquisa. Para melhor visualização dos resultados, a Figura 6.9 apresenta a distribuição das medianas sem os outliers, tendo em vista que alguns periódicos possuem FI muito acima da média.

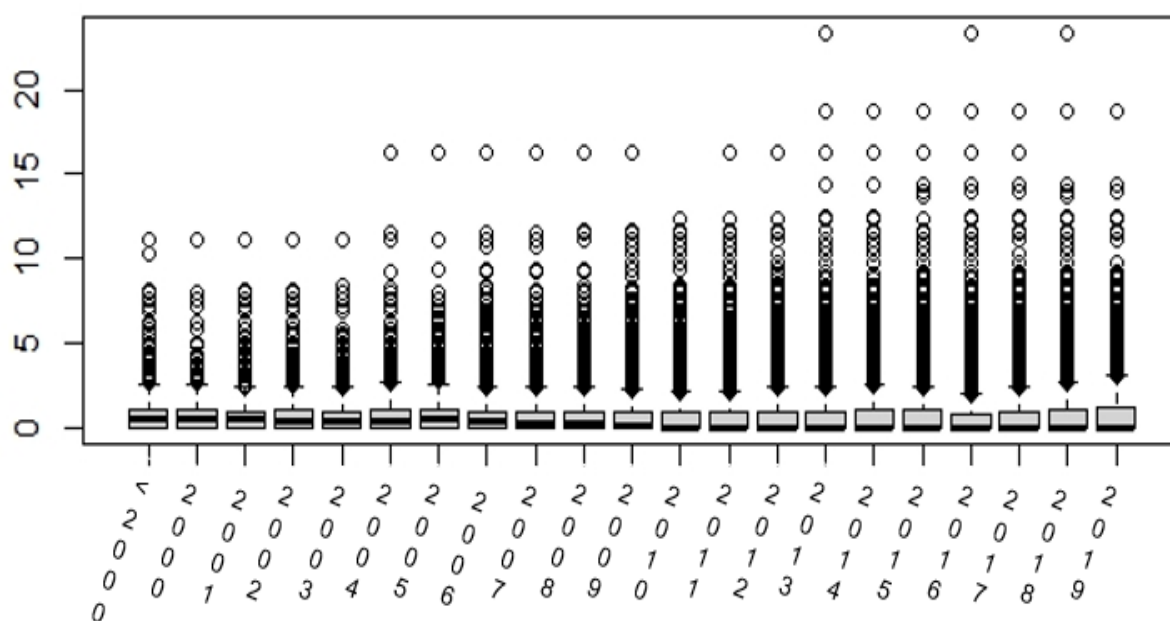


Figura 6.8 – Distribuição da Mediana do Fator de Impacto ao Longo dos Anos.

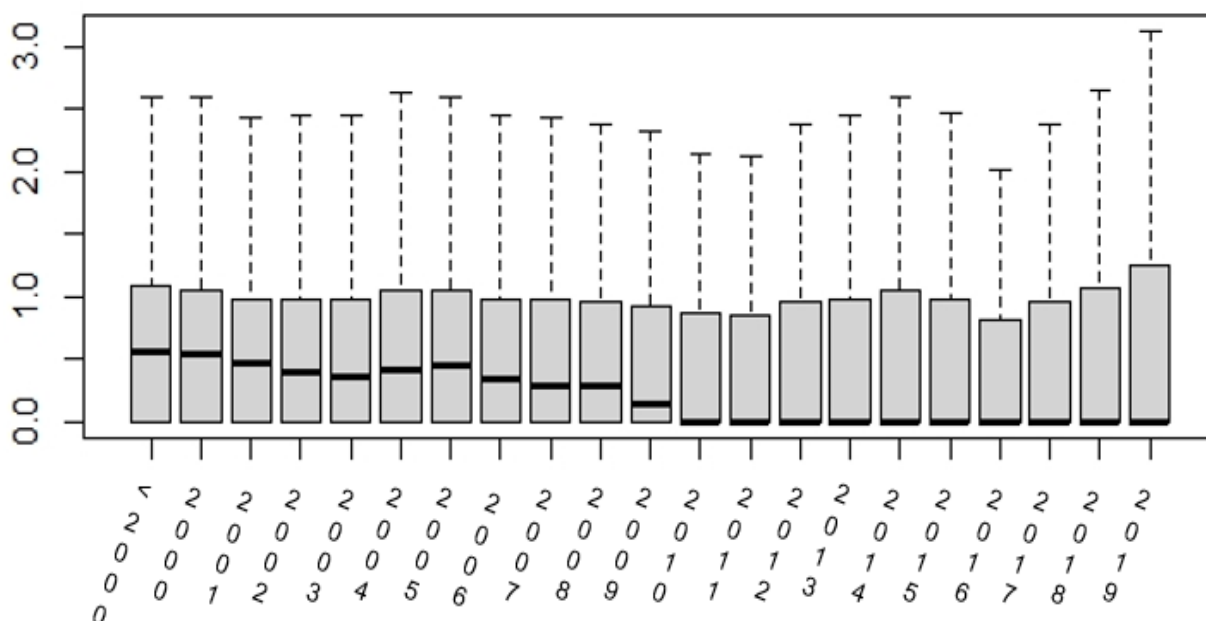


Figura 6.9 – Distribuição da Mediana do Fator de Impacto Excluídos os Outliers.

Logo, é possível perceber pelo comportamento das medianas, que a partir de 2010 a publicação de artigos em periódicos de acesso aberto analisados neste trabalho aconteceu, em sua maioria, em periódicos com FI igual a 0. Vale ressaltar que, mesmo as publicações de artigos em periódicos de acesso aberto estando em ascensão, a maioria dos trabalhos está sendo publicada em periódicos com baixo FI.

Tendo em vista a disparidade entre o valor do FI dos diversos periódicos, a classificação do conjunto foi dividida em quatro estratos (Alto, Médio, Baixo e Nulo). Inicialmente foram selecionados todos os periódicos sem Fator de Impacto, ou seja, com FI igual a 0. Estes, que totalizam 7.576 (82,33%), foram categorizados como Nulo (Estrato Nulo). Posteriormente, cada um dos outros três estratos recebeu 33,3% dos periódicos restantes, conforme ordenação do valor de seus FI, em que o estrato mais baixo contém 33,3% dos periódicos com os menores FIs, e o mais alto contém os 33,3% dos periódicos com os maiores FIs (Tabela 6.14).

Tabela 6.14 – Divisão dos Periódicos em Conjuntos de Estratos.

Estrato	Intervalo de valor do FI	Porcentagem de periódico
Estrato 1 – Alto	23,333 a 2,303	5,89 %
Estrato 2 – Médio	2,300 a 1,151	5,89%
Estrato 3 – Baixo	1,149 a 0,113	5,89%
Estrato 4 – Nulo	0	82,33%

Percebe-se que o estrato categorizado como Nulo contém a maioria dos periódicos, sendo que aproximadamente 82% de todo o conjunto de periódicos em acesso aberto identificados neste trabalho não possuem valor de FI. Posteriormente, destaca-se que os outros estratos possuem a mesma quantidade de periódicos, com uma variação entre os seus valores de FI. O menor

valor identificado é 0,113, pertencente ao estrato Baixo, e o maior valor, que consequentemente está no estrato Alto, é igual a 23,333

Diante da estratificação e classificação dos periódicos por seus valores de FI nos outros três estratos considerados, uma análise importante a ser realizada se dá a respeito do perfil de publicações em cada um destes estratos, considerando as grandes áreas de atuação dos autores das publicação (Tabela 6.15).

Tabela 6.15 – Quantidade de Publicações nos Estratos por Grandes Áreas.

Grande Área	Nulo		Baixo		Médio		Alto		Total
Ciências Agrárias	119.351	11,49%	217.540	41,35%	20.973	9,78%	11.887	6,14%	369.751
	32,28%		58,83%		5,67%		3,21%		100%
Ciências Biológicas	75.222	7,24%	71.118	13,52%	46.822	21,84%	84.534	43,66%	277.696
	27,09%		25,61%		16,86%		30,44%		100%
Ciências da Saúde	316.275	30,46%	150.455	28,60%	115.566	53,89%	56.259	29,06%	638.555
	49,53%		23,56%		18,10%		8,81%		100%
Ciências Exatas e da Terra	62.458	6,01%	38.021	7,23%	22.272	10,39%	32.488	16,78%	155.239
	40,23%		24,49%		14,35%		20,93%		100%
Ciências Humanas	218.541	21,04%	14.673	2,79%	2.383	1,11%	2.298	1,19%	237.895
	91,86%		6,17%		1,00%		0,97%		100%
Ciências Sociais Aplicadas	146.726	14,13%	9.360	1,78%	819	0,38%	551	0,28%	157.456
	93,19%		5,94%		0,52%		0,35%		100%
Engenharias	42.489	4,09%	24.502	4,66%	5.513	2,57%	5.516	2,85%	78.020
	54,46%		31,40%		7,07%		7,07%		100%
Linguística, Letras e Artes	57.425	5,53%	364	0,07%	86	0,04%	65	0,03%	57.940
	99,11%		0,63%		0,15%		0,11%		100%
Total	1.038.487	100%	526.033	100%	214.434	100%	193.598	100%	1.972.552

É possível verificar que a grande área de Ciências da Saúde é responsável pela maioria dos artigos que foram publicados nos periódicos com FI igual a zero (estrato Nulo), e que aproximadamente 50% das publicações dessa grande área estão nesse estrato. Esse quantitativo é fortemente influenciado pela área de Medicina, com 96.674 publicações no Estrato Nulo.

Outra grande área que se destaca pelo número significativo de publicações no Estrato Nulo é a grande área de Ciências Humanas. Observa-se que 91,86% de suas publicações não estão em periódicos com valor de FI. Destaca-se, nessa grande área, a área de Educação, com 74.365 publicações em periódicos com FI igual a zero. Ressalta-se, ainda, a grande área de Linguística, Letras e Artes que, apesar de possuir a segunda menor quantidade de publicações no Estrato Nulo, tem aproximadamente 99% de suas publicações nesse estrato.

Já no Estrato Baixo, duas grandes áreas se destacam por possuírem juntas aproximadamente 70% das publicações em periódicos que foram classificados neste estrato. São elas as Grandes Áreas de Ciências Agrárias e Ciências da Saúde. Na Grande Área de Ciências Agrárias, que possui

58,83% de suas publicações nesse estrato, a área de Agronomia se destaca por ser responsável por 86.765 publicações. Já na Grande Área de Ciências da Saúde, com 23,56% de suas publicações no Estrato Baixo, há significativa participação de publicações da área de Medicina, com 75.565 trabalhos publicados.

No Estrato Médio, observa-se que a grande área de Ciências da Saúde possui aproximadamente 54% das publicações do estrato. Novamente, destaca-se a área de Medicina, com a maior participação nessas publicações, apresentando um quantitativo que de aproximadamente 57% do total de publicações da grande área de Ciências da Saúde.

Por fim, o Estrato Alto contém uma menor quantidade de publicações, tendo em vista que esse é o estrato com os periódicos com maiores valores de FI, e que, conseqüentemente, possui um processo de publicação mais competitivo. Nesse contexto, a grande área de Ciências Biológicas se destaca pela significativa quantidade de publicação em periódicos no Estrato Alto. Além de possuir aproximadamente 44% das publicações desse estrato, ela ainda possui 30,44% de suas publicações classificadas no estrato mais alto. Ressalta-se, ainda, que esse é o maior percentual de publicações dentre todos os outros, apresentando como uma tendência de suas áreas em publicar nos periódicos de maior impacto. Em análises que considerem as áreas mais representativas dessa grande área, três de suas áreas possuem mais de 10.000 publicações, sendo essas as áreas de Bioquímica (14.856), Genética (12.356) e Microbiologia (11.571).

Além do FI, que frequentemente é utilizado como indicador para qualificar um periódico, outro indicador amplamente utilizado para avaliar a qualidade de um determinado trabalho se refere à quantidade de citações que o trabalho recebe. Para análise da qualidade individual dos trabalhos considerados nesta tese, foi extraído da Web of Science (WOS), Scielo e Scopus o número de citações de cada um dos trabalhos publicados em periódicos de acesso aberto. A Figura 6.10 apresenta a média de citações dos artigos publicados em cada um dos anos nas três bases de dados consideradas.

