



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

JÉSSICA BILAC GASPARETO

**O USO DE TÉCNICAS DE CIÊNCIA DE DADOS PARA ANALISAR A
COMUNICAÇÃO E COLABORAÇÃO CIENTÍFICA DOS PROFESSORES DOS
PROGRAMAS DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO DAS
UNIVERSIDADES FEDERAIS BRASILEIRAS**

Brasília

2024

JÉSSICA BILAC GASPARETO

**O USO DE TÉCNICAS DE CIÊNCIA DE DADOS PARA ANALISAR A
COMUNICAÇÃO E COLABORAÇÃO CIENTÍFICA DOS PROFESSORES DOS
PROGRAMAS DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO DAS
UNIVERSIDADES FEDERAIS BRASILEIRAS**

Dissertação de mestrado apresentada à banca examinadora como requisito parcial para a obtenção do título de Mestre em Ciência da Informação pela Faculdade de Ciência da Informação da Universidade de Brasília.

Orientador: Dr. Márcio de Carvalho Victorino

Brasília

2024

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

Bu Bilac Gaspareto, Jéssica
O uso de técnicas de ciência de dados para analisar a comunicação e colaboração científica dos professores dos programas de pós-graduação em ciência da informação das universidades federais brasileiras / Jéssica Bilac Gaspareto; orientador Márcio de Carvalho Victorino. -- Brasília, 2024.
100 p.

Dissertação(Mestrado em Ciência da Informação) -- Universidade de Brasília, 2024.

1. Comunicação Científica. 2. Ciência de Dados. 3. Colaboração Científica. 4. Comunidade Científica. 5. Organização da Informação. I. de Carvalho Victorino, Márcio, orient. II. Título.

UNIVERSIDADE DE BRASÍLIA

PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Ata Nº:65

Aos **vinte** dias do mês de **setembro** do ano de dois mil **e vinte e quatro**, instalou-se a banca examinadora de Dissertação de Mestrado do(a) aluno(a) **Jéssica Bilac Gaspareto**, matrícula **220004145**. A banca examinadora foi composta pelos professores Dr. Dalton Lopes Martins (UnB), Dra. Daniela Lucas da Silva Lemos (UFES), Dr(a). Dr. Fernando César Lima Leite (UnB), (Suplente) e Dr(a). Dr. Marcio de Carvalho Victorino (UnB), orientador(a)/presidente. O(A) discente apresentou o trabalho intitulado “O USO DE TÉCNICAS DE CIÊNCIA DE DADOS PARA ANALISAR A COMUNICAÇÃO E COLABORAÇÃO CIENTÍFICA DOS PROFESSORES DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO DAS UNIVERSIDADES FEDERAIS BRASILEIRAS ”.

Concluída a exposição, procedeu-se a arguição do(a) candidato(a), e após as considerações dos examinadores o resultado da avaliação do trabalho foi:

(X) Pela aprovação do trabalho;

() Pela aprovação do trabalho, com revisão de forma, indicando o prazo de até 30 (trinta) dias para apresentação definitiva do trabalho revisado;

() Pela reformulação do trabalho, indicando o prazo de **(Nº DE MESES)** para nova versão;

() Pela reprovação do trabalho, conforme as normas vigentes na Universidade de Brasília.

Conforme os Artigos 34, 39 e 40 da Resolução 0080/2021 - CEPE, o(a) candidato(a) não terá o título se não cumprir as exigências acima.

Dr. Marcio de Carvalho Victorino (UnB)
(Presidente)

Dr. Dalton Lopes Martins (UnB)
(Membro interno - PPGCINF)

Dra. Daniela Lucas da Silva Lemos (UFES)
(Membra externa ao PPGCINF)

Dr. Fernando César Lima Leite
(Suplente / UnB)

Jéssica Bilac Gaspareto
(Mestranda)



Documento assinado eletronicamente por **Marcio de Carvalho Victorino, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 07/10/2024, às 17:14, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Jessica Bilac Gaspareto, Usuário Externo**, em 08/10/2024, às 15:46, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Dalton Lopes Martins, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 08/10/2024, às 17:56, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Fernando César Lima Leite, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 08/10/2024, às 19:11, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Daniela Lucas da Silva Lemos, Usuário Externo**, em 09/10/2024, às 11:08, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **11717804** e o código CRC **13800190**.

RESUMO

Este trabalho tem como principal propósito analisar a dinâmica da comunicação e colaboração científica entre os professores dos programas de pós-graduação em Ciência da Informação das Universidades Federais Brasileiras, buscando integrar os dados abertos disponibilizados pelo portal da CAPES. A pesquisa adota uma abordagem metodológica quantitativa e qualitativa, fazendo uso de técnicas estatísticas, bem como de métricas de bibliometria e cientometria, para mensurar os fenômenos de colaboração e comunicação científica entre os professores desses programas, e também utiliza da dinâmica qualitativa para conceituação e entendimento das aplicações das leis da bibliometria. A amostra escolhida para o trabalho foi formada pelos professores dos programas de pós-graduação em Ciência da Informação das Universidades Federais Brasileiras. Para atingir esse objetivo, é essencial avaliar o intercâmbio de conhecimento, as parcerias acadêmicas e as redes de colaboração estabelecidas. Os dados coletados são devidamente armazenados em um banco de dados relacional, criado por meio da linguagem SQL, permitindo a organização, manipulação e limpeza eficazes das informações para análise. Além disso, são utilizados gráficos e outras representações visuais para facilitar a compreensão dos padrões e tendências na produção e colaboração científica por meio dos *softwares* Neo4j e PowerBi. Através da manipulação dos dados e da utilização de estatísticas descritivas, foram identificados padrões de colaboração e comunicação científica na área de Ciência da Informação. Métricas de bibliometria e cientometria foram empregadas para avaliar impacto e produtividade, revelando que programas de pós-graduação com melhores conceitos CAPES tendem a ter maior produção e colaboração científica.

Palavras-chaves: Comunicação Científica. Ciência de Dados. Colaboração Científica. Comunidade Científica. Organização da Informação. Dados Abertos.

ABSTRACT

This study aims to analyze the dynamics of communication and scientific collaboration among faculty members of graduate programs in Information Science at Brazilian Federal Universities, by integrating open data available from the CAPES portal. The research employs both quantitative and qualitative methodological approaches, utilizing statistical techniques as well as bibliometric and scientometric metrics to measure collaboration and scientific communication among faculty members of these programs. It also uses qualitative dynamics for the conceptualization and understanding of bibliometric laws. The sample for this study was limited to faculty members of graduate programs in Information Science at Brazilian Federal Universities. To achieve this goal, it is essential to assess knowledge exchange, academic partnerships, and established collaboration networks. The collected data is stored in a relational database created using SQL, enabling effective organization, manipulation, and cleaning of the information for analysis. Additionally, graphs and other visual representations are used to facilitate the understanding of patterns and trends in scientific production and collaboration through Neo4j and PowerBI software. Through data manipulation and descriptive statistics, patterns of collaboration and scientific communication in the field of Information Science were identified. Bibliometric and scientometric metrics were used to evaluate impact and productivity, revealing that graduate programs with better CAPES ratings tend to have higher scientific production and collaboration.

Keywords: Scientific Communication. Data Science. Scientific Collaboration. Scientific community. Organization of Information. Open Data.

LISTA DE SIGLAS

(ETL) - Extração, Transformação e Carregamento

(TI) - Tecnologia da Informação

(FID) - Federação Internacional de Informação e Documentação

(IoT) - Internet das Coisas

(FN) - Fator de Impacto

(LGPD) - Lei Geral de Proteção de Dados

(LAI) - Lei de Acesso à Informação

(OGP) - *Open Government Partnership*

(PDBA) - Portal Brasileiro dos Dados Abertos

(CAPES) - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

(PDA) - Plano de Dados Abertos

(UFs) - Unidades Federativas

LISTA DE TABELAS

Tabela 1 - Amostra das principais colaborações entre professores	75
Tabela 2 - Ranking de maiores colaborações científicas por universidades.....	80

LISTA DE FIGURAS

Figura 1 - Corpos de Conhecimentos da Informação Ligados à Ciência da Informação	21
Figura 2 - Especialidades da Ciência da Informação	22
Figura 3 - Processo helicoidal do conhecimento	25
Figura 4 - Partes do Processo Informativo	27
Figura 5 - Ciclo da Comunicação Científica	30
Figura 6 - Objetivo da Comunidade Científica e da Comunicação Científica	32
Figura 7 - Representação de um grafo de colaboração	42
Figura 8 - Representação das novas características do big data.....	47
Figura 9 - Modelo Conceitual Inicial	55
Figura 10 - Modelo Final Lógico	56
Figura 11 - Fluxo de dados.....	60
Figura 12 - Fluxo de ETL para tratamento dos dados	61
Figura 13 - Filtro de IDs para filtrar os códigos de programa de pós-graduação em CI	62
Figura 14 - Quantidade de produções historicamente.....	63
Figura 15 - Quantidade de produções historicamente por professores.....	64
Figura 16 - Grau comparativo dos dez professores que mais publicam.....	64
Figura 17 - Demonstrativo de dispersão entre publicações por autores	65
Figura 18 - Relação da quantidade de professores publicando x número de publicações	66
Figura 19 - Perfil etário dos professores	67
Figura 20 - Nacionalidade dos professores.....	67
Figura 21 - Distribuição de publicações por região	68
Figura 22 - Distribuição de médias de notas de conceito CAPES por região.....	69
Figura 23 - Amostra das 18 principais notas conceito CAPES por entidades de ensino.....	70
Figura 24 - Distribuição de conceito CAPES por entidade de ensino.....	71
Figura 25 - Distribuição de conceito CAPES por região	72
Figura 26 - Distribuição de colaboração científica por dispersão	78
Figura 27 - Grafo de colaboração científica entre universidades	79
Figura 28 - Regressão linear de conceito CAPES e o total de colaborações.....	81
Figura 29 - Distribuição de colaboração científica entre universidades.....	82

LISTA DE QUADROS

Quadro 1 - Convergência entre Ciência da Informação e Comunicação Científica...	29
Quadro 2 - Cronologia histórica da colaboração científica	37
Quadro 3 - Motivações para Colaboração Científica.....	41
Quadro 4 - Modelo Cronológico dos Princípios Norteadores da Atual Ciência de Dados.....	45
Quadro 5 - Atributos da tabela universidade	57
Quadro 6 - Atributos da tabela relacionamento escreve	57
Quadro 7 - Atributos da tabela relacionamento credencia	55
Quadro 8 - Atributos da tabela professores.....	58
Quadro 8 - Atributos da tabela tabela programa-pos	58
Quadro 9 - Atributos da tabela publicador	59
Quadro 10 - Atributos da tabela artigos.....	59
Quadro 11 - Código utilizado para criar relacionamento de colaboração na ferramenta Neo4j.....	77

SUMÁRIO

1. INTRODUÇÃO	1
1.1 Contextualização e Definição do Problema	1
1.2 Justificativa	5
1.3 Pergunta da pesquisa	7
1.4. Objetivo Geral	7
1.5. Objetivos Específicos	7
2. PROCEDIMENTO METODOLÓGICO	9
2.1 Hipótese Principal	9
2.2 Procedimento Teórico	9
2.3 Passos da Pesquisa	10
2.3.1 Primeira Fase: Levantamento	10
2.3.2. Segunda Fase: Modelagem	10
2.3.3. Terceira Fase: Extração, Transformação e Carga de Dados	11
2.3.4. Quarta Fase: Geração de Gráficos Analíticos (Dashboard)	11
2.3.5. Quinta Fase: Validação	11
3. TRABALHOS RELACIONADOS	13
4. FUNDAMENTAÇÃO TEÓRICA	16
4.1 Ciência da Informação e sua contextualização histórica	16
4.1.1 Objeto de estudo da CI: a informação	22
4.2. Comunicação Científica	26
4.2.1. Indicadores de Comunicação Científica	33
4.3. Colaboração Científica	36
4.4. Ciência de Dados	44
4.5 Dados Governamentais Abertos	51
5. ESTUDO DE CASO	54
5.1 Proposta do trabalho	54
5.2 Fontes de dados para o estudo desenvolvido	54
5.3 Resultados e discussões	62
5.4. Considerações finais	82
REFERÊNCIAS	84

1. INTRODUÇÃO

1.1 Contextualização e Definição do Problema

A ascendência de novas descobertas e avanços científicos foram fundamentais para transformações em todas as épocas, por meio de mudanças de padrões de comportamento e do acesso à informação na sociedade. Juntamente com o advento da Revolução Industrial e a chegada do século XX – marcado pelo impulso sem precedentes do conhecimento e pelo desenvolvimento tecnológico –, a ciência ganhou maior dimensão juntamente com a produção de novas informações, fazendo com que diversos pesquisadores necessitassem publicar seus trabalhos, nascendo assim a comunicação científica (Valeiro; Pinheiro, 2008). Este contexto evidencia a crescente importância da comunicação científica, objeto central de estudo desta dissertação, pois ela não apenas facilita a disseminação do conhecimento, mas também promove a colaboração e a inovação contínua.

A comunicação científica é essencial para o progresso da ciência, concentrando-se na transferência de informações em ciência, tecnologia e inovações para especialistas em áreas específicas (Bueno, 2010). Com o aumento do volume de dados e publicações científicas, surgem novas formas de publicação e colaboração, acompanhadas por indicadores métricos que mensuram citações e coautorias. Esses indicadores são fundamentais para avaliar a qualidade do trabalho acadêmico e a produtividade dos pesquisadores, influenciando diretamente a reputação e os recursos alocados a programas de pós-graduação.

Este estudo visa analisar a produção científica nas universidades federais brasileiras, com foco nos professores de pós-graduação. Por meio da utilização de indicadores que medem coautoria e produção científica, de acordo com a unidade de pós-graduação de cada professor. Segundo Weitzel (2006), a comunicação científica é um processo que relaciona comunicação, construção e uso do conhecimento científico em prol da evolução. Para que uma pesquisa atinja seu objetivo, é necessária a divulgação dos resultados por meio de canais de ampla audiência, destacando a importância da propagação do conhecimento científico para evitar esforços redundantes e fomentar colaborações eficientes.

Neste cenário, a comunidade científica se destaca como um elemento central, tratando de assuntos que contemplam a autoridade em produção científica. Análises bibliométricas permitem o reconhecimento de autoridade e auxiliam na análise da produção científica de um pesquisador, sendo utilizadas como parâmetro para concessão de recursos e indicadores de qualidade na comunicação científica. A relevância da comunidade científica neste estudo é fundamental, pois serão analisados os professores de ciência da informação nos programas de pós-graduação das universidades federais brasileiras.

Katz e Martin (1997) afirmam que a colaboração científica é definida pelo trabalho conjunto de pesquisadores com um objetivo comum: a produção de conhecimento científico. Balancieri *et al* (2005) apontam que, por meio da colaboração científica, é possível maximizar o potencial de produção e elevar os resultados devido à troca de conhecimentos. Este trabalho apresenta indicadores quantitativos sobre a colaboração entre professores de um mesmo programa de pós-graduação e de programas distintos. A análise da colaboração científica é um ponto central nesta pesquisa.

Por meio da Ciência de Dados e das técnicas de ETL (Extração, Transformação e Carregamento), é possível realizar análises de redes de colaborações científicas para aprimorar o entendimento dos dados. A Ciência de Dados manifesta-se como uma ciência altamente interdisciplinar, utilizando técnicas e teorias de diversos campos. Com seu viés interdisciplinar, emerge como um fator chave em setores como saúde, finanças, segurança pública e comunicação (Porto; Ziviani, 2014). No contexto deste estudo, a Ciência de Dados proporciona a identificação de padrões e *insights* sobre as colaborações científicas, utilizando estatísticas e técnicas de ETL para transformar dados brutos em análises valiosas.

Assim, pode-se inferir que as técnicas utilizadas pela Ciência de Dados são favoráveis para transformar dados brutos sobre colaborações científicas entre pesquisadores dos programas de pós-graduação em Ciência da Informação das universidades federais brasileiras. Esses dados, obtidos em bases de dados, repositórios e sites de revistas científicas, serão transformados em análises valiosas e gráficos para inferências estatísticas, exibindo a colaboração e comunicação científica de forma detalhada e significativa.

Em suma, este estudo pretende fornecer uma análise da produção e colaboração científica entre professores de pós-graduação das universidades federais brasileiras, utilizando técnicas de Ciência de Dados para extrair *insights* significativos e contribuir para a compreensão e aprimoramento da comunicação científica no país.

A relevância da comunicação científica não se restringe apenas à disseminação do conhecimento entre especialistas, mas também impacta diretamente a inovação e o desenvolvimento tecnológico. À medida que novas descobertas são feitas e publicadas, outras pesquisas podem se basear nesses achados, criando um ciclo contínuo de progresso científico. A acessibilidade e a qualidade das publicações são, portanto, fatores críticos para garantir que a informação correta chegue aos pesquisadores de maneira eficiente. Este ciclo de produção e disseminação do conhecimento é especialmente importante em um mundo cada vez mais interconectado, onde a rapidez na troca de informações pode determinar o sucesso de inovações tecnológicas e científicas.

Além disso, a comunicação científica desempenha um papel fundamental na educação e na formação de novos pesquisadores. As publicações científicas são frequentemente utilizadas como material didático em programas de pós-graduação, e a participação dos alunos na produção de artigos científicos é incentivada como parte de sua formação acadêmica. Este processo não só contribui para a capacitação dos estudantes, mas também para a ampliação do conhecimento disponível. A análise da produção científica nas universidades federais brasileiras, focada nos professores de pós-graduação, também nos permite compreender como essas instituições estão contribuindo para a formação de novos pesquisadores e para o avanço do conhecimento científico no país.

Outro aspecto relevante da comunicação científica é a sua influência na política científica e na alocação de recursos. Indicadores de produção científica e métricas de impacto são frequentemente utilizados por agências de fomento para decidir quais projetos receberão financiamento. Dessa forma, a capacidade de publicar em revistas de alta qualidade e de alcançar um alto número de citações pode determinar o sucesso de um pesquisador em obter os recursos necessários para continuar suas investigações. A análise das métricas de coautoria e produção científica entre os professores de pós-graduação nas universidades federais brasileiras fornecerá

insights valiosos sobre a eficácia das políticas de incentivo à pesquisa e sobre as áreas que podem necessitar de maior apoio.

A integração das técnicas de Ciência de Dados neste estudo não só facilita a análise de grandes volumes de dados, mas também permite a identificação de padrões e tendências que seriam difíceis de detectar de outra forma. Por exemplo, ao mapear as redes de colaboração científica, podemos identificar quais pesquisadores e instituições estão mais interconectados e quais são os *hubs* de produção de conhecimento. Esses *insights* podem ser utilizados para fomentar novas colaborações e para otimizar a distribuição de recursos, promovendo um ambiente de pesquisa mais eficiente e colaborativo.

Ademais, as análises podem revelar disparidades na produção científica entre diferentes regiões e instituições, destacando áreas que podem estar sub-representadas ou que enfrentam desafios específicos. Com essas informações, políticas mais direcionadas e eficazes podem ser desenvolvidas para apoiar essas áreas, garantindo uma distribuição mais equitativa dos recursos e oportunidades. Este aspecto é particularmente importante no contexto brasileiro, onde as disparidades regionais podem afetar significativamente a capacidade de produção científica e o impacto das pesquisas realizadas.

Por fim, a utilização de dados obtidos de diversas fontes, como bases de dados acadêmicas, repositórios institucionais e o site da CAPES, proporciona uma análise abrangente e detalhada das colaborações científicas. Esses dados serão transformados em gráficos e análises estatísticas que ilustram claramente a dinâmica da produção científica e das redes de colaboração entre os pesquisadores de pós-graduação das universidades federais brasileiras. A visualização dessas informações permite a compreensão dos resultados proporcionando a identificação de áreas de melhoria, contribuindo para o fortalecimento da comunicação científica no Brasil.

Em conclusão, este estudo não apenas busca analisar a produção e colaboração científica entre professores de pós-graduação das universidades federais brasileiras, mas também pretende fornecer insights que possam orientar políticas e práticas de incentivo à pesquisa. Por meio do uso de técnicas avançadas de Ciência de Dados, espera-se revelar padrões e tendências que contribuam para o

aprimoramento da comunicação científica, promovendo um ambiente de pesquisa mais colaborativo e eficiente. A disseminação de conhecimento e a colaboração entre pesquisadores são pilares essenciais para o avanço da ciência e, conseqüentemente, para o desenvolvimento social e tecnológico do país.

1.2 Justificativa

Na atualidade, é possível se deparar com diversas fontes de dados que disponibilizam informações sobre pesquisadores, programas de pós-graduação pertencentes à Ciência da Informação das universidades federais brasileiras e a produção científica desses pesquisadores. Esses dados são encontrados em bases de dados, repositórios, sites de periódicos, Plataforma Lattes, Portal de Periódicos da Capes. No entanto, essas fontes de dados normalmente não estão integradas, inviabilizando a percepção clara das informações sobre os casos de colaboração científica entre os pesquisadores que publicam seus trabalhos científicos em diversas áreas do conhecimento.

A integração e análise desses dados são de extrema importância para a Ciência da Informação, bem como para as áreas de cientometria e bibliometria. Esses indicadores fornecem uma compreensão detalhada de como a ciência se comporta e como funcionam os meios de comunicação científica. Através desses dados, é possível mapear colaborações, identificar tendências emergentes, e avaliar o impacto e a disseminação do conhecimento científico em contextos específicos.

Diante desse cenário, mesmo havendo dados relacionados aos programas de pós-graduação, aos pesquisadores e à comunidade científica, a falta de integração entre as fontes citadas acaba dificultando ou impedindo a utilização desses dados como um recurso estratégico para a análise de indicadores relacionados à colaboração científica. Essa fragmentação de informações impede uma visão holística do panorama científico, limitando a capacidade de enxergar de forma macro e até mesmo micro o comportamento de uma comunidade científica ou até mesmo dos meios de comunicação científica.

Por meio da Ciência de Dados e suas técnicas de extração, limpeza, carga e apresentação de gráficos estatísticos, é possível amplificar a capacidade exploratória. A Ciência de Dados fornece métodos de pesquisa inovadores para as ciências

naturais e sociais, auxiliando na conduta da presente pesquisa (Zhu; Xiong, 2015). Essas técnicas permitem integrar e analisar grandes volumes de dados, revelando padrões ocultos e oferecendo *insights* que seriam inacessíveis através de métodos tradicionais devido a volumetria de dados.

Além disso, a Ciência da Informação se beneficia grandemente dessa integração de dados ao permitir uma análise mais aprofundada e precisa de diversos aspectos relevantes, como a produtividade científica, o impacto das publicações, a rede de colaborações entre pesquisadores, e a eficiência dos programas de pós-graduação. Essas análises são essenciais para o desenvolvimento de políticas científicas mais eficazes, para a melhoria contínua dos programas acadêmicos e para o fortalecimento da comunidade científica.

A cientometria e a bibliometria, subáreas da Ciência da Informação, utilizam esses indicadores para avaliar o desempenho científico, identificar áreas emergentes de pesquisa, e mapear as dinâmicas de colaboração entre pesquisadores e instituições. Esses estudos são fundamentais para entender como o conhecimento científico é produzido, disseminado e utilizado, além de fornecer subsídios para a gestão e a tomada de decisões em instituições acadêmicas e de pesquisa.

Finalmente, com o objetivo de reduzir esse óbice, a presente pesquisa visa usar as técnicas da Ciência de Dados para proporcionar análises detalhadas e robustas, em busca de conhecimentos que possam contribuir para o entendimento da colaboração científica entre os pesquisadores vinculados aos programas de pós-graduação da área da Ciência da Informação das Universidades Federais Brasileiras. Ao integrar e analisar esses dados, espera-se não apenas mapear as colaborações existentes, mas também identificar oportunidades de novas parcerias, otimizar recursos e fomentar uma cultura de colaboração científica mais efetiva e estratégica.

Essa pesquisa também tem o potencial de contribuir significativamente para a Ciência da Informação ao desenvolver metodologias e ferramentas que possam ser aplicadas em outros contextos e áreas de estudo. A abordagem proposta pode servir como um modelo para outras iniciativas que busquem integrar e analisar dados científicos, promovendo uma maior transparência e acessibilidade das informações e

fortalecendo a base de conhecimento disponível para pesquisadores, gestores e formuladores de políticas.

1.3 Pergunta da pesquisa

Como os recursos de Ciência de Dados e da Ciência da Informação podem ser utilizados para gerar indicadores quantitativos que possibilitem a análise da relação entre a colaboração científica e a comunicação em programas de pós-graduação em Ciência da Informação, considerando o conceito da CAPES?

1.4. Objetivo Geral

Analisar a relação entre a avaliação dos programas de pós-graduação em Ciência da Informação das universidades federais brasileiras e seus níveis de colaboração e comunicação científica, utilizando dados integrados dessas instituições.

1.5. Objetivos Específicos

- Realizar uma revisão da literatura para compreender o contexto teórico e metodológico que relaciona a produção e colaboração científica à comunicação nos programas de pós-graduação em Ciência da Informação, considerando o conceito da CAPES.
- Coletar dados abrangentes sobre os programas de pós-graduação em Ciência da Informação das universidades federais brasileiras, incluindo informações sobre os professores, suas áreas de pesquisa e publicações, para identificar potenciais indicadores de desempenho.
- Extrair dados sobre a produção e colaboração científica dos docentes dos programas de pós-graduação em Ciência da Informação, utilizando bases de dados abertas e fontes relevantes, visando quantificar a colaboração e sua relação com a comunicação científica.
- Desenvolver uma base de dados relacional que integre as informações coletadas, estruturada para facilitar consultas e análises detalhadas, possibilitando a criação de indicadores quantitativos para avaliação das relações científicas.

- Analisar a produção científica e as colaborações utilizando a linguagem SQL para realizar consultas complexas, permitindo a identificação de padrões que revelem a relação entre colaboração e comunicação em função do conceito CAPES.
- Visualizar os dados por meio de gráficos e outras representações visuais que evidenciem as relações entre a colaboração científica e a comunicação nos programas de pós-graduação em Ciência da Informação, facilitando a compreensão dos indicadores quantitativos gerados.

2. PROCEDIMENTO METODOLÓGICO

A pesquisa é predominantemente quantitativa, empregando estatísticas para mensurar o fenômeno de colaboração científica entre os professores dos programas de pós-graduação em Ciência da Informação das universidades federais brasileiras. A análise estatística permite avaliar a importância da colaboração científica e seu impacto na comunicação científica (Sampieri *et al.*, 2013).

Além dos aspectos quantitativos, a pesquisa também integra elementos qualitativos. Investigaremos o fenômeno da comunicação científica, incluindo suas características, influenciadores (como a comunidade científica e os meios de comunicação) e os impactos dessa comunicação. Utilizaremos cientometria, leis da bibliometria e informetria para fornecer uma análise abrangente.

A metodologia adotada é uma abordagem mista, combinando elementos quantitativos e qualitativos. Utilizou-se de estatísticas e métricas quantitativas para analisar a colaboração científica e a influência do conceito CAPES, enquanto os aspectos qualitativos permitem uma compreensão mais profunda dos fatores que influenciam a comunicação científica e as redes de colaboração. A pesquisa segue uma estrutura detalhada, desde o levantamento e modelagem dos dados até a validação dos resultados, com o objetivo de testar e confirmar a hipótese principal.

2.1 Hipótese Principal

H1: Entidades de ensino com maior conceito CAPES produzem e colaboram mais cientificamente?

2.2 Procedimento Teórico

O método de pesquisa bibliográfica é utilizado para fundamentar a parte teórica. O levantamento é realizado através de diversas fontes, incluindo livros e bases de dados de acesso restrito e público, como BRAPCI, Google Acadêmico, Portal de Periódicos da CAPES e repositórios institucionais de universidades federais.

A principal ênfase é dada aos artigos científicos, predominantes nas fontes BRAPCI, Portal de Periódicos da CAPES e Google Acadêmico.

Os principais termos buscados foram: “comunicação científica”, “comunidade científica”, “ciência de dados AND ciência da informação”, “colaboração científica”, “data science”, “indicators scientific communication”, “data science and information science”.

