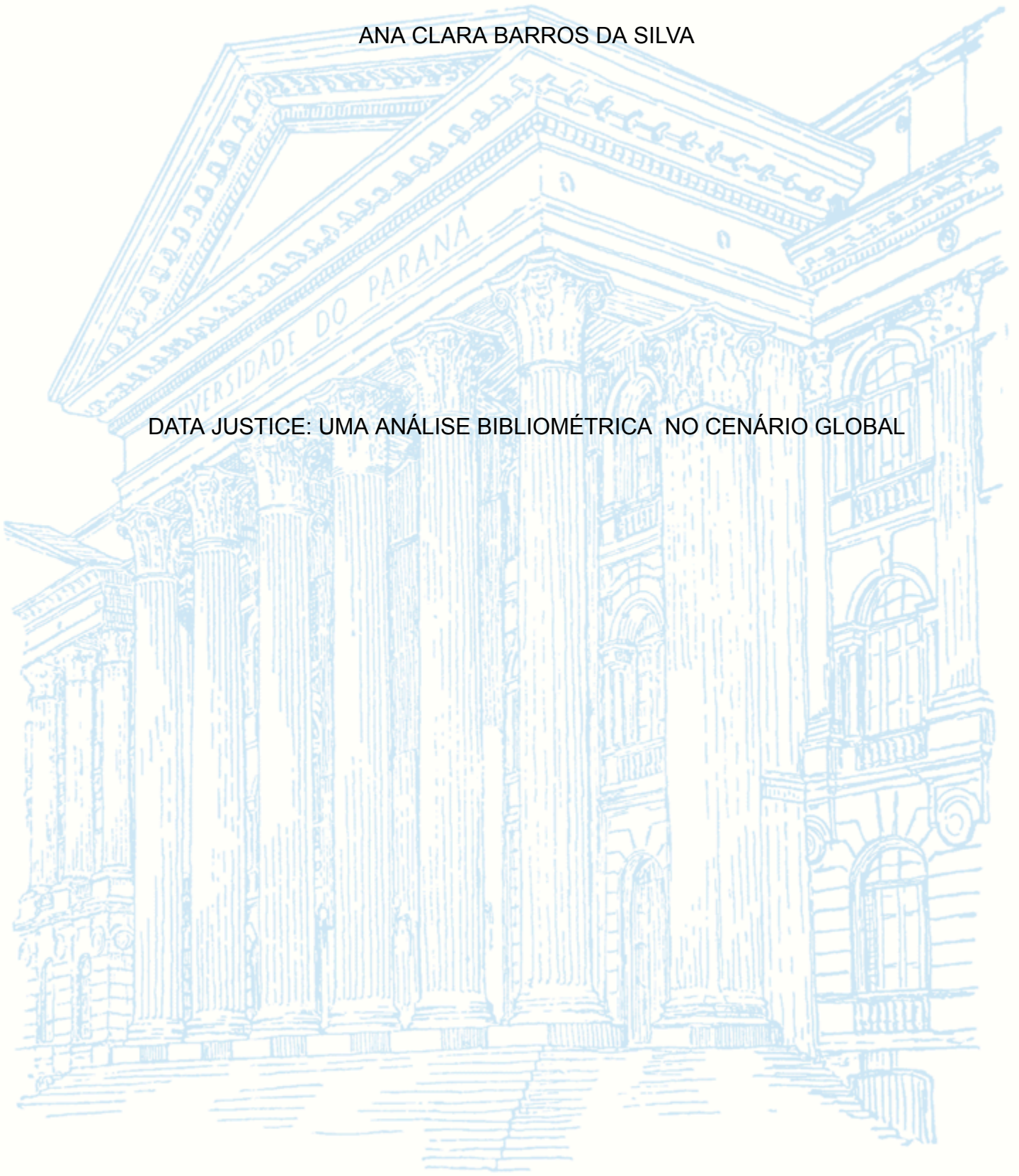


UNIVERSIDADE FEDERAL DO PARANÁ

ANA CLARA BARROS DA SILVA

DATA JUSTICE: UMA ANÁLISE BIBLIOMÉTRICA NO CENÁRIO GLOBAL



CURITIBA

2024

ANA CLARA BARROS DA SILVA

DATA JUSTICE: UMA ANÁLISE BIBLIOMÉTRICA NO CENÁRIO GLOBAL

Trabalho de Conclusão de Curso apresentado ao curso de Graduação em Gestão da Informação, Setor de Sociais Aplicadas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Bacharel em Gestão da Informação.