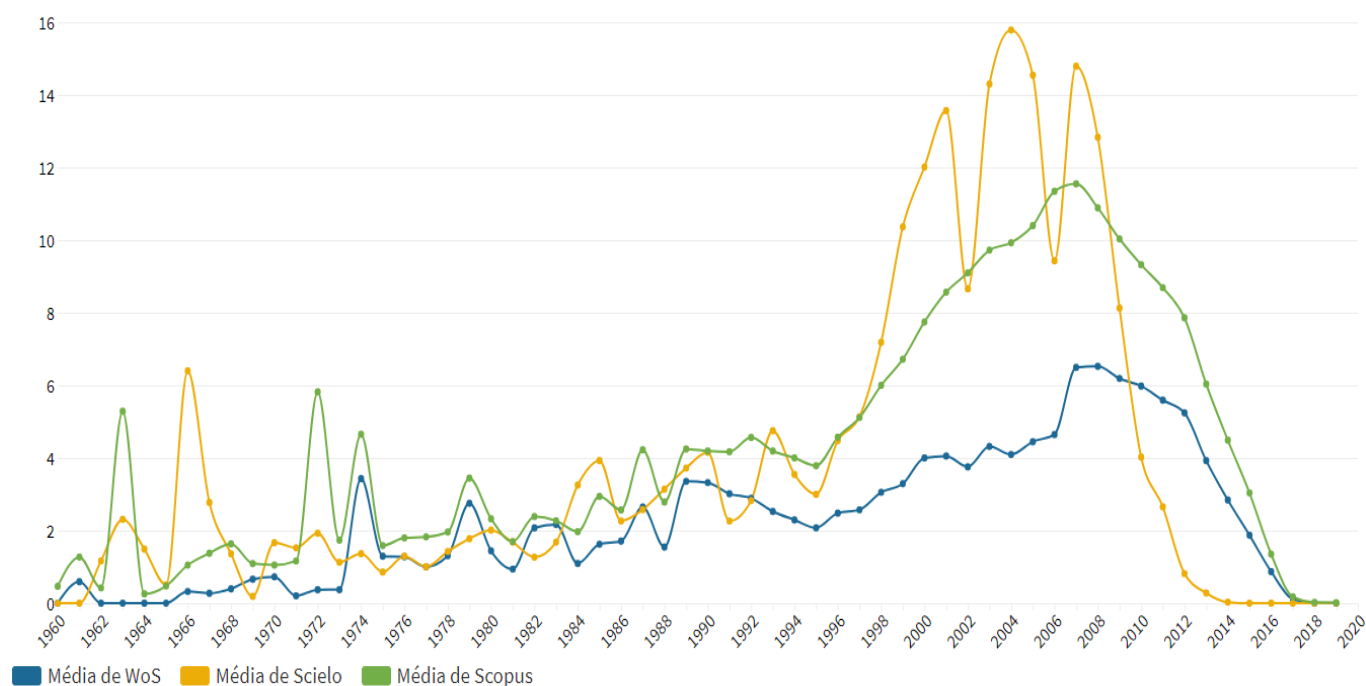


Figura 6.10 – Média de Citações dos Artigos Analisados.

Observa-se que a análise se inicia em 1960, ano da primeira publicação do conjunto selecionado. Em geral, a média de citações dos artigos possui uma variabilidade nos anos posteriores nas três bases de dados de citações consultados. Percebe-se um aumento constante nas médias de citações dos artigos a partir do final de década de 1970. De 1996 em diante, o crescimento da média de citações da Scielo se torna bem mais representativo que dos outros diretórios, e a média da WoS se mantém com crescimento mais modesto que os outros diretórios. Destaca-se que, nos artigos publicados nos últimos anos da análise, a média de citações tende a sofrer queda significativa, passando a ter valores médios muito próximos de zero a partir de 2017 nas três bases de dados, tendo em vista que trabalhos recentemente publicados tendem a não ter recebido citações.

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 o seu tempo de carreira ou a rede de colaboração na qual ele está inserido (DIAS, 2016). 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).

A grande discrepância na quantidade de citações recebidas pelos artigos publicados em periódicos de acesso aberto nas diversas áreas do conhecimento torna necessária uma análise mais detalhada, que possibilite compreender essa distinção. Logo, a Tabela 6.16 apresenta a média, a mediana e o desvio padrão desse indicador em cada uma das grandes áreas, considerando tanto os valores obtidos na WoS e Scielo, como também na Scopus. Além disso, os dados apresentados para cada grande área também levam em consideração a divisão dos periódicos em que os artigos foram publicados, conforme estratificação feita pelos seus FI apresentada anteriormente.

Tabela 6.16 – Mapa de Calor com as Médias, Medianas e Desvio Padrão de Citações.

Grande Área		MÉDIA					MEDIANA					DESVIO PADRÃO			
		NULO	BAIXO	MÉDIO	ALTO		NULO	BAIXO	MÉDIO	ALTO		NULO	BAIXO	MÉDIO	ALTO
Ciências Agrárias	WoS	0,02	1,77	1,00	5,34	WOS	0,00	0,00	1,00	0,00	WOS	0,46	4,23	8,90	15,00
	SCIELO	0,19	5,25	0,00	1,36	SCIELO	0,00	0,00	0,00	0,00	SCIELO	1,91	13,78	17,63	6,69
	SCOPUS	0,72	3,27	2,00	7,00	SCOPUS	0,00	1,00	2,00	1,00	SCOPUS	2,24	6,13	11,82	17,36
Ciências Biológicas	WoS	0,09	2,13	4,30	6,82	WOS	0,00	0,00	1,00	2,00	WOS	2,97	5,60	9,56	16,61
	SCIELO	0,27	4,28	2,93	1,23	SCIELO	0,00	0,00	0,00	0,00	SCIELO	2,73	11,98	10,27	7,18
	SCOPUS	1,32	4,14	6,50	8,66	SCOPUS	0,00	1,00	3,00	3,00	SCOPUS	5,90	8,78	13,38	19,05
Ciências da Saúde	WoS	0,03	1,83	2,58	5,41	WOS	0,00	0,00	0,00	1,00	WOS	0,56	5,67	7,73	13,87
	SCIELO	1,18	5,53	3,34	1,00	SCIELO	0,00	0,00	0,00	0,00	SCIELO	6,38	15,01	12,29	6,58
	SCOPUS	1,04	4,00	5,27	7,47	SCOPUS	0,00	1,00	1,00	2,00	SCOPUS	10,25	9,07	11,51	17,32
Ciências Exatas e da Terra	WoS	0,04	2,84	4,98	12,64	WOS	0,00	0,00	1,00	3,00	WOS	0,78	8,19	14,74	62,09
	SCIELO	0,24	2,69	1,71	0,17	SCIELO	0,00	0,00	0,00	0,00	SCIELO	2,52	9,72	8,27	2,75
	SCOPUS	0,54	4,34	6,49	13,56	SCOPUS	0,00	1,00	2,00	4,00	SCOPUS	2,59	10,59	17,02	56,09
Ciências Humanas	WoS	0,01	1,12	3,85	5,09	WOS	0,00	0,00	0,00	0,00	WOS	0,21	3,20	8,36	12,17
	SCIELO	0,60	6,22	5,02	1,97	SCIELO	0,00	0,00	0,00	0,00	SCIELO	4,68	16,25	16,48	9,16
	SCOPUS	0,25	2,73	6,02	7,49	SCOPUS	0,00	0,00	1,00	2,00	SCOPUS	1,39	6,15	11,18	15,20
Ciências Sociais Aplicadas	WoS	0,00	0,79	3,14	6,35	WOS	0,00	0,00	0,00	0,00	WOS	0,00	2,91	7,45	16,07
	SCIELO	0,00	2,66	5,59	0,70	SCIELO	0,00	0,00	0,00	0,00	SCIELO	0,00	10,79	17,50	5,19
	SCOPUS	0,00	1,66	4,90	7,61	SCOPUS	0,00	0,00	1,00	1,00	SCOPUS	0,00	5,07	9,83	18,09
Engenharias	WoS	0,02	1,70	2,81	5,96	WOS	0,00	0,00	0,88	0,00	WOS	0,36	4,52	4,24	72,16
	SCIELO	0,44	1,57	0,00	0,24	SCIELO	0,00	0,00	0,00	0,00	SCIELO	3,27	5,98	1,00	3,18
	SCOPUS	0,56	3,02	9,27	7,25	SCOPUS	0,00	0,00	5,71	0,00	SCOPUS	2,20	6,98	11,47	63,18
Linguística, Letras e Artes	WoS	0,00	0,58	2,15	5,62	WOS	0,00	0,00	0,00	1,00	WOS	52,58	2,37	7,55	11,92
	SCIELO	0,00	4,16	0,85	0,40	SCIELO	0,00	0,00	0,00	0,00	SCIELO	1,04	12,72	3,90	2,56
	SCOPUS	0,00	2,10	4,02	8,35	SCOPUS	0,00	0,00	0,00	1,00	SCOPUS	56,10	6,59	10,02	15,98
GLOBAL	WoS	0,03	1,85	3,33	7,19	WOS	0,00	0,00	0,00	1,00	WOS	11,96	5,16	9,19	31,1
	SCIELO	0,57	4,71	3,33	0,95	SCIELO	0,00	0,00	0,00	0,00	SCIELO	4,37	13,31	12,06	6,32
	SCOPUS	0,62	3,57	5,67	8,9	SCOPUS	0,00	1,00	2,00	2,00	SCOPUS	14,03	7,77	12,50	29,9

Como poder 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 em WoS, Scielo e Scopus, sendo que a média da Scopus é em geral superior nos estratos e nas grandes áreas analisadas.

De modo geral, percebe-se que a média de citações é maior nos estratos mais altos, com valores sendo, em geral, bem superiores em relação a todos os outros. Já os estratos mais baixos recebem um número muito baixo de citações. Um fato identificado é que no Estrato Nulo, que concentra os periódicos com valor de FI igual a zero, há valores médios de citações abaixo de um, em praticamente toda a sua totalidade.

A grande área com melhor desempenho de citações no Estrato Alto é a de Ciências Exatas e da Terra, e em seguida a grande área de Ciências Biológicas, com valores médios um pouco superiores ao da grande área de Linguística, Letras e Artes. É válido destacar, ainda, no Estrato Alto os valores do desvio padrão das grandes áreas de Engenharias e Ciências Exatas e da Terra, que possuem os maiores valores. Uma hipótese para explicar tal fato está relacionada a artigos altamente citados.

Já ao analisar as médias do Estrato Nulo, observa-se que as grandes áreas de Ciências Sociais Aplicadas e Linguística, Letras e Artes possuem média igual a zero em todos os diretórios considerados. As grandes áreas de Ciências da Saúde e Ciências Biológicas são as que possuem as melhores médias do estrato.

Foi possível perceber que vários trabalhos possuem uma quantidade muito grande de citações, o que impacta diretamente nas médias dos artigos em diversas das grandes áreas analisadas. Logo, ao considerar a mediana da quantidade de citações, foi possível identificar que aproximadamente 73% das medianas — de todas as grandes áreas, em todos os estratos e considerando os três diretórios — possuem valor igual a zero. Destaca-se, ainda, a grande área de Engenharia, com mediana igual a 5,71 no Estrato Médio.

O principal motivo da grande quantidade de medianas igual a zero se deve ao fato de que uma grande quantidade dos artigos publicados em periódicos de acesso aberto não possui citações, resultando em medianas com valores nulos. A Tabela 6.17 apresenta o percentual de artigos sem citações em cada grande área.

Tabela 6.17 – Percentual de Publicações sem Citações.

Grande Área	WOS	Scielo	Scopus	WOS+Scielo+Scopus
Ciências Agrárias	73,53%	81,72%	57,48%	54,31%
Ciências Biológicas	63,38%	88,19%	50,82%	48,87%
Ciências da Saúde	80,44%	86,43%	61,10%	58,73%
Ciências da Exatas e da Terra	66,91%	93,29%	58,47%	56,45%
Ciências Humanas	97,07%	94,66%	89,12%	87,53%
Ciências Sociais Aplicadas	98,17%	97,46%	93,04%	91,93%
Engenharias	82,42%	92,76%	69,72%	67,93%
Linguística, Letras e Artes	99,61%	99,46%	97,51%	97,17%

É possível observar que em todas as grandes áreas analisadas e em todos os diretórios o percentual de publicações sem nenhuma citação está acima de 50%, corroborando os valores das medianas apresentados na Tabela 6.16. O menor percentual de publicações sem nenhuma citação é da Scopus, sendo os outros diretórios detentores de percentuais bem mais representativos. A grande área de Linguística, Letras e Artes se destaca com os maiores percentuais de publicações sem citações, com ênfase para os diretórios WoS e Scielo, em que os percentuais são acima de 99% de suas publicações. Em seguida, também pode-se destacar as grandes áreas de Ciências Sociais Aplicadas e Ciências Humanas com percentuais bem representativos.

A grande área com a menor quantidade de publicações sem citação é a de Ciências Biológicas. Ela possui o menor percentual na WoS e Scopus, e o terceiro menor na Scielo. Ao se considerar conjuntamente os três diretórios (coluna: WOS+Scielo+Scopus), é verificado que 48,87% de suas publicações não possuem citações em nenhum dos diretórios analisados, única grande área com percentual abaixo de 50% nesta avaliação que considera os três diretórios em conjunto.

7. CONSIDERAÇÕES FINAIS

Neste capítulo são apresentadas as considerações finais da tese, destacando-se a sumarização das principais contribuições (Seção 7.1). Tendo em vista essas contribuições, são apresentadas as publicações resultantes do trabalho realizado, bem como, as submissões em andamento (Seção 7.2). Por fim, são discutidas as possíveis extensões para trabalhos futuros (Seção 7.3).