2.3 Passos da Pesquisa

2.3.1 Primeira Fase: Levantamento

Nesta primeira fase, foi realizado um levantamento sistemático dos programas de pós-graduação em Ciência da Informação nas universidades brasileiras, incluindo a coleta de dados sobre os professores participantes e suas publicações científicas. É importante frisar que o recorte temporal para seleção dos dados foi do período de 2013 até 2021.

- Identificação dos programas de pós-graduação (mestrado e doutorado) em Ciência da Informação das universidades brasileiras e dos professores participantes.
- Pesquisa das publicações científicas dos professores nas diversas fontes de dados, como repositórios, bases de dados, sites de revistas científicas e a base de currículos Lattes.
- Análise qualitativa das publicações e dos perfis dos professores para compreender o contexto da produção científica.
- Principais fontes de dados: Plataforma Sucupira, BRAPCI e Portal de Periódicos da CAPES.

2.3.2. Segunda Fase: Modelagem

A segunda fase, foi realizada a modelagem dos dados, criando um banco de dados relacional que integra informações sobre a produção e a colaboração científica dos professores dos programas de pós-graduação identificados.

- Criação de um banco de dados relacional com dados de produção e colaboração científica dos professores dos programas de pós-graduação identificados.
- Inclusão de informações sobre as universidades, os professores e os programas.
- Modelagem qualitativa para entender as redes de colaboração e a influência das instituições na produção científica.

2.3.3. Terceira Fase: Extração, Transformação e Carga de Dados

Já terceira fase, foram realizadas as etapas de extração, transformação e carga de dados, visando armazenar informações relevantes no sistema gerenciador de banco de dados MySQL e prepará-las para análise.

- Extração e armazenamento dos dados relevantes no sistema gerenciador de banco de dados MySQL.
- Limpeza e integração dos dados para garantir a pertinência para visualização e análise.
- Processamento de dados qualitativos para identificar padrões e temas recorrentes na colaboração científica.

2.3.4. Quarta Fase: Geração de Gráficos Analíticos (Dashboard)

Nesta quarta fase, foram gerados gráficos analíticos e dashboards utilizando ferramentas como Power BI e Neo4j, com dados previamente tratados para facilitar a análise.

- Modelagem e carregamento dos gráficos com dados tratados, utilizando ferramentas como PowerBI e o Neo4j.
- Criação de dashboards que integrem visualizações quantitativas e qualitativas, facilitando a análise dos padrões de colaboração e comunicação científica.

2.3.5. Quinta Fase: Validação

O último passo, foram construídas e aplicadas métricas quantitativas e qualitativas para verificar a confirmação da hipótese principal, seguidas de uma

análise crítica dos resultados para entender a relação entre o nível de conceito CAPES e a produção e colaboração científica dos professores.

- Construção e aplicação de métricas quantitativas e qualitativas para verificar se a hipótese principal pode ser confirmada.
- Análise crítica dos resultados para compreender a relação entre o nível de conceito CAPES e a produção e colaboração científica dos professores.

3. TRABALHOS RELACIONADOS

O estudo realizado pelos autores Balencieri *et al* (2005) constrói uma linha histórica das pesquisas envolvendo as redes de colaboração científica por meio de uma evolução cronológica e suas principais abordagens de estudo. No artigo apresentado, realiza-se uma construção através de rede de nós do âmbito de conhecimento da Plataforma Lattes com os seguintes elementos: Egressos, colaboradores, redes-GP. O principal ponto dos autores é mostrar que, por meio da construção dessas redes de conexão e pelas TICs, é possível experienciar análises por meio desses volumes de dados.

Cajazeira e Silva (2021) abordam discussões sobre a colaboração científica e a coautoria de artigos de periódicos publicados por doutores com vínculo a grupos de pesquisas em Comunicação da Universidade Federal do Piauí. O problema de pesquisa abordado pelos autores se alinha a esta pesquisa de mestrado pois busca-se estabelecer indicadores de produção científica das redes sociais de coautoria formado por esses pesquisadores e entender suas relações sociais, demanda entender padrões de crescimento, verificação de tendências, e compreensão de quais causas relacionais influenciam no aumento ou diminuição da produção dos artigos analisados. Um dos pontos de destaque da relação do trabalho de Cajazeira e Silva (2021) com este trabalho é a análise da construção dos gráficos de nós para entender a produção em geral e, do mesmo modo, de outros dados para entender causalidades entre as variáveis analisadas.

O nicho tratado pelos autores Hayashi *et al* (2012) trata-se da rede de colaboração científica de Educação Especial e Fonoaudiologia da Universidade Federal de São Carlos. A metodologia utilizada pelos autores consiste na análise de redes sociais que possibilitou a construção de uma rede de colaboração científica por meio das relações de autoria e coautoria. Neste trabalho, os autores buscaram exibir por meio de *cluster* um tipo de relação para provarem que a análise de redes de colaboração científica no campo da Educação Especial e Fonoaudiologia pode contribuir para o desenvolvimento de futuras pesquisas.

Os autores Hilário, Caires e Castanha (2014) realizaram um estudo bibliométrico utilizando a temática 'colaboração científica' na Brapci em determinado período, para a construção de uma rede de colaboração científica. Por meio dos dados coletados, os autores criaram vários grafos exibindo as redes de colaboração científica entre os pesquisadores, e conseguiram exibir conclusões de análises inferenciais como o aumento de produção do tema escolhido em determinado período, o aumento do número de pesquisadores do tema e diversos outros elementos.

Maia e Caregnato (2008) realizaram análises por meio de estudos bibliométricos sobre redes de colaboração científica envolvendo os professores do programa de pós-graduação em Epidemiologia da Universidade Federal de Pelotas. Neste trabalho também é elaborado o estudo de redes sociais entre os professores e, por meio de coautorias dos artigos publicados, são construídos os grafos de colaboração científica. No estudo apresentado, os autores testam algumas hipóteses de causalidade, porém, não há comprovação entre certas variáveis.

Oliveira e Grácio (2009) produziram um trabalho com a proposta de analisar a rede de colaboração científica na área da Ciência da Informação envolvendo o tema de 'estudos métricos', por meio de coautorias dos periódicos publicados pela SciELO. Neste trabalho, os autores constroem a rede de colaboração científica envolvendo as instituições escolhidas para analisar a ocorrência de colaboração científica, e como resultado, os autores apresentam alguns agregados de comunidade de forma isolado e outra vezes formando uma relação de colaboração científica.

Danuello e Oliveira (2012) identificaram os programas de pós-graduação em Fonoaudiologia no Brasil, realizando uma apresentação dos docentes e sua relação com as tipologias documentais, demonstrando evolução da produção científica até o ano de 2011. O estudo exhibe a construção de uma rede de colaboração científica entre os docentes e a nível de instituição. Os resultados demonstraram uma descrição quantitativa da quantidade de docentes em determinadas regiões e demonstrou um grafo inicial de rede colaborativa significativa.

Mugnaini, Leite e Leta (2011), utilizam da base Lattes para realizar a análise de desempenho dos pesquisadores brasileiros no cenário internacional, mapeando quais publicações foram declaradas tanto quanto no Lattes e na Web Of Science. É

um estudo que evidencia a Lei de Lotka exibindo o nível de produtividade dos professores escolhidos.

Fuhr, Araújo e Pereira (2016), realizam um estudo com um nível de recorte do ano de 1978 até 2014 para reconhecimento das relações de coautoria entre os servidores do Sistema de Bibliotecas da Universidade Federal do Paraná por meio de sua produção intelectual em todas as edições presentes no Seminário Nacional de Bibliotecas Universitárias.

Rodrigues, Viera e Godoy (2016) realizaram um estudo sobre a produção científica na área de Tecnologia da Informação e Comunicação em bibliotecas. O objetivo foi identificar os autores mais produtivos utilizando a Lei de Lotka e analisar a evolução histórica das publicações, além dos periódicos mais frequentes sobre o tema. Como resultado, o estudo destacou os pesquisadores mais citados, evidenciando o nível de inserção e o impacto desses autores na temática abordada.

4. FUNDAMENTAÇÃO TEÓRICA

4.1 Ciência da Informação e sua contextualização histórica

Miranda (2002) afirma que a Ciência da Informação teve seu advento situado após a 2ª Guerra Mundial com o surgimento do fenômeno denominado como “explosão informacional”. Esse período pós-guerra foi fundamental para um grande progresso científico e como consequência aumentou significativamente o material bibliográfico no mundo devido a necessidade de registro de novas descobertas provenientes da guerra (Silva; Sampaio, 2017).

Como consequência da Segunda Guerra Mundial, houve um enorme aumento em atividades de investigações científicas para fins militares. A comunidade científica estadunidense participou de forma extremamente ativa devido à sua alta participação na guerra. Esse esforço acabou gerando uma gama de documentos científicos e a Segunda Guerra Mundial ficou então conhecida como “a maior explosão de atividades bibliográficas que o mundo havia conhecido” (Fernández Molina, 2002).

Durante os anos 50 e 60, nos Estados Unidos, o congresso e agências governamentais aprovaram diversos programas estratégicos que custearam os esforços de cientistas e engenheiros para o controle da explosão informacional no âmbito da ciência e tecnologia e, eventualmente, em todos os outros campos (Saracevic, 1992).

Vannevar Bush (1945) em seu artigo, apresenta um sistema de informação de forma ideal (Memex), que possui uma capacidade associativa e que poderia ser utilizado por cientistas em seus laboratórios ou em locais de trabalho. Porém, ele notou que os sistemas de informação do seu tempo eram de estruturas hierárquicas, e mesmo que as informações pudessem ter a capacidade de duplicação, poderiam apenas ser utilizadas de forma local.

Fernández Molina (2002) afirma que após as duas Guerras Mundiais, foi criado o *Science Service*, cujo objetivo básico era a polarização da ciência. Watson Davis, em 1928, acaba assumindo a direção do *Science Service* e um ponto nos ideais de Davis deve ser evidenciado: ele possuía ideias parecidas com Otlet em sua base

filosófica, positivista e tinha uma devoção pela ciência chamada de quase “espiritual”. Para ele, a difusão da ciência deveria seguir três passos:

1. A publicação dos resultados de estudos;
2. A incorporação das referências em uma bibliografia utilizável;
3. A popularização da ciência por meio de periódicos, revistas, rádios e outros meios.

Refletindo neste raciocínio, foi necessário pensar em uma resposta para controle de todo este material bibliográfico decorrente de novas tecnologias provenientes da Revolução Científica e, de acordo com Silva e Freire (2012, p. 2), “é no século XX que acontece o advento da Ciência da Informação emergindo como ciência e prática com um papel de destaque na sociedade contemporânea que, teve a partir da Revolução Científica do século XVI”. Além disso, os autores reiteram que a Ciência da Informação surgiu com a finalidade de diminuir a insegurança e incerteza crescentes da época de uma sociedade pós-moderna em especial, o que tangia as questões informacionais.

A respeito da terminologia, embora o termo seja recente, estudos já vinham sendo realizados, no final do século XIX e no início do século XX, por Paul Otlet e La Fontaine no estudo documental (Carvalho; Crippa, 2013), ambos completam ressaltando que a Ciência da Informação passou por uma crise de identidade, e essa afirmação se baseia no estado que a Ciência da Informação sempre se encontrou em estudos prévios realizados por diversos autores clássicos, em estudos da contextualização da Ciência da Informação e, ainda assim, é uma questão discutida até os tempos contemporâneos sobre os acontecimentos que influenciaram o surgimento da Ciência da Informação.

Belkin (1978) traz em sua obra a afirmativa que a Ciência da Informação ou *Informatics*, desde o início, tem sua caracterização uma autoconsciência que o autor considera como desordenada. O autor afirma que isto ocorre porque a Ciência da Informação é uma ciência interdisciplinar e isso acaba dificultando que estudiosos consigam descrevê-la como algo simples.

Ademais, o autor também reitera que parte dos problemas de caracterização da Ciência da Informação baseiam-se em fatores sociais e políticos que são

enfrentados por quaisquer disciplinas que necessitam de fundos para campos de investigação, fato que pode ser analisado nos dias atuais com a lacuna de pesquisas no campo da Ciência da Informação no Brasil (Belkin, 1978).

Brookes (1980) sugere que qualquer atividade social que tenha o objetivo de se tornar uma ciência, precisa seguir esses pontos: ser teórico assim como prático; começar a analisar ou medir racionalizações de senso comum de fenômenos que podem ser observáveis; a estrutura teórica de uma ciência nunca é completa ou fechada, e por fim; cada aspecto permanece “sempre aberto”, oferecendo a possibilidade de novas pesquisas.

O argumento de Brookes (1980) demonstra aspectos que são encontrados em outros textos sobre a caracterização da Ciência da Informação, que para ser considerada ciência — ou quando envolve a criação de um campo teórico —, não ocorrerá de forma “espontânea” sem motivações, mas sim um processo em que deve ser levado em consideração todo um novo contexto. Além disso, também é exposto que há uma dificuldade em um consenso geral acerca do conceito definitivo da Ciência da Informação e quais são suas delimitações de objetos de estudos.

Elópez Yepes (1995) traz em sua obra uma comparação entre o advento da Ciência da Informação tanto na Europa quanto nos Estados Unidos. O autor também aponta que, para fundamentar a Ciência da Informação, foi considerado como um dos maiores esforços no *American Documentation Institute*, e que muitos tradicionalistas afirmaram que a origem da Ciência da Informação estaria em textos de autores que abordassem problemas causados pela informação científica e na explosão informacional.

Para Fernández Molina (2002), houve uma explicação que o movimento documental liderado por Otlet e La Fontaine não teve o êxito necessário do alcance desejado nos Estados Unidos e, de forma geral, no mundo anglo saxônico. Isto ocorreu devido à existência de uma “rejeição” desses países para se integrarem ao movimento documental. Conseqüentemente, nesses países o estudo e desenvolvimento de técnicas documentais acabaram tomando outro rumo, e como efeito, houve a aparição da ‘*Information Science*’, que rapidamente se tornou foco de estudos científicos.

Corroborando com a afirmação do autor, ao contrário da Europa, o movimento documental europeu iniciado por Otlet e La Fontaine não tomou força, pois seus ideais tiveram pouca disseminação na América do Norte e na Grã-Bretanha. Um dos grandes motivos pelos quais a Ciência da Informação deteve força insuficiente, se dá pela firme oposição por parte da *Royal Society*, que na época era completamente contrária à utilização da classificação decimal de Dewey.

Além disso, por volta de 1962, um grupo de pesquisadores reunidos no *Georgia Institute of Technology* declarou que a Ciência da Informação era a ciência que investigava o comportamento da informação, suas propriedades, forças que regem o fluxo da informação e, finalmente, seus meios de processamento visando o melhor uso da informação. É uma área que se relaciona e deriva de outras áreas, por isso é considerada uma área emergente de novas disciplinas interdisciplinares (Robredo, 2003).

No decorrer da década de 1990, a Ciência da Informação é definida como um campo dedicado às questões científicas, com foco em práticas profissionais voltadas aos problemas em relação à comunicação do conhecimento registrado entre os seres humanos (Saracevic, 1992).

Em 1997, a *Royal School of Librarianship* atualiza seu nome para *Royal School of Library and Information Science*, demonstrando uma nova tendência que era a utilização do termo 'informação' nas bibliotecas (Hjørland, 2000).

1. O autor exemplifica vários termos utilizados na área como por exemplo: Biblioteconomia no século XIX, que pode ser substituído por 'biblioteca e ciência da informação';
2. O termo 'documentação' é considerado um neologismo inventado por Otlet para designar o que chamamos de 'armazenamento e recuperação da informação'.

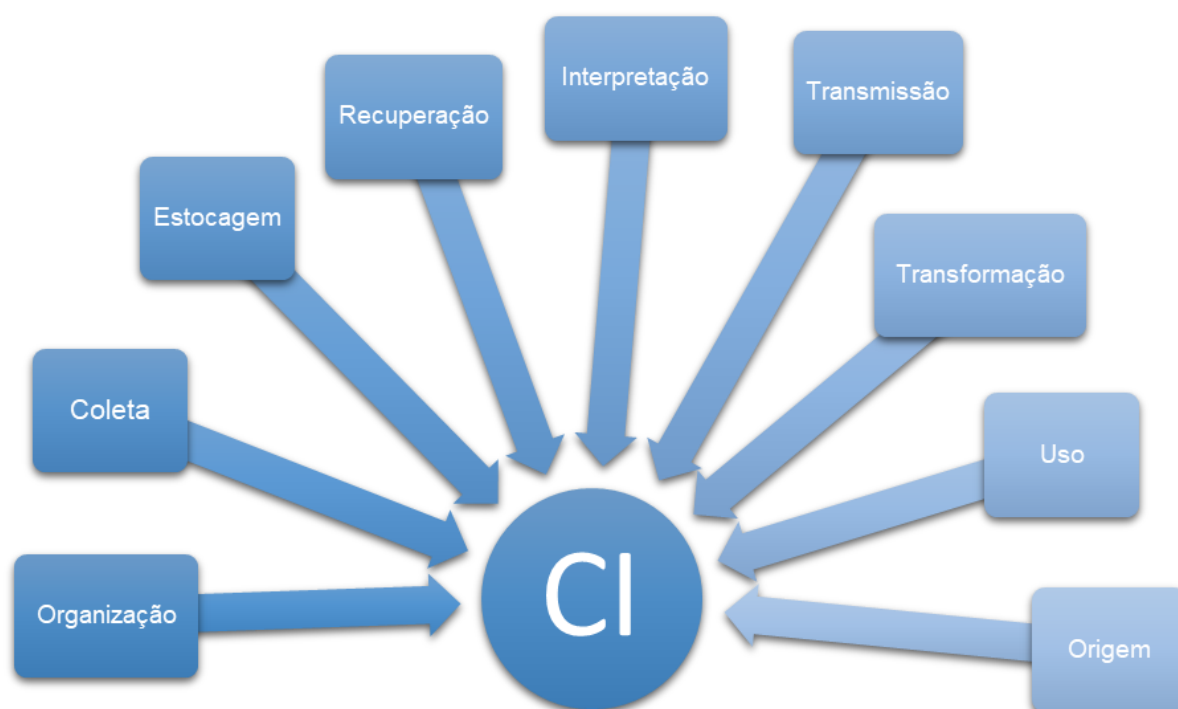
Houve ainda certa dificuldade para diferenciar documentalistas e bibliotecários tradicionais, no entanto, havia um consenso que os documentalistas estavam preocupados não apenas no manuseio físico dos documentos — o que era mais presente em documentalistas do que nos bibliotecários tradicionais —, mas também na exploração da informação.

O autor faz a analogia de que a definição da Ciência da Informação acaba se tornando um caos, uma vez que uma ciência deve ser definida por seus objetos de estudo e não por suas ferramentas (Hjorland, 2000, n.p).

Borko (1968) afirma que a Ciência da Informação é definida como uma disciplina que investiga o comportamento da informação por meio de suas propriedades. A Ciência da Informação possui caráter puro — por meio de pesquisas de fundamentos —, assim como também possui componentes de ciência aplicada — ao desenvolver produtos e serviços. Essas são partes dos comportamentos e propriedades da Ciência da Informação. Além disso, a Ciência da Informação tem como propriedade em seus corpos de conhecimento os seguintes elementos: organização da informação; coleta da informação; estocagem da informação; recuperação da informação; interpretação da informação; transmissão da informação; transformação da informação; uso e origem da informação.

Na Figura 1 é proposto um modelo com base em Borko (1968), este modelo representa parte dos processos e características atreladas à Ciência da Informação, comprovando o seu caráter de ciência interdisciplinar:

Figura 1 - Corpos de Conhecimentos da Informação Ligados à Ciência da Informação



Fonte: Elaborado pelo autor (2021) com base em Borko (1968).

Para Taylor Robert (1966), a Ciência da Informação é uma ciência que investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo e os seus meios de processamento. Ademais, o autor também aponta que a Ciência da Informação se relaciona com diversas áreas como: matemática, lógica, linguística, psicologia, tecnologia de computadores, entre outras.

Com base em Dias (2007), a Ciência da Informação, por se tratar de uma ciência interdisciplinar, é um campo científico que abrange diversas áreas do conhecimento — biblioteconomia, documentação, arquivologia e outras —, sendo subdividida em especialidades e, dentro dessas especialidades a Figura 2 representa algumas das principais voltadas para o campo da Biblioteconomia:

Figura 2 - Especialidades da Ciência da Informação



Fonte: Elaborado pelo autor, com base em Dias (2007).

4.1.1 Objeto de estudo da CI: a informação

Visto que a Ciência da Informação é considerada uma ciência que investiga o comportamento da informação e suas características, a informação se torna o principal objeto de estudo desta área.

Para Goffman (1966), o conceito de informação é utilizado em tantos contextos diferentes que uma única definição não seria suficiente, necessitando de um significado que aborde todos os aspectos que não foram formulados. Além disso, a Ciência da Informação seria uma ciência que estuda fenômenos relacionados à informação e não da informação em si.

Belkin (1978) também afirma que os conceitos de informação são necessários para a Ciência da Informação com base na utilidade potencial, ou seja, a CI deve se preocupar com um complexo sistema de comunicação que contém vários níveis. Desta forma, os conceitos de informação serão úteis para ajudar a isolar e compreender esses vários níveis no processo de comunicação.

De acordo com Le Codic (1970, p. 5), “a informação é um conhecimento inscrito (gravado) sob a forma escrita (impressa ou numérica), oral ou audiovisual”, que tem como objetivo comportar um elemento de sentido. Tem como finalidade comportar um componente que transmita sentido por meio de uma mensagem inscrita em um suporte.

Conforme Brookes (1980), a informação pode ser considerada subjetiva e ao mesmo tempo objetiva, tendo em vista que, para alguns a informação pode ser um conceito claro, mas simultaneamente o senso comum pode ser diferente, sendo considerada uma entidade que permeia toda a atividade humana. Paralelamente a isto, é extremamente difícil de observar os fenômenos da informação.

Dessa forma, a informação pode ser inferida como efeito do processo cognitivo, ou seja, a informação acaba atuando como um intermediário entre o emissor e o receptor de uma mensagem transmitida. Em síntese, a forma como uma mensagem é recebida contribui para a produção do saber, sendo em ciências e humanidades, atuando em parte do fluxo informacional (Queirós, 1999).

Além do entendimento do que se trata a informação, se faz igualmente necessária a compreensão das possibilidades de uso da informação. Para seu uso efetivo, é fundamental que a informação seja acessível, isto é, o processo de organização de informação baseia-se no acesso do conhecimento por meio de métodos de organização (Brascher; Café, 2008).

Tekfo Saracevic (1992) resumiu de forma básica a Recuperação da Informação (*Retrieval Information*) e os novos sistemas de informação:

1. Surgem mediante a explosão informacional;
2. Seus objetivos básicos são a relevância e a orientação do usuário;
3. Utilizar formas coordenadas para representar e recuperar informação;
4. Utilizar a linguagem natural como base da representação da informação;
5. A tecnologia da informação deve ser vista como um aspecto básico da recuperação da informação;
6. O enfoque sistemático será utilizado como método para analisar o projeto e atuação dos sistemas de recuperação da informação.

Referindo-se especificamente ao caso da correlação entre a Ciência da Informação e a Organização da Informação, o autor Guimarães (2014) afirma que há uma preocupação proveniente da necessidade de sistematização e consolidação do conhecimento — seja ele histórico — dentro da sociedade ao qual seu principal objetivo é de transmissão desse conhecimento em nível mais pragmático, bem como a possibilidade de resgate desse conhecimento, buscando como principal objetivo: organizar para recuperar.

De acordo com Brascher e Café (2008), um dos aspectos fundamentais em buscas eficazes é a organização da informação. Segundo Svenonius (2000), para que a informação seja considerada organizada, é essencial que apresente uma descrição minuciosa fundamentada em um enunciado das propriedades de um objeto específico, ou ainda das relações às quais esse objeto está vinculado.

Taylor (2004) já denominava o objeto de estudo da Ciência da Informação, isto é, a informação como uma unidade de informação organizável, e que essa informação deveria estar registrada com textos, registros sonoros, registros de imagens e dentre outros materiais. Esses registros informacionais são denominados por Taylor como '*information package*' (Brascher; Café, 2008).

Um dos principais objetivos da Organização da Informação é a realização da descrição física e, principalmente, do conteúdo de um objeto. Basicamente, para Brascher e Café (2008), a organização da informação trata-se de um processo envolvendo a descrição física do conteúdo dos objetos informacionais abordados por Taylor.

Guimarães (2014) realiza um apontamento no trabalho de Borko (1968), ao qual foi possível observar um conjunto de macroprocessos interligando à produção, organização e ao uso da informação em processos mais específicos, cujo produto se dá por meio de um conhecimento gerado pela sociedade.

Quando a sociedade percebe a concretização ou registro desse conhecimento, ocorre a aplicação do conceito que Michael Buckland (1991) descreve em sua obra como "informação como coisa". No entanto, para o uso social desse conhecimento — que é produzido em um contexto social —, surge a necessidade de adotar um processo mediador conhecido como "substitutos do conhecimento", permitindo assim

que os contextos de produção e utilização desse conhecimento possam ser interligados (Guimarães, 2014; Buckland, 1991).

Para enfatizar os “substitutos do conhecimento”, Guimarães (2008) explica um processo que chama-se processo helicoidal, propondo que muitos acreditavam que esse processo era um ciclo. A Figura 3 mostra que este processo inicia-se com o ‘Conhecimento A’ sendo produzido, em seguida uma ‘Organização B’ é requerida para que haja o ‘Uso C’, em contrapartida, após uso e apropriação do ‘Uso C’, gera-se um ‘Conhecimento D’, que por fim, recebe uma ‘Organização E’, e finalmente, ‘Uso F’, seguindo assim o fluxo deste ciclo.

Figura 3 - Processo helicoidal do conhecimento



Fonte: Elaborado pelo autor (2023), com base em Guimarães (2008).

Parte desses macroprocessos, como indicado na Figura 3, entre a mediação de um conhecimento socialmente produzido e seu uso posterior socialmente, estabelece uma área de estudo conhecida como organização do conhecimento. Guimarães (2008) afirma que "(...) transcende a Ciência da Informação, mas que hoje nela ocupa um dos mais significativos espaços de reflexão teórica, metodológica e, mais recentemente, vem sendo abordada a partir de seu contexto cultural." A organização do conhecimento é um ponto chave desta pesquisa, visto que um dos principais temas do seu estudo é a comunicação científica, à qual esses macroprocessos citados contribuem significativamente para a propagação do conhecimento científico.

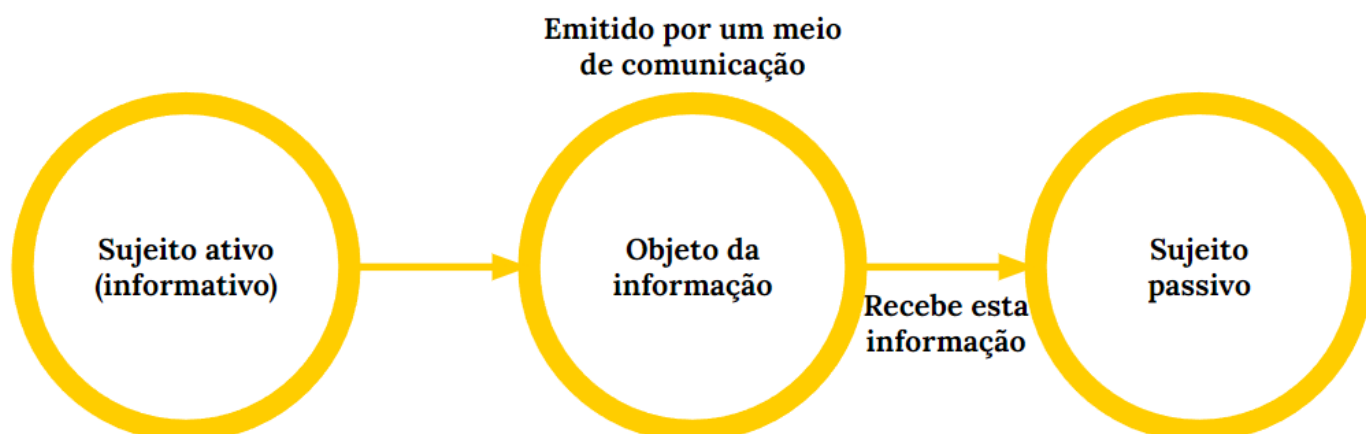
4.2. Comunicação Científica

O princípio da comunicação científica remete ao que considera-se como definição de pesquisa, visto que por meio de pesquisas se dá a necessidade da comunicação de resultados a um determinado público (Alves, 2011). Dessa forma, a comunicação dispõe de um papel central na ciência em razão de que, para ser considerado científico, um determinado conhecimento necessita da aprovação de outros pesquisadores.