Orientador(a): Prof(a). Dr(a). Taiane Ritta Coelho

CURITIBA

2024

## **AGRADECIMENTOS**

Agradeço à Universidade Federal do Paraná (UFPR), instituição que me proporcionou o ambiente acadêmico e as oportunidades necessárias para o desenvolvimento pessoal e profissional ao longo desta jornada.

À minha orientadora, professora Taiane Ritta, expresso meu profundo agradecimento pela paciência, pelo conhecimento compartilhado e pelo constante apoio. Sua orientação foi essencial para que este trabalho pudesse se concretizar, e sou grata pela confiança depositada em mim durante esse processo.

Às minhas amigas Isis Guazina, Luana Horze, Helena Marques e ao meu amigo Marcos Zellner, minha sincera gratidão pelo companheirismo, pelas conversas que aliviaram os momentos de tensão e pelas risadas que tornaram esta caminhada mais leve.

Um agradecimento especial à Paula Melo, cuja motivação, apoio e ajuda constante foram fundamentais para que eu conseguisse superar os desafios deste percurso.

Faço aqui uma menção honrosa ao meu querido colega Patrick Fernandes que me guiou por muitas vezes ao longo deste trajeto, foi um enorme privilégio aprender com você.

À minha família, que sempre acreditou em mim e me deu força para continuar, minha eterna gratidão. Especialmente a minha prima Kacilda Barros, minha querida vó Marline Barros e minha incansável mãe Simone Barros, obrigada pelo apoio incondicional.

## RESUMO

Este trabalho apresenta uma análise bibliométrica da produção científica global sobre o conceito de justiça de dados (*data justice*), com foco nas discussões éticas, sociais e políticas que envolvem o uso e a governança de dados na era digital. A pesquisa justifica-se pela relevância crescente do tema, dado seu impacto na transformação digital e nos debates sobre vigilância, privacidade e desigualdade. Utilizando dados da base *Web of Science* e ferramentas como o VOSviewer, foram mapeados os principais autores, palavras-chave e clusters temáticos relacionados ao tema. Os resultados destacaram a centralidade de tópicos como vigilância algorítmica, governança de dados e transformação digital, bem como lacunas na representação de contextos regionais e culturais. Este estudo contribui para a representação do campo, ao mesmo tempo em que sugere novas direções de pesquisa para aprofundar a compreensão do impacto social e ético da justiça de dados.

Palavras-chave: *Data Justice*; análise bibliométrica; justiça de dados; governança de dados; vigilância, privacidade; transformação digital.

## **ABSTRACT**

This article presents a bibliometric analysis of global scientific production on the concept of data justice, focusing on the ethical, social and political discussions surrounding the use and governance of data in the digital age. The research is justified by the growing relevance of the topic, given its impact on digital transformation and debates on surveillance, privacy and inequality. Using data from Web of Science database and tools such as VOSviewer, the main authors, keywords and thematic clusters related to the topic were mapped. The results highlighted the centrality of topics such as algorithmic surveillance, data governance and digital transformation, as well as gaps in the representation of regional and cultural contexts. This study contributes to the representation of the field, while suggesting new research directions to deepen the understanding of the social and ethical impact of data justice.

**Keywords:** Data Justice; bibliometric analysis; data governance; surveillance; privacy; digital transformation.

## LISTA DE ILUSTRAÇÕES

|  |    |
|--|----|
| FIGURA 1 – ANÁLISE RELACIONAL DOS PRINCIPAIS AUTORES.....                                | 33 |
| FIGURA 2 – VISUALIZAÇÃO DA REDE DAS PRINCIPAIS REVISTAS.....                             | 39 |
| FIGURA 3 – VISUALIZAÇÃO DA REDE DE COCORRÊNCIAS DE<br>PALAVRAS-CHAVE (ALL KEYWORDS)..... | 44 |
| FIGURA 4 – REDE DOS PRINCIPAIS ARTIGOS BASEADO EM CITAÇÕES.....                          | 46 |
| FIGURA 5 – REDE DE CONTRIBUIÇÃO POR PAÍSES.....  | 49 |