7.1 Considerações

A publicação de artigos em periódicos de acesso aberto tem obtido o interesse de toda a comunidade científica, tendo em vista o crescimento da quantidade de periódicos nesse formato, bem como a reconhecida relevância que os trabalhos nesse meio de publicação têm adquirido. Aliado a isso, acrescenta-se o interesse da comunidade científica em verificar os esforços realizados para impulsionar a publicação de artigos nesse meio de publicação, principalmente devido aos custos financeiros envolvidos no acesso a pesquisas que geralmente são financiadas com recursos públicos. Por esses motivos, é importante compreender como estão evoluindo as publicações nesse formato de divulgação.

Com o intuito de traçar um retrato sobre a publicação de artigos em periódicos de acesso aberto pelos pesquisadores brasileiros, foi necessário o desenvolvimento de componentes que, incorporados ao *LattesDataXplorer*, pudessem viabilizar as análises realizadas neste trabalho. Assim, todo o repositório de dados curriculares da Plataforma Lattes foi analisado, possibilitando um estudo amplo e inédito sobre a produção brasileira de artigos em periódicos de acesso aberto, utilizando também dados do DOAJ como validador.

Ao analisar o perfil dos pesquisadores que têm publicado artigos em periódicos de acesso aberto, percebe-se que a maioria desses indivíduos tem elevados níveis de formação acadêmica, com mestrado e/ou doutorado concluídos. Destaca-se que um percentual significativo de indivíduos se concentra em determinadas áreas de atuação, como Medicina, Educação, Agronomia e Enfermagem, o que pode representar uma tendência na forma de divulgação dos resultados de pesquisas nessas áreas.

Observou-se, ainda, que os autores de artigos em periódicos de acesso aberto estão concentrados nos estados das regiões Sudeste e Sul do país, e que preferencialmente publicam os resultados de suas pesquisas em anais de congressos e periódicos. Esse resultado corrobora a importância do presente estudo, que visa compreender a evolução da publicação de artigos em periódicos de acesso aberto pela comunidade científica brasileira.

Além disso, foi possível identificar como tem evoluído temporalmente, ao longo dos últimos anos, a quantidade de trabalhos publicados em periódicos de acesso aberto, podendo ser observado como a produção de artigos nesse meio de publicação vem aumentando de forma significativa, principalmente nos últimos anos. Considerando o último ano analisado nesta tese, foi possível identificar, diante de todo o quantitativo de artigos em periódicos registrados nos

currículos da Plataforma Lattes, que aproximadamente 27% dos artigos estão em periódicos de acesso aberto.

Quando se analisa individualmente a produção científica dos pesquisadores que integram o conjunto estudado, percebe-se que os pesquisadores com elevada produção científica em periódicos de acesso aberto são vinculados a universidades públicas, e que em sua maioria são pesquisadores experientes, com significativa quantidade de orientações concluídas nos níveis mais altos de formação acadêmica. Estes contam, ainda, com reconhecida relevância científica entre seus pares, sendo em geral pesquisadores de Produtividade em Pesquisa do CNPq.

Como resultado da produção científica elevada desses autores altamente produtivos, algumas áreas de atuação a que eles estão vinculados acabam por ter sua produtividade aumentada. Destaca-se que em alguns casos os pesquisadores mais produtivos são responsáveis por alavancar a produção científica de todo um conjunto analisado, já que a maioria dos indivíduos possui poucas publicações: conforme apresentado, cerca de 70% dos autores com artigos em periódicos de acesso aberto possuem no máximo três publicações nesse formato.

Tendo em vista os autores altamente produtivos, o que exerce influência sobre a produtividade de algumas áreas, e considerando a disparidade de indivíduos que compõem cada uma dessas áreas, uma análise sobre a produção *per capita* foi realizada. Nessa análise, considerou-se a quantidade de publicações dos autores de cada área, pela quantidade de autores que ela possui. Logo, foi possível perceber algumas áreas com valores de publicações *per capita* significativos, como as áreas de Zootecnia (11,95) e Engenharia Agrícola (10,66), ambas vinculadas à grande área de Ciências Agrárias. Esta encontra-se com a segunda maior quantidade de publicações, e é apenas a quarta com a maior quantidade de autores. Por outro lado, é possível destacar as áreas de Engenharia Naval e Oceânica (2,14) e Artes (2,74) com os menores valores de produção *per capita*.

Foi possível, ainda, perceber que a colaboração científica do conjunto analisado tem evoluído constantemente ao longo dos últimos anos, principalmente a partir do ano 2000, e que essa colaboração é influenciada principalmente por fatores como área de atuação, localização geográfica e tópicos de pesquisa dos autores.

Tendo em vista análises que considerem a adoção de métricas bibliométricas no conjunto de dados selecionados, foi possível avaliar a relevância dos periódicos com a Lei de Bradford, a produtividade dos autores com a Lei de Lotka, os principais tópicos de pesquisa com a Lei de Zip. Além disso, realizou-se uma análise considerando fatores qualitativos dos periódicos e seus artigos publicados.

No intuito de melhor compreender o conjunto de periódicos em acesso aberto em que os autores brasileiros têm publicado os resultados de suas pesquisas, uma análise com a adoção da Lei de Bradford foi realizada. Foram considerados inicialmente todos os 9.202 periódicos encontrados nos currículos da Plataforma Lattes e validados pelo DOAJ. Esse conjunto foi dividido em zonas, conforme propõe a lei. Após a divisão dos periódicos em zonas, identificou-se que o núcleo, composto pelos principais periódicos do conjunto, possui um total de apenas seis periódicos altamente representativos. Esse conjunto, por si só, possui um total de 134.946

artigos publicados. Destaca-se a PLoS One, que possui um total de 35.508 artigos. Considerando os periódicos que compõem a Zona 1, destaca-se que em sua maioria são periódicos da grande área de Ciências da Saúde, a exemplo do que também acontece com os periódicos da Zona 2.

Levando em conta os trabalhos que foram publicados nos periódicos que compõem o Núcleo, é possível destacar alguns autores com uma grande quantidade de publicações. Ao considerar o conjunto com os autores mais representativos com publicações no Núcleo, destacam-se autores da área de Medicina.

Ao realizar uma análise sobre a produtividade dos autores com a Lei de Lotka, vista como uma função de probabilidade da produção, tal lei foi aplicada para analisar a produtividade dos autores em periódicos de acesso aberto de todo conjunto selecionado. Observou-se que apenas um autor possui 544 publicações em periódicos de acesso aberto, e que aproximadamente 63% do conjunto de autores possuía no máximo três contribuições.

De acordo com o modelo proposto no trabalho de Alvarado (2006), utilizado nesta tese, foi calculada a distribuição teórica da produtividade dos autores, e em seguida, testou-se a qualidade do ajuste dos dados selecionados com o Teste de Kolmogorov-Smirnov (K-S), para determinar a aceitação ou rejeição da distribuição. Enfim, após fazer o teste de K-S, foi possível concluir que a distribuição da produção dos autores que publicam em periódicos de acesso aberto não se aplica à Lei de Lotka.

A respeito de uma análise sobre os principais tópicos de pesquisa do conjunto de publicações selecionadas, extraíram-se as palavras dos títulos de cada um dos artigos. A razão da seleção da estratégia de se obter as palavras dos títulos das publicações se deu em razão da baixa quantidade de artigos com palavras-chave vinculadas. Foi possível observar que apenas 17% dos artigos publicados em periódicos de acesso aberto nos currículos da Plataforma Lattes possuíam palavras-chave. Logo, todos os títulos foram analisados e, dessa forma, um conjunto de termos foi caracterizado para a aplicação da Lei de Zipf.

Após o tratamento de todo o conjunto de dados, identificou-se um quantitativo global de 28.636.958 de tópicos. Posteriormente, removeram-se as duplicatas, reduzindo o conjunto a 423.364 tópicos únicos. Em seguida foram removidas todas as *StopWords*, diminuindo o conjunto para um total de 393.896 palavras para objeto de estudo. Logo, foi possível observar que dentre os tópicos mais frequentes, a maioria está nos idiomas português e inglês. O tópico mais frequente é o Brasil, com uma frequência de 88.485 citações. Apesar de todo o tratamento realizado, também foi possível identificar diversos tópicos que possuem apenas uma citação, caracterizando possíveis registros inconsistentes.

Segundo as referências utilizadas nesta tese, ao classificar o conjunto de tópicos em três zonas textuais de distribuição, identificou-se que 2,21% dos tópicos representam informação trivial, 13,30% informação interessante e 84,49% ruído. Importante destacar que as palavras não passaram por um processo de tradução e radicalização, o que poderia influenciar os resultados obtidos.

Por fim, ao realizar-se uma análise que considerou métricas tradicionais de qualidade, os artigos publicados em periódicos de acesso aberto cadastrados nos currículos da Plataforma Lattes foram considerados, tendo em vista as suas citações na WoS, Scielo e Scopus, bem como o FI dos periódicos. O FI é considerado informação valiosa para análises qualitativas, portanto, ao avaliar esse índice dos periódicos em acesso aberto, foi observada grande distinção em seus valores, com variação entre 0 e 23,333. Diante disso, ao realizar uma análise temporal dos artigos, considerando-se a mediana dos FI's de todos os periódicos, identificou-se que o ápice se deu no ano 2000, com valor de mediana igual a 0,5340. Destaca-se que em 2010 o valor da mediana foi igual a zero, e permaneceu o mesmo até o último ano analisado nesta tese (2019). Uma hipótese para esse valor de mediana igual a zero a partir do ano de 2010 pode estar relacionada à grande quantidade de artigos registrados nos currículos a partir deste ano, os quais, na maioria das publicações, foram publicados em periódicos de FI nulo, ou seja, com valor igual a zero. Do conjunto total de periódicos analisados, aproximadamente 82% possui FI igual a zero.

Categorizando os periódicos por estratos que considerem os valores de FI, percebe-se que as publicações de artigos da grande área de Ciências da Saúde possuem aproximadamente 50% de suas publicações no estrato Nulo, que contém periódicos com FI igual a zero. Por outro lado, no estrato mais alto, que contém os periódicos com os maiores valores de FI, destaca-se a grande área de Ciências Biológicas, com aproximadamente 30% de suas publicações nesse estrato. Destacam-se ainda as áreas de Bioquímica, Genética e Microbiologia com quantitativos significativos nesse estrato.

Com relação às citações recebidas pelos artigos analisados, percebe-se um aumento na média de citações dos trabalhos publicados a partir do final da década de 1970, sendo tal crescimento intensificado no final da década de 1990. Posteriormente, destaca-se uma queda significativa na média de citações a partir de 2008, em todos os diretórios considerados. Importante ainda destacar que, a partir de 2017, em todos os diretórios os valores médios de citações estão próximos de zero, tendo em vista que trabalhos recentemente publicados podem ainda não terem recebido quantidades significativas de citações.

A análise de média das citações sofre influência direta de artigos altamente citados, o que acaba por elevar tais médias, principalmente em determinadas áreas. Isso se dá porque, mesmo tendo muitos artigos sem citações, esses trabalhos altamente citados acabam influenciando nos resultados. Conforme apresentado, mais de 50% das publicações em todas as grandes áreas e nos três diretórios consultados não possuem citações.

Ao considerar uma análise comparativa das citações dos artigos nas diversas áreas do conhecimento, fica evidente uma grande discrepância entre as áreas. Além disso, foi possível observar que os artigos publicados em periódicos com maior valor de FI tendem a receber uma maior quantidade de citações, resultado da credibilidade e visibilidade que tais periódicos proporcionam. Em geral, ao se publicar em periódicos com valor de FI significativo, o trabalho tende a possuir maior repercussão, em detrimento da maioria das publicações que são publicadas em periódicos com FI nulo.

Nesse contexto, a partir da metodologia utilizada — a saber, o cruzamento entre dados do DOAJ e da Plataforma Lattes —, foi possível mapear e conhecer de maneira inédita as características dos autores brasileiros que estão publicando artigos em periódicos de acesso aberto nas diferentes áreas do conhecimento. Além disso, foi possível identificar a distribuição geográfica e os níveis de formação dos autores, a distribuição das publicações por áreas de atuação e a maneira como tem ocorrido a colaboração científica nesses artigos.

Além disso, com a adoção das principais métricas bibliométricas, foi possível traçar um retrato dos periódicos de maior relevância no conjunto analisado, da produtividade dos autores com publicações em periódicos de acesso aberto, dos principais tópicos de pesquisa analisados e de alguns indicadores qualitativos.

Por fim, a validade dos resultados da abordagem aqui proposta permite sua utilização em diversos tipos de novas contribuições, que poderão servir-se dos resultados aqui apresentados como motivação e fonte para novas análises.

7.2 Publicações

A seguir, são apresentados os artigos que foram produzidos a partir dos primeiros resultados de pesquisa deste projeto.