De acordo com Alves (2011), as primeiras atividades de pesquisa são prestadas pelos gregos antigos, porém, uma das consequências do aumento explosivo do volume bibliográfico é que um numeroso material passou a se desenvolver, e também, diversos avanços científicos foram impulsionados. Para mais, os autores Pecegueiro e Jesus (2003) apontam que a sociedade já possui compreensão da força da ciência no desenvolvimento do cotidiano e das próprias nações mundiais.

Para Valois *et al* (1989), não existe uma asserção única de conduta do método científico, e bem como a Ciência da Informação, não existe uma teoria única e completamente consensual que englobe todo o fenômeno da comunicação. Porém, um dos métodos de comunicação é apresentado na Figura 4, ao qual um sujeito ativo emite por um meio de comunicação um objeto, e por fim, um sujeito passivo recebe essa informação:

Figura 4 - Partes do Processo Informativo



Fonte: Elaborado pelo autor (2022).

Além de ser um objeto de estudo da Ciência da Informação, o processo informativo também exemplifica um fluxo fundamental na comunicação científica: a transmissão da informação.

A comunicação e a ciência estão fortemente atreladas, em razão da própria comunicação estar inserida na essência da própria ciência, pois é a partir da comunicação que os membros de uma determinada comunidade científica dialogam e transmitem informações de interesse, isto significa que, é graças à comunicação que a ciência se propaga (Maroldi *et al.*, 2019; Meadows, 1999).

A ciência possui um papel fundamental na resolução de problemas por parte dos pesquisadores, uma vez que esses problemas dão origem às pesquisas científicas. Os trabalhos efetuados por cientistas e pesquisadores acabam formulando novos paradigmas e reforçando os existentes, encarregando-se de estimular a propagação científica por meio da comunicação científica (Pecegueiro; Jesus, 2003).

Parte do processo de institucionalização da ciência se atrela em três eixos principais: o universalismo, a objetividade e a neutralidade. Ainda que esses três elementos não sejam valores consensuais e inquestionáveis dentro da produção científica, até o presente momento são fatores que influenciam no pensamento científico (Valois *et al.*, 1989).

No meio da comunicação científica, a validade dos resultados das pesquisas efetuadas por esses pesquisadores e cientistas passam por uma aprovação da comunidade científica. Essas avaliações podem ser entendidas como discussões, análises e estudos por meio de outros pesquisadores na forma de pares (Pecegueiro; Jesus, 2003).

A aprovação se dá em dois momentos: o primeiro ocorre antes da publicação, por meio de um teste de qualidade denominado “avaliação prévia¹”; o segundo ocorre após a publicação quando aprovado na avaliação prévia, sendo publicado como artigo científico e exposto à crítica de todos. Ao ser publicado e acessível aos demais pesquisadores, esse conhecimento pode contribuir para outras pesquisas gerando novos conhecimentos (Mueller, 2007).

Uma vez que esses conhecimentos são publicados, a comunicação científica nasce mediante a inevitabilidade do registro dos quais os avanços científicos e tecnológicos produzidos pelo ser humano sucedem, isto ocorre para que os precursores de diversas áreas do conhecimento possam prosperar daquela bagagem científica, contribuindo para pesquisas pelo bem da humanidade.

Afora a publicação dos conhecimentos científicos, outro fator é de extrema importância para a efetivação de uma ciência como uma disciplina, visto que além de firmes bases filosóficas e conceituais, é imprescindível a existência da confiabilidade do conhecimento produzido. A confiabilidade é uma das características para a distinção da ciência do senso comum, pois se atinge a confiabilidade por meio da divulgação de resultados e a avaliação das pesquisas por meio de pares (Alves, 2011).

A ciência vem sendo difundida por diversos canais de comunicação e os meios eletrônicos forneceram facilidade para que pesquisadores se comunicassem e manifestassem seus resultados com maior rapidez. Graças ao avanço tecnológico, a troca de informação por parte dos pesquisadores acontece de forma imediata, possibilitando não somente a troca de informações com seus pares, mas

¹ É o processo de julgamento que um manuscrito é submetido antes de uma publicação realizada pelos pares (Mueller, 2007).

possibilitando melhorias em seus trabalhos ou contribuições para futuros trabalhos na mesma linha de pesquisa (Pecegueiro; Jesus, 2003).

A Ciência da Informação enquanto área do conhecimento com foco nas questões científicas, se alinha diretamente com a comunicação científica. Isso acontece, pois, ao tratar dos centros de interesses e de ações da Ciência da Informação, destaca-se no que se refere à comunicação científica – conhecer os tipos de publicações, características e formas –, mas havendo a necessidade de compreender também as características próprias da informação científica – sua estrutura de processos e seus sistemas de comunicação –, alinhando ambas em prol do avanço científico (Mueller, 2007).

Dessa forma, ao analisar os pontos de especialidades que a Ciência da Informação abrange e a comunicação científica, destaca-se pontos essenciais de convergência entre ambas, sendo proposto pelo modelo representado no Quadro 1:

Quadro 1 - Convergência entre Ciência da Informação e Comunicação Científica

Ciência da Informação	Comunicação Científica
Processos de Comunicação	<i>Emissor (divulgar avanços científicos) enviar uma mensagem para o receptor (comunidade científica)</i>
Representação da Informação	<i>Representar a informação para que usuários possam recuperar artigos de periódicos ou trabalhos acadêmicos, facilitando a comunidade científica</i>

Fonte: Elaborado pelo autor (2021) com base em Baptista *et al* (2007) e Dias (2007).

A comunicação científica é considerada parte essencial nos estudos da Ciência da Informação, na qual compõe uma disciplina cujo encargo central é atribuído a questões relacionadas – direta ou indiretamente – com o compartilhamento do conhecimento na sociedade (Baptista *et al.*, 2007). Além disso, há dois tipos de comunicação dentro da comunicação científica (Alves, 2011; Le Coadic 2004):

1. Comunicação escrita (canal formal): publicações primárias, resultados de pesquisas;

2. Comunicação oral (canal informal): comunicação oral por meio de conferências, seminários, eventos.

Segundo Targino (2000), a comunicação científica torna-se indispensável para a atividade científica, pois permite a conexão de esforços individuais de membros da comunidade científica, permitindo uma troca contínua de informações e difundindo conhecimento para sucessores ou auferidos de seus predecessores.

O processo da comunicação científica pode ser considerado como um sistema cíclico, pois precisa passar por algumas etapas que devem se repetir para a garantia do avanço científico, bem como a retroalimentação deste ciclo. Na Figura 5 é proposto um modelo para representação do ciclo da comunicação científica:

Figura 5 - Ciclo da Comunicação Científica



Fonte: Elaborado pelo autor (2021) com base em Schweitzer, Rodrigues e Varvakis (2011).

De acordo com Mueller (2006), o sistema de comunicação científica pode ser considerado como a infraestrutura da comunidade científica, pois neste sistema há a garantia da qualidade científica por meio de trabalhos que são validados por pares,

sendo uma forma de evitar redundâncias e preservar autorias de pesquisas, reforçando o ciclo da comunicação científica (Schweitzer; Rodrigues; Varvakis, 2011).

Com a solidificação da propagação da ciência, as comunidades científicas estão cada vez mais solidificadas, estáveis e auto reprodutivas, sendo consideradas independentes e por meio desse ecossistema científico a ciência passa cada vez mais a ser divulgada, garantindo a autenticidade científica ao conhecimento (Maroldi *et al.*, 2019; Deus, 1979).

Um dos meios formais de propagação da ciência é o periódico científico, que assume uma postura de categoria de cientificidade ao passar pela análise de pares entre membros da comunidade científica, desta forma, as comunidades científicas possuem a autoridade do controle do que seria relevante para ser publicado no meio científico, trazendo mérito e reconhecimento aos membros da comunidade científica (os cientistas) (Brito; Lima, 2015; Maroldi *et al.*, 2019).

É importante ressaltar que há uma relação íntima entre “comunidade científica” e “comunicação científica”, uma vez que, uma funciona como pilar para a outra (Mueller, 2006). Infere-se que, para haver uma comunidade científica, é imprescindível a ocorrência da comunicação científica a fim de alimentar o ciclo da comunicação científica e reforçar esse pilar científico.

Conforme Leite e Costa (2007), as universidades são consideradas parte da comunidade científica – como comunidades acadêmicas – por constituírem os elementos do sistema científico e também “são consideradas ainda como o cerne da produção do conhecimento, e os processos de comunicação científica permeiam boa parte de suas atividades, o que permite tanto as trocas internas de conhecimento quanto externas, em interação com comunidades científicas” (Leite; Costa, 2007, p. 94). Nesta linha de raciocínio, a comunidade científica é constituída por produtores de conhecimento científico que buscam novas bagagens acadêmicas, alimentando a comunicação científica.

Quando se abordam os elementos dentro da comunidade científica, constrói-se uma imagem de que os cientistas são os únicos e principais agentes dentro da comunicação científica, levando à visão de que eles são os únicos usuários, porém,

deve-se considerar parte do público geral para ter conhecimento daquele produto científico gerado (Valois *et al.*, 1989).

Desta forma, é importante lembrar que há uma diferença clara entre a comunicação científica e a comunidade científica. Independente da relação íntima entre os dois, a comunicação científica trata-se do veículo e forma de propagação do conhecimento científico, enquanto a comunidade científica designa os tipos de pessoas ou instituições que produzem uma linha de pesquisa. Ressalta-se como o principal objetivo de ambos: produzir e propagar o conhecimento científico, como mostra a Figura 6.

Figura 6 - Objetivo da Comunidade Científica e da Comunicação Científica



Fonte: Elaborado pelo autor, 2021.

Um dos principais resultados da produção e propagação entre a comunidade científica e sua inserção na comunicação científica é o periódico científico. De acordo com Cronin (1984), as quatro principais funções ofertadas pelo periódico científico baseiam-se nos seguintes pontos:

1. Meio de comunicação entre os pares;
2. Exercício do controle de qualidade por meio da avaliação e consenso dos pares;
3. Demonstração da originalidade;
4. Valoração dos autores e pesquisadores.

4.2.1. Indicadores de Comunicação Científica

Desde a década de 1960, as três terminologias - bibliometria, cientometria e informetria - têm aparecido nos campos de estudo da biblioteconomia e, principalmente, da Ciência da Informação. Esses termos são tratados como métricas e, apesar de compartilharem uma origem comum e utilizarem princípios, métodos e ferramentas similares, possuem objetos de estudo e finalidades de pesquisa distintas (Yang; Yuan; Dong, 2020).

De acordo com Beira *et al* (2020), a bibliometria teve seu advento no início do século XX, por meio de estudos desenvolvidos por Cole e Eales em 1917, sendo compreendida naquela época como uma disciplina destinada à avaliação de livros. Conforme Fonseca (1986), a bibliometria é definida como uma técnica quantitativa e estatística que visa medir os índices de produção e disseminação do conhecimento científico. Embora suas raízes estejam na bibliografia estatística, sua aplicação original se deu dentro da Biblioteconomia.

Os métodos bibliométricos vêm sendo aplicados ao longo de séculos, utilizando técnicas matemáticas e estatísticas, com Pritchard contribuindo significativamente para ampliar seu escopo. Broadus também define a bibliometria como um estudo quantitativo de unidades publicadas, tanto físicas quanto digitais (Yang; Yuan; Dong, 2020).

Para Araújo (2016), a bibliometria consiste em técnicas estatísticas e matemáticas com a finalidade de descrever aspectos da literatura e outros meios de comunicação – análise quantitativa da informação. Originalmente conhecida por Hulme em 1923 como “bibliografia estatística” e por Otlet em 1934 como “bibliometria”, a disciplina era inicialmente voltada para a medida de livros – quantidade de edições, exemplares, palavras contidas em um livro, espaços ocupados por livros em bibliotecas – e gradualmente expandiu-se para outros formatos de produção bibliográfica (Araújo, 2006).

A bibliometria é utilizada em pesquisas e pode atuar em conjunto com a ciência de dados, facilitando a análise de dados para obter conclusões científicas. Alvarado (1984, p. 91) afirma que a bibliometria se fundamenta por meio de três leis básicas:

1. A Lei de Bradford: para descrever a distribuição da literatura periódica em uma área específica;
2. A Lei de Lotka: para descrever a produtividade de autores;
3. A Lei de Zipf: para descrever a frequência do uso de palavras em um determinado texto.

De acordo com Araújo (2006), a primeira lei aborda o conjunto de periódicos – Lei de Bradford –, com objetivo de exibir a extensão na qual artigos científicos com assuntos específicos acabavam aparecendo em periódicos destinados a assuntos não-correlatos, estudando a distribuição de artigos levando em consideração as variáveis de proximidade ou afastamento.

A Lei de Lotka, foi formulada em 1926, e construída a partir de um estudo sobre a produtividade dos cientistas, envolvendo a contagem de autores que estavam presentes no Chemical Abstracts. Por meio desta análise, Lotka se deu conta que uma grande proporção da literatura científica era produzida por um pequeno número de autores, e um grande número de pequenos produtores se igualaria, em meios de proporção, ao número reduzido de grandes produtores (Araújo, 2006).

Por fim, Araújo (2006) afirma que a terceira lei clássica da bibliometria – Lei de Zipf – foi formulada em 1949 e descreve a relação entre palavras inseridas em um texto determinado suficientemente grande e a ordem de série destas palavras.

Essas três leis clássicas compõem a bibliometria e suas implicações são usadas em vários contextos acadêmicos e científicos, a fim de auxiliar a análise bibliométrica, provendo valiosos resultados para o avanço científico.

Retomando ao ponto do acelerado desenvolvimento da tecnologia da informação (TI) e da popularização da Internet, é afirmado que informação, materiais e energia se tornaram os três pilares da sociedade humana. Nesse contexto, a informetria atraiu considerável atenção.

O termo “informetria” fez sua primeira aparição com Otto Nacke, que introduziu o termo “informetrie” em alemão, correspondente ao inglês “informetrics”. Nacke expandiu o conceito de Informetria durante o primeiro seminário de Informetria em Frankfurt, em 1980.

A informetria rapidamente se disseminou e foi reconhecida pela Federação Internacional de Informação e Documentação (FID), marcando o surgimento de um novo ramo da disciplina. Segundo Otto Nacke, a informetria foi definida como um estudo que aplica métodos matemáticos aos objetos da Ciência da Informação.

No entanto, Yang, Yuan e Dong (2020) argumentam que essa definição é algo unilateral, pois limita o escopo da informetria à Ciência da Informação. Eles propõem que a informetria deve ser estendida ao estudo quantitativo de qualquer forma de informação, não se restringindo apenas a registros bibliográficos.

Já a cientometria tem parte de suas raízes no trabalho do historiador Derek de Solla Price, além de se desenvolver paralelamente aos índices de citação utilizados por Eugene Garfield. A cientometria é classificada como uma disciplina que, assim como as outras métricas mencionadas, utiliza métodos matemáticos para quantificar as pesquisas científicas. Isso inclui a análise de citações e outros métodos para classificar a atividade científica (Yang; Yuan; Dong, 2020).

O ponto é: se as três métricas utilizam quase o mesmo objeto de estudo e possuem propósitos semelhantes, por que são ramificadas de forma separada? Yang, Yuan e Dong (2020) justificam que esses três termos diferem em seu atributo disciplinar. A bibliometria pertence ao ramo da biblioteconomia e ciência documental, a cientometria pertence à ciência das ciências e, por fim, a informetria pertence à ciência da informação como um todo.

Os autores também afirmam que as três métricas têm sido estudadas em relação a suas origens e inter-relações, contudo, nota-se uma grande falta de consenso, um fenômeno comum na ciência da informação. O uso do termo “bibliometria” como sinônimo das outras métricas gera confusão, apesar de haver um grande consenso sobre a terminologia específica de cada uma.

Com o passar do tempo, é notável que os objetos, objetivos e métodos dessas três métricas sofreram modificações devido à evolução e ao surgimento de novos

elementos, como a Internet das Coisas (IoT). Novos ramos, como a webmetria e a altimetria, surgiram, assim como novos índices e medidas de avaliação. A altimetria, por exemplo, emergiu como um indicador crucial na análise da atividade acadêmica, fundamentada na premissa de quantificar diversos aspectos, como o número de salvamentos, citações em artigos científicos, menções, respostas, entre outros. Sua origem remonta à década de 1990, com o início da webmetria, centrada no estudo quantitativo da web, derivada de métodos bibliométricos (Priem *et al.*, 2010; Thelwall *et al.*, 2005; Torres-Salinas, Cabezas-Clavijo, & Jiménez-Contreras, 2013).

O interesse na mediação, monitoramento e avaliação de atividades científicas e tecnológicas por meio de indicadores gerados por publicações remonta ao século XX, com trabalhos de pesquisadores como Lotka, Zipf, Bradford, Nalimov, Solla Price e Eugene Garfield. Esses autores foram pioneiros na construção da base teórica de áreas importantes dentro da ciência da informação, como a bibliometria e a cientometria (de Melo Maricato; Martins, 2018).

De acordo com Maricato e Martins (2018), grande parte da prática da cientometria e da bibliometria baseia-se no uso de citações para a avaliação e o monitoramento de pesquisas científicas. A análise de citações está diretamente relacionada ao conceito de Fator de Impacto (FI), uma metodologia desenvolvida por Eugene Garfield em 1995, com o objetivo de qualificar revistas científicas por meio de um índice calculado a partir da média do número de citações de artigos publicados nos dois anos anteriores.

Além disso, a evolução da informática e a popularização da internet foram fatores cruciais para a digitalização das produções científicas e a implantação de bases de dados, facilitando o acesso a esses materiais e promovendo a difusão da informação científica.

4.3. Colaboração Científica

De acordo com Grácio (2018), a coautoria no ramo científico, ainda que em pequenas quantidades, já ocorria no século XVII, tendo o primeiro registro de artigo escrito contando com coautoria entre pesquisadores em 1665.

Meadows (1999) também já apontava casos de colaboração científica entre autores por meio de um exemplo clássico de colaboração entre os trabalhos

matemáticos de 1939 de Bourbaki: um pseudônimo. Esse pseudônimo nada mais era que um nome adotado por vários estudiosos da área que realizavam colaborações científicas entre si (Cajazeira; Silva, 2021).

No Quadro 2 é apresentado de forma cronológica um histórico breve de acontecimentos pertinentes dentro da colaboração científica, sendo apresentada por períodos (Balancieri *et al.*, 2005):

Quadro 2 - Cronologia histórica da colaboração científica

Período até a década de 1960	Smith (1958) realiza uma das primeiras reflexões sobre o crescimento de artigos em coautoria, notando ainda que esses artigos poderiam ser utilizados como métricas para analisar a colaboração entre grupos específicos. Há também a percepção da publicação de resultados finais por meio de publicações documentadas por esses pesquisadores envolvidos.
Em 1963	Observações realizadas por Smith (1958) exibem evidências do aumento de autorias múltiplas dentro da ciência. Há o surgimento do fenômeno descrito pelo autor como “colégios invisíveis”, aos quais participavam comunidades informais de pesquisadores que trocavam informações e ideias, resultando em publicações formais de resultados do campo científico.
Em 1965	Hagstrom (1965) observa que alguns professores não incluíam seus orientados em suas relações de colaboradores, e então, por meio de entrevistas Hagstrom questionava os professores se esses trabalhos haviam sido realizados por colaborações com outros e grande parte das respostas consistia em “não”.
Em 1966	Os autores Price e Beaver (1966) apresentam um estudo mostrando que pesquisadores que mantinham encontros por meio de conferências, congressos e eventos acabavam formando uma relação informal, conseqüentemente resultando em futuras colaborações científicas.
Em 1967	Milgram (1967) por meio do trabalho “problema do mundo pequeno” apresenta que cada autor dentro de uma rede pode acabar encontrando um outro autor com pontos semelhantes por meio dos seis graus de separação.
Em 1970	Norman Storer (1970) salienta que grande parte do que se sabia sobre a colaboração científica foi estudado por parte da Bibliometria.
Em 1971	Os autores Meadows e O’Connor (1971) fazem o apontamento que independentemente das razões escolhidas por autores para colaborar, os resultados publicados deverão identificar os autores envolvidos, trazendo a definição de colaboração científica sendo um conjunto de trabalhos de forma cooperativa desenvolvidos por mais de dois pesquisadores e identificado por artigos co-assinados.
Em 1972	Diana Crane (1972) faz a caracterização dos colégios invisíveis por sua alta produtividade, o compartilhamento de pesquisas e treinamento de estudantes. Além disso, destaca-se o trabalho de Nuldeman e Landers (1972), que faz uma correlação entre o número de co-autores e o impacto que um artigo pode trazer.
Em 1978 e 1979	Para os autores Beaver e Rosen (1978), o primeiro artigo científico produzido de forma colaborativa ocorreu em 1678. Porém, em 1979 os autores apontam que a necessidade de contribuição dos autores nasce do objetivo de alcançar o resultado de uma pesquisa e a vontade de desenvolver novas habilidades e conhecimentos

	tácitos.
Em 1980	Goffman e Warren (1980) apresentam que pesquisas realizadas por grandes grupos tendem a ter maior influência dentro do meio científico.
Em 1982	Subramanyam (1983) aponta que algumas vezes um pesquisador pode ser considerado colaborador ou até mesmo co-autor por apenas disponibilizar um material que possui controle para uma pesquisa.
Em 1990	Narin e Whitlow (1990) evidenciam um fator de peso para que artigos com co-autoria sejam mais citados por se tratarem de artigos internacionais.
Em 1992	Leclerc <i>et al.</i> (1992) usa a contagem de co-autoria dentro de artigos como uma métrica para colaboração científica.
Em 1993 e 1994	Katz (1993) evidencia que a proximidade regional acaba encorajando a colaboração científica. Em 1994 é demonstrado ainda que a colaboração científica também é motivada pelo compartilhamento do uso de equipamentos mais caros e pelos padrões de financiamento das agências (KATZ, 1994).
Em 1997	Katz e Martin (1997) evidenciam que experimentalistas tendem a colaborar mais que em produções teóricas. Também é evidenciado que existem colaborações científicas em diferentes níveis.

Fonte: Elaborado pelo autor (2023) com base em Balancieri *et al* (2005).

Dessa forma, pensando o produto da ciência em prol do avanço da sociedade, a colaboração científica é definida por um esforço cooperativo que busca alcançar metas em comum entre os pesquisadores, esforço coordenado e resultados – os trabalhos científicos –, por meio de méritos e responsabilidades compartilhadas (Balancieri *et al.*, 2005). Nesta linha de pensamento, cientistas que trabalham em conjunto podem acelerar suas pesquisas por trabalharem em alinhamento com outros pesquisadores ao buscarem o mesmo objetivo, resultando no enriquecimento da ciência.

A colaboração científica acaba potencializando o crescimento profissional de um pesquisador, uma vez que permite o trabalho conjunto de pesquisadores renomados com pesquisadores iniciantes (Grácio, 2018). Este crescimento profissional gerado pode ser estipulado com base em novas produções científicas e publicações em revistas que assim, eventualmente, servirão de base para contribuição de novos precursores a fim de renovar e manter atualizado o conhecimento científico.

Um exemplo válido para esta constatação pode ser através de: um trabalho conjunto de um professor ao convidar um de seus alunos para escrever um artigo

científico e/ou participar de projetos de iniciação científica em universidades. Uma colaboração dessa não só contribui com o professor, em razão de aumentar sua produção acadêmica, assim como proporciona prestígio a um aluno que ainda não dispõe de produções científicas, oriunda da pouca bagagem acadêmica.

Segundo Katz e Martin (1997), a colaboração científica ocorre com diferentes públicos — nações, instituições, grupos de pesquisas —, potencializando os resultados quando um grupo põe esforço para chegar no mesmo objetivo. Um dos grandes exemplos para a colaboração científica em nações pode ser apresentado durante a pandemia do Covid-19, iniciada no Brasil em 2020, onde cientistas de todo o mundo se mobilizaram para estudar formas de combate à doença, expondo o potencial da colaboração científica.

De acordo com Vilan Filho *et al* (2008), a pesquisa em colaboração científica é impulsionada cada vez mais por meio do governo, agências de fomento, universidades e instituições de pesquisa. Este fenômeno ocorre pela crença da comunidade científica e dos indivíduos responsáveis por políticas científicas de que as colaborações científicas são um ponto positivo para a ciência e precisam ser encorajadas (Katz; Martin, 1997).

Katz e Martin (1997) afirmam que a colaboração científica pode reduzir custos e aumentar os benefícios da pesquisa, por se tratar de um projeto que conta com pessoas diferentes trabalhando em conjunto em prol de um objetivo comum, mesclando e difundindo conhecimentos para o avanço de uma pesquisa.

Os tipos de colaboração científica podem acabar impactando nas motivações entre pesquisadores a efetuarem sua participação, visto que a coautoria é uma dessas facetas mais estudadas dentro da Bibliometria mesmo que seja apenas um espectro (Cajazeira; Silva, 2021).

Os autores Cajazeira e Silva (2021) apontam que alguns pesquisadores costumam escolher projetos de pesquisa baseados nas habilidades que possuem como profissionais e acadêmicos, ou seja, cada pesquisador possui um conjunto único de competências e por meio dessas aptidões busca-se resoluções de problemas. Na busca dessa resolução de problemas científicos, Zollman (2018) infere que cada pesquisador irá encarar um problema de pesquisa de forma diferenciada, sendo que

uma de suas estratégias é a utilização de esquemas conceituais de seus pares. Dessa forma ocorre o que o autor caracteriza como “colaboração unidirecional”, ao qual um pesquisador dispõe um esquema conceitual pertencente a outro pesquisador por meio de algum custo (Cajazeira; Silva, 2021). O custo desse empréstimo de conhecimento pode ser exemplificado como um dos vários motivos que levam pesquisadores a colaborarem e publicarem soluções científicas minimizando esforços e potencializando resultados.

Wagner e Leydesdorff (2005) indicam que a colaboração científica é capaz de ser uma rede de comunicação diferenciada das convencionais, isto ocorre porque essas redes de colaboração científica realizam-se de forma interna e possuem sua própria dinâmica pessoal, de tal forma que, desenvolvem-se como um sistema organizado e formado por redes de pesquisadores que atuam em forma de grupo para produzir conhecimento, tendo como resultado principal as publicações em conjunto (Hilário; Caires; Castanha, 2014).

Uma estrutura de redes de colaborações científicas, geralmente, forma um ecossistema próprio que leva pesquisadores a buscarem e colaborarem em objetivo a um propósito comum. Os autores Hilário, Caires e Castanha (2014, p. 1) afirmam que “(...) a estrutura das redes de coautoria é fruto das escolhas da própria comunidade científica, que pode ser motivada por inúmeros fatores externos”.

Corroborando essa linha de pensamento, os autores Hayashi *et al* (2012, p. 286) afirmam que:

“No âmbito da ciência, o pesquisador pode ser entendido como um ator socialmente interligado, pois o processo de produção de conhecimento exige interações, associações, e união das competências em função de metas comuns, visando o compartilhamento das informações. Essas relações entre pesquisadores fortalecem grupos de pesquisas e ampliam as parcerias existentes criando novas oportunidades para a pesquisa e incrementa a produção científica, especialmente a realizada em colaboração”.