## LISTA DE QUADROS

|  |    |
|--|----|
| QUADRO 1 – CRONOGRAMA DE EXECUÇÃO.....   | 29 |
| QUADRO 2 – CARACTERÍSTICAS GERAIS DAS PUBLICAÇÕES.....                             | 32 |
| QUADRO 3 – RESUMO GERAL DA REDE DOS PRINCIPAIS AUTORES.....                        | 36 |
| QUADRO 4 – RESUMO GERAL DE CO-OCORRÊNCIAS DE PALAVRAS-CHAVE<br>(ALL KEYWORDS)..... | 45 |

## LISTA DE GRÁFICOS

|   |    |
|---|----|
| GRÁFICO 1 – DISTRIBUIÇÃO TEMPORAL DAS PUBLICAÇÕES.....  | 32 |
| GRÁFICO 2 – VISUALIZAÇÃO DA DISTRIBUIÇÃO DE PRODUTIVIDADE<br>CIENTÍFICA ENTRE OS AUTORES..... | 37 |
| GRÁFICO 3 – PERIÓDICOS MAIS RELEVANTES PARA O TEMA BASEADO EM<br>CITAÇÕES.....                | 41 |
| GRÁFICO 4 – FONTES DE INFORMAÇÃO EM ZONAS DE PRODUTIVIDADE.....                               | 42 |



## SUMÁRIO

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>INTRODUÇÃO.....</b>   | <b>10</b> |
| 1.1      | PROBLEMÁTICA DA PESQUISA.....  | 11        |
| 1.2      | JUSTIFICATIVA.....   | 12        |
| 1.3      | OBJETIVOS.....   | 12        |
| 1.3.1    | Objetivo geral.....  | 13        |
| 1.3.2    | Objetivos específicos.....   | 13        |
| <b>2</b> | <b>REFERENCIAL TEÓRICO.....</b>  | <b>14</b> |
| 2.1      | DATA JUSTICE.....  | 14        |
| 2.2      | DATIFICAÇÃO.....   | 16        |
| 2.3      | GOVERNANÇA DE DADOS.....   | 17        |
| 2.4      | FRAMEWORKS E MODELOS.....  | 18        |
| 2.5      | DADOS.....   | 18        |
| 2.6      | COLETA DE DADOS.....   | 20        |
| 2.6.1    | Fonte dos dados.....   | 20        |
| 2.6.2    | Desafios na coleta de dados.....   | 20        |
| 2.7      | ANÁLISE DE DADOS.....  | 22        |
| 2.7.1    | Técnicas e ferramentas de análise.....   | 22        |
| 2.8      | TRANSFORMAÇÃO DIGITAL.....   | 23        |
| 2.8.1    | Tecnologias emergentes.....  | 24        |
| <b>3</b> | <b>METODOLOGIA.....</b>  | <b>25</b> |
| 3.1      | TIPO DE PESQUISA.....  | 25        |
| 3.2      | COLETA DE DADOS.....   | 26        |
| 3.3      | ANÁLISE DE DADOS.....  | 27        |
| 3.4      | CRONOGRAMA DE EXECUÇÃO.....  | 29        |
| <b>4</b> | <b>RESULTADOS.....</b>   | <b>31</b> |
| 4.1      | ANÁLISE GERAL DAS PUBLICAÇÕES.....   | 31        |
| 4.2      | ANÁLISE DOS PRINCIPAIS AUTORES.....  | 33        |
| 4.2.1    | Análise da produtividade científica por autor segundo a Lei de Lotka.....        | 36        |
| 4.3      | ANÁLISE DAS PRINCIPAIS REVISTAS.....   | 38        |
| 4.3.1    | Fontes de informação em zonas de produtividade segundo a Lei de<br>Bradford..... | 42        |
| 4.4      | ANÁLISE DE COCORRÊNCIA DE PALAVRAS-CHAVE.....                                    | 43        |
| 4.5      | ANÁLISE DOS PRINCIPAIS ARTIGOS.....  | 46        |
| 4.6      | ANÁLISE DA REDE DE CONTRIBUIÇÃO POR PAÍS.....                                    | 48        |
| <b>5</b> | <b>CONSIDERAÇÕES FINAIS.....</b>   | <b>51</b> |
|          | <b>REFERÊNCIAS.....</b>  | <b>53</b> |