Como resultado das pesquisas iniciais realizadas a partir dos dados curriculares da Plataforma Lattes e do DOAJ, foram produzidos os seguintes artigos que foram apresentados na IX Conferência Luso-Brasileira sobre Acesso Aberto – CONFOA 2018, realizada em outubro de 2018 em Lisboa:

DIAS, T. M. R.; LUCAS, E. R. O.; DIAS, P. M. Análise da Produção Científica Brasileira em Periódicos de Acesso Aberto (CONFOA 2018).

DIAS, T. M. R. *et al.* Adoção de Repositórios de Dados Governamentais para Obtenção de Dados Científicos (CONFOA 2018).

Como consequência dos artigos apresentados na CONFOA 2018, os autores receberam um convite para envio de uma versão estendida dos artigos para o periódico Cadernos BAD. Os artigos foram publicados no n. 1 (2018):

DIAS, T. M. R. *et al.* Obtenção de dados científicos a partir de repositórios de dados curriculares. **Cadernos de Biblioteconomia, Arquivística e Documentação**, Lisboa, v. 1, n. 1, p.86-93, out. 2018. (Qualis: Comunicação e Informação – B2, Interdisciplinar – B3)

DIAS, T. M. R. *et al.* Uma visão geral da produção em periódicos de acesso aberto no Brasil. **Cadernos de Biblioteconomia, Arquivística e Documentação**, Lisboa, v. 1, n. 1, p.326-333, out. 2018. (Qualis: Comunicação e Informação – B2, Interdisciplinar – B3)

Posteriormente, um novo trabalho descrevendo o perfil dos autores com publicações em periódicos de acesso aberto foi enviado e publicado no v. 13, n. 2 (2018) do periódico Pesquisa Brasileira em Ciência da Informação e Biblioteconomia:

DIAS, P. M. *et al.* Produção científica em periódicos de acesso aberto no Brasil: uma caracterização dos autores. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, João Pessoa, v. 13, n. 2, p.136-145, nov. 2018. (Qualis: Comunicação e Informação – B1, Interdisciplinar – B2).

Com a inclusão de uma análise preliminar dos artigos publicados em periódicos de acesso aberto no Brasil e das colaborações científicas de seus autores, outros estudos foram apresentados no IX Encuentro Ibérico de la Asociación de Educación e Investigación en Ciencia de la Información de Iberoamérica y el Caribe (EDICIC), realizado em Barcelona, em julho de 2019, cujo tema foi “*Datos abiertos e inclusión digital en la era del Big Data*”. Foram apresentadas seguintes contribuições:

DIAS, P. M.; *et al.* Publicações em Periódicos de Acesso Aberto no Brasil: Uma Análise Temporal (Comunicación EDICIC - 2019).

DIAS, P. M.; *et al.* Uma Visão da Comunidade Científica Brasileira com Publicações em Periódicos de Acesso Aberto (Comunicación EDICIC - 2019).

DIAS, P. M.; *et al.* Uma Plataforma para a Extração, Integração e Análises de Dados Científicos em Acesso Aberto (Comunicación EDICIC - 2019).

Completando os resultados iniciais obtidos na pesquisa e apresentados no EDICIC foram aceitos para apresentação na X Conferência Luso-Brasileira de Ciência Aberta (CONFOA 2019) realizada em Manaus em outubro de 2019, os seguintes trabalhos:

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Uma Análise da Colaboração Científica dos Autores com Publicação em Periódicos de Acesso Aberto (CONFOA 2019).

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Publicações em Acesso Aberto no Brasil: Uma Caracterização do Conjunto de Autores (CONFOA 2019).

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Evolução da Produção Científica em Periódicos de Acesso Aberto no Brasil: Uma Análise Temporal (CONFOA 2019).

Como consequência da apresentação dos trabalhos na CONFOA 2019, versões estendidas dos trabalhos foram publicadas no periódico Ciência da Informação (ISSN 1518-8353) - (Qualis: Comunicação e Informação – B1, Interdisciplinar – B1):

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Evolução da Produção Científica em Periódicos de Acesso Aberto no Brasil: Uma Análise Temporal. **Ciência da Informação**, [S. l.], v. 48, n. 3, 2020.

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Publicações em Acesso Aberto no Brasil: Uma Caracterização do Conjunto de Autores. **Ciência da Informação**, [S. l.], v. 48, n. 3, 2020.

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Uma Análise da Colaboração Científica dos Autores com Publicações em Periódicos de Acesso Aberto. **Ciência da Informação**, [S. l.], v. 48, n. 3, 2020.

Com a inclusão de novas análises aos resultados já alcançados, foram publicados no II Congresso de Altmatria e Ciência Aberta na América Latina (LATmetrics 2019) outros artigos:

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. PUBLICAÇÕES EM PERIÓDICOS DE ACESSO ABERTO NO BRASIL: uma análise bibliométrica. In: II Congresso de Altmatria e Ciência Aberta na América Latina (LATmetrics 2019) – Cusco, 2019.

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. UM RETRATO DA COMUNIDADE CIENTÍFICA BRASILEIRA COM PUBLICAÇÕES EM ACESSO ABERTO. In: II Congresso de Altmatria e Ciência Aberta na América Latina (LATmetrics 2019) – Cusco, 2019.

Com o avanço do estudo, em especial sobre o processo de identificação de publicações em periódicos de acesso aberto, foi publicado no 7º Encontro Brasileiro de Bibliometria e Cientometria (7º EBBC 2020) o seguinte trabalho:

DIAS, P. M.; *et al.* Uma estratégia para identificação de artigos em periódicos de acesso aberto na plataforma lattes. In: 7º Encontro Brasileiro de Bibliometria e Cientometria (7º EBBC 2020), p. 71-78, Salvador, 2020.

Como consequência do trabalho apresentado no 7º EBBC, uma versão estendida do artigo foi submetida para a edição especial do periódico Encontros Bibli (ISSN 1518-2924) - (Qualis: Comunicação e Informação – A2, Interdisciplinar – B3):

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Uma estratégia para identificação e análise de publicações científicas em periódicos de acesso aberto. (Edição Especial 7º EBBC). **Em avaliação**.

Com a realização de novas análises no conjunto de dados e avanço na pesquisa, novos trabalhos foram realizados e aceitos para apresentação na XI Conferência Luso-Brasileira de Ciência Aberta (CONFOA 2020) realizada em outubro de 2020, sendo eles:

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Perfil dos Autores Brasileiros com Publicações Científicas em Periódicos de Acesso Aberto (CONFOA 2020).

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Tendência de Publicação em Periódicos de Acesso Aberto no Brasil: Uma Abordagem Quantitativa (CONFOA 2020).

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Periódicos de Acesso Aberto: Uma Análise Bibliométrica da Produção Científica Brasileira (CONFOA 2020).

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. Uma Estratégia para a Identificação de Publicações em Periódicos de Acesso Aberto Cadastradas na Plataforma Lattes (CONFOA 2020).

Como consequência da apresentação dos trabalhos na CONFOA 2020, versões estendidas dos trabalhos foram publicadas no periódico Páginas A&B. Arquivos & Bibliotecas (ISSN 0873-5670) - (Qualis: Comunicação e Informação – B4, Interdisciplinar – B4):

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. PERFIL DOS AUTORES BRASILEIROS COM PUBLICAÇÕES CIENTÍFICAS EM PERIÓDICOS DE ACESSO ABERTO. **Páginas a&b: arquivos e bibliotecas**, p. 150-154, 2021.

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. TENDÊNCIA DE PUBLICAÇÃO EM PERIÓDICOS DE ACESSO ABERTO NO BRASIL: uma abordagem quantitativa. **Páginas a&b: arquivos e bibliotecas**, p. 159-163, 2021.

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. PERIÓDICOS DE ACESSO ABERTO: uma análise bibliométrica da produção científica brasileira. **Páginas a&b: arquivos e bibliotecas**, p. 253-254, 2021.

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. UMA ESTRATÉGIA PARA A IDENTIFICAÇÃO DE PUBLICAÇÕES EM PERIÓDICOS DE ACESSO ABERTO CADASTRADAS NA PLATAFORMA LATTES. **Páginas a&b: arquivos e bibliotecas**, p. 238-239, 2021.

Um detalhamento sobre todo o ferramental proposto para a identificação e análise dos dados utilizados nesta tese foi apresentado no 2º EAI International Conference on Data and Information in Online Environments (DIONE 2021):

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. A Strategy for the Identification of Articles in Open Access Journals in Scientific Data Repositories (DIONE 2021).

O trabalho apresentado no DIONE 2021, foi publicado como um capítulo do livro Data and Information in Online Environme, 2021, ISBN: 978-3-030-77417-2:

DIAS, P. M.; DIAS, T. M. R.; MOITA, G. F. A Strategy for the Identification of Articles in Open Access Journals in Scientific Data Repositories. In: International Conference on Data and Information in Online. Springer, Cham, 2021. p. 125-135.

Além disso, uma versão estendida do trabalho que foi apresentado no DIONE 2021, foi a convite, submetido a edição especial do periódico Transinformação (ISSN 0103-3786) - (Qualis: Comunicação e Informação – A1, Interdisciplinar – A2):

DIAS, P. M.; *et al.* Extraction and Validation of Scientific Data for the Identification of Publications in Open Access Journals. (Edição Especial DIONE 2021). **Em avaliação.**

Além destes trabalhos já apresentados e publicados no âmbito desta tese, ainda estão aceitos para publicação em eventos que irão ocorrer ao longo de 2021:

XII CONFERÊNCIA LUSO-BRASILEIRA DE CIÊNCIA ABERTA (CONFOA 2021) – Híbrido, 2021:

- Uma Análise Bibliométrica de Periódicos em Acesso Aberto: Abordagem Baseada em Áreas do Conhecimento.
- Um Estudo Baseado em Citações dos Artigos Publicados em Periódicos de Acesso Aberto.
- Avaliando o Fator de Impacto de Periódicos em Acesso Aberto: Uma Abordagem Quantitativa.

XII Encuentro EDICIC – Costa Rica, 2021:

- Colaboração Científica Institucional: Uma Análise Baseada em Publicações de Periódicos em Acesso Aberto.

Latmétricas 2021 (LATmetrics y Simposio Latinoamericano sobre Estudios Métricos en Ciencia y Tecnología):

- Identificando os Principais Tópicos de Pesquisa em Periódicos de Acesso Aberto: Uma Análise com Métricas Bibliométricas.

XXI Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação - XXI ENANCIB:

- Indicadores da produção científica brasileira em periódicos de acesso aberto.

7.3 Trabalhos Futuros

Com a proposta de explorar todo o potencial do conjunto de dados analisados nesta tese e dos resultados até então inéditos aqui apresentados, são sugeridos os seguintes trabalhos futuros:

- Comparação dos resultados obtidos nesta tese, considerando o conjunto de artigos publicados em periódicos de acesso aberto, com os artigos publicados nos outros periódicos que não estão no formato aberto. Tal comparação poderia avaliar a similaridade da evolução temporal, as características dos autores e sua produtividade, os periódicos mais relevantes, além de indicadores qualitativos.
- Adoção da Lei de Lotka no conjunto de dados analisados, considerando janelas temporais de análise, com o objetivo de verificar se a partir de um determinado período a Lei se aplica. Tal análise poderia evitar possíveis ruídos que o conjunto de dados pode apresentar, principalmente nos primeiros anos de análise.
- Verificação do conjunto de artigos publicados em periódicos de acesso aberto e sua representatividade em repositórios de publicações institucionais. Tal análise poderia verificar a visibilidade dos artigos que já estão disponibilizados de forma livre nos repositórios das instituições em que os autores estão vinculados.
- Avaliação da utilização de identificadores persistentes nas publicações, como o DOI (*Digital Object Identifier*), para que, dessa forma, seja possível avaliar o percentual de artigos com tais identificadores, bem como o perfil dos periódicos e dos autores que têm adotado tais elementos no registro de suas publicações. Esse tipo de análise seria importante em repositório de dados como a Plataforma Lattes, tendo em vista que viabiliza a recuperação automática do texto do trabalho.
- Identificação dos conjuntos de dados públicos disponíveis em repositórios de dados vinculados aos artigos analisados. No contexto da Ciência Aberta, em que os dados deveriam ser disponibilizados de forma livre junto aos resultados alcançados na pesquisa, espera-se que os autores que publicam artigos em periódicos de acesso aberto também forneçam os dados utilizados na pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, E. C. E.; GUIMARÃES, J. A.; ALVES, I. T. G. Dez anos do Portal de Periódicos da CAPES: histórico, evolução e utilização. **Revista Brasileira de Pós-graduação**, Brasília, v. 7, n. 13, p.218-246, 2010. Quadrimestral. Disponível em: <<http://ojs.rbpg.capes.gov.br/index.php/rbpg/article/view/194/188>>. Acesso em: 12 dez. 2018.
- ALVARADO, R. U. A Produtividade Dos Autores Na Literatura De Enfermagem Um Modelo De Aplicação Da Lei De Lotka. **Informação & Sociedade: Estudos**, João Pessoa, v. 16, n. 1, p. 63–78, 2006.
- ALVARADO, R. U. Elistismo na literatura sobre a produtividade. **Ciência da Informação**, Brasília, v. 2, n. 38, p. 69-79, ago. 2009. Quadrimestral.
- ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. LattesMiner. **Proceedings Of The Compilation Of The Co-located Workshops On Dsm'11, Tmc'11, Agere!'11, Aoopes'11, Neat'11, & Vmil'11 - Splash '11 Workshops**, [s.l.], p.85-92, 2011. ACM Press. <http://dx.doi.org/10.1145/2095050.2095065>.
- ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. LattesMiner: a multilingual DSL for information extraction from Lattes Platform. **Proceedings Of The Compilation Of The Co-located Workshops On Dsm'11, Tmc'11, Agere!'11, Aoopes'11, Neat'11, & Vmil'11 - Splash '11 Workshops**, [s.l.], p.85-92, 2011. ACM Press. <http://dx.doi.org/10.1145/2095050.2095065>.
- ANDRADE, R. M.; MURIEL-TORRADO, E. Declarações de Acesso Aberto e a Lei de Direitos Autorais brasileira. **RECIIS: Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, [s.l.], v. 11, p.1-5, 30 nov. 2017. Instituto de Comunicação e Informação Científica e Tecnológica em Saúde. <http://dx.doi.org/10.29397/reciis.v11i0.1374>.
- ANTELMAN, K. Do Open-Access Articles Have a Greater Research Impact? **College & Research Libraries**, [s.l.], v. 65, n. 5, p.372-382, 1 set. 2004. American Library Association. <http://dx.doi.org/10.5860/crl.65.5.372>. Disponível em: <<https://crl.acrl.org/index.php/crl/article/view/15683/17129>>. Acesso em: 05 jan. 2019.
- ANDRÉ, F. **Libre Accès aux Savoirs/Open Access to Knowledge**. Paris: Futuribles, 2005. 72 p.
- ARAÚJO, C. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, Porto Alegre, v. 79, n. 3, p.11-32, 2006.
- ARRUDA, D. *et al.* Brazilian computer science research: Gender and regional distributions. **Scientometrics**, [s.l.], v. 79, n. 3, p.651-665, 31 jan. 2009. Springer Nature. <http://dx.doi.org/10.1007/s11192-007-1944-0>.
- BJÖRK, B.; SOLOMON, D. Article processing charges in OA journals: relationship between price and quality. **Scientometrics**, [s.l.], v. 103, n. 2, p.373-385, 13 mar. 2015. Springer Nature. <http://dx.doi.org/10.1007/s11192-015-1556-z>.
- BOAI (**Budapest Open Access Initiative**). Dez anos da Iniciativa de Budapeste em Acesso Aberto: a abertura como caminho a seguir. 2012. Disponível em: <http://www.budapestopenaccessinitiative.org/boai-10-translations/portuguese-brazilian-translation>. Acesso em: 20 ago. 2021.

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, Brasília. **Anais...** . Brasília: Brasnam, 2014. p. 1 - 12.

BONGIOVANI, P.; GÓMEZ, N. D.; MIGUEL, S. Opiniones y hábitos de publicación en acceso abierto de los investigadores argentinos. Un estudio basado en los datos de la encuesta SOAP. **Revista Española de Documentación Científica**, [s.l.], v. 35, n. 3, p.453-467, 1 ago. 2012. Departamento de Publicaciones del CSIC. <http://dx.doi.org/10.3989/redc.2012.3.903>. Disponível em: <<http://redc.revistas.csic.es/index.php/redc/article/view/752/833>>. Acesso em: 18 out. 2018.

BRITO, A. G. C. de; QUONIAM, L.; MENA-CHALCO, J. P. Exploração da Plataforma Lattes por assunto: proposta de metodologia. **Transinformação**, [s.l.], v. 28, n. 1, p.77-86, abr. 2016. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/2318-08892016002800006>.

BRODY, T. *et al.* The effect of Open Access on Citation Impact. In: SOUTHAMPTON UNIVERSITY, 1., 2004, Southampton Uk. **Anais...** . Southampton Uk: Southampton University, 2004. p. 1 - 2. Disponível em: <<http://opcit.eprints.org/feb19prog.html>>. Acesso em: 10 dez. 2018.

BRUNI, A. L. **Estatística aplicada à gestão empresarial**. 4. ed. [s.l.] : Atlas, 2013.

CAPARELLI, L. L.; DIGIAMPIETRI, L. A. Combinando Análise Bibliométrica e Análise de Redes Sociais para a Avaliação de Grupos Acadêmicos. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING (BRASNAM_CSBC), 7. , 2018, 1/2018. 7º Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2018). Porto Alegre: Sociedade Brasileira de Computação, July 2018 . ISSN 2595-6094.

CAPES: O Portal de Periódicos da Capes. Disponível em: <www.periodicos.capes.gov.br/index.php?option=com_pcontent&view=pcontent&alias=missao-objetivos&Itemid=109>. Acesso em: 12 nov. 2018.

CAPES. **Orçamento - Evolução em reais**. Disponível em: <<https://www.gov.br/capes/pt-br/acesso-a-informacao/institucional/orcamento-evolucao-em-reais>>. Acesso em: 20 jul. 2021.

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**, Rio de Janeiro, v. 42, n. 3, p.495-528, 2008.

CASSETTARI, R. R. B. *et al.* Comparação da Lei de Zipf em conteúdos textuais e discursos orais. **Profesional de la Información**, v. 24, n. 2, p. 157-167, 2015.

CASTELLS, M. **A sociedade em rede**. Paz e Terra, 2010.

CHALHUB, T.; PINHEIRO, L. V. R. Mapa dos repositórios das universidades públicas do Brasil:: estudo do acesso livre à informação científica. In: CONFERÊNCIA LUSO-BRASILEIRA DE ACESSO ABERTO, 1., 2010, Braga, Portugal. **Anais...** . Braga: Confoa, 2010. p. 1 - 10.

CHALHUBE, T.; PINHEIRO, L. V. R.. Publicações de acesso livre: tendências entre pesquisadores de universidades públicas do estado do Rio de Janeiro. In: ENCONTRO

NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011a, Brasília. **Anais...**
. Brasília: Ancib : Unb.ppgci, 2011. p. 2225 - 2241.

CLAUSET, A.; SHALIZI, C. R.; NEWMAN, M. EJ. Power-law distributions in empirical data. **SIAM review**, v. 51, n. 4, p. 661-703, 2009.

CNPQ. 2021: A Plataforma Lattes. A Plataforma Lattes. Disponível em:
<<http://memoria.cnpq.br/web/portal-lattes/sobre-a-plataforma>>. Acesso em: 12 fev. 2021.

COELHO, G. B. Ciência, sociedade e complexidade: da disciplinarização do conhecimento à emergência de programas de pós-graduação interdisciplinares no Brasil. **Revista Brasileira de Pós-graduação**, [s.l.], v. 14, p.1-22, 24 nov. 2017. CAPES. <http://dx.doi.org/10.21713/2358-2332.2017.v14.1455>.

COILE, R. C. Lotka's frequency distribution of scientific productivity. **Journal of the american society for information science**, v. 28, n. 6, p. 366-370, 1977.

COSTA, E. H. S. **Periódicos de acesso aberto: um canal de disseminação dos pesquisadores bolsistas PQ1A do CNPq?** 2018. 127 f. Tese (Doutorado) - Programa de Pós-graduação em Ciência da Informação, Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

CROW, R. The case for institutional repositories: a SPARC position paper. **Arl**, Washington, v. 1, n. 223, p.1-4, ago. 2002. Disponível em: <<http://www.arl.org/sparc/IR/ir.html>>. Acesso em: 9 nov. 2018.

CRUZ, C. *et al.* Teoria das Restrições: Um Estudo Bibliométrico da Produção Científica Apresentada no Congresso Brasileiro de Custos. **ABCustos**, v. 5, n. 1, p. 132-153, 2010.

CUNHA, M. V.; *et al.* Redes de títulos de artigos científicos variáveis no tempo. **II Brazilian Workshop on Social Network Analysis and Mining** (BraSNAM 2013). Maceió - AL: SBC. p. 01-10. 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., 2013, Maceió. **Anais...** . Maceió: Brasnam, 2013. p. 1 - 6.

DIAS, T. M. R. **Um Estudo Sobre a Produção Científica Brasileira a partir de dados da Plataforma Lattes.** 2016. 181 f. Tese (Doutorado) - Programa de Pós-graduação, Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2016.

DIAS, T. M. R.; MOREIRA, T. H. J.; DIAS, P. M. Caracterização e análise das redes de colaboração científica dos bolsistas de produtividade em pesquisa do CNPq. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 7., 2018, Natal. **Anais**. Natal: Sbc, 2018. p. 64 - 75.

DIAS, T. M. R.; MOITA, G. F.; DIAS, P. M. Um estudo sobre a rede de colaboração científica dos pesquisadores brasileiros com currículos cadastrados na Plataforma Lattes. **Em Questão**, [s.l.], v. 25, n. 1, p.83-86, 1 jan. 2019. Faculdade de Biblioteconomia Comunicacao. <http://dx.doi.org/10.19132/1808-5245251.83-86>.

DIGIAMPIETRI, L. A. *et al.* BraX-Ray: An X-Ray of the Brazilian Computer Science Graduate Programs. **Plos One**, [s.l.], v. 9, n. 4, p.94541-94551, 11 abr. 2014. Public Library of Science (PLOS). <http://dx.doi.org/10.1371/journal.pone.0094541>.

DIGIAMPIETRI, L. A. *et al.* Minerando e caracterizando dados de currículos Lattes. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 2., 2012, Curitiba. **Anais...**. Curitiba: Brasnam, 2012.

DIGIAMPIETRI, L. A. **Análise da Rede Social Acadêmica Brasileira**. 2015. 160 f. Tese (LIVRE DOCÊNCIA) - Informação e Tecnologia, Escola de Artes Ciências e Humanidades da Universidade de São Paulo, São Paulo, 2015.

EYSENBACH, G. Citation Advantage of Open Access Articles. **Plos Biology**, [s.l.], v. 4, n. 5, p.157-164, 16 maio 2006. Public Library of Science (PLOS). Disponível em: <journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040157>. Acesso em: 06 jan. 2019.

FERRAZ, R. R. N. *et al.* Extração e disponibilização on line de indicadores de desempenho e prospecção dos resultados das pesquisas em dengue com a utilização da ferramenta computacional Scriptlattes. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [s.l.], v. 20, n. 43, p.93-114, 9 ago. 2015. Universidade Federal de Santa Catarina (UFSC). <http://dx.doi.org/10.5007/1518-2924.2015v20n43p93>. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2015v20n43p93/29952>>. Acesso em: 05 fev. 2019.

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

FREIRE, J. D. **CNPq e o acesso aberto à informação científica**. 2011. 275 f. Tese (Doutorado) - Programa de Pós-graduação, Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2011.

FRY, J. *et al.* **PEER behavioural research: authors and users vis-à-vis journals and repositories**. Leicestershire: Lisu/loughborough University, 2011. Disponível em: <http://www.peerproject.eu/fileadmin/media/reports/Final_revision_-_behavioural_baseline_report_-_20_01_10.pdf>. Acesso em: 01 nov. 18.

FUKUZAWA, N. Characteristics of papers published in journals: an analysis of open access journals, country of publication, and languages used. **Scientometrics**, [s.l.], v. 112, n. 2, p.1007-1023, 23 maio 2017. Springer Nature. <http://dx.doi.org/10.1007/s11192-017-2414-y>.

FURNIVAL, A. C. M.; SILVA-JREZ, N. S. Percepções de pesquisadores brasileiros sobre o acesso aberto à literatura científica. **Percepções de Pesquisadores Brasileiros Sobre O Acesso Aberto à Literatura Científica**, João Pessoa, v. 27, n. 2, p.153-166, maio/agosto 2017.

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

GARFIELD, E. The use of JCR and JPI in measuring short and long term journal impact. In: **Council of Scientific Editors Annual Meeting**. 2000.

GAVRON, E. M.; CANTO, F. L. ANÁLISE DA UTILIZAÇÃO DOS PERIÓDICOS DE ACESSO ABERTO DE UMA BASE DE DADOS ASSINADA PELA BIBLIOTECA UNIVERSITÁRIA DA UFSC. In:

CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 27., 2017, Fortaleza. **Anais...** . Fortaleza: Febab, 2017. v. 26, p. 1 - 6.

GOFFMAN, W.; NEWILL, V. A. Generalization of Epidemic Theory: An Application to the Transmission of Ideas. **Nature**, [s.l.], v. 204, n. 4955, p.225-228, 17 out. 1964.