Para Hilário, Grácio e Guimarães (2017), as razões que impulsionam a ocorrência de pesquisas colaborativas são diversas e podem variar de acordo com a área do conhecimento e pesquisadores de um mesmo campo. Dentre esses pontos, destaca-se as 17 possíveis motivações, levantadas por Vanz e Stumpf (2010), fundamentadas por literaturas científicas internacionais e nacionais, apresentadas no Quadro 3:

Quadro 3 - Motivações para Colaboração Científica

1. Desejo de aumentar a popularidade científica, a visibilidade e o reconhecimento pessoal;
2. Aumento da produtividade;
3. Racionalização do uso da mão-de-obra científica e do tempo dispensado à pesquisa;
4. Redução da possibilidade de erro;
5. Obtenção e/ou ampliação de financiamentos, recursos, equipamentos especiais, materiais;
6. Aumento da especialização na Ciência;
7. Possibilidade de "ataque" a grandes problemas de pesquisa;
8. Crescente profissionalização da ciência;
9. Desejo de aumentar a própria experiência através da experiência de outros cientistas;
10. Desejo de realizar pesquisa multidisciplinar;
11. União de forças para evitar a competição;
12. Treinamento de pesquisadores e orientandos;
13. Necessidade de opiniões externas para confirmar ou avaliar um problema;
14. Possibilidade de maior divulgação da pesquisa;
15. Como forma de manter a concentração e a disciplina na pesquisa até a entrega dos resultados ao resto da equipe;
16. Compartilhamento do entusiasmo por uma pesquisa com alguém;
17. Necessidade de trabalhar fisicamente próximo a outros pesquisadores, por amizade e desejo de estar com quem se gosta.

Fonte: VANZ; STUMPF, 2010, p. 50-51.

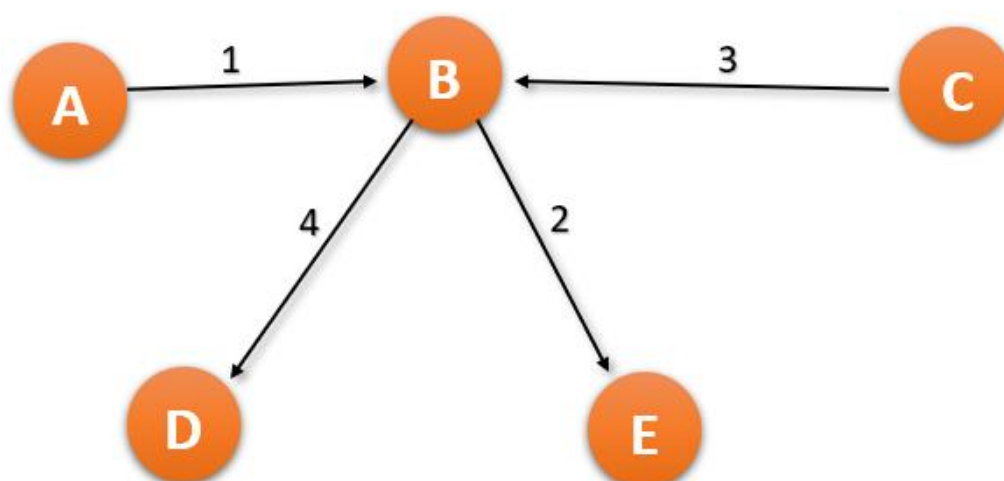
Após seguirem suas motivações e colaborarem, os autores e suas colaborações científicas acabam construindo o esqueleto de uma rede de colaboração científica, que pode ser exibida por meio da análise de redes sociais (ARS). Conforme

Hayashi *et al* (2008), a análise de redes no contexto da colaboração científica é uma abordagem que realiza o estudo das relações entre o produtor de conhecimento e suas ligações, ou melhor, seus laços. Por meio da análise de redes, é plausível encontrar padrões dentro da análise desses vetores e nós (Hayashi *et al.*, 2012).

Através do mapeamento das atividades científicas, proposto por meio de estudos métricos, são representados gráficos de nós e vetores que exibem de forma organizada um panorama de produção científica entre autores (Hilário; Caires; Castanha, 2014). Hilário e Grácio (2013) alegam que a colaboração científica, quando analisada sob a luz da filosofia, acaba trazendo um sistema organizado ao qual para os estudos métricos são de extrema importância para se obter informações valiosas.

Na Figura 7 é proposto um modelo conceitual para exibir como um grafo utilizado em uma ARS é formado. Fazendo uso desses grafos é possível reproduzir de forma visual para analisar uma rede social de professores. Os pontos principais são representados por nós ("A", "B", "C", "D" e "E"), e esses nós podem ser substituídos por outras representações de acordo com a necessidade. Os vetores são representados por setas que indicam a direção e peso entre nós para o cálculo de interação.

Figura 7 - Representação de um grafo de colaboração



Fonte: Elaborado pelo autor (2023).

Dentre os elementos cruciais na análise de redes de colaboração científica, destacam-se as autorias, coautorias e citações. Esses três elementos desempenham

um papel fundamental na representação dos grafos de colaboração científica, sendo também peças-chave na Bibliometria (Hayashi *et al.*, 2012).

Por meio da análise de um grafo de colaboração científica pode-se construir um entendimento da forma como uma determinada rede social está se comportando, e também como evidenciam os autores Oliveira e Grácio (2009):

As análises métricas oferecem subsídios e instrumentação para o estudo das redes sociais na medida em que, a partir de tratamentos quantitativos, torna possível a avaliação de alguns aspectos dessas relações, através de gráficos, densidades, proximidades, similaridades, vetores, intensidades, centralidades e homogeneidades. Assim, a ligação entre dois pontos pode significar não só a existência da colaboração científica entre autores e instituições científicas, mas também a intensidade dessa colaboração na forma de co-autorias (Oliveira; Grácio, 2009, p. 39).

Uma consideração relevante em um estudo de redes sociais, conforme ilustrado na Figura 7, é a localização dos nós em relação aos demais, enfatizando-se a noção de centralidade. A posição central ocupada pelo nó principal, frequentemente localizado no centro, está correlacionada ao poder que esse nó detém sobre os outros pontos (Maia; Caregnato, 2008).

Outro ponto de destaque entre o estudo das redes sociais, avalia-se o grau de proximidade entre nós. No exemplo da Figura 7 observa-se que o nó “B” além de possuir maior centralidade, também possui uma grande proximidade em relação aos outros nós, isto significa que caso esse grafo representasse uma rede de colaboração científica entre professores de uma determinada universidade, o nó “B” apresentaria maior proximidade com outros professores determinando que provavelmente eles iriam colaborar mais ainda em um período futuro.

Por último, nas análises de redes há um último grau chamado de grau de intermediação, determinando a posição que um nó ocupa em relação a outros nós. Exemplificando no contexto abordado da Figura 7, para que os nós “D” e “E” tivessem contado com “A” e “C” seria necessário ter contato primeiro com “B” (Maia; Caregnato, 2008).

4.4. Ciência de Dados

Apesar do termo “Ciência de Dados” ser relativamente novo, a busca pela compreensão desses dados por meio do trabalho de estatísticos, cientistas e profissionais da informação, já vinha sendo abordada em um espaço de discussões antigos (Press, 2013; Rolim, 2018).

A Ciência de Dados teve sua ascensão com o advento do desenvolvimento das tecnologias de informação e também das possibilidades de busca, por meio de mecanismos mais aprimorados como buscas avançadas (Reis, 2019).

A ciência de dados (*data science*), pode ser considerada como um reflexo do ambiente interconectado com sua enorme quantidade de dados disponíveis aos quais se tem conhecimento. Trata-se de uma área interdisciplinar que consegue abordar áreas como as: ciências exatas e engenharias (Comarela *et al.*, 2019).

Reis (2019) afirma que a ciência de dados é considerada a ciência que reúne múltiplos aspectos de informações por meio de dados e conta como uma equipe multidisciplinar envolvendo profissionais de diversas áreas como: matemáticos, estatísticos, programadores, analistas de dados e bibliotecários.

Press (2013) afirma que em 1962, John W. Tukey — estatístico norte-americano — defendia a precisão da realização de análises por meio de dados. Em sua obra *The future of data analysis*, o autor explica que por muito tempo acreditou estar interessado apenas em inferências realizadas do particular para o geral, viabilizando apenas métodos provenientes da estatística clássica. Posteriormente ele percebeu que seu interesse se destacava na área de análise de dados, nascendo então a ideia de que partes da estatística deviam passar por uma análise de dados, assumindo características de ciência, e deixando de ser vista apenas como dados brutos no ramo da matemática (Rolim, 2018).

Os dados possuem uma grande relevância nos meios científicos, tendo participação direta na ciência e em organizações. Essa importância se dá por meio do tratamento de insumos essenciais para conduta e andamento de pesquisas – desde o processo de tomada de decisões até o avanço científico. Todavia, apenas nos

tempos atuais a ciência de dados tem sido tratada como educação formal, destacando seus profissionais e sua importância no mercado de trabalho (Curty; Serafim, 2016).

Analisando uma evolução histórica elaborada por Press (2013), é possível perceber como a ciência de dados teve sua evolução ao longo das últimas décadas. No Quadro 4 é proposto um modelo fundamentado nessa cronologia, apresentando os princípios norteadores da atual ciência de dados

Quadro 4 - Modelo Cronológico dos Princípios Norteadores da Atual Ciência de Dados

Década de 1960	Surgimento de livros, publicações seriadas, artigos, <i>workshops</i> , entidades e encontros de especialistas que abordam análise de dados e <i>big data</i> ;
Em 1966	Lançamento do livro <i>from data mining to knowledge Discovery in databases</i> de Usama Fayyad, Gregory Piatetsky-Shapiro e Padhraic Smyth; o periódico <i>The Journal Data Mining and Knowledge Discovery</i> , lançado em 1967;
Em 1977	O livro <i>Exploratory data analysis</i> , em 1977, de John W. Tukey;
Em 2002	O periódico <i>Data Science Journal</i> ;
Em 2005	O livro <i>Competing on analytics</i> de Thomas H. Davenport, Don Cohen e Al Jacobson;
Em 2009	O artigo <i>Rise of the data scientist</i> ;
Em 2010	O artigo <i>What is Data Science?</i> de Mike Loukides;
Em 2012	O artigo <i>Data scientist: the sexiest job of the 21st century</i> de Thomas H. Davenport e D. J. Patil;

Fonte: Elaborado pelo autor (2021), com base em Press (2013) e Rolim (2018, p. 38).

A partir do fenômeno do *big data*, a ciência de dados surge para utilizar métodos científicos por meio do volume de dados gerado por esse fenômeno para produzir conhecimento (Reis; Sena, 2021). É importante salientar que há uma distinção conceitual entre o *big data* e a ciência de dados, onde a ciência de dados tratará de problemas envolvendo extração de informações por meio de bases de dados complexas e dinâmicas e o *big data* tratará do fenômeno desse volume de dados que

cresceu de forma indescritível e dos avanços recentes derivados da utilização de tecnologias novas (Rautenberg; Carmo, 2019; Bugnion; Manivannan; Nicolas, 2017).

Por meio das cinco dimensões existentes no fenômeno do *big data* (volume, valor, velocidade, variedade e veracidade), é viável a análise dos dados quando se encaixam sob os cinco fenômenos, tornando-se uma fonte de dados confiável para essas análises (Reis; Sena, 2021).

Os três Vs característicos foram denominados por Laney em 2001 e são definidos da seguinte forma (Rautenberg; Carmo, 2019):

- 1) **Volume:** recorrendo ao uso de recursos computacionais de forma abundante, a consequência direta é a formação de grandes volumes de dados, das quais, com o uso e evolução das mídias sociais os usuários estão produzindo um abundante nível de dados como fotos, *tweets*, vídeos, textos e vários outros tipos de dados.
- 2) **Velocidade:** além do exorbitante volume de dados, eles também são gerados em uma enorme velocidade, ou seja, quanto mais recursos computacionais forem evoluindo em forma de produção, captura e processamento, mais dados com maiores velocidades serão gerados.
- 3) **Variedade:** com a enorme massa de dados e a velocidade de geração desses dados, a maior consequência será a origem dessas fontes de dados, atribuindo variedade a esses dados.

Com a evolução das tecnologias de comunicação e informação, outros V's foram adicionados aos três V's citados acima: veracidade, valor e variabilidade (Rautenberg; Carmo, 2019). Além disso, o autor Akhtar (2018) descreve da seguinte forma a existência de 6 V's na Figura 8:

Figura 8 - Representação das novas características do big data



Fonte: Elaborado pelo autor (2024).

Os novos Vs são definidos por Akhtar (2018) da seguinte forma (Rautenberg; Carmo, 2019):

- 4) **Veracidade:** é definido pela integridade e a precisão dos dados, evitando ruídos e incertezas na hora de armazenamento desses dados, de modo que esse processo não deve interferir na análise de tais dados no processo de tomada de decisões.
- 5) **Variabilidade:** apesar de se parecer com variedade em termos de nomenclatura do conceito, trata-se da análise de eventos que podem refletir em padrões de comportamento observáveis presentes nos dados.
- 6) **Valor:** caracteriza-se como um dos pontos mais importantes nos termos dos dados, a valoração dos dados é dada mediante a análise com dados que serão considerados precisos e úteis na tomada de decisão.

Os autores Maneth e Poulouvassilis (2017) afirmam que as principais dificuldades encontradas durante pesquisas em ciência de dados são as seguintes: desenvolver técnicas computacionais capazes de ordenar a gama de variedade e volume dos dados gerados por tecnologias em *web*, móveis e difusas; a proporção de dados produzidos por empresas de grande porte; a aplicação científica e das mídias sociais; a capacidade de desenvolvimento de ferramentas capazes de limpar, transformar, modelar, analisar, e trabalhar com esses dados de forma que cientistas de dados possam mediar com os produtos gerados por meio do *big data*, a fim de garantir segurança, privacidade dos dados de organizações e usuários (Rolim, 2018).

Mediante este contexto, Rolim (2018) sugere a necessidade de um profissional capaz de apresentar soluções e extrair valor dessa quantidade de dados, colaborando

com o surgimento do novo campo de estudo chamado de ciência de dados. Neste ponto, notam-se pontos de habilidades e competências em bibliotecários para lidar com o fenômeno do *big data*, tornando-se relevantes pontos de estudo dentro da Ciência da Informação (Reis; Sena, 2021).

O cientista de dados é considerado como a carreira mais “sexy” do século 21 (Davenport; Patil, 2012), essa afirmação tem como base a grande demanda existente no mercado por profissionais capazes de lidar com a *big data* e sua recorrente necessidade de obtenção de resultados a partir dos dados (Rolim, 2018), que geram vantagens competitivas para as organizações pois auxiliam na prevenção de cenários e facilitam na tomada de decisões.

De acordo com Curty e Serafim (2016), a formação deste perfil profissional surgiu devido a demanda por profissionais que possuíam capacidade analítica e técnica a fim de lidar com grandes volumes de dados. Atualmente, é notável no mercado o crescimento desta área e o interesse por diversas instituições em profissionais capacitados na ciência de dados.

A “era” da *big data* juntamente com a ciência de dados demanda que os profissionais da informação tenham competências fundamentais para lidar com os dados: a organização, a representação de dados, a recuperação e a visualização de dados (Rautenberg; Carmo, 2019).

Por fim, evidenciando esse contexto de procura por especialistas em ciência de dados, foi realizado uma análise pelo *Google Trends* em 2012, mostrando uma crescente busca de usuários de diversos países por informações acerca de termos como “*data scientist*” e “*data science*”, essas buscas foram continuamente associadas com formação profissional envolvendo: cursos, salários, habilidades requeridas e certificação profissional (Curty; Serafim, 2016; Rolim, 2018). Desta forma, fica evidente em como organizações e profissionais buscam oportunidades emergentes desta nova área.

Os autores Smith (2006), Reis e Sena (2021) esquematizam que tanto a ciência de dados quanto os cientistas de dados possuem um ponto em comum com a biblioteconomia: capturar dados, metadados e recuperar. Ao longo de diversos

materiais científicos é comprovado como a biblioteconomia ao longo de sua história já desenvolveu inúmeras técnicas para realizar os processos citados.

O autor Gray (2007) em seu trabalho afirma que as ciências tradicionais acabam se juntando à ciência de dados por meio de técnicas e práticas aplicadas em diferentes campos do conhecimento, logo, há a possibilidade de associação da Ciência da Informação juntamente com a ciência de dados pela volatilidade de adaptação da CI (Reis; De Carvalho, 2022).

Ademais, conforme delineado por Wang em seu estudo de 2018, as instituições educacionais de Ciência da Informação são instadas a adotar uma postura ambidestra, caracterizada pela capacidade de adaptação e pela propensão a engajar-se em diálogos interdisciplinares. Neste contexto, a Ciência da Informação, por sua natureza interdisciplinar, estabelece conexões com a Ciência de Dados, manifestando-se em diversas interfaces. Por exemplo, a perspectiva histórica e pragmática adotada na análise de dados pela Ciência de Dados reflete uma similaridade conceitual com a abordagem da Ciência da Informação em relação à informação, considerada como objeto de estudo central (Fernandes, 2020).

Outro ponto de convergência entre a Ciência da Informação e a Ciência de Dados está fundamentado no desenvolvimento da teoria da informação por renomados cientistas como Alan Turing e Norbert Wiener. Esse avanço levou à adoção dos computadores analógicos e digitais na década de 1940 (Marchionini, 2016).

A gestão da qualidade dos dados, em paralelo com os princípios de governança de dados, delineia um processo de controle que visa não apenas salvaguardar os dados, mas também facilitar sua recuperação. Este aspecto guarda semelhança com a preocupação da Ciência da Informação em relação à organização da informação, visando não só sua preservação, mas também a eficácia na sua acessibilidade. Tanto a Ciência de Dados quanto a Ciência da Informação convergem na busca por estratégias de controle e utilização de metadados, visando facilitar o processo de recuperação desses recursos informacionais.

A disciplina da Ciência de Dados assume uma posição de destaque na capacidade de extrair insights através do processo de extração, transformação e

carregamento (ETL) de dados. Este domínio, empregando ferramentas estatísticas, técnicas de aprendizado de máquina e análise de dados, emerge como um suporte crucial para uma variedade de setores, desde negócios até ciências sociais, na tomada de decisões. Seu escopo inclui a busca por padrões intrincados dentro dos dados, consolidando-se como uma disciplina essencial na era da sobrecarga informacional.

Por outro lado, a Ciência da Informação concentra-se primordialmente na sistematização da informação em suas diversas manifestações e contextos. Abordando tanto aspectos teóricos quanto práticos, ela abrange a produção, organização e recuperação da informação. A Ciência da Informação desempenha um papel significativo na implementação de sistemas de informação, garantindo a acessibilidade aos recursos disponíveis.

Enquanto a Ciência de Dados se concentra na análise desses dados, a Ciência da Informação pode contribuir organizando-os, facilitando sua recuperação e, até mesmo, auxiliando na definição de regras conceituais através de taxonomias e vocabulários controlados. Desta forma, ambas as disciplinas desempenham papéis complementares e cruciais na sociedade contemporânea, especialmente no contexto da big data, e têm potencial para sinergias produtivas, abordando de forma abrangente os aspectos positivos mencionados.

Existe uma variedade de aspectos que podem estabelecer conexões intrínsecas entre a disciplina da Ciência de Dados e os domínios abrangidos pela Ciência da Informação, delineando uma intersecção potencialmente significativa. Essa convergência pode impulsionar o interesse e o envolvimento de profissionais que operam dentro do campo da Ciência da Informação na prática e na investigação em Ciência de Dados. Tanto os profissionais da Ciência de Dados quanto os da Ciência da Informação podem desempenhar papéis como cientistas da informação, preocupando-se com o ciclo de vida dos dados e abordando questões relacionadas à coleta e à utilização desses dados. É pertinente ressaltar que o profissional da Ciência da Informação agrega um elemento distintivo: a ênfase na valorização e importância da informação (Marchionini, 2016).

Os possíveis domínios nos quais um profissional da área da informação pode exercer suas competências no contexto dos dados incluem:

1. Avaliação de questões relacionadas à Lei Geral de Proteção de Dados (LGPD) durante o tratamento de dados;
2. Avaliação, verificação de qualidade e depuração de dados;
3. Catalogação de metadados em conformidade com o contexto dos documentos;
4. Armazenamento e preservação de dados, considerando o ciclo de vida dos dados e enfocando na preservação digital desses elementos;
5. Utilização dos dados para embasar processos de tomada de decisão.

É relevante ressaltar que a Ciência de Dados pode representar um complemento para a Ciência da Informação, e que a Ciência de Dados pode se valer das abordagens da Ciência da Informação para desenvolver programas de treinamento, visando integrar esses dois campos e maximizar os benefícios derivados de suas interseções, com vistas a alcançar resultados substanciais (Marchionini, 2016).

Em conclusão, a interseção entre a Ciência de Dados e a Ciência da Informação revela uma sinergia poderosa e complementar. Enquanto a Ciência de Dados se concentra na análise e extração de insights dos dados, a Ciência da Informação oferece expertise na organização, recuperação e gestão eficiente da informação. O profissional da informação, portanto, desempenha um papel crucial como um mediador entre essas duas disciplinas, aproveitando seu conhecimento em sistemas de informação, taxonomias, vocabulários controlados e metodologias de organização da informação para enriquecer os processos de análise de dados e maximizar seu valor. Ao integrar essas duas perspectivas, os profissionais da informação podem criar soluções mais abrangentes e eficazes para lidar com os desafios da era da informação, contribuindo significativamente para o avanço e inovação em uma ampla gama de campos e setores.

4.5 Dados Governamentais Abertos

A Lei de Acesso à Informação (LAI), sancionada em 2011, representa um marco significativo na promoção da transparência das ações públicas no Brasil, conforme destacado por Victorino *et al* (2020). Este avanço legislativo se insere em

um contexto mais amplo de iniciativas globais voltadas para a melhoria da transparência e da participação cidadã, influenciado pelo movimento de Dados Abertos.

Desde 2009, países como Estados Unidos e Reino Unido têm adotado políticas que visam aumentar a visibilidade das ações governamentais e estimular a participação ativa da sociedade civil nos processos administrativos (Santarém Segundo, 2013; Victorino *et al.*, 2020).

A implementação da LAI no Brasil está alinhada com a adesão do país ao movimento internacional de Governo Aberto, formalmente conhecido como *Open Government Partnership* (OGP), que começou em 2011. A adesão ao OGP é uma estratégia internacional que busca garantir que os governos promovam práticas de transparência e responsabilidade pública.

Desde 2012, o Brasil tem se engajado em diversas iniciativas para fomentar a abertura e o acesso a dados governamentais, o que é evidenciado pela criação de plataformas e ferramentas que facilitam a disponibilização de informações ao público (Victorino *et al.*, 2020).

Um exemplo notável dessas iniciativas é o Portal Brasileiro dos Dados Abertos (PDBA), que foi estabelecido como uma ferramenta essencial para a promoção da transparência. O PDBA oferece acesso irrestrito a uma vasta gama de dados governamentais em formato bruto, permitindo que qualquer cidadão possa consultar e utilizar essas informações de acordo com suas necessidades e interesses. Este portal representa um esforço concreto para cumprir o objetivo da LAI, que é não apenas a disponibilização de dados, mas também a facilitação de sua utilização para aumentar o conhecimento e a participação pública (Pereira; Marcondes, 2014).

Assim, a LAI e as políticas associadas ao movimento de Dados Abertos no Brasil são fundamentais para a construção de uma administração pública mais transparente e responsável. Essas medidas são parte de um esforço contínuo para fortalecer a democracia e promover a participação cidadã, através do acesso aberto e eficiente às informações governamentais.

No contexto atual é possível notar diversos *sites* da esfera pública que contam com o acesso aos dados de forma aberta visando a transparência aos seus usuários. A CAPES disponibiliza uma série de conjuntos de dados buscando promover essa transparência para quem quiser acessá-los, modificá-los e compartilhá-los, sendo uma importante fonte utilizada para pesquisas científicas.

A CAPES conta com uma iniciativa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) para a publicação dos seus dados abertos promovendo a transparência foi por meio do PDA (Plano de Dados Abertos) tendo como propósito a publicação de ações, estratégias a nível organizacional para nortear e promover a abertura desses dados².

² Disponível em: <https://dadosabertos.capes.gov.br/about>. Acesso em: 06 agosto 2024.

5. ESTUDO DE CASO

5.1 Proposta do trabalho

O presente trabalho tem como objetivo localizar bases de dados sobre produção científica, extrair dados dessas bases, limpar, transformar e integrar esses dados em um banco de dados relacional projetado para a construção de uma rede social de colaboração científica entre os professores de pós-graduação em Ciência da Informação das Universidades Federais Brasileiras, buscando exibir redes de conexão de universidades de forma isolada para entendimento do comportamento colaborativo entre os professores de nichos específicos, e ainda exibir de forma macro como esses professores colaboram com outros professores de outras universidades. Este estudo visa também mostrar as variáveis que mais influenciaram professores a colaborarem, e analisar quais programas de pós-graduação e quais universidades possuem maiores índices de colaboração científica.

Para o desenvolvimento deste estudo de caso foram desenhados os seguintes passos:

1. Localizar bases de dados sobre produção científica, extrair dados dessas bases, limpar, transformar e integrar esses dados em um banco de dados relacional;
2. Escolher os softwares para a apresentação dos dados por meio de gráficos e para a criação das redes sociais de colaboração científica;
3. Utilização da estatística descritiva para análises inferenciais;
4. Construção dos gráficos para análises das variáveis escolhidas;
5. Realização das conclusões por meio da análise dos artefatos criados.

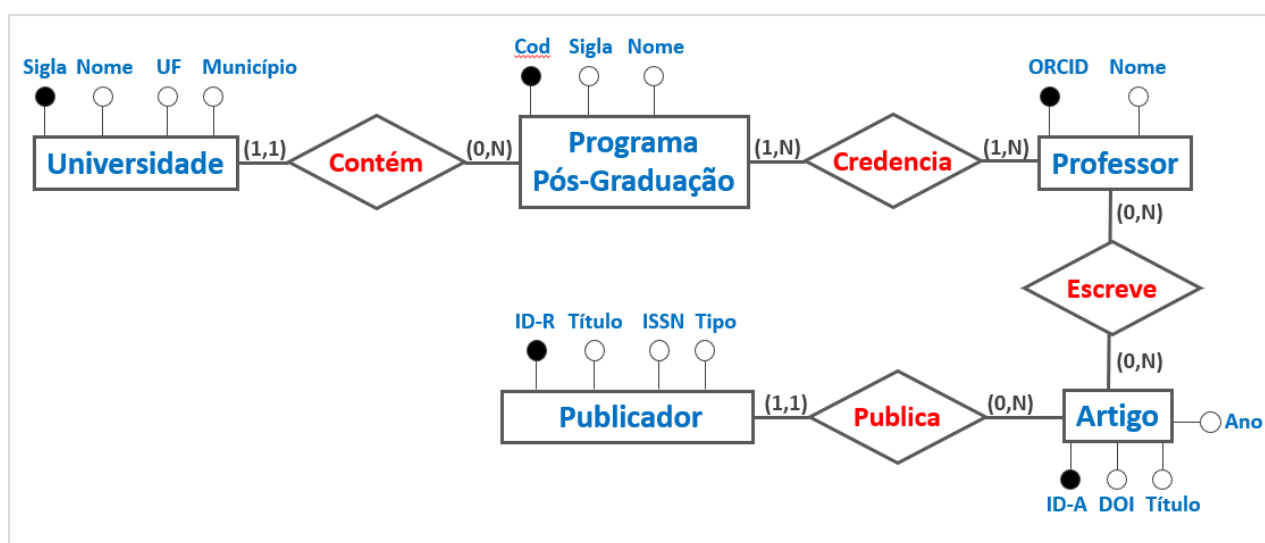
5.2 Fontes de dados para o estudo desenvolvido

Para o desenvolvimento desta pesquisa, foi implementado o modelo de dados conceitual apresentado na Figura 9, baseado na técnica do modelo de entidade-relacionamento proposto por Peter Chen (1976). Os diagramas de ER utilizam símbolos como retângulos e losangos conectados por linhas para representar a conectividade entre as entidades e os relacionamentos entre elas, juntamente com seus atributos (Franck; Pereira; Filho, 2021). O modelo conceitual inicial, conforme

apresentado na Figura 9, define as entidades Universidade, Programa de Pós-Graduação, Professor, Artigo e Publicador. A figura permite associar os tipos de relacionamentos entre cada entidade e suas cardinalidades, onde os números em parênteses indicam como esses relacionamentos se comportam. Por exemplo, uma cardinalidade de (1,1) significa que uma entidade deve existir obrigatoriamente em outra entidade apenas uma vez, enquanto (0,N) indica que uma entidade pode não existir em outra entidade, mas pode existir em vários outros relacionamentos.