## 1 INTRODUÇÃO

Na era digital, os dados emergiram como um recurso essencial e estratégico, frequentemente comparado ao petróleo em termos de valor. Segundo *The Economist* (2017), "o recurso mais valioso do mundo não é mais o petróleo, mas os dados", o que evidencia sua centralidade na modelagem da vida contemporânea. Desde a personalização de serviços até a formulação de políticas públicas, a capacidade de coletar, analisar e utilizar grandes volumes de dados possibilitou avanços significativos em áreas como saúde, educação e segurança. Contudo, esses benefícios vêm acompanhados de desafios substanciais (MAHANTI, 2019).

As implicações sociais do uso de dados incluem questões relacionadas à privacidade, vigilância e a possibilidade de discriminação e exclusão. Zuboff (2019) alerta sobre os riscos da coleta de dados em massa, que pode promover formas de controle social e vigilância. Ao mesmo tempo, a Autoridade Europeia para a Proteção de Dados (2016) enfatiza a necessidade de encontrar um equilíbrio entre o uso de dados e a proteção dos direitos individuais e da privacidade. O contexto atual, marcado pela crescente integração de tecnologias digitais, como inteligência artificial e aprendizado de máquina, amplia esses desafios. Crawford e Calo (2016) argumentam que essas tecnologias podem, inadvertidamente, ampliar as desigualdades e levantar dilemas éticos.

Nesse contexto, o conceito de justiça de dados ganha relevância. Ele envolve a aplicação dos princípios de justiça social ao gerenciamento de dados, garantindo que sua coleta, uso e distribuição sejam feitos de forma ética e justa, sem reforçar a desigualdade ou a discriminação (DENCIC et al., 2019). Para Taylor (2017), a justiça de dados consiste em garantir o acesso igualitário aos benefícios da dataficação (processo de transformação de informações em dados digitais), promovendo a inclusão e protegendo as comunidades dos impactos negativos.

Segundo Dencic et al. (2019) o conceito de justiça de dados busca enquadrar a coleta e o uso de dados como poder, política, inclusão e interesses, ao invés de meramente se preocupar com eficiência ou privacidade. Esse conceito propõe uma abordagem crítica e inclusiva para garantir que os benefícios da era digital sejam compartilhados equitativamente entre todos os segmentos da sociedade.

Entretanto, a maneira como dados são coletados e analisados pode reforçar desigualdades e criar novas formas de exclusão. O'Neil (2016) adverte que algoritmos e modelos preditivos, quando projetados sem critérios éticos claros, podem institucionalizar discriminações. Da mesma forma, Eubanks (2018) aponta que a tecnologia, ao ser implementada sem considerar o contexto social, pode agravar exclusões, intensificando desigualdades já existentes. Esses argumentos reforçam a necessidade de adotar uma abordagem ética, capaz de integrar os dados nas práticas sociais e políticas de forma a promover o bem comum.

## 1.1 PROBLEMÁTICA DA PESQUISA

A crescente digitalização das interações sociais, econômicas e políticas impulsionou o desenvolvimento de tecnologias capazes de capturar, processar e analisar vastas quantidades de dados. Este processo tem transformado diversos setores, desde a saúde e a educação até a segurança pública e o setor financeiro (MASIERO; BUDDHA, 2021). Entretanto, a expansão do uso de dados em escala global levanta importantes questões éticas e sociais, que muitas vezes são negligenciadas em favor dos avanços tecnológicos e dos benefícios econômicos. Nesse contexto, a ausência de uma governança efetiva e de princípios que orientem a coleta e o uso dos dados tem potencial para ampliar desigualdades sociais, afetando desproporcionalmente grupos vulneráveis e minorias (MASIERO; DAS, 2019).

Isso destaca a importância de investigar de maneira crítica como a ciência está abordando o conceito de justiça de dados para garantir uma sociedade mais justa e equitativa. Ao entender como os estudos sobre justiça de dados estão sendo desenvolvidos, é possível avaliar se as ciências contemporâneas estão acompanhando as demandas éticas emergentes da sociedade digital.