GOFFMAN, W.; NEWILL, V. A. Communication and epidemic processes. **Proceedings Of The Royal Society Of London. Series A. Mathematical And Physical Sciences**, [s.l.], v. 298, n. 1454, p.316-334, 2 maio 1967. The Royal Society. <http://dx.doi.org/10.1098/rspa.1967.0106>. Disponível em: <<http://rspa.royalsocietypublishing.org/content/298/1454/316.abstract>>. Acesso em: 15 nov. 2018.

GOMES, J. O. **Uma análise temporal dos principais tópicos de pesquisa da ciência brasileira a partir das palavras-chave de publicações científicas**. 127 p. Tese (Doutorado) Programa de Pós-graduação em Modelagem Matemática e Computacional — Centro Federal de Educação Tecnológica de Minas Gerais, Dezembro 2018.

GOMES, J. O.; DIAS, T. M. R.; MOITA, G. F. Uma análise dos principais tópicos de pesquisas investigados pelos pesquisadores doutores brasileiros. **Em Questão**, [s.l.], v. 24, n. 2, p.55-82, 19 abr. 2018. Faculdade de Biblioteconomia Comunicacao. <http://dx.doi.org/10.19132/1808-5245242.55-82>.

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, 6., 2005, Florianópolis. **Anais...** . Florianópolis: Ancib, 2005. p. 1 - 6.

GUMPENBERGER, C.; OVALLE-PERANDONES, M.; GORRAIZ, J. On the impact of Gold Open Access journals. **Scientometrics**, [s.l.], v. 96, n. 1, p.221-238, 24 nov. 2012. Springer Nature. <http://dx.doi.org/10.1007/s11192-012-0902-7>.

GUPTA, B.; KARISIDDIPPA, C. Collaboration and author productivity: A study with a new variable in Lotka's law. **Scientometrics**, v. 44, n. 1, p. 129-134, 1999.

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, 4., 2012, São Paulo. **Anais...** . São Paulo: Episted, 2012. p. 1 - 10.

HILÁRIO, C. M.; GRÁCIO, M. C. C. Scientific collaboration in Brazilian researches: a comparative study in the information science, mathematics and dentistry fields. **Scientometrics**, [s.l.], v. 113, n. 2, p.929-950, 4 set. 2017. Springer Nature America, Inc. <http://dx.doi.org/10.1007/s11192-017-2498-4>.

KLEINUBING, L. S. Análise bibliométrica da produção científica em gestão da informação na base de dados lisa. **Rdbci: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 8, n. 1, p.1-11, 2010.

KURAMOTO, H. Informação científica: proposta de um novo modelo para o Brasil. **Ciência da Informação**, Brasília, v. 35, n. 2, p.91-102, maio 2006.

KURTZ, M. J. Restrictive access policies cut readership of electronic research journal articles by a factor of two. In: NATIONAL POLICIES ON OPEN ACCESS (OA) PROVISION FOR UNIVERSITY RESEARCH OUTPUT: AN INTERNATIONAL MEETING, 1., 2004, Cambridge, Ma. **Anais...**

. Cambridge: Eprints.org - Self-archiving And Open Archives, 2004. p. 1 - 2. Disponível em: <<http://opcit.eprints.org/feb19oa/kurtz.pdf>>. Acesso em: 19 dez. 2018.

LANE, J. Let's make science metrics more scientific. **Nature**, [s.l.], v. 464, n. 7288, p.488-489, mar. 2010. Springer Nature. <http://dx.doi.org/10.1038/464488a>.

LAWRENCE, S. Online or Invisible? **Nature**, [s.l.], v. 411, n. 6837, p.521-523, nov. 2001. Disponível em: <<http://www.m-hikari.com/online.pdf>>. Acesso em: 10 nov. 2018.

LEITE, F. C. L. **Gestão do conhecimento científico no contexto acadêmico : proposta de um modelo conceitual**. 2016. 240 f. Dissertação (Mestrado) - Curso de Mestrado em Ciência da Informação, Departamento de Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2016.

LIMA, H. *et al.* Assessing the profile of top Brazilian computer science researchers. **Scientometrics**, [s.l.], v. 103, n. 3, p.879-896, 26 mar. 2015. Springer Nature. <http://dx.doi.org/10.1007/s11192-015-1569-7>.

LONE, F.; RATHER, R.; SHAH, Gh. J. Indian contribution to open access literature: A case study of DOAJ & OpenDOAR. **Chinese Librarianship: An International Electronic Journal**, [s.l.], v. 26, p.1-10, 29 out. 2008.

MACHADO JÚNIOR, C. *et al.* Análise de viabilidade de utilizar as leis da bibliometria em diferentes bases de pesquisa. In: ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM ADMINISTRAÇÃO, 38., 2014, Rio de Janeiro. **Anais...** . Rio de Janeiro: ANPAD, 2014.

MALTRÁS, B. B. **Los indicadores bibliométricos : fundamentos y aplicación al análisis de la ciencia**. Trea: Biblioteconomía y Administración Cultural, 2003. 288 p.

MARTÍN-MARTÍN, A. *et al.* Evidence of open access of scientific publications in Google Scholar: A large-scale analysis. **Journal Of Informetrics**, [s.l.], v. 12, n. 3, p.819-841, ago. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.joi.2018.06.012>.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999. 268 p.

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**, [s.l.], v. 15, n. 4, p.31-39, dez. 2009. Springer Nature America, Inc. <http://dx.doi.org/10.1007/bf03194511>.

MENA-CHALCO, J. P.; 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. **Em Questão**, Porto Alegre, v. 18, n. 3, p.215-229, dez. 2012.

MINNITI, S.; SANTORO, V.; BELL, S. Mapping the development of Open Access in Latin America and Caribbean countries. An analysis of *Web of Science* Core Collection and SciELO Citation Index (2005–2017). **Scientometrics**, [s.l.], v. 117, n. 3, p.1905-1930, 1 nov. 2018. Springer Nature. <http://dx.doi.org/10.1007/s11192-018-2950-0>.

MOORE, G. Survey of University of Toronto Faculty awareness, attitudes and practises regarding scholarly communication: a preliminary report. **Tspace**: University of Toronto, Toronto, mar. 2011. Disponível em: <<https://tspace.library.utoronto.ca/handle/1807/26446>>. Acesso em: 13 dez. 2018.

- MORRISON, H. *et al.* Open Access Article Processing Charges: DOAJ Survey May 2014. **Publications**, [s.l.], v. 3, n. 1, p.1-16, 5 fev. 2015. MDPI AG. <http://dx.doi.org/10.3390/publications3010001>.
- MRYGLOD, O.; *et al.* Quantifying the evolution of a scientific topic: reaction of the academic community to the Chornobyl disaster. **Scientometrics**, v. 106, n. 3, p. 1151- 1166, 2016. Disponível: < <http://link.springer.com/article/10.1007/s11192-015-1820-2>>.
- MUELLER, S. P. M. O círculo vícios o que prende os periódicos nacionais. **Datagramazero**, Brasília, v. 0, n. 4, p.1-8, dez. 1999.
- MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, Brasília, v. 35, n. 2, p.27-38, 22 ago. 2006.
- MUGNAINI, R.; LEITE, P.; LETA, J. FONTES DE INFORMAÇÃO PARA ANÁLISE DE INTERNACIONALIZAÇÃO DA PRODUÇÃO CIENTÍFICA BRASILEIRA. **Pontodeacesso**, [s.l.], v. 5, n. 3, p.87-102, 1 fev. 2012. Universidade Federal da Bahia. <http://dx.doi.org/10.9771/1981-6766rpa.v5i3.5684>.
- NAVARRO, G. A guided tour to approximate string matching. **Acm Computing Surveys**, [s.l.], v. 33, n. 1, p.31-88, 1 mar. 2001. Association for Computing Machinery (ACM). <http://dx.doi.org/10.1145/375360.375365>.
- NEUBERT, P. S.; RODRIGUES, R. S.; GOULART, L. H. Periódicos da Ciência da Informação em acesso aberto: uma análise dos títulos listados no DOAJ e indexados na Scopus | Open access journals in information Science. **Liinc em Revista**, [s.l.], v. 8, n. 2, p.389-401, 18 dez. 2012. Liinc em Revista. <http://dx.doi.org/10.18617/liinc.v8i2.497>.
- OCDE (2004) - ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. Declaration on Access to Research Data from Public Funding. Anexo 1 do Science, Technology and Innovation for the 21st Century. **Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level**, 29-30 January 2004. Disponível em: <http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html>. Acesso em: 19 outubro 2018.
- OLIVEIRA, E. C. P.; CHALHUB, T. O movimento de acesso livre à informação e repercussões nas revistas científicas Ibero-americanas. In: FORO IBERO-AMERICANO DE COMUNICAÇÃO E DIVULGAÇÃO CIENTÍFICA, 1., 2009, Campinas. **Anais...** Campinas: Ibict, 2009. p. 1 - 7.
- PACHECO, R. C. S. *et al.* Uma análise da pesquisa em engenharia e ciências mecânicas no Brasil a partir dos dados da Plataforma Lattes. **Abcm Engenharia**, Rio de Janeiro, v. 12, n. 1, p.16-22, dez. 2007.
- PAO, M. L. Lotka's law: A testing procedure. **Information Processing and Management**, [S. l.], v. 21, n. 4, p. 305–320, 1985. DOI: 10.1016/0306-4573(85)90055-X.
- PARK, J.; QIN, J. Exploring the Willingness of Scholars to Accept Open Access: A Grounded Theory Approach. **Journal of Scholarly Publishing**, [s.l.], v. 38, n. 2, p.55-84, jan. 2007. University of Toronto Press Inc. (UTPress). <http://dx.doi.org/10.3138/c972-1321-8720-314m>.
- PAVAN, C.; BARBOSA, M. C. B. Financiamento público no Brasil para a publicação de artigos em acesso aberto: alguns apontamentos. **Em Questão**, Porto Alegre, v. 23, n. 2, p.120-145, 26 abr. 2017.

PAVAN, C.; BARBOSA, M. C. Article processing charge (APC) for publishing open access articles: the Brazilian scenario. **Scientometrics**, [s.l.], v. 117, n. 2, p.805-823, 31 ago. 2018. Springer Nature. <http://dx.doi.org/10.1007/s11192-018-2896-2>.

PAVÃO, C. G.; ROCHA, R. P.; GABRIEL JUNIOR, R. F. Proposta de criação de uma rede de dados abertos da pesquisa brasileira. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 16, n. 2, p.329-343, 19 abr. 2018.

PESSOA-JUNIOR, G. J. *et al.* On interdisciplinary collaborations in scientific coauthorship networks: the case of the Brazilian community. **Scientometrics**, v. 124, n. 3, p. 2341-2360, 2020.

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

QUONIAM, L. Bibliométrie sur des référence bibliographiques: methodologie. In: DESVALS H.; DOU, H. (Org.). **La veille technologique**. Paris : Dunod, 1992. p. 244-262.

RIDKER, P. M.; RIFAI, N. Expanding Options for Scientific Publication. **Circulation**, [s.l.], v. 127, n. 2, p.155-156, 15 jan. 2013. Ovid Technologies (Wolters Kluwer Health). <http://dx.doi.org/10.1161/circulationaha.112.155952>.

RODRIGUES, R. S.; OLIVEIRA, A. B. Periódicos Científicos na America Latina: títulos em Acesso Aberto indexados no ISI e SCOPUS. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 17, n. 4, p.76-99, dez. 2012.

RONDA-PUPO, G. A. Knowledge map of Latin American research on management: Trends and future advancement. **Social Science Information**, v. 55, n. 1, p. 3 – 27, 2016.

Disponível em: <<http://journals.sagepub.com/doi/abs/10.1177/0539018415610225>>.

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**, [s.l.], v. 98, n. 1, p.415-427, 17 maio 2013. Springer Nature. <http://dx.doi.org/10.1007/s11192-013-1017-5>.

SÁNCHEZ-GIL, S.; GORRAIZ, J.; MELERO-FUENTES, D. Reference density trends in the major disciplines. **Journal Of Informetrics**, [s.l.], v. 12, n. 1, p.42-58, fev. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.joi.2017.11.003>.

SANTIAGO, M. O.; AFFONSO, F.; DIAS, T. M. R. Scientific production of women in Brazil. **Transinformação**, v. 32, 2020.

SANTOS, J. C. F.; MONTEIRO, M. S. A. Estudo da percepção de pesquisadores da área de ciências agrárias sobre Acesso Aberto. **Incid: Revista de Ciência da Informação e Documentação**, [s.l.], v. 4, n. 2, p.34-53, 20 dez. 2013.

SARVO, D. D. O. **Avaliação da Plataforma Lattes como fonte de coleta de metadados para povoamento automatizado em repositórios institucionais**. 2018. 96 f. Dissertação (Mestrado) - Programa de Pós-graduação em Ciência, Tecnologia e Sociedade, Centro de Educação e Ciências Humanas, Universidade Federal de São Carlos, São Carlos, 2018.