Assim, foi definido que um programa de pós-graduação deve estar obrigatoriamente inserido em uma universidade (1,1) e pode existir apenas em uma universidade. Uma universidade pode não ter nenhum programa de pós-graduação, mas pode ter vários programas de pós-graduação (0,N). Professores devem estar inseridos em um programa de pós-graduação e podem estar vinculados a mais de um programa (1,N), e cada programa de pós-graduação deve ter pelo menos um professor, podendo ter vários simultaneamente. Em relação à entidade Professor e à entidade Artigo, o relacionamento se comporta da seguinte forma: um professor pode não escrever nenhum artigo ou pode escrever vários artigos científicos, assim como um artigo tem que ter pelo menos um autor. Por fim, um artigo científico deve ter obrigatoriamente um publicador e será divulgado apenas por esse publicador, enquanto um publicador pode não publicar nenhum artigo ou pode publicar vários artigos.

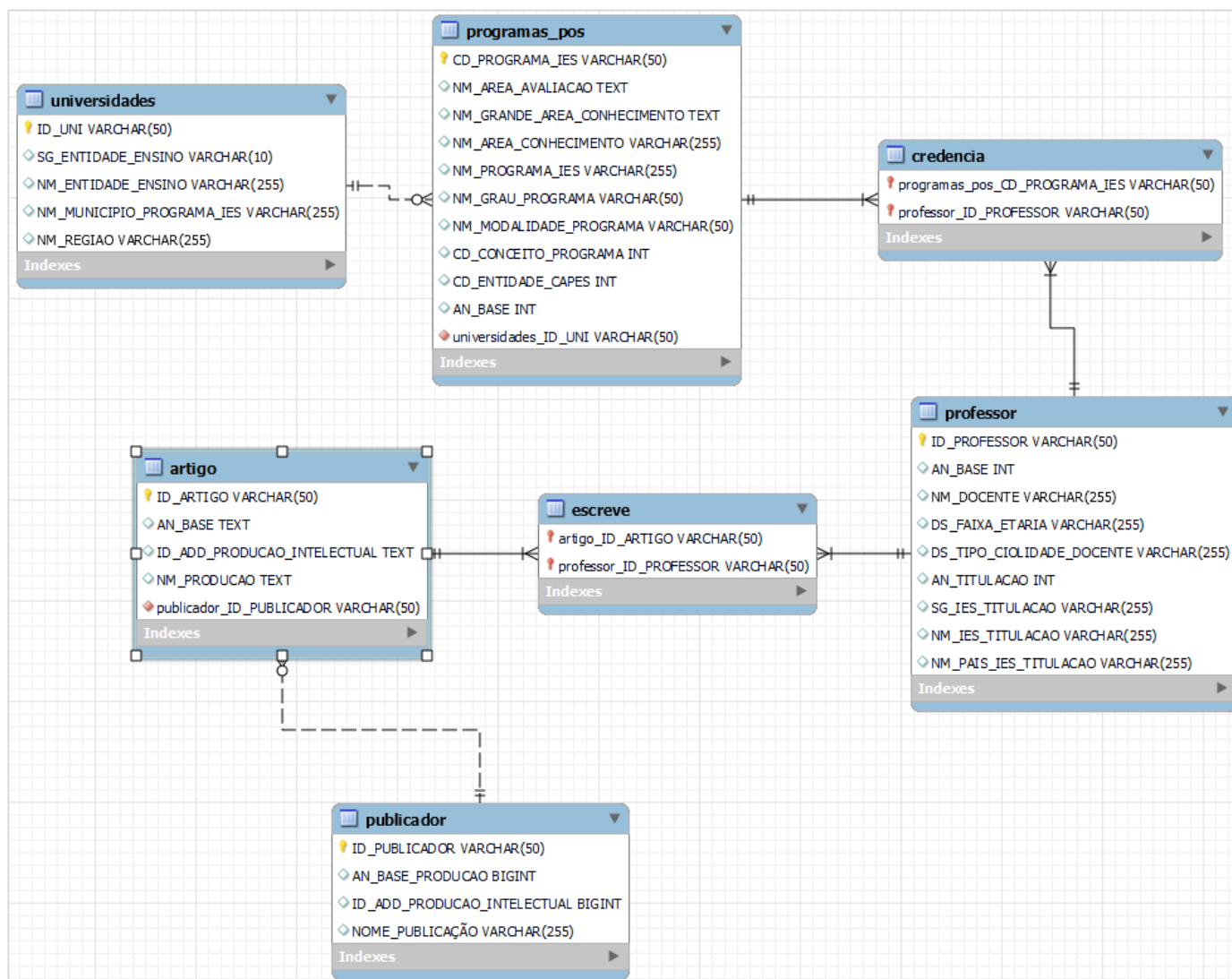
Figura 9 - Modelo Conceitual Inicial



Fonte: Elaborado pelo autor, 2023.

O mapeamento do modelo conceitual apresentado na Figura 9 para o modelo lógico relacional é apresentado na Figura 10 sendo construído utilizando o SGBD.

Figura 10 - Modelo Final Lógico



Por meio dos quadros apresentados serão explicados cada tipo de atributo para cada tipo de tabela diferente:

No Quadro 5 é descrito a tabela universidade que conta com cinco tipos de atributos e totaliza 36 registros dentro da tabela.

Quadro 5 - Atributos da tabela universidade

ID_UNI	Código do programa de pós-graduação existente
SG_ENTIDADE_ENSINO	Sigla da universidade
NM_ENTIDADE_ENSINO	Nome da entidade de ensino
NM_MUNICIPIO_PROGRAMA_IES	Nome do município
NM_REGIAO	Nome da região

Fonte: Elaborado pelo autor (2024).

O Quadro 6 descreve os atributos da tabela relacionamento escreve, essa tabela foi realizada para registrar todos os docentes que produziram artigos científicos. Essa tabela conta com dois atributos e totaliza 29585 registros.

Quadro 6 - Atributos da tabela relacionamento escreve

ID_ARTIGO	Código identificador
ID_PROFESSOR	Código identificador

Fonte: Elaborado pelo autor (2024).

O Quadro 7 descreve os atributos da tabela relacionamento credencia, que foi criada para registrar todos os docentes credenciados por algum programa de pós-graduação. Esta tabela possui dois atributos e totaliza 3.996 registros.

Quadro 7 - Atributos da tabela relacionamento credencia

CD_PROGRAMA_IES	Código identificador
ID_PROFESSOR	Código identificador

Fonte: Elaborado pelo autor (2024).

O Quadro 8 é a representação da tabela de professores incluídos em programas de pós-graduação na área da Ciência da Informação das universidades federais brasileiras. Nesta tabela algumas informações adicionais foram levantadas para mapear melhor o nível de detalhamento do perfil desses professores que contribuem para a comunicação científica. É uma tabela que conta com nove atributos e totaliza 803 registros.

Quadro 8 - Atributos da tabela professores

ID_PROFESSOR	Código identificador
AN_BASE	Ano de exercício
NM_DOCENTE	Nome do docente
DS_FAIXA_ETARIA	Faixa etária
DS_TIPO_CIOLIDADE_DOCENTE	Nacionalidade
AN_TITULACAO	Ano de titulação
SG_IES_TITULACAO	Sigla da universidade de titulação
NM_IES_TITULACAO	Nome da universidade de titulação
NM_PAIS_IES_TITULACAO	País de titulação

Fonte: Elaborado pelo autor (2024).

No Quadro 9 é apresentado a tabela dos programas de pós-graduação totalizando 13 atributos que também tiveram intuito de serem mais detalhados em níveis descritivos para as análises dos perfis de produtores de conhecimento científico a nível de entidades educacionais contando com um total de 143 registros.

Quadro 9 - Atributos da tabela programa-pos

CD_PROGRAMA_IES	Código do programa de pós-graduação existente
AN_BASE	Ano da produção
ID_ADD_PRODUCAO_INTELECTUAL	ID da produção
NM_PRODUCAO	Nome da produção
NM_AREA_AVALIACAO	Nome da área
NM_GRANDE_AREA_CONHECIMENTO	Nome da grande área do conhecimento
NM_AREA_CONHECIMENTO	Nome área do conhecimento
CD_PROGRAMA_IES	Código do programa de pós-graduação existente
NM_PROGRAMA_IES	Nome do programa de pós-graduação
NM_GRAU_PROGRAMA	Grau do programa de pós-graduação
NM_MODALIDADE_PROGRAMA	Modalidade do programa
CD_CONCEITO_PROGRAMA	Nível do conceito CAPES

AN_BASE	Ano base de avaliação
---------	-----------------------

Fonte: Elaborado pelo autor (2024).

O Quadro 10 representa a tabela de publicador, esta tabela é responsável por demonstrar quais os publicadores mais detêm produtores científicos podendo ser eventos científicos ou revistas científicas. A tabela contém quatro diferentes tipos de atributos e totaliza 19712 registros.

Quadro 10 - Atributos da tabela publicador

AN_BASE_PRODUCAO	Ano da produção
ID_PUBLICADOR	Código identificador
ID_ADD_PRODUCAO_INTELCTUAL	Código da produção
NOME_PUBLICAÇÃO	Nome do publicador

Fonte: Elaborado pelo autor (2024).

Por fim, o Quadro 11 apresenta a tabela mais importante para as análises montadas, que é a tabela de produções. A tabela conta com 4 atributos e totaliza 27741 registros.

Quadro 11 - Atributos da tabela artigos

ID_ARTIGO	Código identificador
AN_BASE	Ano da produção
ID_ADD_PRODUCAO_INTELCTUAL	Código da produção
NM_PRODUCAO	Nome da produção

Fonte: Elaborado pelo autor, 2024.

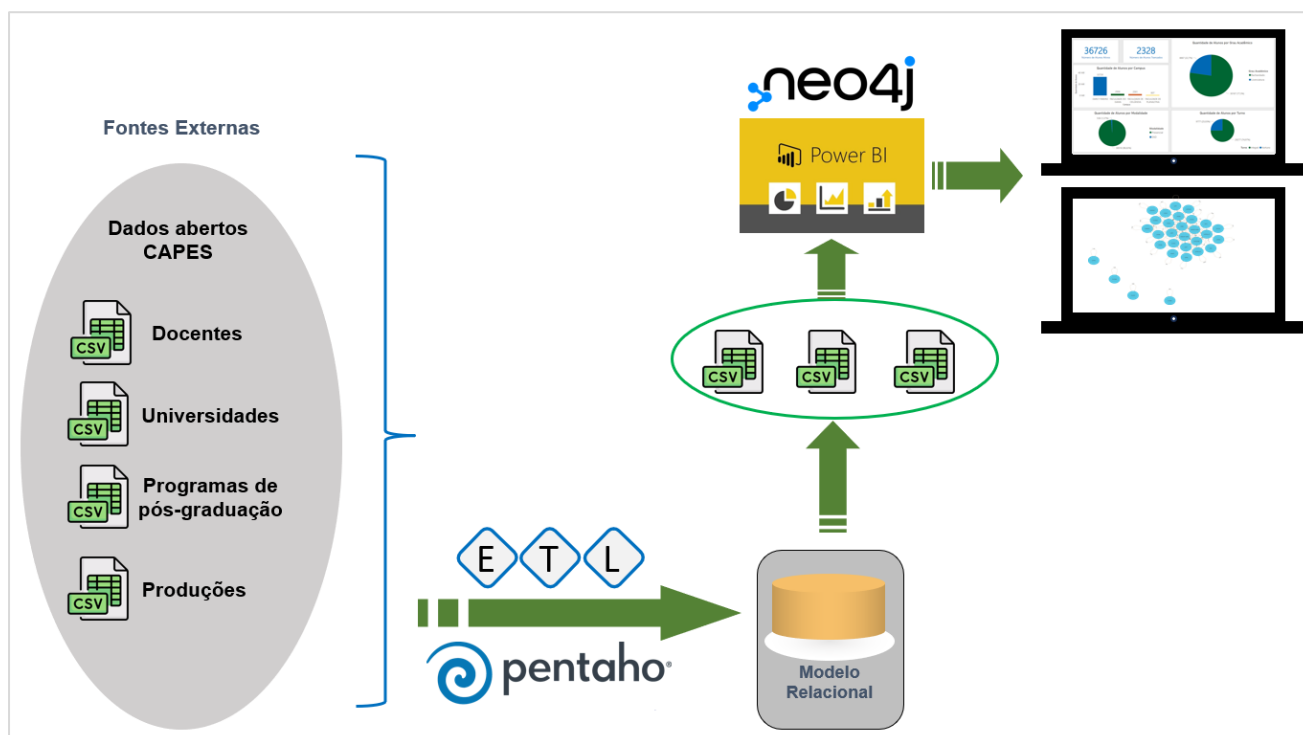
O fluxo da Figura 11 teve seus dados extraídos de fontes governamentais abertas, especificamente dos Dados Abertos da CAPES por meio de CSVs disponibilizados. As categorias selecionadas para a construção desta pesquisa incluem: detalhamento da produção intelectual dos programas de pós-graduação Stricto Sensu no Brasil, docentes da pós-graduação Stricto Sensu no Brasil, produção intelectual dos programas de pós-graduação Stricto Sensu no Brasil e autores da produção intelectual de programas de pós-graduação Stricto Sensu no Brasil. É

importante destacar que a margem temporal definida para a extração dos dados foi de 2013-2021.

A principal dificuldade inicial encontrada no tratamento dos dados foi a volumetria dos arquivos CSV, que continham mais de 70 mil linhas. Este grande volume de dados tornou o trabalho inviável sem o uso de uma ferramenta de ETL (Extração, Transformação e Carga).

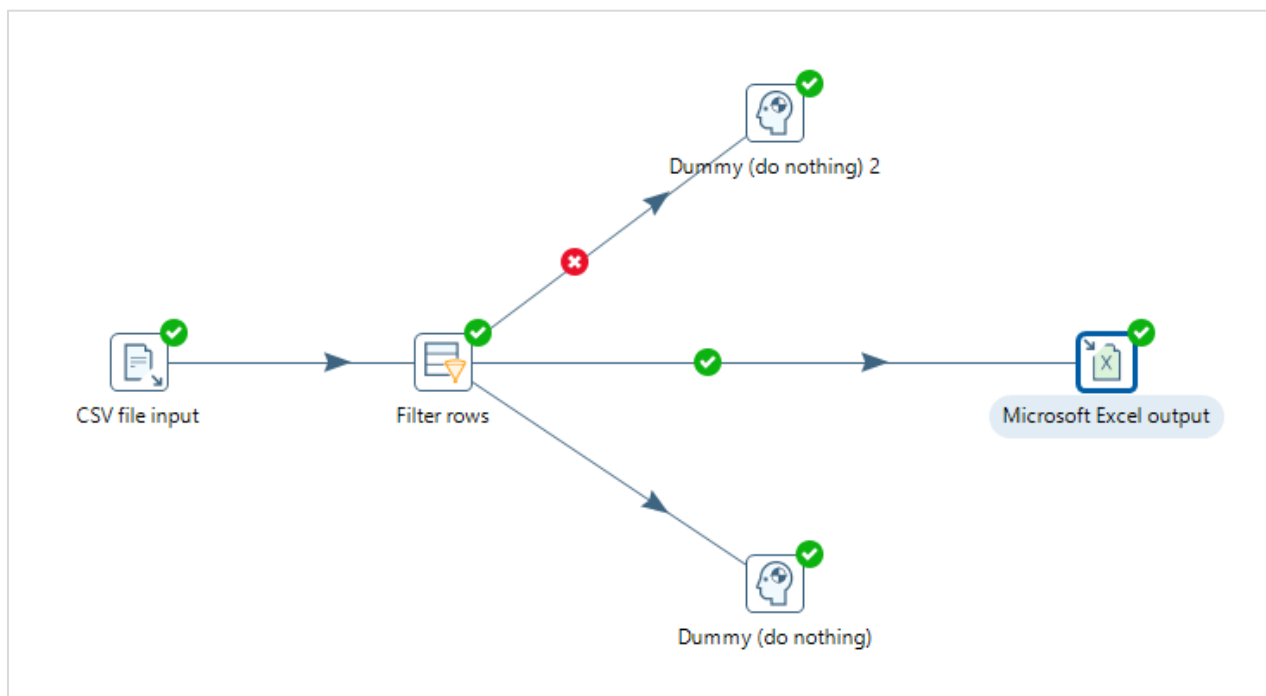
Uma ferramenta de grande ajuda para o tratamento de dados deste trabalho foi o uso do *Pentaho Data Integration*³ que possibilitou a construção de um fluxo para transformação desses dados. O fluxo é apresentado na Figura 12 que funcionou da seguinte forma: a ferramenta lê o CSV escolhido, após isso é acrescentado um componente que realiza o filtro de IDs repassados em forma de filtro e quando ele encontra esses IDs ele gera um novo CSV disponibilizando apenas os dados dentro do escopo (dados referentes a Ciência da Informação).

Figura 11 - Fluxo de dados



³ Disponível em: <https://pentaho.com/products/pentaho-data-integration/>. Acesso em 15 de jul. de 2024.

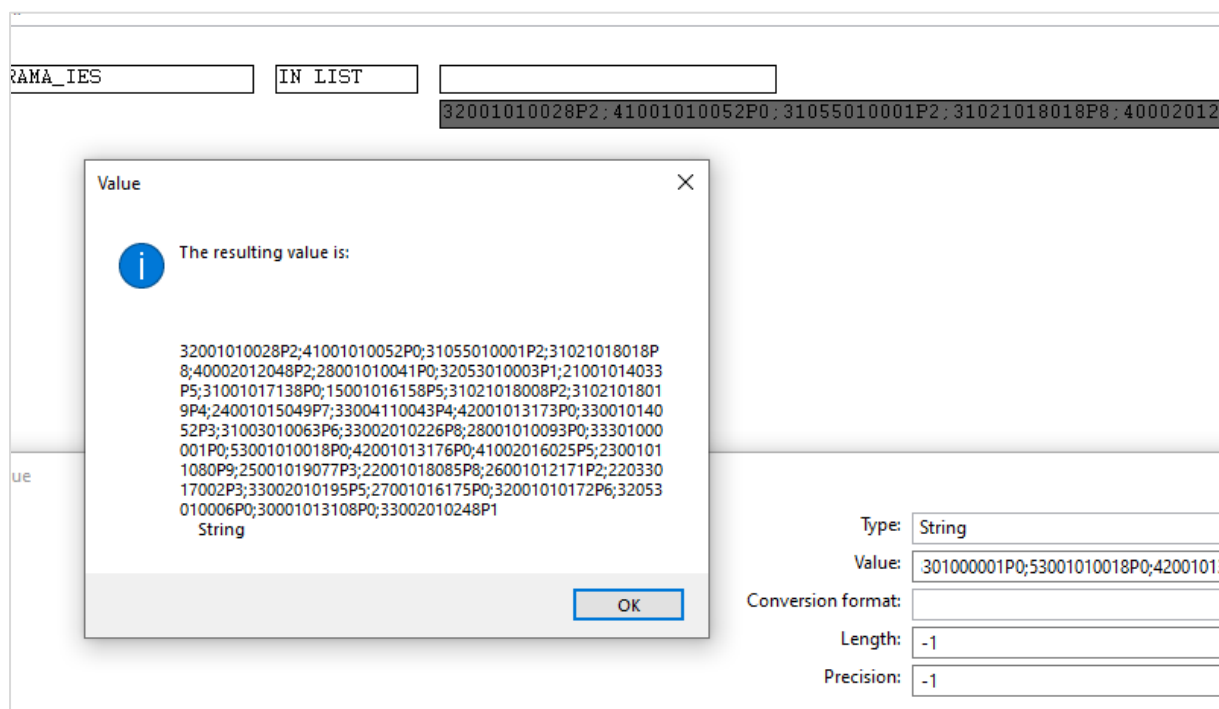
Figura 12 - Fluxo de ETL para tratamento dos dados



Fonte: Elaborado pelo autor, 2024.

A Figura 13 representa a lista fornecida dentro do componente de filtro como *string* para filtrar apenas os IDs dos programas de pós-graduação em Ciência da Informação das universidades federais brasileiras. Dessa forma, a ferramenta de ETL realiza a filtragem desses dados em segundos, eliminando o processo oneroso de abertura de um CSV com milhares de linhas.

Figura 13 - Filtro de IDs para filtrar os códigos de programa de pós-graduação em CI



Fonte: Elaborado pelo autor, 2024.

Após a filtragem dos dados, conforme apresentado nas Figuras 12 e Figura 13, foi realizada uma etapa de limpeza e tratamento. Esses dados foram então utilizados para popular o modelo relacional criado, permitindo a realização de consultas e transformações dentro desse modelo. Posteriormente, um CSV tratado foi gerado e importado para a ferramenta Power BI, onde foram construídos gráficos exploratórios. Por fim, os dados foram transferidos para a ferramenta Neo4j, permitindo a criação dos grafos de redes sociais.

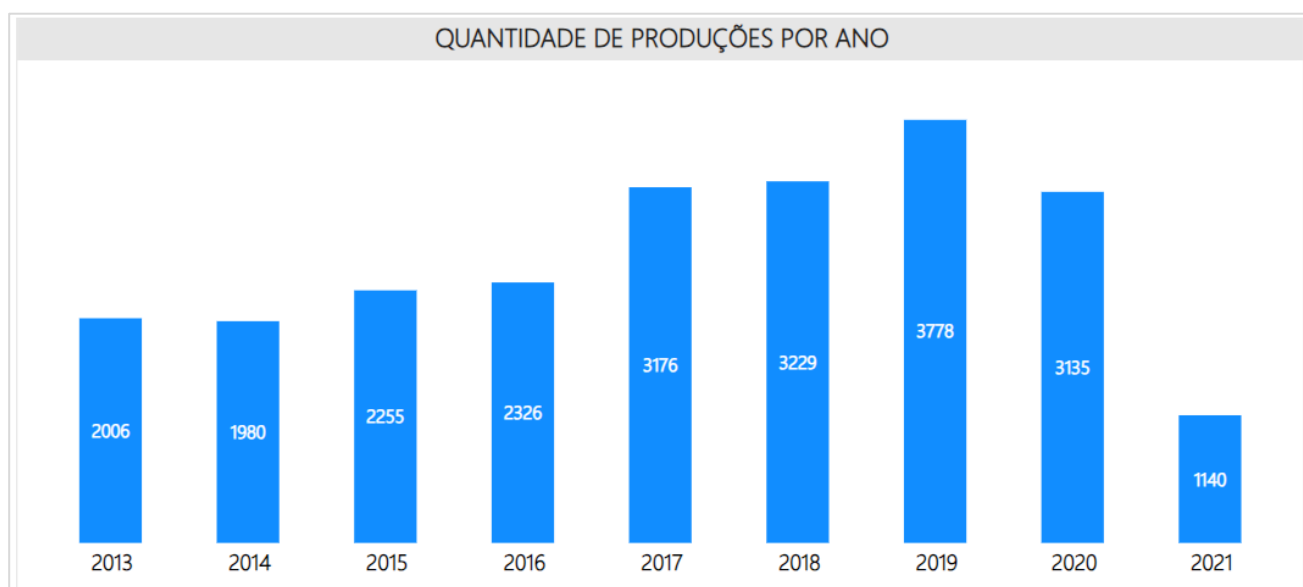
5.3 Resultados e discussões

Nos gráficos a seguir, serão apresentados dados descritivos do banco de dados construído, com o objetivo de explicar o perfil dos dados tratados e representá-los visualmente. O perfil dos dados é de suma importância, pois fornece uma visão detalhada sobre a natureza, distribuição e características dos dados, permitindo identificar padrões, tendências e possíveis anomalias. Essa compreensão aprofundada é essencial para assegurar a qualidade dos dados, facilitar a análise precisa e informar decisões baseadas em evidências. Visualizar esses perfis ajuda a

comunicar insights de maneira clara e eficaz, tornando os dados mais acessíveis e compreensíveis para todos os envolvidos no projeto.

A Figura 14 demonstra a Lei de Bradford aplicada à bibliometria, destacando a evolução histórica das publicações em Ciência da Informação pelas universidades federais brasileiras. O gráfico revela um crescimento constante na produção ao longo dos anos, seguido por uma queda acentuada em 2021.

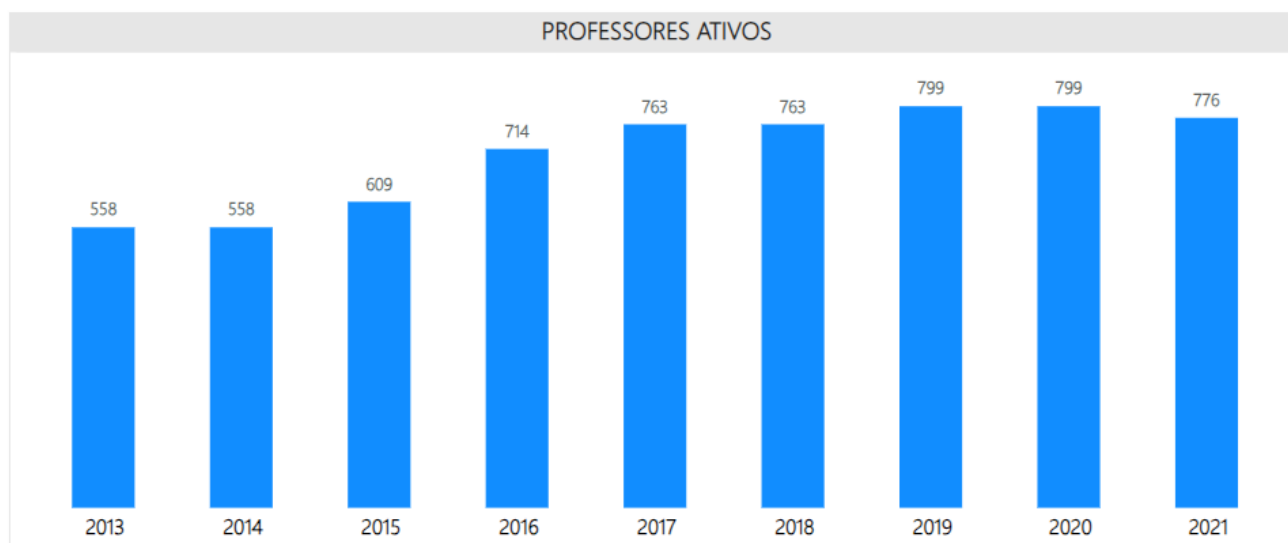
Figura 14 - Quantidade de produções historicamente



Fonte: Elaborado pelo autor, 2024.

Em comparação, a Figura 15 detalha a quantidade de professores ativos em publicações por ano. É possível notar que há um aumento em professores ativos durante os períodos de 2013 até 2016, porém, os números se mantêm estáveis nos próximos anos.

Figura 15 - Quantidade de produções historicamente por professores



Fonte: Elaborado pelo autor, 2024.

Olhando sob o contexto da Lei de Lokta da bibliometria podemos notar quais são os autores mais produtivos entre o comparativo do *ranking* 10 apresentado na Figura 16, havendo uma grande disparidade entre o primeiro colocado e o último que possuem diferença de quase o dobro de publicações.