A problemática central desta pesquisa consiste em compreender a relevância do estudo sobre justiça de dados (data justice) no contexto das ciências contemporâneas, a partir da análise de publicações científicas. Nesse sentido, a investigação busca responder à seguinte questão: como a literatura científica sobre justiça de dados está organizada globalmente?

## 1.2 JUSTIFICATIVA

A escolha do tema Data Justice é justificada pela necessidade de abordar as implicações sociais, políticas e econômicas do uso intensivo de dados na era digital. Na prática, a crescente dependência de dados para a personalização de serviços e a elaboração de políticas públicas revisitou a questão crítica de como os dados são geridos e utilizados (VERA et al., 2019). Em uma sociedade onde a coleta massiva de dados pode levar a formas de vigilância e controle social, é fundamental garantir que esses dados sejam utilizados de maneira ética e equitativa (MASIERO; BUDDHA, 2021).

No âmbito prático, a justiça de dados visa assegurar que todos os grupos sociais tenham acesso igualitário aos benefícios proporcionados pelas tecnologias digitais. O manejo inadequado de dados pode exacerbar desigualdades existentes e criar formas de exclusão social (EUBANKS 2018). Tecnologias emergentes, como a inteligência artificial e o aprendizado de máquina, estão cada vez mais presentes em nosso dia a dia. Sem uma consideração adequada dos princípios de justiça social, essas tecnologias podem amplificar desigualdades e introduzir novos desafios éticos, afetando negativamente grupos vulneráveis. Garantir a justiça de dados é, portanto, uma questão de promover a inclusão digital e a equidade no acesso aos benefícios tecnológicos, prevenindo a discriminação e a marginalização (HEEKS; RENKEN, 2018).

Do ponto de vista acadêmico, a justiça de dados é um campo de estudo novo e totalmente relevante para a Gestão da Informação. O foco deste trabalho está trazer a representação de um enquadramento teórico que transcende as tradicionais preocupações com privacidade e segurança, incorporando princípios de justiça social.

Por fim, analisar o desenvolvimento e a produção científica sobre justiça de dados permite mapear as abordagens e as lacunas existentes na literatura, oferecendo uma base sólida para fomentar políticas e práticas que promovam uma sociedade digital mais equitativa e inclusiva.

## 1.3 OBJETIVOS

Os objetivos desta pesquisa estão divididos em objetivo geral e específicos.

### 1.3.1 Objetivo geral

O objetivo principal desta pesquisa é analisar como a produção científica existente sobre o conceito de justiça de dados está organizada globalmente, compreendendo suas implicações e aplicações nas discussões sobre justiça social no uso de dados.

### 1.3.2 Objetivos específicos

- Mapear a produção científica sobre *Data Justice*;
- Avaliar a representação global das pesquisas sobre *Data Justice*;
- Explorar a distribuição de produtividade científica dos autores no campo de *Data Justice*.

Ao atingir estes objetivos será possível mostrar as principais publicações, autores e periódicos que têm contribuído para o desenvolvimento desse campo. Além disso, analisar como os estudos sobre justiça de dados têm sido desenvolvidos em diferentes regiões do mundo, destacando os que países mais contribuem e apresentar como a produtividade científica sobre o tema da justiça de dados está distribuída entre os autores.

## 2 REFERENCIAL TEÓRICO

O referencial teórico deste trabalho aborda a justiça de dados como um conceito emergente e interdisciplinar, fundamentado em princípios de justiça social aplicados ao uso, coleta e distribuição de dados. Além disso, o referencial traz o processo de dataficação e os desafios éticos e sociais associados, oferecendo uma base conceitual para compreender o papel das políticas de proteção e dos *frameworks* que visam assegurar um uso justo e inclusivo dos dados na sociedade digital.

### 2.1 DATA JUSTICE

A justiça de dados, ou *data justice*, é um conceito emergente que se refere à aplicação de princípios de justiça social na coleta, uso e distribuição de dados. Segundo Dencic, Hintz e Cable (2016), a justiça de dados visa garantir que as práticas de dataficação respeitem os direitos humanos e promovam a equidade social, abordando questões de poder, inclusão e transparência.