SIDONE, O. J. G.; HADDAD, E. A.; MENA-CHALCO, J. P. Scholarly publication and collaboration in Brazil: The role of geography. **Journal Of The Association For Information Science and Technology**, [s.l.], v. 68, n. 1, p.243-258, 11 jan. 2016. Wiley. <http://dx.doi.org/10.1002/asi.23635>.

- SILVA, T. E.; ALCARÁ, A. R. Políticas de acesso aberto à informação científica: iniciativas governamentais. In: IX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. **Anais...**. São Paulo: Ancib, 2008. p. 1 - 14.
- SILVA, T. H. P. *et al.* The Impact of Academic Mobility on the Quality of Graduate Programs. **D-lib Magazine**, [s.l.], v. 22, n. 9/10, p.1-12, set. 2016.
- SUBER, P. Open Access Overview: focusing on open access to peer-reviewed research articles and their preprints. **Creative Commons** 2007. Disponível em: <<http://legacy.earlham.edu/~peters/fos/overview.htm>>. Acesso em: 9 de jan. de 2019.
- SUBER, P. **Open Access**. Cambridge, Ma: Mit Press, 2012. 225 p.
- SWAN, A. Why Open Access for Brazil? | Por que acesso livre no Brasil? **Liinc em Revista**, [s.l.], v. 1, n. 1, p.158-171, 19 set. 2008. Liinc em Revista. <http://dx.doi.org/10.18617/liinc.v4i2.279>.
- TAGA, V. **ACESSO ABERTO: CLASSIFICAÇÃO DOS ARTIGOS INDEXADOS NA SCOPUS**. 2016. 179 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Programa Pós-graduação em Ciência da Informação, Universidade Federal de Santa Catarina, Florianópolis, 2016.
- TAGUE-SUTCLIFFE, J. An introduction to informetrics. **Information Processing & Management**, [s.l.], v. 28, n. 1, p.1-3, jan. 1992. Elsevier BV. [http://dx.doi.org/10.1016/0306-4573\(92\)90087-g](http://dx.doi.org/10.1016/0306-4573(92)90087-g).
- TRUCOLO, C. C.; DIGIAMPIETRI, L. A. Análise de tendências da produção científica nacional na área de Ciência da Informação: estudo exploratório de mineração de textos. **Atoz: novas práticas em informação e conhecimento**, [s.l.], v. 3, n. 2, p.87-94, 31 dez. 2014. Universidade Federal do Paraná. <http://dx.doi.org/10.5380/atoz.v3i2.41341>. Disponível em: <<https://revistas.ufpr.br/atoz/article/view/41341/25336>>. Acesso em: 10 fev. 2019.
- TUESTA, E. F. *et al.* Análise da participação das mulheres na ciência: um estudo de caso da área de Ciências Exatas e da Terra no Brasil. **em Questão**, [s.l.], v. 25, n. 1, p.37-62, 1 jan. 2019. Faculdade de Biblioteconomia Comunicacao. <http://dx.doi.org/10.19132/1808-5245251.37-62>.
- TUFFANI, M. **Capes negocia redução de US\$ 20 milhões em contratos e mantém Portal de Periódicos**. 2016. Disponível em: <<http://www.diretodaciencia.com/2016/05/18/capes-negocia-reducao-de-us-20-milhoes-em-contratos-e-mantem-portal-de-periodicos/>>. Acesso em: 15 jan. 2019.
- VANDI, 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**, Brasília, v. 31, n. 2, p.152-162, ago. 2002.
- VINKERS, C. H.; TIJDINK, J. K.; OTTE, W.M., J. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. **BMJ**, 2015. Disponível em: <<http://www.bmj.com/content/351/bmj.h6467>>
- ZHANG, L.; WATSON, E. The prevalence of green and grey open access: Where do physical science researchers archive their publications?. **Scientometrics**, [s.l.], v. 117, n. 3, p.2021-2035, 5 out. 2018. Springer Nature. <http://dx.doi.org/10.1007/s11192-018-2924-2>.

ANEXOS

ANEXO A - Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica⁸



Manifesto Brasileiro de Apoio ao Acesso Livre à Informação Científica

Contextualização

A informação científica é o insumo básico para o desenvolvimento científico e tecnológico de uma nação. Trata-se de um processo contínuo em que a informação científica contribui para o desenvolvimento científico, e este, por sua vez, gera novos conteúdos realimentando todo o processo.

No entanto, a comunidade científica enfrenta dificuldades no acesso à informação científica, se considerado o modelo tradicional de publicação científica. Tradicionalmente, os artigos ou trabalhos científicos são publicados em revistas especializadas, e a forma de acesso a esses trabalhos dá-se mediante assinatura das publicações pelas bibliotecas ou pelo pesquisador.

Com o surgimento das novas tecnologias da informação e da comunicação, diversos paradigmas estão mudando. Isso porque essas tecnologias facilitam o acesso à informação científica, promovendo o surgimento de novas alternativas para a comunicação científica. A Open Archives Initiative (OAI) é um exemplo disso. Esta iniciativa estabelece, além de padrões de interoperabilidade, alguns princípios e ideais, como o uso de *software open source* e o acesso livre à informação. Surge, a partir dessa iniciativa, o **paradigma do acesso livre à informação**.

A OAI constitui, portanto, um marco na área do tratamento e disseminação da informação em geral e na área da comunicação científica em especial. Essa iniciativa proporcionou a construção, implantação e manutenção de diversos repositórios de acesso livre, assim como o surgimento de diversas ferramentas de *software* para a construção e manutenção de repositórios, como o E-Prints, o Open Journal Systems (OJS), o DSPACE, entre outros.

O movimento de apoio aos *open archives* e ao acesso livre à informação surge em consequência das dificuldades encontradas pela comunidade científica mundial no acesso à informação científica. Concretizou-se, de fato, por meio de diversos manifestos, como as declarações de Bethesda, Budapeste e Berlim, além de manifestações de organizações não-governamentais e internacionais, como a IFLA e a OCDE, entre outras.

É importante observar que o paradigma do acesso livre à informação provocará otimização nos custos de registro e acesso à informação, além de promover maior rapidez no fluxo da informação científica e no desenvolvimento científico e tecnológico. Esse cenário aponta para a necessidade de o Brasil manifestar-se favoravelmente ao acesso livre à informação, promovendo, por conseguinte, o aumento significativo da visibilidade de suas pesquisas, de seus pesquisadores e de suas instituições. Para tanto, é necessário aderir ao movimento mundial e estabelecer uma **política nacional de acesso livre à informação científica**, mediante o apoio de toda a comunidade científica, com o envolvimento não apenas das suas organizações, mas, **obrigatoriamente**, dos pesquisadores e das agências de fomento.

O estabelecimento do acesso livre como um procedimento vantajoso requer o empenho ativo de todo e qualquer indivíduo que produza conhecimento científico ou seja, de todo detentor de patrimônio cultural.

Objetivos:

- promover o registro da produção científica brasileira em consonância com o paradigma do acesso livre à informação;
- promover a disseminação da produção científica brasileira em consonância com o paradigma do acesso livre à informação;
- estabelecer uma política nacional de acesso livre à informação científica;
- buscar apoio da comunidade científica em prol do acesso livre à informação científica.

Paradigma do Acesso Livre à Informação

O modelo que se preconiza para o acesso livre à informação e que nesse documento é denominado de Paradigma do Acesso Livre à Informação, baseia-se nos termos da **Declaração de Berlim**, na parte relativa à "Definição de uma contribuição em acesso livre...", que é aqui reescrita conforme os dois itens abaixo:

1. Contribuições em acesso livre incluem resultados de pesquisas científicas originais, dados não processados, metadados, fontes originais, representações digitais de materiais pictóricos, gráficos e material acadêmico multimídia.
2. As contribuições em acesso livre devem satisfazer duas condições:
 - I. os(s) autor(es) e o(s) detentores dos direitos de tais contribuições concede(m) a todos os usuários:
 - a. direito gratuito, irrevogável e irrestrito de acessá-las;
 - b. licença para copiá-las, usá-las, distribuí-las, transmiti-las e exibi-las publicamente;

- c. licença para realizar e distribuir obras derivadas, em qualquer suporte digital para qualquer propósito responsável, em obediência à correta atribuição da autoria (as regras da comunidade continuarão a fornecer mecanismos para impor a atribuição e uso responsável dos trabalhos publicados, como acontece no presente) e com a garantia de fazer cópias;

Recomendações à Comunidade Científica

Torna-se necessário, nesse momento, o compromisso por parte da comunidade científica brasileira de apoiar o movimento mundial em favor do acesso livre à informação científica. Nesse sentido, os principais atores do sistema de comunicação científica, nomeadamente autores, editores, agências de fomento e as instituições acadêmicas, devem se comprometer a colaborar para que os resultados de pesquisas realizadas no país estejam disponíveis livremente para acesso. Para isso, portanto, de acordo com o que especifica a **Declaração de Berlim**, recomenda-se:

- A. É **imperativo** que as **instituições acadêmicas** brasileiras se comprometam a:

1. criar repositórios institucionais e temáticos, observando o paradigma do acesso livre;
2. requerer que seus pesquisadores depositem uma cópia de todos os seus trabalhos publicados em pelo menos um repositório de acesso livre;
3. encorajar seus pesquisadores a publicar seus resultados de pesquisa em periódicos de acesso livre, onde houver um periódico apropriado para isso. Deve-se, além disso, prover o apoio necessário para que isso ocorra;
4. reconhecer a publicação em ambiente de acesso livre para efeito de avaliação e progressão acadêmica;
5. ter disponíveis, em ambiente de acesso livre, os periódicos editados pela instituição ou seus órgãos subordinados.

- B. É **primordial** que os **pesquisadores** (autores):

1. contribuam para o incremento de conteúdos em repositórios institucionais ou temáticos, depositando o maior número possível de seus trabalhos, publicados ou não, inclusive *pré e post prints*, material de aula, quando for o caso, entre outros materiais.
2. depositar, obrigatoriamente, em um repositório de acesso livre publicações que envolvam resultados de pesquisas financiadas com recursos públicos.

- C. É **necessário** que as **agências de fomento**:

1. reconheçam a publicação científica em repositórios de acesso livre para efeito de avaliação da produção científica dos pesquisadores e de concessão de auxílios e financiamentos para pesquisa;
2. recomendem aos pesquisadores a quem concedem auxílio financeiro para suas pesquisas que depositem uma cópia dos resultados publicados em um repositório de acesso livre e/ou que publiquem prioritariamente em periódicos eletrônicos de acesso livre;
3. recomendem aos pesquisadores a quem concedem auxílio financeiro para participação em eventos que depositem uma cópia do seu trabalho em um repositório de acesso livre;
4. promovam e apoiem a construção e manutenção de repositórios institucionais e temáticos;
5. apoiem, prioritariamente, a edição de publicações científicas eletrônicas de acesso livre;
6. requerer que toda publicação científica financiada com recursos públicos tenha uma versão disponível eletronicamente em ambiente de acesso livre.

- D. É **imprescindível** que as **editoras comerciais** de publicações científicas:

- A. concordem que os trabalhos por elas publicados com autoria de pesquisadores que obtiveram recursos públicos para suas pesquisas tenham uma cópia depositada em repositório de acesso livre;
- B. tenham disponível uma versão eletrônica, em ambiente de acesso livre, das publicações impressas por elas editadas cuja autoria seja de pesquisadores que obtiveram recursos públicos para suas pesquisas.

- E. É **recomendável** que **editoras não comerciais**:

1. tenham disponíveis uma versão eletrônica, em conformidade com o paradigma do acesso livre à informação, das publicações impressas por elas editadas;
2. adotem os padrões que estejam em conformidade com aqueles estabelecidos pela Open Archives Initiative (OAI).

⁸ <http://livroaberto.ibict.br/Manifesto.pdf>

PORTARIA Nº 13, DE 15 DE FEVEREIRO DE 2006.

Institui a divulgação digital das teses e dissertações produzidas pelos programas de doutorado e mestrado reconhecidos.

O PRESIDENTE DA FUNDAÇÃO COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR - Capes, no uso das atribuições conferidas pelo artigo 20, inciso II, do Estatuto aprovado pelo Decreto nº 4.631, de 21 de março de 2003, e considerando as manifestações do Conselho Técnico - Científico verificadas no ano de 2005, indicando que a produção científica discente é um relevante indicador da qualidade dos programas de mestrado e doutorado, não aferível apenas através da publicação seletiva nos periódicos especializados, resolve:

Art. 1º Para fins do acompanhamento e avaliação destinados à renovação periódica do reconhecimento, os programas de mestrado e doutorado deverão instalar e manter, até 31 de dezembro de 2006, arquivos digitais, acessíveis ao público por meio da Internet, para divulgação das dissertações e teses de final de curso.

§1º Os programas de pós-graduação exigirão dos pós-graduandos, a entrega de teses e dissertações em formato eletrônico, simultânea à apresentação em papel, para atender ao disposto neste artigo.

§2º Os arquivos digitais disponibilizarão obrigatoriamente as teses e dissertações defendidas a partir de março de 2006.