Figura 16 - Grau comparativo dos dez professores que mais publicam

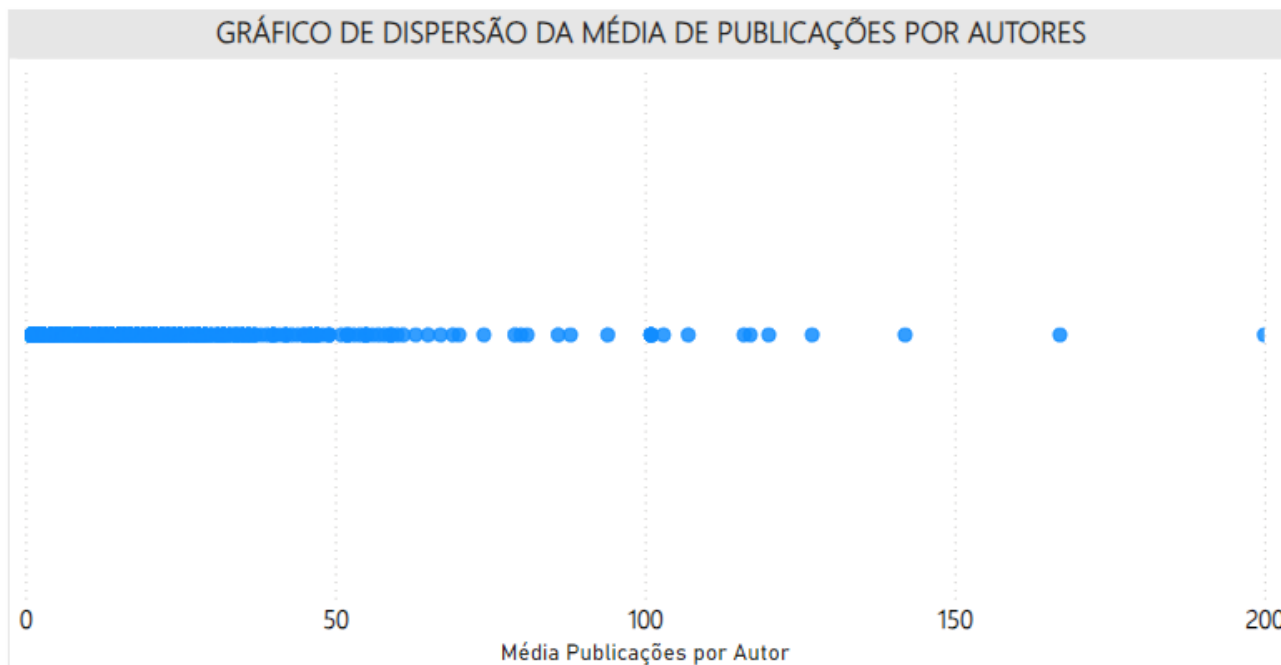


Fonte: Elaborado pelo autor, 2024.

Na Figura 17, observamos uma dispersão na média de publicações por autor, onde a maioria dos professores se encontra no intervalo de 0 a 60 artigos publicados.

Além disso, identificamos um *outlier* com 199 artigos publicados, o que também é evidenciado na Figura 16.

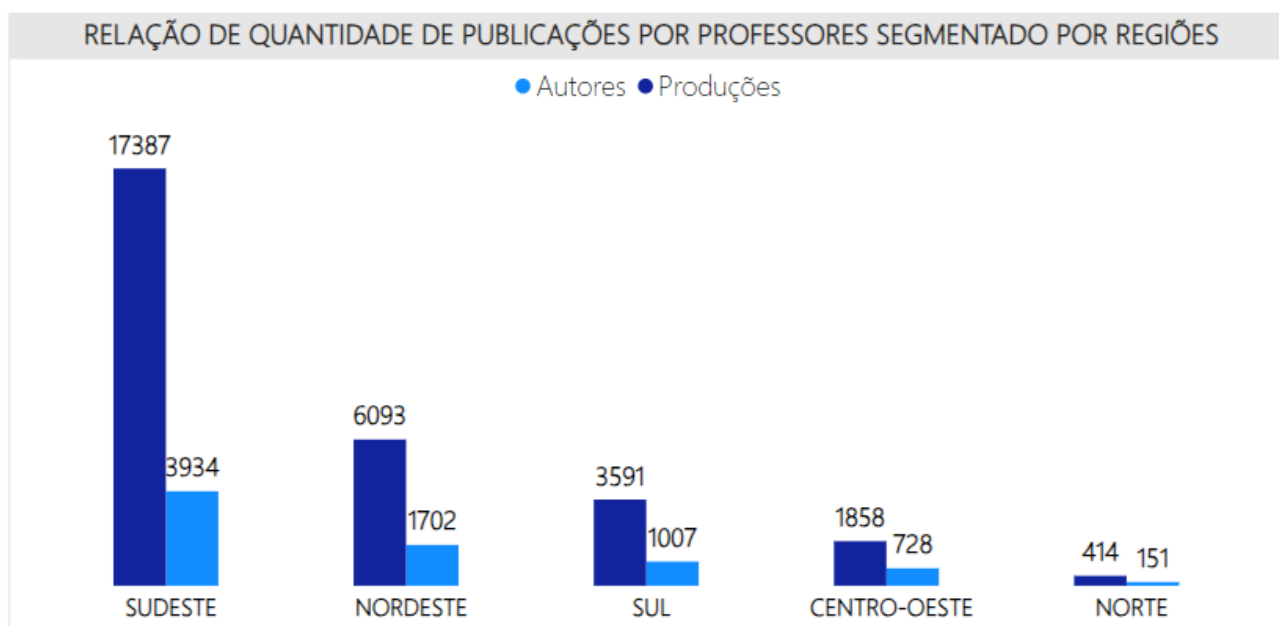
Figura 17 - Demonstrativo de dispersão entre publicações por autores



Fonte: Elaborado pelo autor, 2024.

Destacando os professores mais prolíficos em publicações, a Figura 18 apresenta uma relação entre a quantidade de produções, o número de professores envolvidos e as unidades federativas (UFs) às quais esses professores estão associados. Observa-se que o Sudeste é a região com o maior volume de publicações, totalizando 17387 produções por 3934 professores ativos. Em seguida, o Nordeste contribui com 6093 publicações envolvendo 1702 professores.

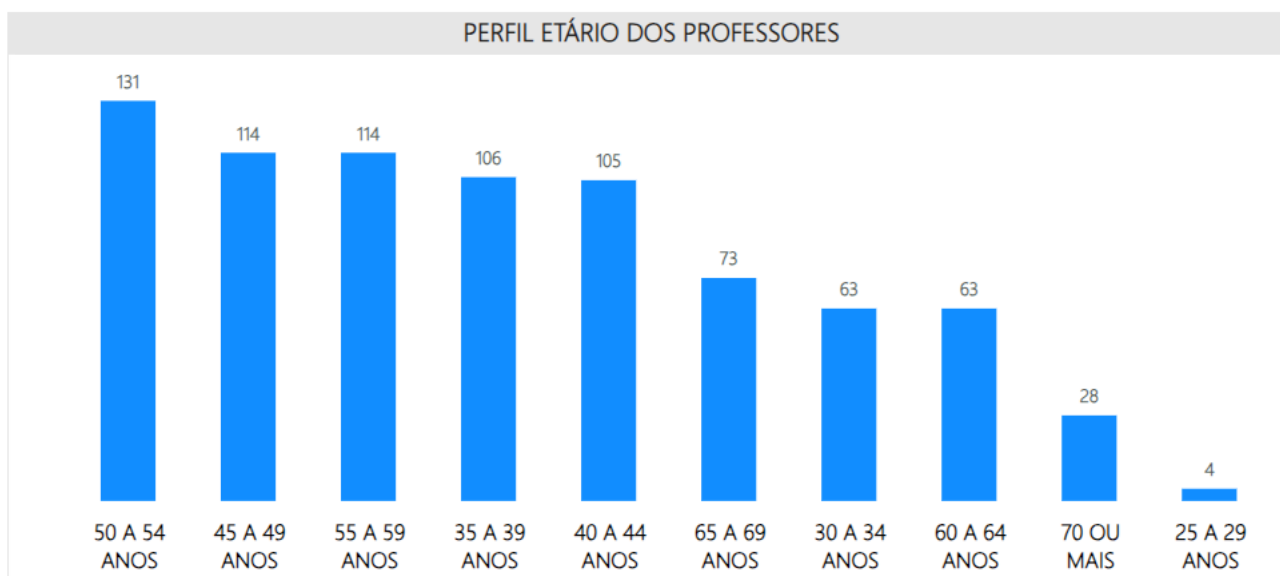
Figura 18 - Relação da quantidade de professores publicando x número de publicações



Fonte: Elaborado pelo autor, 2024.

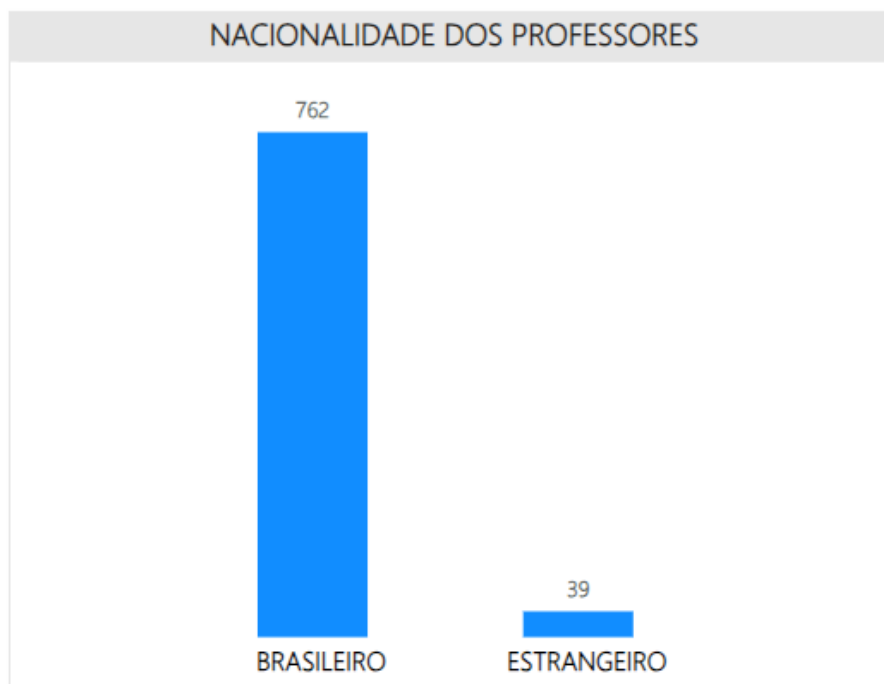
Um ponto importante a destacar é o perfil etário e a nacionalidade dos professores, conforme descrito nas Figuras 19 e Figura 20. É perceptível que a maioria dos professores se concentra na faixa etária entre 40 e 60 anos, com uma distribuição bastante homogênea entre as categorias. Já na Figura 20, observa-se que a maioria dos professores possui nacionalidade brasileira. Esses dados destacam a diversidade etária e nacional dos professores, fornecendo *insights* sobre a composição demográfica e cultural do corpo docente.

Figura 19 - Perfil etário dos professores



Fonte: Elaborado pelo autor, 2024.

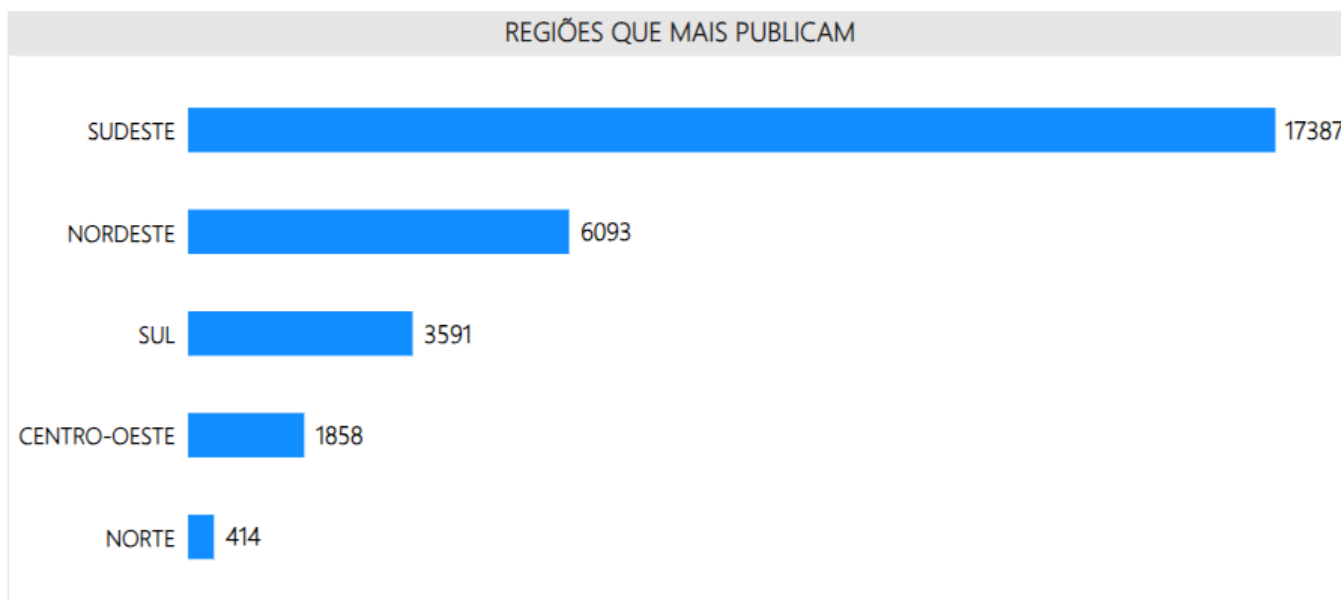
Figura 20 - Nacionalidade dos professores



Fonte: Elaborado pelo autor, 2024.

A Figura 21 mostra um comparativo entre o nível de publicação das regiões evidenciando que o Sudeste destaca-se de forma notória como a região que mais publica entre todos seguido pela região Nordeste.

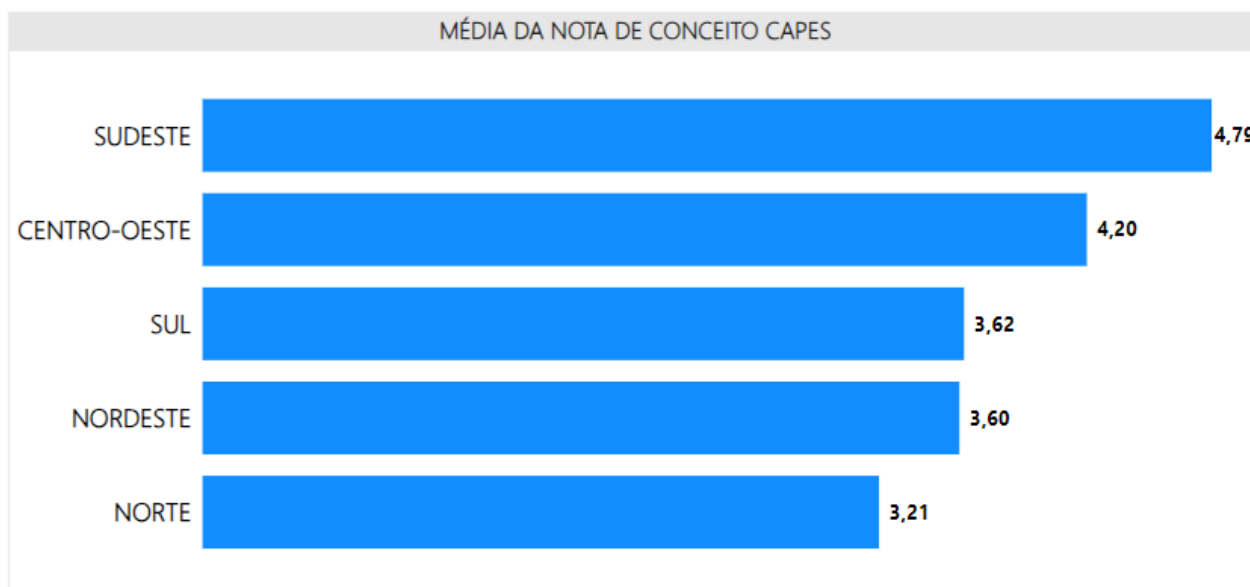
Figura 21 - Distribuição de publicações por região



Fonte: Elaborado pelo autor, 2024.

Ao segmentar pelo nível médio da nota conceito CAPES, observa-se que a região Sudeste continua se destacando, possuindo o maior número de médias de notas conceito CAPES. No entanto, o Centro-Oeste, apesar de ter poucas publicações conforme apresentado na Figura 22, destaca-se em segundo lugar com a maior nota média de conceito CAPES. Esses dados destacam a qualidade e o desempenho diferenciado das instituições de ensino superior na região Centro-Oeste, apesar do menor volume de publicações em comparação com outras regiões, como o Sudeste.

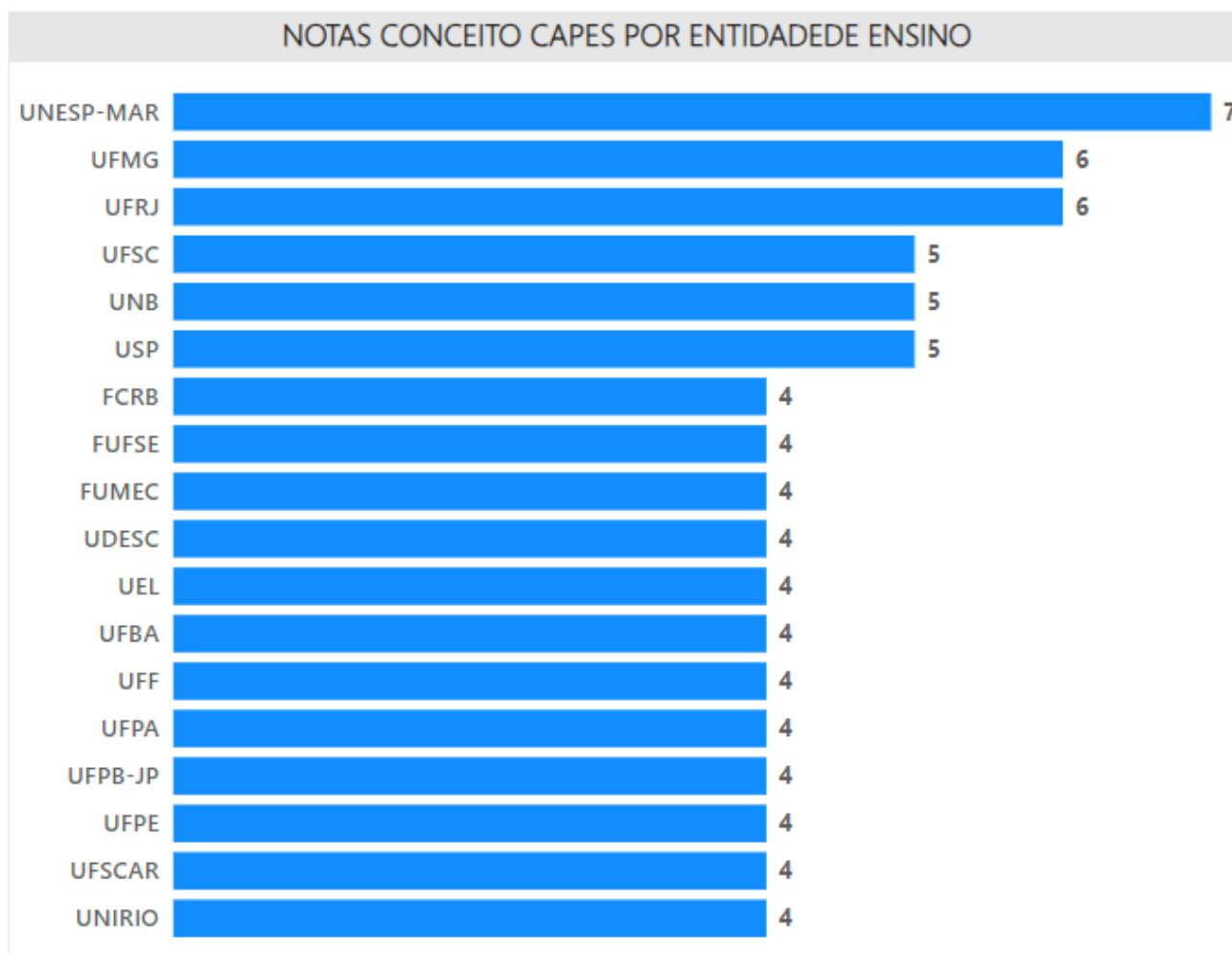
Figura 22 - Distribuição de médias de notas de conceito CAPES por região



Fonte: Elaborado pelo autor, 2024.

A Figura 23 demonstra de forma distribuída como as notas são distribuídas por entidades de ensino. Destaca-se a UNESP-MAR com a maior nota do conceito CAPES (nota 7). Além disso, podemos observar que a UFMG e a UFRJ possuem notas 6, enquanto a UFSC, UNB e USP têm nota 5. Várias outras entidades de ensino possuem diferentes notas de conceito CAPES. É importante ressaltar que o gráfico abaixo não representa todas as entidades de ensino, mas sim as 18 primeiras como exemplo ordenada de forma decrescente.

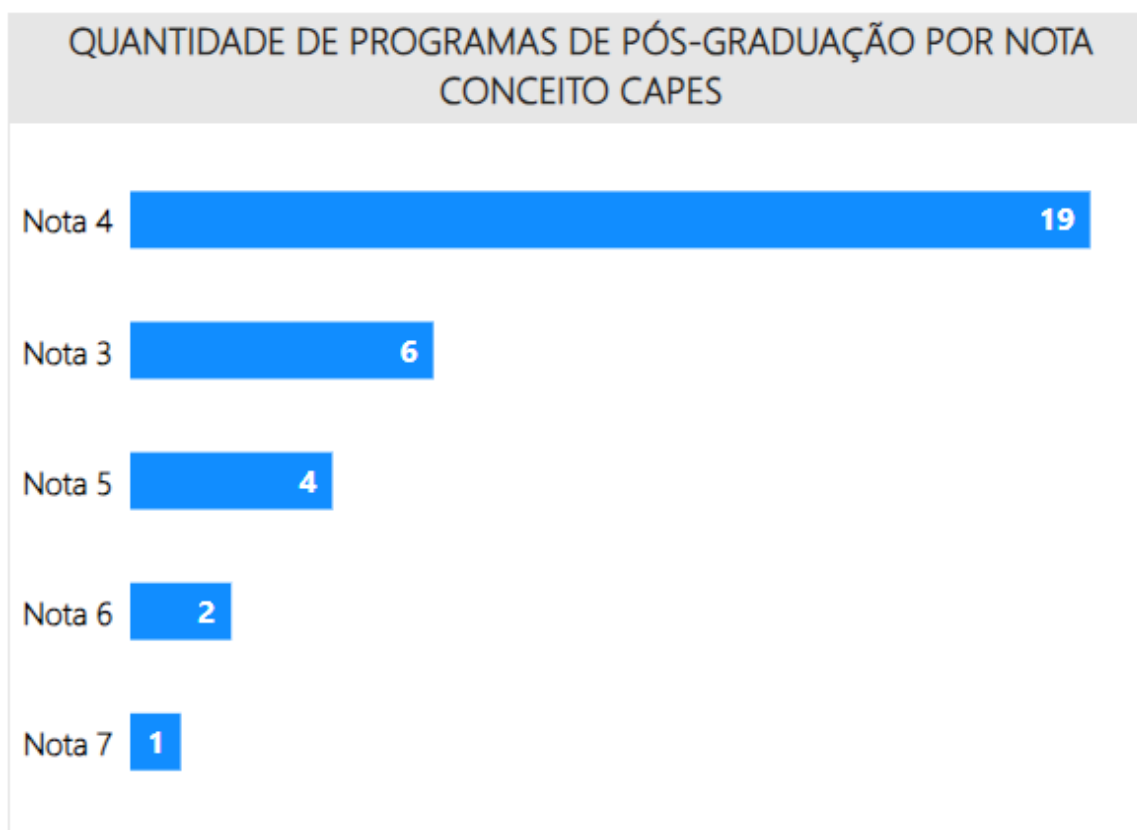
Figura 23 - Amostra das 18 principais notas conceito CAPES por entidades de ensino



Fonte: Elaborado pelo autor, 2024.

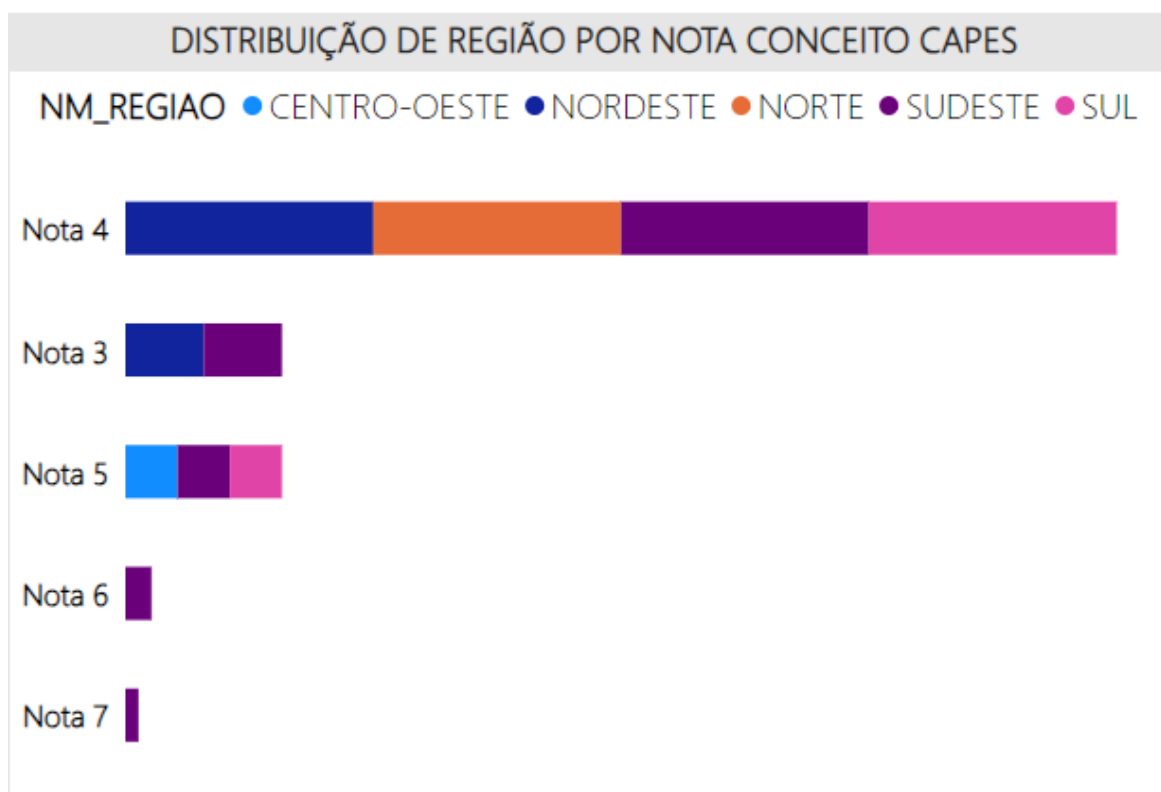
Figura 24 e na Figura 25 temos uma visão macro da distribuição dessas de conceito CAPES evidenciando que a maioria das entidades de ensino possuem nota 4 e temos apenas uma entidade de ensino como nota 7 como evidenciado na Figura 23.

Figura 24 - Distribuição de conceito CAPES por entidade de ensino



Fonte: Elaborado pelo autor, 2024.

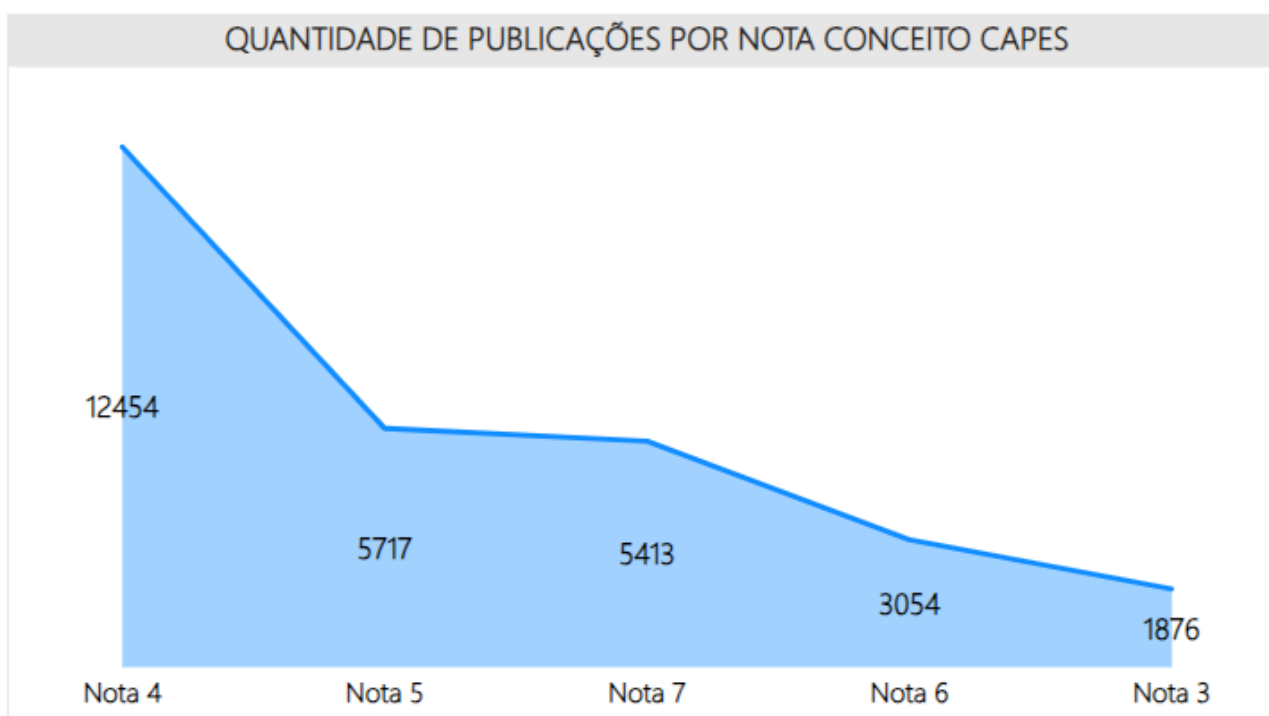
Figura 25 - Distribuição de conceito CAPES por região



Fonte: Elaborado pelo autor, 2024.