Heeks e Renken (2018) definem a justiça de dados como um esforço para assegurar que os dados sejam utilizados de maneira ética e justa, evitando a perpetuação de desigualdades e discriminações estruturais. A relevância da justiça de dados está em seu potencial para promover um uso mais responsável e equitativo das tecnologias de informação, assegurando que todos os grupos sociais possam se beneficiar dos avanços tecnológicos.

Conforme Taylor (2017), o aumento na disponibilidade de dados digitais, especialmente aqueles gerados como subprodutos de atividades humanas em dispositivos tecnológicos, traz implicações significativas sobre como as pessoas são vistas e tratadas pelo estado e pelo setor privado. Esse fenômeno, denominado "dataficação", introduz desafios de visibilidade, engajamento digital e discriminação, que precisam ser enfrentados com um arcabouço ético sólido. Ao destacar que os sistemas de dados têm o poder de categorizar, intervir e reforçar desigualdades sociais, Taylor propõe pilares para a justiça de dados, incluindo (in)visibilidade, (des)engajamento com tecnologia e combate à discriminação, como alicerces para promover um uso mais equitativo e responsável da informação digital.

Além disso, Taylor (2017) argumenta que o modelo tradicional de proteção de dados, baseado em direitos individuais, é insuficiente frente às dinâmicas globais e estruturais da economia de dados. A autora aponta que a interface público-privada na coleta e uso de dados intensifica as assimetrias de poder, dificultando a responsabilização e o acesso a mecanismos de reparação. Nesse contexto, a justiça de dados não se limita a proteger a privacidade, mas também visa garantir que indivíduos e grupos possam optar por não participar de bases de dados comerciais e governamentais. Ao integrar princípios de justiça social, o conceito busca equilibrar a necessidade de representação adequada com a preservação da autonomia e dignidade das pessoas em um mundo cada vez mais datafocado.

Segundo Gangadharan (2020), as práticas de coleta e análise de dados muitas vezes exacerbam desigualdades sociais existentes, especialmente quando envolvem grupos marginalizados. Questões como a vigilância em massa, discriminação algorítmica e a perda de privacidade são preocupações centrais no debate sobre justiça de dados.

Somado a isso, Heeks e Renken (2018) destacam que a implementação de políticas de justiça de dados é crucial para proteger os direitos dos cidadãos, especialmente em um contexto em que as desigualdades sociais são profundas e históricas. As questões éticas envolvem garantir a transparência nos processos de coleta e análise de dados e a responsabilidade das organizações em proteger os dados pessoais e sensíveis dos indivíduos.

A criação de políticas e regulamentações adequadas é fundamental para assegurar a justiça de dados. A União Europeia, por exemplo, implementou o Regulamento Geral sobre a Proteção de Dados (GDPR), que estabelece padrões rigorosos para a proteção de dados pessoais e impõe penalidades significativas para violações. Nos Estados Unidos, o *California Consumer Privacy Act* (CCPA) oferece direitos similares aos residentes da Califórnia. No Brasil, a Lei Geral de Proteção de Dados (LGPD) foi sancionada em 2018 criando um marco legal para a proteção de dados pessoais e garantindo direitos aos titulares dos dados (MARTINS, 2020). Estas regulamentações visam proteger a privacidade dos indivíduos e garantir que os dados sejam utilizados de maneira ética e responsável.

## 2.2 DATIFICAÇÃO

O termo dataficação refere-se ao processo de transformar aspectos da vida cotidiana em dados quantificáveis, que podem ser coletados, analisados e utilizados para diversos fins. Mejias e Couldry (2019) definem *datafication* como a capacidade de capturar, quantificar e analisar dados de praticamente qualquer atividade ou processo, desde as interações sociais nas redes sociais até os movimentos físicos registrados por dispositivos *IoT* (do inglês *Internet of Things*).

A dataficação tem impactos profundos tanto sociais quanto econômicos. Do ponto de vista social, a capacidade de coletar e analisar dados em larga escala pode melhorar serviços públicos, como saúde e educação, através da personalização e eficiência dos serviços prestados (SOUTHERTON, 2022). No entanto, também levanta preocupações sobre privacidade e vigilância, uma vez que a coleta constante de dados pode resultar em monitoramento excessivo e perda de autonomia pessoal (ZUBOFF, 2015).