§3º A publicidade objeto deste artigo poderá ser assegurada mediante publicação através de sítio digital indicado pela CAPES, quando o programa não dispuser de sítio próprio.

Art. 2º Por ocasião do envio dos relatórios para acompanhamento e avaliação o programa deverá apresentar a justificativa para a eventual ausência de depósito de obra, na forma disciplinada por esta Portaria, motivada pela proteção ao sigilo industrial ou ético.

Art. 3º No acompanhamento e a valiação dos programas de pós-graduação serão ponderados o volume e a qualidade das teses e dissertações publicadas, além de dados confiáveis sobre a acessibilidade e possibilidade de download.

⁹ http://repositorio.unb.br/documentos/Portaria_N13_CAPES.pdf

Art. 4º A CAPES divulgará em seu sítio digital a lista dos arquivos utilizados para os fins do disposto nesta Portaria, classificada por Área do Conhecimento.

Art. 5º O financiamento de trabalho com verba pública, sob forma de bolsa de estudo ou auxílio de qualquer natureza concedido ao Programa, induz à obrigação do mestre ou doutor apresentá-lo à sociedade que custeou a realização, aplicando-se a ele as disposições desta Portaria.

JORGE ALMEIDA GUIMARÃES
(DOU Nº 35, 17/2/2006, SEÇÃO 1, P. 15)

PROJETO DE LEI Nº 1120, DE 2007
(Do Sr. Rodrigo Rollemberg)

Dispõe sobre o processo de disseminação da produção técnico-científica pelas instituições de ensino superior no Brasil e dá outras providências.

O Congresso Nacional decreta:

Art. 1º. As instituições de ensino superior de caráter público, assim como as unidades de pesquisa, ficam obrigadas a construir os seus repositórios institucionais, nos quais deverão ser depositados o inteiro teor da produção técnico-científica conclusiva do corpo discente, com grau de aprovação, dos cursos de mestrado, doutorado, pós-doutorado ou similar, a produção técnico-científica conclusiva do corpo docente dos níveis de graduação e pós-graduação, assim como a produção técnico-científica, resultado das pesquisas realizadas pelos seus pesquisadores e professores, financiadas com recursos públicos, para acesso livre na rede mundial de computadores – INTERNET.

§ 1º. Os repositórios institucionais deverão ser compatíveis com padrões de interoperabilidade adotados internacionalmente.

§ 2º. Fica o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) responsável pela integração, consolidação e disseminação, em seu sítio na Internet, de todos os repositórios institucionais.

§ 3º. Os pesquisadores, que receberem apoio financeiro proveniente do governo federal, estadual ou municipal para suas pesquisas, são obrigados a depositar uma cópia das publicações dos resultados dessas pesquisas.

§ 4º. No caso em que tais publicações sejam protegidas por contratos de “copyright” que as impeçam de serem depositadas em seu completo teor, os pesquisadores se obrigarão a pelo menos depositar os seus metadados, com o compromisso de disponibilizar o acesso ao completo teor a partir do momento de sua liberação.

§ 5º. O mesmo se aplica em casos em que as publicações contiverem objetos passíveis de serem patenteados.

§ 6º. No que tange aos padrões de interoperabilidade, estes deverão ser estabelecidos pelo Instituto Brasileiro de Informação em Ciência e Tecnologia que, inclusive, terá a atribuição de orientar tecnicamente e dar total assistência às instituições de ensino superior e às unidades de pesquisa para a construção dos repositórios.

¹⁰

https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra;jsessionid=F86D3B71F1E4FCA4F9135827EBF4063D.node2?codteor=572586&filename=Avulso+-PL+1120/2007

§ 7º. A inobservância do disposto no presente artigo por parte dos pesquisadores, das instituições de ensino superior ou das unidades de pesquisa torná-los-ão inelegíveis para obtenção de qualquer apoio financeiro para suporte às suas pesquisas.

Art. 2º. Com o propósito de dar suporte e estimular os pesquisadores a observarem o disposto nesta Lei, o Ministério da Ciência e Tecnologia deverá constituir um Comitê de Alto Nível, coordenado pelo IBICT, composto pelos principais segmentos da Comunidade Científica envolvidos na cadeia produtiva da pesquisa científica, para discutir e propor uma política nacional de acesso livre à informação.

Parágrafo único. O Comitê proposto deverá ser criado e instalado em um prazo máximo de 45 (quarenta e cinco) dias a partir da entrada em vigor desta lei.

Art. 3º. Esta Lei entra em vigor na data da sua publicação oficial.

JUSTIFICAÇÃO

Atualmente, um dos temas mais candentes no meio acadêmico é o acesso aberto à produção científica através da Internet. A sociedade do conhecimento, forjada por aqueles que criaram e animam o espírito da Internet – muito antes de sua popularização, é caracterizada pela busca do saber com base na coletivização, na distribuição e na difusão das descobertas nas comunidades de interesse. O compartilhamento em todas esferas do conhecimento em escala global é uma tendência que choca, desde sua essência, com velhas práticas que obram – até por inércia – no sentido contrário: pela concentração do saber, pela hierarquização, pelo poder auto-legitimado e auto-concedido e pelo prevalecimento do ganho econômico sobre o interesse social. Trata-se de matrizes ideológicas e comportamentais bastante distintas e altamente conflitivas que afetam não apenas o campo acadêmico, mas também os campos cultural, social e econômico.

A disponibilização pública de conteúdos digitais, sua proteção legal e a garantia de acesso aos seus produtos derivados são fundamentais para alimentar as cadeias culturais, artísticas, educativas e científicas. Devem ser consideradas também as questões da proteção aos direitos autorais e da garantia do domínio público, no caso do resultado de pesquisas financiadas pelo erário. Nesse sentido, as declarações internacionais ajudaram a estabelecer os primeiros marcos para o início de uma discussão madura sobre tais temas.

A universidade hoje dispõe de meios acessíveis, criativos e baratos para superar os muros invisíveis que a separam de parte da sociedade. Iniciativas como o estímulo ao acesso aberto, a criação de um repositório de conteúdos digitais e a elaboração de uma política consistente com respeito à divulgação de conteúdos digitais vêm ao encontro da demanda dos docentes de difundir sua produção intelectual e de ter acesso facilitado às pesquisas dos colegas. Além disso, a promoção de mecanismos de acesso aberto dá mais visibilidade e transparência àquilo que é produzido pela universidade, reforçando sua função de servir à sociedade ao promover o conhecimento científico e a difusão cultural.

A construção dos repositórios e o arquivamento digital da produção técnico-científica proporcionará maior visibilidade dos investimentos do governo em C&T, além de dar subsídios, ao governo, para a elaboração da política de fomento de C&T para o Brasil.

É importante ressaltar o impacto da aplicação do conhecimento científico no desenvolvimento social, econômico, científico e tecnológico de um país.

O governo, em especial, com a iniciativa de criação dessa rede de repositórios institucionais passa a ter possibilidade de obter indicadores que orientem o planejamento da ciência e tecnologia no país.

Considerando a concentração de conhecimento gerados, em países localizados no hemisfério norte e os altos custos de manutenção das publicações periódicas, um novo fenômeno aparece que é a “exclusão cognitiva”. A ação proposta neste projeto de lei contribui para reduzir essa exclusão, uma vez que registra e dissemina a produção científica brasileira, assim como a estrangeira de forma livre e integrada.

Pelo exposto, conto com o apoio dos nobres Pares para aprovação da proposição em epígrafe.

Sala das Sessões, em 21/05/2007.

Deputado Rodrigo Rollemberg
PSB/DF

Institui a “Política para Acesso Aberto às Publicações Resultantes de Auxílios e Bolsas FAPESP”.

O DIRETOR PRESIDENTE DO CONSELHO TÉCNICO ADMINISTRATIVO DA FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, no uso das atribuições que lhe são conferidas pelo Estatuto aprovado pelo Decreto nº 40.132/62, e com base nas decisões do Conselho Superior, em reunião realizada em 11 de junho de 2008, e do Conselho Técnico Administrativo, em reunião realizada em 21 de fevereiro de 2019, expede a seguinte Portaria:

Art. 1º Esta Portaria tem por objetivo instituir a Política para Acesso Aberto às Publicações Resultantes de Auxílios e Bolsas FAPESP, Anexo I desta Portaria.

Art. 2º As normas estabelecidas pela Política para Acesso Aberto às Publicações Resultantes de Auxílios e Bolsas FAPESP se aplicam à publicação de qualquer artigo ou outro tipo de comunicação científica, que contenha resultados originados de pesquisas apoiadas, parcial ou totalmente, pela FAPESP, em qualquer modalidade de apoio.

Art. 3º Para efeitos desta Portaria consideram-se as seguintes definições:

I - Modalidades de Apoio: São as modalidades de Bolsas, Auxílios à Pesquisa e Programas concedidas pela FAPESP;

II - Acesso Aberto: disponibilização em repositórios de acesso universal e aberto de artigos ou outros tipos de comunicação científica publicados;

III - Termo de Outorga: Documento que oficializa a concessão de recursos pela FAPESP ao(s) Outorgado(s), para o desenvolvimento do projeto de pesquisa previamente aprovado, e estabelece as condições e obrigações das partes nele qualificadas.

Art. 4º As obrigações estabelecidas na Política para Acesso Aberto às Publicações Resultantes de Auxílios e Bolsas FAPESP, Anexo I desta Portaria, serão disciplinadas nos Termos de Outorga dos processos concedidos pela FAPESP.

¹¹ <https://fapesp.br/12632/portaria-cta-no-012019>

Art. 5º Demais orientações quanto à forma de encaminhamento de informações referentes às publicações de qualquer artigo ou outro tipo de comunicação científica, que contenham resultados originados de pesquisas apoiadas pela FAPESP, serão incluídas nas normas das modalidades de apoio.

Art. 6º Esta Portaria entra em vigor 30 dias após a data de sua assinatura.

São Paulo, 21 de fevereiro de 2019.

Carlos Américo Pacheco
Diretor Presidente
Conselho Técnico-Administrativo

ANEXO E: Tabela Kolmogorov-Smirnov de uma amostra¹².

A tabela fornece os valores críticos D_n, α conforme descrito no Teste de Kolmogorov-Smirnov.

$n \backslash \alpha$	0.001	0.01	0.02	0.05	0.1	0.15	0.2
1		0.99500	0.99000	0.97500	0.95000	0.92500	0.90000
2	0.97764	0.92930	0.90000	0.84189	0.77639	0.72614	0.68377
3	0.92063	0.82900	0.78456	0.70760	0.63604	0.59582	0.56481
4	0.85046	0.73421	0.68887	0.62394	0.56522	0.52476	0.49265
5	0.78137	0.66855	0.62718	0.56327	0.50945	0.47439	0.44697
6	0.72479	0.61660	0.57741	0.51926	0.46799	0.43526	0.41035
7	0.67930	0.57580	0.53844	0.48343	0.43607	0.40497	0.38145
8	0.64098	0.54180	0.50654	0.45427	0.40962	0.38062	0.35828
9	0.60846	0.51330	0.47960	0.43001	0.38746	0.36006	0.33907
10	0.58042	0.48895	0.45662	0.40925	0.36866	0.34250	0.32257
11	0.55588	0.46770	0.43670	0.39122	0.35242	0.32734	0.30826
12	0.53422	0.44905	0.41918	0.37543	0.33815	0.31408	0.29573
13	0.51490	0.43246	0.40362	0.36143	0.32548	0.30233	0.28466
14	0.49753	0.41760	0.38970	0.34890	0.31417	0.29181	0.27477
15	0.48182	0.40420	0.37713	0.33760	0.30397	0.28233	0.26585
16	0.46750	0.39200	0.36571	0.32733	0.29471	0.27372	0.25774
17	0.45440	0.38085	0.35528	0.31796	0.28627	0.26587	0.25035
18	0.44234	0.37063	0.34569	0.30936	0.27851	0.25867	0.24356
19	0.43119	0.36116	0.33685	0.30142	0.27135	0.25202	0.23731
20	0.42085	0.35240	0.32866	0.29407	0.26473	0.24587	0.23152
25	0.37843	0.31656	0.30349	0.26404	0.23767	0.22074	0.20786
30	0.34672	0.28988	0.27704	0.24170	0.21756	0.20207	0.19029
35	0.32187	0.26898	0.25649	0.22424	0.20184	0.18748	0.17655
40	0.30169	0.25188	0.23993	0.21017	0.18939	0.17610	0.16601
45	0.28482	0.23780	0.22621	0.19842	0.17881	0.16626	0.15673
50	0.27051	0.22585	0.21460	0.18845	0.16982	0.15790	0.14886
OVER 50	1.94947	1.62762	1.51743	1.35810	1.22385	1.13795	1.07275
	$\sqrt{n}$	$\sqrt{n}$	$\sqrt{n}$	$\sqrt{n}$	$\sqrt{n}$	$\sqrt{n}$	$\sqrt{n}$

¹² <https://www.real-statistics.com/statistics-tables/kolmogorov-smirnov-table/>