A Figura 26 apresenta a distribuição das publicações de acordo com o agrupamento de conceito CAPES no eixo x. Observa-se que a nota 4 é a que mais publica, o que se justifica pela maior quantidade de instituições de ensino com essa classificação. As notas 5 e 7 também possuem um número significativo de publicações, enquanto as notas 6 e 3 são as que menos publicam. Esta análise destaca as tendências de produtividade conforme o nível de avaliação acadêmica.

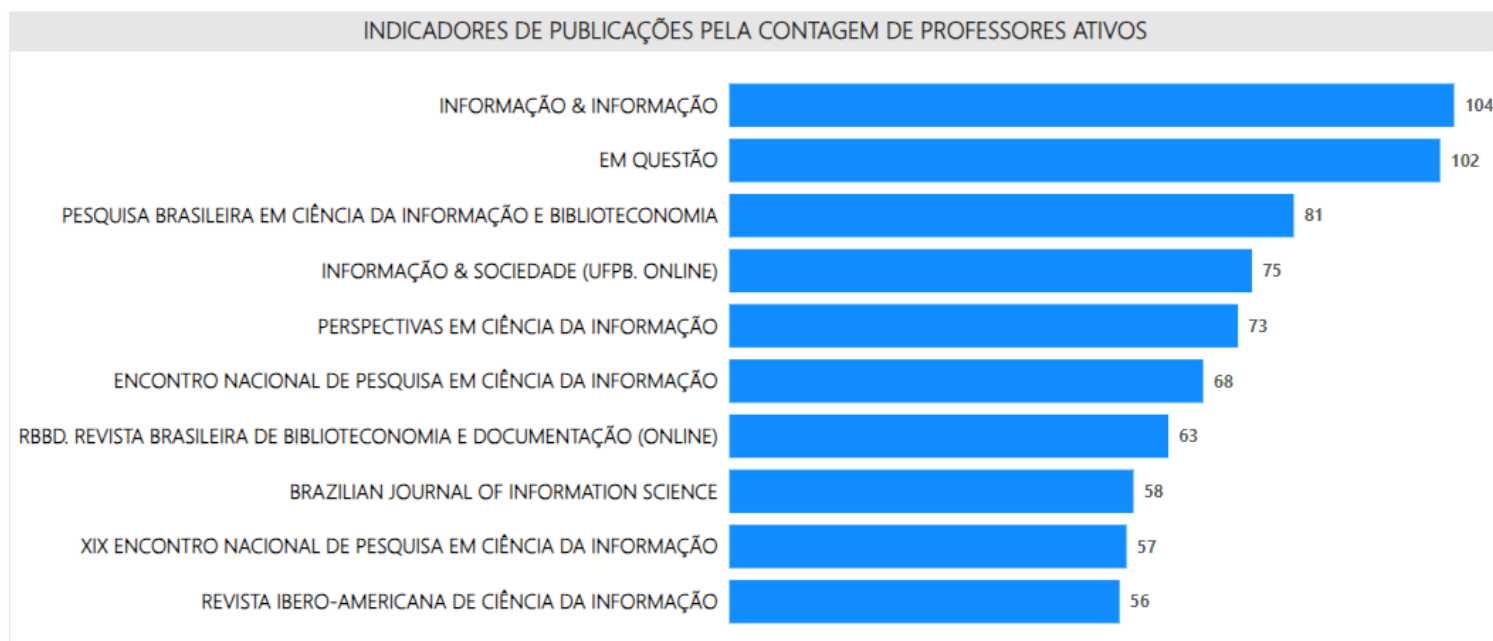
Figura 26 - Distribuição de publicações por nível de nota conceito CAPES



Fonte: Elaborado pelo autor, 2024.

Por fim, na Figura 27 são apresentados os indicadores de publicações, mostrando quantos professores ativos já publicaram em diferentes tipos de publicadores, sejam revistas científicas ou eventos científicos.

Figura 27 - Distribuição de publicações por publicadores



Fonte: Elaborado pelo autor, 2024.

Um dos resultados amplamente utilizados na cienciometria é a construção de grafos de redes para representar interações entre pontos. A cienciometria permite a exibição de aspectos quantitativos da ciência, mostrando atividades científicas através de grafos de colaboração. A informetria, por sua vez, permite analisar as características das comunicações formais, englobando todos os estudos métricos relacionados à ciência da informação.

Esses tipos de comportamento são objetos principais de estudo dentro das métricas da cienciometria e informetria, pois analisam o grau de colaboração no ciclo da comunicação científica entre diferentes professores. Na Tabela 1, destacamos uma amostra das principais colaborações entre professores dos programas de pós-graduação das universidades federais brasileiras, listadas em ordem decrescente. Observa-se que professores como Ziviani e Corrêa aparecem em múltiplas relações de colaboração, tanto com outros professores quanto entre si.

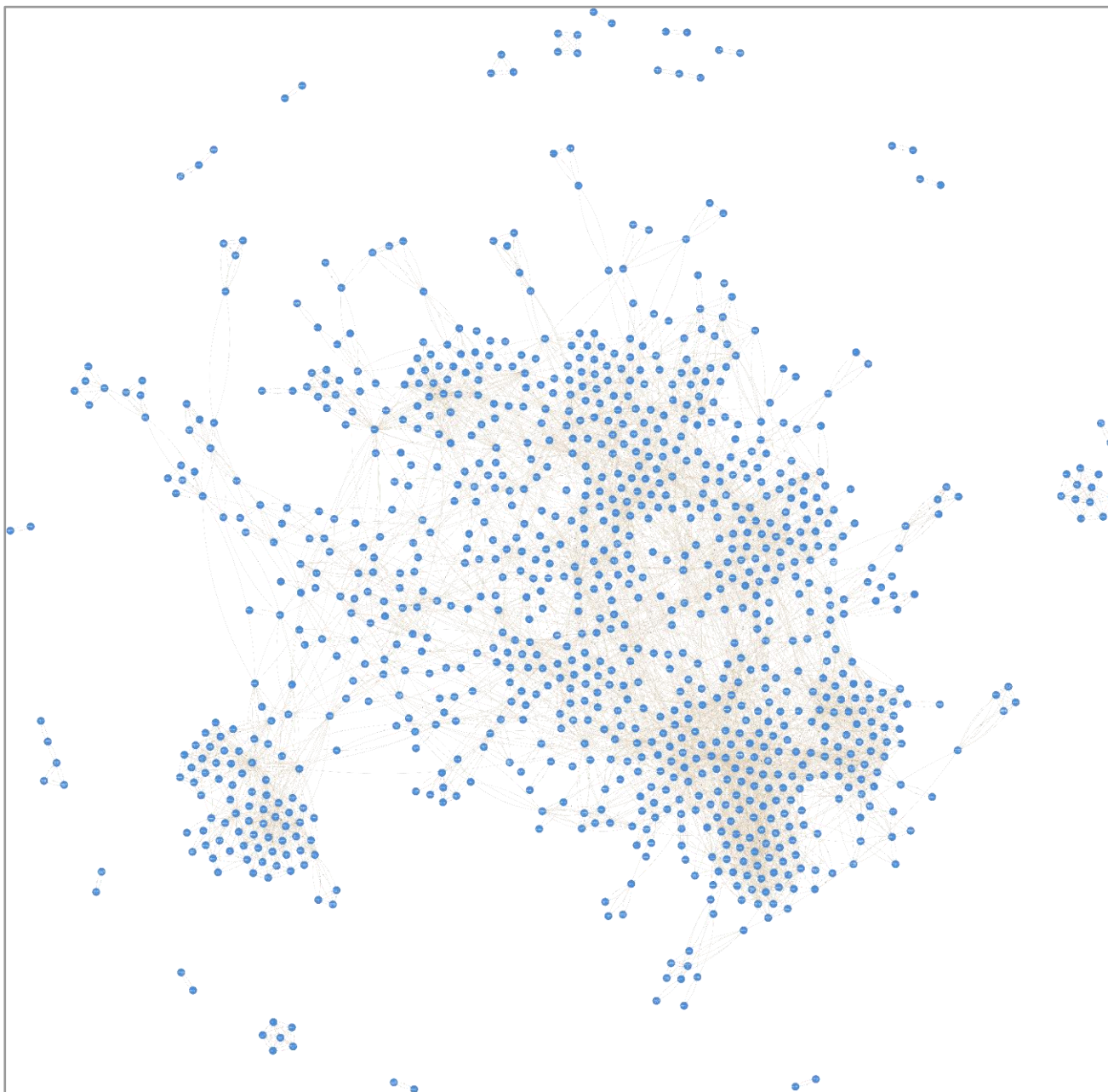
Tabela 1 - Amostra das principais colaborações entre professores

Professor 1	Professor 2	Num. Colaborações
DAMIAN, I. P. M.	SANTOS, B. R. P.	46
CORRÊA, F.	ZIVIANI, F.	42
BUFREM, L. S.	FREITAS, J. L.	34
LIMA, G. Â. B. O.	MACULAN, B. C. M. S.	30
CORRÊA, F.	RIBEIRO, J. S. A. N.	30
RIBEIRO, J. S. A. N.	ZIVIANI, F.	30
FRANCA, R. S.	ZIVIANI, F.	30
VECHIATO, F. L.	VIDOTTI, SILVANA APARECIDA BORSETTI GREGORIO	28
CORRÊA, F.	FRANCA, R. S.	27
GRACIO, M. C. C.	OLIVEIRA, ELY FRANCINA TANNURI DE	26

Fonte: Elaborado pelo autor, 2024.

Na Figura 28 é demonstrado um grafo geral de todos os professores de pós-graduação das universidades federais brasileiras e suas formas de interação formando uma grande rede de colaboração científica.

Figura 28 - Grafo de colaboração entre todos os professores



Fonte: Elaborado pelo autor, 2024.

Esse grafo foi construído por meio da ferramenta Neo4j com base na linguagem Cypher, linguagem de consulta do banco de dados de grafos Neo4j, ele carrega dados de um arquivo CSV e cria ou atualiza nós e relações no banco de dados.

Quadro 7 - Código utilizado para criar relacionamento de colaboração na ferramenta Neo4j

```
LOAD CSV WITH HEADERS FROM 'file:///sgcolab.csv' AS row
FIELDTERMINATOR ','
MERGE (p1:Professor {nome: row.professor1})
MERGE (p2:Professor {nome: row.professor2})
MERGE (p1)-[c:COLABOROU_COM]->(p2)
ON CREATE SET c.num_colaboracoes = toInteger(row.num_colaboracoes)
ON MATCH SET c.num_colaboracoes = c.num_colaboracoes +
toInteger(row.num_colaboracoes);
```

Fonte: Elaborado pelo autor, 2024.

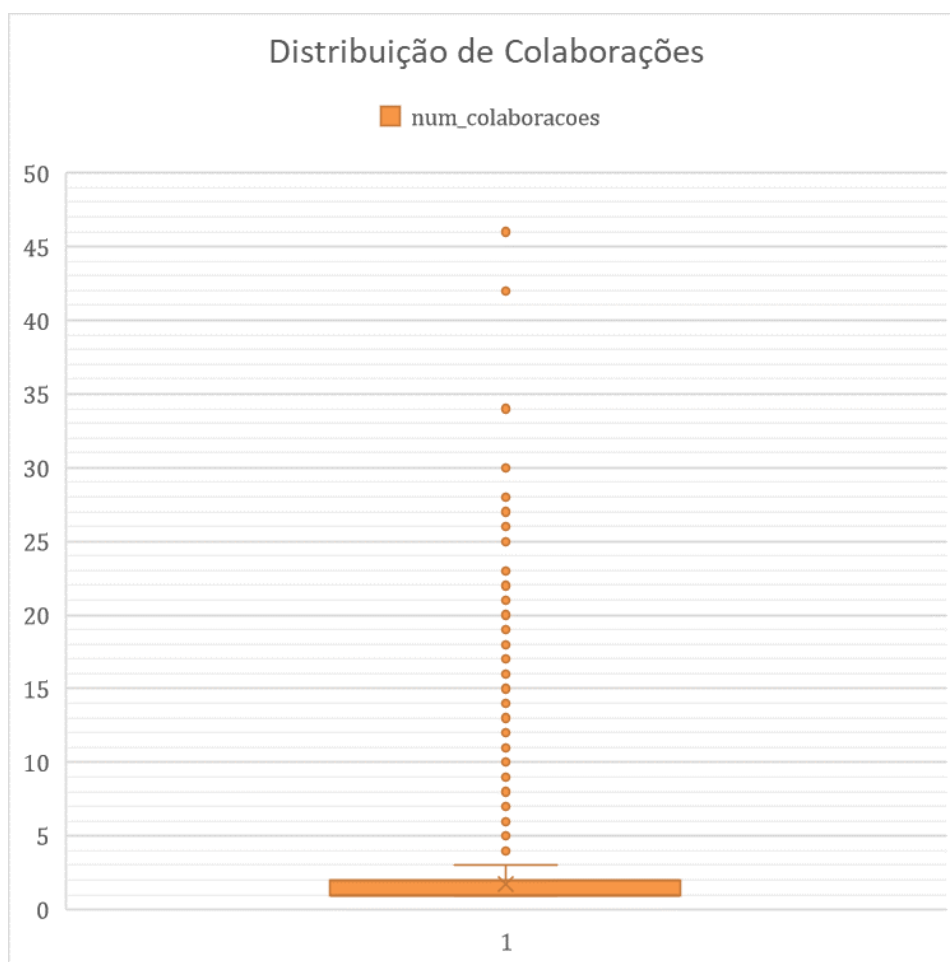
O código no Quadro 11 é detalhado da seguinte maneira:

- **LOAD CSV WITH HEADERS:** Carrega um arquivo CSV que contém cabeçalhos (nomes das colunas).
- **FROM 'file:///sgcolab.csv':** Especifica o caminho do arquivo CSV. O prefixo file:/// indica que o arquivo está no sistema de arquivos local do servidor onde o Neo4j está rodando.
- **AS row:** Cada linha do CSV será tratada como um mapa de valores, onde os nomes das colunas do CSV são as chaves.
- **FIELDTERMINATOR ',':** Especifica que o delimitador de campo no CSV é um ponto e vírgula (;).
- **MERGE:** Tenta encontrar um nó que corresponda aos critérios especificados. Se o nó não existir, ele será criado.
- **(p1{nome: row.professor1}):** Procura ou cria um nó com o rótulo Professor e o atributo nome igual ao valor da coluna professor1 da linha atual do CSV. O nó é referenciado como p1.
- Similar ao anterior, mas para um segundo professor (p2) com o nome igual ao valor da coluna professor2.

- **MERGE:** Procura ou cria uma relação entre p1 e p2 com o rótulo COLABOROU_COM. A relação é referenciada como c.
- **ON CREATE SET:** Se a relação foi criada (não existia anteriormente), define o atributo num_colaboracoes da relação c com o valor da coluna num_colaboracoes
- **ON MATCH SET:** Se a relação já existia, atualiza o atributo num_colaboracoes da relação c somando o valor existente com o valor da coluna num_colaboracoes convertido para inteiro.

A Figura 26 ilustra a distribuição das colaborações entre todos os professores mostrados no grafo de colaboração. Observa-se a presença de alguns outliers, conforme indicado na Tabela 1. No entanto, a maior parte dos dados se concentra entre 25 e 5 colaborações, com média em torno de 3 colaborações por autor.

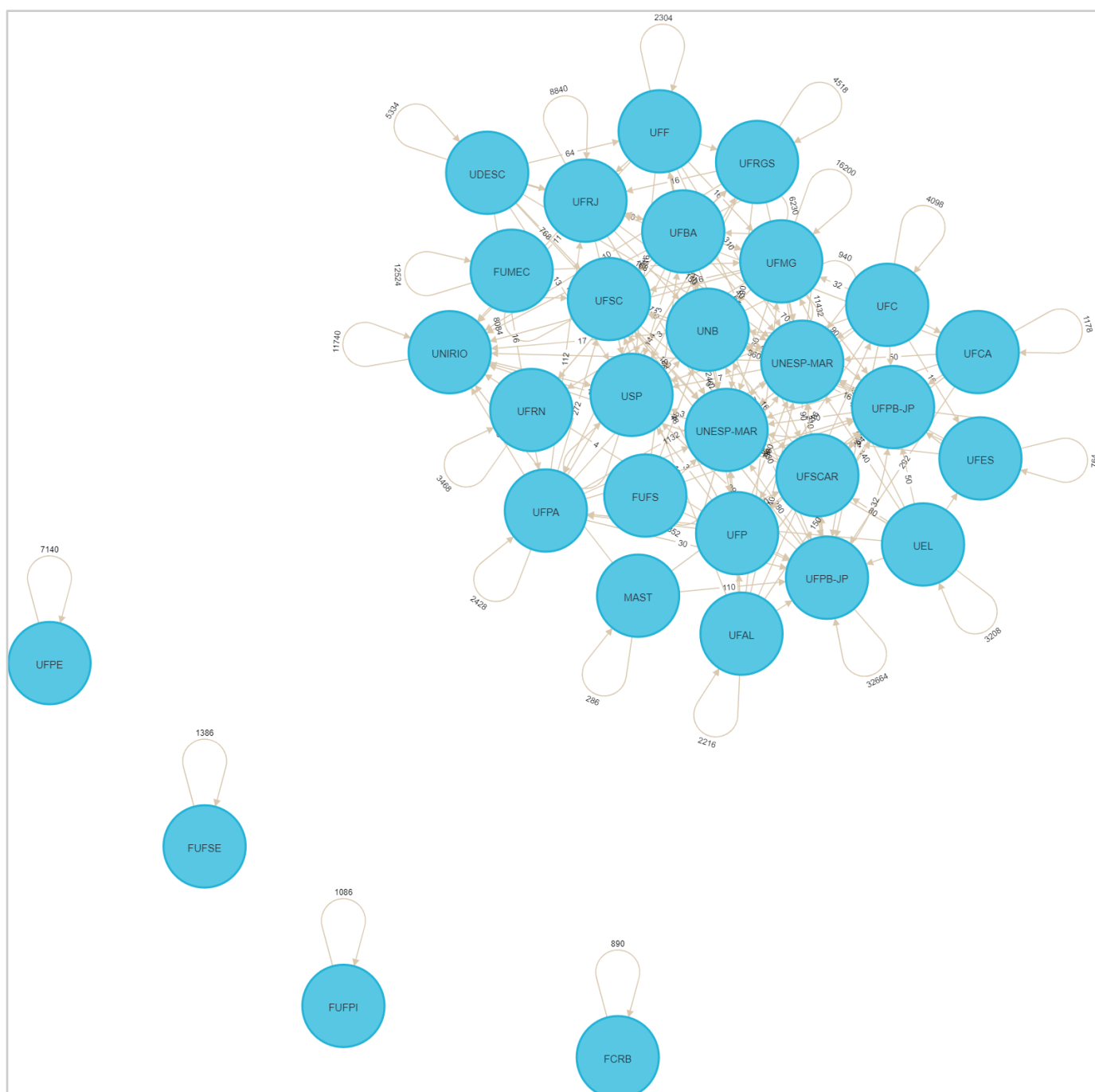
Figura 26 - Distribuição de colaboração científica por dispersão



Fonte: Elaborado pelo autor, 2024.

A Figura ilustra o nível de colaboração científica entre diferentes instituições de ensino. Observa-se um alto grau de colaboração entre a maioria das instituições, exceto por algumas, onde os professores colaboram predominantemente dentro de suas próprias instituições, conforme indicado pelas setas que apontam para o próprio nó. No centro, a UNESP-MAR se destaca, liderando em colaborações com 421 parcerias, e possuindo o maior conceito CAPES.

Figura 27 - Grafo de colaboração científica entre universidades



Fonte: Elaborado pelo autor, 2024.

A comparação entre a Figura 23 e a Tabela 2 confirma que as instituições com os maiores níveis de colaboração também possuem as maiores notas CAPES, validando uma das hipóteses iniciais. É importante destacar que a Tabela 2 demonstra o total de colaborações entre diferentes universidades.

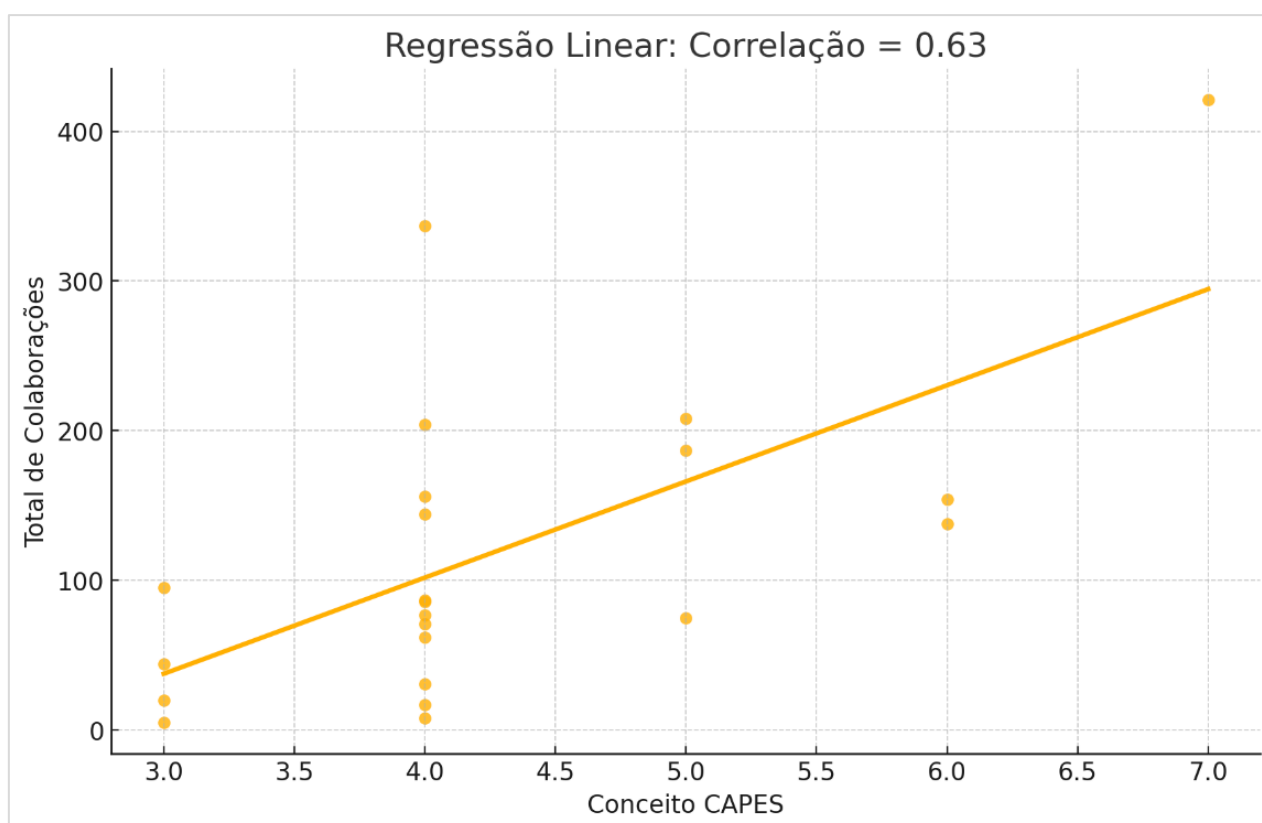
Tabela 2 - Ranking de maiores colaborações científicas por universidades.

UNIVERSIDADE	CONCEITO CAPES	TOTAL DE COLABORAÇÕES
UNESP-MAR	7	421
UFPB-JP	4	337
USP	5	208
UFPA	4	204
UFSC	5	187
UDESC	4	156
UFRJ	6	154
UNIRIO	4	144
UFMG	6	138
UFC	3	95
FUMEC	4	87
UFF	4	86
UFBA	4	77
UNB	5	75
UFRGS	4	71
UEL	4	62
UFCA	3	44
UFSCAR	4	31
MAST	3	20
FUFS	4	17
UFRN	4	8
UFES	3	5

Fonte: Elaborado pelo autor, 2024.

A correlação calculada entre o conceito CAPES e o total de colaborações é de 0,63, o que indica uma correlação positiva moderada entre as duas variáveis. O gráfico de regressão linear na Figura 28 ilustra essa relação, apresentando uma tendência ascendente. Isso sugere que, em geral, à medida que o conceito CAPES aumenta, o número de colaborações também tende a aumentar, apesar da certa dispersão nos dados. No entanto, é importante ressaltar que essa correlação não implica uma relação causal entre as variáveis. Esses resultados, portanto, fortalecem as evidências apresentadas ao longo do trabalho, corroborando a hipótese formulada.

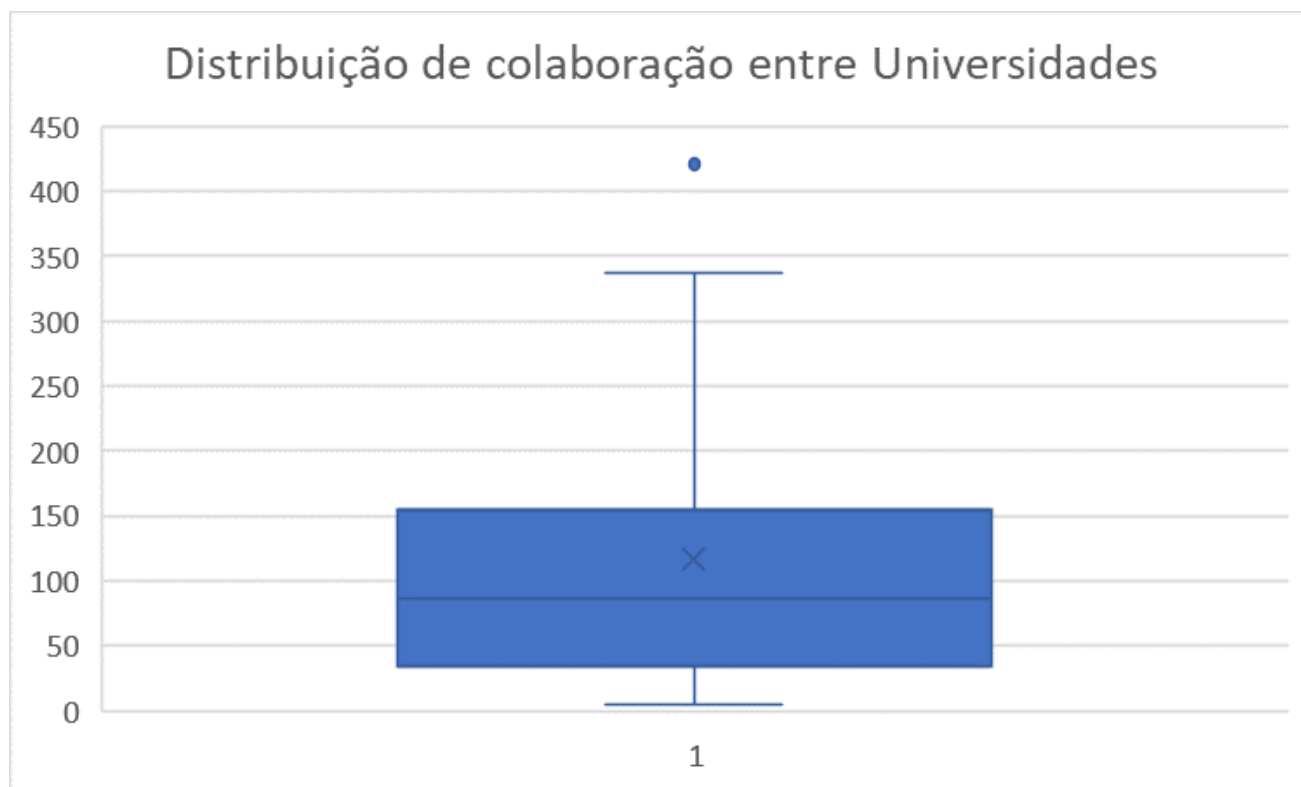
Figura 28 - Regressão linear de conceito CAPES e o total de colaborações



Fonte: Elaborado pelo autor, 2024.

A Figura 29 ilustra a distribuição das colaborações entre entidades de ensino. Observa-se a presença de apenas um *outlier*, sendo a UNESP-MAR, e a maioria da distribuição dos dados varia entre 30-160 colaborações científicas a nível de entidades de ensino sendo a média por volta de 110 colaborações.

Figura 29 - Distribuição de colaboração científica entre universidades



Fonte: Elaborado pelo autor, 2024.

5.4. Considerações finais

Em conclusão, esta dissertação oferece uma visão geral das pesquisas em materiais científicos, com o objetivo de apoiar teoricamente esta pesquisa e construir uma narrativa bibliográfica que contextualiza o problema apresentado por meio de uma revisão de literatura linear e narrativa. Além disso, foi realizada uma amostra seletiva com professores dos programas de pós-graduação em Ciência da Informação das universidades federais brasileiras, utilizando a plataforma de dados abertos da CAPES.

A plataforma de dados abertos da CAPES foi fundamental para a coleta desses dados, pois permitiu extrair informações sobre programas, universidades, professores e suas produções. Esses dados foram tratados utilizando ferramentas de ETL e carregados em um banco de dados relacional, facilitando a consulta e a geração de arquivos CSV, que alimentaram os programas usados para a criação de artefatos necessários à realização das análises estatísticas quantitativas, permitindo responder à pergunta de pesquisa formulada ao longo do trabalho.

Os principais resultados encontrados nesta pesquisa corroboram a hipótese de que a nota conceito da CAPES influencia a produção e a comunicação científica dos professores de pós-graduação em Ciência da Informação das universidades federais brasileiras. Isso foi evidenciado pelo cálculo da correlação, que resultou em um valor positivo de 0,63, sugerindo a existência de uma relação. Além disso, os números e gráficos apresentados demonstram que as universidades federais com notas mais altas colaboraram e produziram mais.

As principais limitações deste trabalho estavam relacionadas ao tratamento dos dados, mas foram superadas com o uso da ferramenta Pentaho PDI, que facilitou o processamento rápido das informações. Isso foi crucial, pois os arquivos CSV extraídos da plataforma da CAPES eram grandes e volumosos, tornando inviável o tratamento direto no Excel, além de garantir a precisão dos dados e evitar análises imprecisas.

Este trabalho contribuirá significativamente para a Ciência da Informação, mostrando como a Ciência de Dados pode interagir de maneira convergente com essa área, permitindo que a Ciência da Informação aproveite as oportunidades oferecidas pela ciência de dados. Isso facilitará a compreensão da produção científica, além de ajudar na criação de indicadores de colaboração científica e no entendimento do comportamento científico, entre outros aspectos.

Esta pesquisa visa inspirar profissionais da Ciência da Informação a se especializarem em técnicas de ciência de dados. Além disso, o principal desafio reside no tratamento adequado dos dados, para garantir a precisão das informações. É importante que pesquisas nessa área continuem a ser realizadas para promover um diálogo cada vez mais próximo entre essas duas disciplinas.

REFERÊNCIAS

AKHTAT, S. M. F. **Big Data Architect's Handbook**. Birmingham: Pack Publishing, 2018.

ALVES, L. Informação e os sistemas de comunicação científica na ciência da informação. **DataGramZero**, v. 12, n. 3, 2011. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/7379>. Acesso em: 29 ago. 2023.

BALANCIERI, R. *et al.* A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, [s. l.], v. 34, n. 1, 2005. Disponível em: <https://revista.ibict.br/ciinf/article/view/1103>. Acesso em: 4 set. 2023.

BALANCIERI, R.; BOVO, A. B.; KERN, V. M.; PACHECO, R. C. dos S.; BARCIA, R. M. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, v. 34, n. 1, 26 out. 2005. DOI [10.18225/ci.inf.v34i1.1103](https://doi.org/10.18225/ci.inf.v34i1.1103). Disponível em: <https://revista.ibict.br/ciinf/article/view/1103>. Acesso em: 4 set. 2023.

BAPTISTA, Ana Alice; COSTA, Sely Maria de Souza; KURAMOTO, Hélio; RODRIGUES, Eloy. Comunicação científica : o papel da Open Archives Initiative no contexto do acesso livre. 2007. Disponível em: <https://repositorio.unb.br/handle/10482/635>. Acesso em: 17 ago. 2021.

BELKIN, N. J. INFORMATION CONCEPTS FOR INFORMATION SCIENCE. **Journal of Documentation**, v. 34, n. 1, p. 55–85, 1 jan. 1978. <https://doi.org/10.1108/eb026653>.

BIBLIOS JOURNAL OF LIBRARIANSHIP AND INFORMATION SCIENCE. [S. l.], [s. d.]. Disponível em: <http://biblios.pitt.edu/ojs/biblios>. Acesso em: 8 maio 2024.

BORKO, H. Information science: What is it? **American Documentation**, v. 19, n. 1, p. 3–5, 1968. Disponível em: <https://doi.org/10.1002/asi.5090190103>. Acesso em: 17 ago. 2021.

BRASCHER, M.; CAFÉ, L. 1835 - Organização da Informação ou Organização do Conhecimento. [s. l.],

BRITO, G. do N.; LIMA, I. F. de. Periódicos Científicos como Fonte de Informação: um estudo na Informação & Sociedade e na Biblionline. Folha de rosto em *Biblioteconomia e Ciência da Informação*, v.1, n.2, p. 48-60, 2015.

BROOKES, B. C. The foundation of information science. Part I. **Philosophical aspects. Journal of Information Science** 2 (1980), 125-133.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science**, [s.n.], v. 42, n. 5, p. 351-360, jun.1991. Disponível em: [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5<351::AID-ASI5>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3). Acesso em: 6 ma. 2021.

BUENO, W. C. Comunicação científica e divulgação científica: aproximações e rupturas conceituais. **Informação & Informação**, v. 15, n. 1esp, p. 1–12, 15 dez. 2010. <https://doi.org/10.5433/1981-8920.2010v15n1espp1>.

Bugnion, P.; Manivannan, A.; Nicolas, P. R. **Scala: Guia para Profissionais de Ciência de Dados**. Birmingham: Packt Publishing, 2017.

CAJAZEIRA, P. E. S. L.; SILVA, H. A. As publicações em coautoria e colaboração científica em Comunicação na Universidade Federal do Piauí (UFPI). **Ciência da Informação**, v. 50, n. 1, 12 maio 2021. DOI [10.18225/ci.inf.v50i1.5209](https://doi.org/10.18225/ci.inf.v50i1.5209). Disponível em: <https://revista.ibict.br/ciinf/article/view/5209>. Acesso em: 4 set. 2023.

CARVALHO, L. A. de; CRIPPA, G. Ciência da informação: histórico, delimitação do campo e a sua perspectiva sobre a área da Comunicação. **Perspectivas em Ciência da Informação**, v. 18, p. 241–251, dez. 2013. <https://doi.org/10.1590/S1413-99362013000400015>.

COCCIA, M.; BOZEMAN, B. Allometric models to measure and analyze the evolution of international research collaboration. **Scientometrics**, [s. l.], v. 108, n. 3, p. 1065–1084, 2016.

COM CIÊNCIA - SBPC/LABJOR. [S. l.], [s. d.]. Disponível em: <https://www.comciencia.br/comciencia/handler.php?section=8&edicao=115&id=1381>. Acesso em: 30 ago. 2023.

COMARELA, Giovanni; FRANCO, Gabriel; TROIS, Celio; LIBERATO, Alextian; MARTINELLO, Magno; CORRÊA, João Henrique; VILLAÇA, Rodolfo. Introdução à Ciência de Dados: Uma Visão Pragmática utilizando Python, Aplicações e Oportunidades em Redes de Computadores. *In*: SCHAEFFER FILHO, Alberto; CORDEIRO, Weverton Luis; CAMPISTA, Miguel Elias (orgs.). **Minicursos do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**. 1. ed. [S. l.]: SBC, 2019. p. 246–295. DOI [10.5753/sbc.6555.9.6](https://doi.org/10.5753/sbc.6555.9.6). Disponível em: <https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/65/289/538-1>. Acesso em: 18 set. 2021.

CRONIN, B. **The Citation process: the rule and significance of citations in scientific communication**. London: Taylor Graham, 1984.

CURTY, Renata Gonçalves; SERAFIM, Jucenir Da Silva. A formação em ciência de dados: uma análise preliminar do panorama estadunidense. **Informação & Informação**, v. 21, n. 2, p. 307, 20 dez. 2016. Disponível em: <https://doi.org/10.5433/1981-8920.2016v21n2p307>. Acesso em: 18 set. 2021.

DANUELLO, J. C.; OLIVEIRA, E. F. T. de. Análise cientométrica: produção científica e redes colaborativas a partir das publicações dos docentes dos programas de pós-graduação em Fonoaudiologia no Brasil. **Em Questão**, [s. l.], v. 18, n. 3, p. 65–79, 2012.

DAVENPORT, Thomas H.; PATIL, D. J. Data Scientist: The Sexiest Job of the 21st Century. **Harvard Business Review**, , seq. Analytics and data science, 1 out. 2012.

Disponível em: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>. Acesso em: 26 set. 2021.

DE MELO MARICATO, J.; MARTINS, D. Altmetrics complexities, challenges and new forms of measuring and comprehending scientific communication in the social. [s. l.], 2018.

DEUS, J. D. de (Org.). **A crítica da ciência**: sociologia e ideologia da ciência. 2. ed. Rio de Janeiro: Zahar, 1979.

DIAS, Eduardo José Wense. Biblioteconomia e Ciência da Informação: natureza e relações. v. 5, n. Perspectivas em Ciência da Informação, 20 nov. 2007. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/556/338>. Acesso em: 18 set. 2021.

FERNANDES, J. H. C. Interloquções bibliográficas e epistemológicas entre a ciência de dados e a ciência da informação. **Ciência da Informação**, [s. l.], v. 49, n. 3, 2020. Disponível em: <https://revista.ibict.br/ciinf/article/view/5655>. Acesso em: 6 maio 2024.

FERNANDEZ-MOLINA, J. C., GUIMARÃES, J. A. C. Ethical aspects of knowledge organization and representation in the digital environment: their articulation in professional codes of ethics. In: LOPEZHUERTAS, M. J. (Org.). **Challenges in knowledge representations and organization for the 21st century: integration of knowledge across boundaries**. Wurzburg: Ergon Verlag, 2002, p. 487-492.

FRANCK, K. M.; PEREIRA, R. F.; FILHO, J. V. D. Diagrama Entidade-Relacionamento: uma ferramenta para modelagem de dados conceituais em Engenharia de Software. **Research, Society and Development**, [s. l.], v. 10, n. 8, p. e49510817776–e49510817776, 2021.

GHOLAMPOUR, B.; GHOLAMPOUR, S.; NORUZI, A. Research Trend Analysis of Information Science in France based on Total, Cited and Uncited Publications: A Scientometric and Altmetric. [s. l.], v. 1, n. 1, 2022.

GOFFMAN, W. Information science: discipline or disappearance. *ASLIB Proceedings*, v. 22 n.12, p. 589-596, 1970 *apud* SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, n. 1, v. 1, p. 41-62, jan./jun., 1996.

GRÁCIO, Maria Claudia Cabrini. Colaboração científica: indicadores relacionais de coautoria. **Brazilian Journal of Information Science: research trends**, v. 12, n. 2, 1 ago. 2018. DOI [10.36311/1981-1640.2018.v12n2.04.p24](https://doi.org/10.36311/1981-1640.2018.v12n2.04.p24). Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/7976>. Acesso em: 19 ago. 2021.

GRAY, Jim. Jim Gray on e-Science: A transformed scientific method based on the transcript of a talk given by Jim Gray to the NRC-CSTB1 in Mountain View, CA, (orgs). **The fourth paradigm. Dataintensive scientific discovery**. Redmond, WA: Microsoft Research, 2007. 284 p.

GUIMARÃES, J. A. C. Análise de domínio como perspectiva metodológica em organização da informação. **Ciência da Informação**, [s. l.], v. 43, n. 1, 2014.

Disponível em: <https://revista.ibict.br/ciinf/article/view/1415>. Acesso em: 17 ago. 2023.

GUIMARÃES, J.A.C.; Ciência da Informação, Arquivologia e Biblioteconomia: em busca do necessário diálogo entre o universo teórico e os fazeres profissionais. *In*: FUJITA, M.S.L.; GUIMARÃES, J.A.C. (Org.). **Ensino e pesquisa em Biblioteconomia no Brasil: a emergência de um novo olhar**. São Paulo: Cultura Acadêmica, 2008. p. 33-44.

GUSMÃO, A. C. dos S.; SANTOS, S. M. dos; MENA-CHALCO, J. P. Análise da longevidade e do tamanho das coautorias acadêmicas: os caminhos na ciência brasileira. **Em Questão**, [s. l.], p. 116156–116156, 2022.

HAYASHI, C. R. M. *et al.* Análise de redes de colaboração científica. [s. l.], v. 35, 2012.

HAYASHI, C. R. M.; HAYASHI, M. C. P. I.; MARCELO, J. F.; BELLO, S. F. Análise de redes de colaboração científica entre educação especial e fonoaudiologia. **Revista Interamericana de Bibliotecología (Colombia)**, v. 35, n. 3, 2012. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/84928>. Acesso em: 25 set. 2023.

HILÁRIO, C. M.; CAIRES, C. M.; CASTANHA, R. C. G. Redes de colaboração científica na perspectiva da auto-organização: um estudo bibliométrico na temática "colaboração científica" (2003- 2012). **Encontro Brasileiro de Bibliometria e Cientometria**, v. 4, 2014. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/45446>. Acesso em: 31 ago. 2023.

HILÁRIO, C. M.; CAIRES, C. M.; CASTANHA, R. C. G. REDES DE COLABORAÇÃO CIENTÍFICA NA PERSPECTIVA DA AUTO-ORGANIZAÇÃO:. [s. l.], 2014.

HILÁRIO, C. M.; GRÁCIO, M. C. C. A ciência como um sistema auto-organizado: a colaboração científica na perspectiva da autopoiese. *In*: **ENCONTRO NACIONAL DE PESQUISA DA ANCIB – ENANCIB**, XIV, Florianópolis, 2013. Anais... Florianópolis: ANCIB, 2013.

HJØRLAND, B. Documents, memory institutions and information science. **Journal of Documentation**, v. 56, n.1 p. 27-41, 2000.

INFORMAÇÃO E OS SISTEMAS DE COMUNICAÇÃO CIENTÍFICA NA CIÊNCIA DA INFORMAÇÃO. [S. l.], [s. d.]. Disponível em: <https://www.brapci.inf.br/index.php/res/v/7379>. Acesso em: 29 ago. 2023.

KATZ, J. Sylvan; MARTIN, Ben R. What is research collaboration? **Research Policy**, v. 26, n. 1, p. 1–18, 1 mar. 1997. Disponível em: [https://doi.org/10.1016/S0048-7333\(96\)00917-1](https://doi.org/10.1016/S0048-7333(96)00917-1). Acesso em: 18 set. 2021.

LE COADIC, Y.-F. **A Ciencia Da Informacao**. Brasília, DF: Brique de Lemos/Livros, 1970.

LEITE, Fernando César Lima; COSTA, Sely Maria de Souza. Gestão do conhecimento científico: proposta de um modelo conceitual com base em processos

de comunicação científica. **Ciência da Informação**, v. 36, p. 92–107, abr. 2007. Disponível em: <https://doi.org/10.1590/S0100-19652007000100007>. Acesso em: 27 ago. 2021.

MAIA, M. de F. S.; CAREGNATO, S. E. Co-autoria como indicador de redes de colaboração científica. **Perspectivas em Ciência da Informação**, [s. l.], v. 13, n. 2, p. 18–31, 2008.

MANETH, S.; POULOVASSILIS, A. Data Science. **The Computer Journal**, v. 60, n. 3, p. 285–286, 1 mar. 2017. <https://doi.org/10.1093/comjnl/bxw073>.

MARCHIONINI, G. Information Science Roles in the Emerging Field of Data Science. **Journal of Data and Information Science**, [s. l.], v. 1, n. 2, p. 1–6, 2016.

MAROLDI, A. M.; LIMA, L. F. M.; HAYASHI, C. R. M.; HAYASHI, M. C. P. I. Comunicação científica. **Folha de Rosto**, v. 5, n. 1, p. 5–15, 9 set. 2019.

MEADOWS, Arthur Jack. **A comunicação científica**. trad. Antonio Agenor Briquet de Lemos Lemos. Brasília: Briquet de Lemos/livros, 1999.

MEDEIROS, M. B. B.; CAFÉ, L. M. A. Organização da informação ou organização do conhecimento?. , . Disponível em: <http://hdl.handle.net/20.500.11959/brapci/176535>. Acesso em: 25 set. 2023.

MUELLER, Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, p. 27–38, ago. 2006. Disponível em: <https://doi.org/10.1590/S0100-19652006000200004>. Acesso em: 27 ago. 2021.

MUELLER, Suzana. Literatura científica, comunicação científica e ciência da informação. **Para entender a ciência da informação**. Salvador: EDUFBA, 2007.

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, 2011.

OLIVEIRA, E. F. T. D.; GRÁCIO, M. C. C. Scientific collaboration network in “Metrical Studies”: a co-authorship study using the SciELO information Science periodicals. **Brazilian Journal of Information Science: research trends**, v. 2, n. 2, 19 out. 2009. DOI [10.36311/1981-1640.2008.v2n2.04.p35](https://doi.org/10.36311/1981-1640.2008.v2n2.04.p35). Disponível em: <https://revistas.marilia.unesp.br/index.php/bjis/article/view/47>. Acesso em: 6 set. 2023.

ORGANIZAÇÃO DA INFORMAÇÃO OU ORGANIZAÇÃO DO CONHECIMENTO? [S. l.], [s. d.]. Disponível em: <https://www.brapci.inf.br/index.php/res/v/176535>. Acesso em: 15 ago. 2023.

OSINSKA, V.; KLIMAS, R. **Mapping science: tools for bibliometric and altmetric studies**. [S. l.], 2021. text. Disponível em: <https://informationr.net/ir/26-4/paper909.html>. Acesso em: 20 maio 2024.

PECEGUEIRO, C. M. P. de A.; JESUS, S. M. de. Comunicação científica dos docentes da Universidade Federal do Maranhão no período de 1998 a 2001. **Transinformação**, [s. l.], v. 15, n. 2, p. 1–12, 2003.

PRESS, Gil. A Very Short History Of Data Science. 2013. **Forbes**. Disponível em: <https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/>. Acesso em: 26 set. 2021.

QUEIRÓS, C. F. G. G. É a ciência da informação um saber humano ou social? In: CASTRO, A. L. S. et al. (Orgs.). *Ciência da informação, ciências sociais e interdisciplinaridade*. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 1999.

RABELLO, R. História dos conceitos e ciência da informação: apontamentos teórico-metodológicos para uma perspectiva epistemológica. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, [s. l.], v. 13, n. 26, p. 17–46, 2008.

RAUTENBERG, S.; CARMO, P. R. V. D. Big data e ciência de dados: complementariedade conceitual no processo de tomada de decisão. **Brazilian Journal of Information Science: research trends**, v. 13, n. 1, p. 56–67, 29 mar. 2019. <https://doi.org/10.36311/1981-1640.2019.v13n1.06.p56>.

REIS, M. de J.; DE CARVALHO, T. CIÊNCIA DE DADOS E CIÊNCIA DA INFORMAÇÃO: evolução e paradigmas da ciência. [s. l.], v. 7, 2022.

REIS, M. de J.; SENA, N. C. dos S. Biblioteconomia de dados e ciência de dados no contexto da e-science. **Revista Fontes Documentais**, v. 4, n. Ed. especial, p. 51–64, 2021.

REIS, Makson de Jesus. Ciência de dados e ciência da informação: guia para alfabetização de dados para bibliotecários. 12 jul. 2019. Disponível em: <https://ri.ufs.br/jspui/handle/riufs/12667>. Acesso em: 1 set. 2021.

ROBREDO, Jaime. **Da Ciência da Informação. Revisitada aos Sistemas Humanos de Informação**. Brasília: Thesaurus, 2003.

RODRIGUES, C.; VIERA, A. F. G. Estudos bibliométricos sobre a produção científica da temática Tecnologias de Informação e Comunicação em bibliotecas. **InCID: Revista de Ciência da Informação e Documentação**, [s. l.], v. 7, n. 1, p. 167–180, 2016.

ROLIM, Mesaque Vidal. Análise do perfil do profissional da informação para a atuação como cientista de dados em ambientes de big data : uma perspectiva a partir das disciplinas do curso de Biblioteconomia da UnB. 3 jul. 2018. Disponível em: <https://bdm.unb.br/handle/10483/20898>. Acesso em: 26 set. 2021.

ROSENDAL, H. E.; GEURTS, P. A. T. M. Forces and functions in scientific communication: an analysis of their interplay. In: CONFERENCE ON “CO-OPERATIVE RESEARCH IN INFORMATION SYSTEMS IN PHYSICS”, 1997. **Anais [...]**. [S. l.: s. n.], 1997. Disponível em:

<https://research.utwente.nl/en/publications/forces-and-functions-in-scientific-communication-an-analysis-of-t>. Acesso em: 8 maio 2024.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações.

Perspectivas em Ciência da Informação, v. 1, n. 1, 1992. Disponível em:

<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>. Acesso em: 27 set. 2021.

SCHWEITZER, Fernanda; RODRIGUES, Rosângela Schwarz; VARVAKIS, Gregório Jean. Comunicação científica e as tecnologias de informação e comunicação.

Comunicação & Sociedade, v. 32, n. 55, p. 83–104, 27 jun. 2011. Disponível em:

<https://doi.org/10.15603/2175-7755/cs.v32n55p83-104>. Acesso em: 25 out. 2021.

SILVA, E. B. F. da; SAMPAIO, D. A. O BOOM INFORMACIONAL: a tecnologia e a gênese da ciência da informação. **BiblioCanto**, v. 3, n. 2, p. 3–16, 11 dez. 2017.

<https://doi.org/10.21680/2447-7842.2017v3n2ID12349>. Acesso em: 30 ago. 2023.

SILVA, J. L. C.; FREIRE, G. H. de A. Um olhar sobre a origem da ciência da informação: indícios embrionários para sua caracterização identitária. **Encontros**

Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 17, n. 33, p. 1–29, 17 abr. 2012. <https://doi.org/10.5007/1518-2924.2012v17n33p1>.

SMITH, J. Data Science as an academic discipline. **Data Science Journal**, v. 5, p. 163-164, 2006. Disponível em:

<https://datascience.codata.org/articles/abstract/10.2481/dsj.5.163/&qt>. Acesso em: 06 set. 2023.

SVENONIUS, E. The intellectual foundations of information organization. Cambridge: The MIT Press, c2000. 255p.

TARGINO, Maria das Graças. COMUNICAÇÃO CIENTÍFICA: uma revisão de seus elementos básicos. **Informação & Sociedade: Estudos**, 30 jan. 2000. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/326>. Acesso em: 17 ago. 2021.

TAYLOR, R. S. Professional aspects of information science and technology. *In*:

CUADRA, C. A. (Ed.). **Annual Review of Information Science and Technology**.

New York: John Willey & Sons, 1966. v.1

TORRES-SALINAS, D.; CABEZAS-CLAVIJO, Á.; JIMÉNEZ-CONTRERAS, E.

Altmetrics: New indicators for scientific communication in Web 2.0. **Comunicar**, [s. l.], v. 21, n. 41, p. 53–60, 2013.

VALEIRO, Palmira Moriconi; PINHEIRO, Lena Vania Ribeiro. Da comunicação científica à divulgação. **Transinformação**, v. 20, p. 159–169, ago. 2008. Disponível em:

<https://brapci.inf.br/index.php/res/v/116012>. Acesso em: 25 out. 2021.

VALOIS, E. C.; RAMOS, M. G.; RODRIGUES, N. S. S.; ESTEVÃO, S. N. N. de M.

Comunicação científica e usuários elementos de discussão. **Ciência da Informação**, v. 18, n. 1, 1989. DOI [10.18225/ci.inf.v18i1.320](https://doi.org/10.18225/ci.inf.v18i1.320). Disponível em:

<https://revista.ibict.br/ciinf/article/view/320>. Acesso em: 31 ago. 2023.

VANZ, Samile Andrea de Souza; STUMPF, Ida Regina Chittó. Colaboração científica: revisão teórico-conceitual. **Perspectivas em Ciência da Informação**, v. 15, p. 42–55, ago. 2010. Disponível em: <https://doi.org/10.1590/S1413-99362010000200004>. Acesso em: 25 out. 2021.

VIGNOLI, R. G.; ALMEIDA, P. O. P. D.; CATARINO, M. E. Folksonomias como ferramenta da organização e representação da informação. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, [s. l.], v. 12, n. 2, p. 120, 2014.

VILAN FILHO, J. L.; SOUZA, H. B.; MUELLER, S. Artigos de periódicos científicos das áreas de informação no Brasil: evolução da produção e da autoria múltipla. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 13, n. 2, p. 2-17, maio/ago. 2008.

WAGNER, C. S.; LEYDESDORFF, L. Network structure, self-organization, and the growth of international collaboration in science. **Research Policy, Amsterdam**, v. 34, p. 1608-1618, 2005.

WEITZEL, S. D. R. **Os repositórios de e-prints como nova forma de organização da produção científica: o caso da área das Ciências da Comunicação no Brasil**. 2006. Doutorado em Cultura e Informação - Universidade de São Paulo, São Paulo, 2006. Disponível em: <http://www.teses.usp.br/teses/disponiveis/27/27151/tde-14052009-133509/>. Acesso em: 30 ago. 2023.

WELLER, K.; PUSCHMANN, C. Twitter for Scientific Communication: How Can Citations/References be Identified and Measured?. [s. l.],

YANG, S.; YUAN, Q.; DONG, J. Are Scientometrics, Informetrics, and Bibliometrics Different?. **Data Science and Informetrics**, [s. l.], v. 01, n. 01, p. 50, 2020b.

YEPES, J. L. **La documentación como disciplina: teoría e historia**. 2ª ediçãoed. Pamplona: EUNSA. EDICIONES UNIVERSIDAD DE NAVARRA, S.A., 1995.

ZIVIANI, A.; PORTO, F.; OGASAWARA, E. Ciência de dados: desafio para a ciência, indústria e governo. **Com Ciência**, 10 jul. 2015. Disponível em: <https://www.comciencia.br/comciencia/handler.php?section=8&edicao=115&id=1381>. Acesso em: 30 ago. 2023.

ZIVIANI, A.; PORTO, F.; OGASAWARA, E. Ciência de dados: desafio para a ciência, indústria e governo. [s. l.], 2015. Disponível em: <https://www.comciencia.br/comciencia/handler.php?section=8&edicao=115&id=1381>. Acesso em: 30 ago. 2023.

ZOLLMAN, K.J.S. Learning to collaborate. In: BOYERKASSEM, T.; MAYO-WILSON, C.; WEISBERG, M. **Scientific collaboration and collective knowledge**. New York: Oxford University Press, 2018. p. 65-77.