Economicamente, a dataficação permite que empresas criem modelos de negócio baseados em dados, como publicidade direcionada e serviços personalizados, aumentando a eficiência e a competitividade (MAYER-SCHÖNBERGER; CUKIER, 2013).

A tomada de decisão baseada em dados (*data-driven decision making*) refere-se ao uso de dados e análises quantitativas para informar e orientar decisões estratégicas e operacionais (WU et al., 2021)

De acordo com Brynjolfsson e McAfee (2014), às organizações que adotam uma abordagem baseada em dados tendem a ser mais eficientes e inovadoras, pois suas decisões são fundamentadas em evidências concretas e análises rigorosas. Silva (2024) destaca que a implementação de uma cultura orientada por dados está se tornando uma prioridade para empresas que buscam aumentar sua competitividade e responder rapidamente às mudanças do mercado. A análise de grandes volumes de dados permite identificar tendências, prever comportamentos e otimizar processos, resultando em decisões mais precisas e eficazes (SOUTHERTON, 2022).

Apesar dos benefícios, a dataficação apresenta diversos desafios. Um dos principais desafios é garantir a qualidade dos dados coletados, pois dados



imprecisos ou incompletos podem levar a conclusões errôneas e decisões inadequadas (FLENSBURG; LOMBORG, 2023). A partir disso, a privacidade e a segurança dos dados são preocupações constantes, especialmente no contexto de regulamentações rigorosas como a Lei Geral de Proteção de Dados (LGPD) no Brasil (MARTINS, 2020). Outro desafio é a necessidade de infraestrutura tecnológica avançada e profissionais qualificados para coletar, armazenar e analisar grandes volumes de dados. No entanto, a dataficação também cria inúmeras oportunidades, como a possibilidade de inovação em produtos e serviços, a melhoria da eficiência operacional e a capacidade de oferecer experiências personalizadas aos clientes (MEJIAS; COULDRY, 2019).

### 2.3 GOVERNANÇA DE DADOS

A governança de dados refere-se ao conjunto de práticas e políticas destinadas a garantir a gestão eficaz dos dados dentro de uma organização. Além disso, a governança de dados envolve a definição de responsabilidades, a implementação de processos padronizados e o estabelecimento de normas e diretrizes para a coleta, armazenamento, uso e proteção dos dados (ALHASSAN; SAMMON; DALY, 2016). A governança eficaz é crucial para assegurar a integridade, a disponibilidade e a confidencialidade dos dados, minimizando riscos e garantindo conformidade com regulamentações legais.

De acordo com Otto e Weber (2011) a governança de dados também promove a transparência e a responsabilidade, facilitando a tomada de decisões informadas e a mitigação de problemas relacionados à qualidade dos dados. A implementação de uma estrutura robusta de governança de dados ajuda as organizações a maximizar o valor dos dados, assegurando que eles sejam usados de forma ética e eficiente, e que atendam às exigências de conformidade regulatória (JANSSEN et al., 2020).

A qualidade dos dados também desempenha um papel central na governança de dados, conforme destacado por Wang e Strong (1996). A garantia de que os dados sejam precisos, consistentes, completos e acessíveis é um dos principais objetivos de qualquer programa de governança. Ferramentas e tecnologias, como *master data management* (MDM) e *data cataloging*, têm sido amplamente utilizadas

para apoiar iniciativas de governança de dados, facilitando a gestão centralizada e a padronização dos dados dentro das organizações (LOSHIN, 2010).

## 2.4 FRAMEWORKS E MODELOS

Conforme ressaltado por Khatri e Brown (2010), um componente essencial da governança de dados é a criação de um modelo de governança que alinhe estratégias organizacionais com as operações de gestão de dados. Esse modelo deve considerar aspectos como a definição de papéis e responsabilidades (por exemplo, *data stewards* e *data owners*), a formalização de processos de tomada de decisão relacionados a dados e o monitoramento contínuo para garantir que as políticas sejam seguidas de forma eficaz.

Usando como base a governança de dados, para implementar a justiça de dados, diversos frameworks e modelos têm sido propostos. Heeks e Renken (2018) propõem um modelo de justiça de dados que abrange três dimensões principais: acesso, uso e impacto. Este modelo sugere que a justiça de dados deve garantir que todos tenham acesso igualitário aos benefícios da dataficação, que os dados sejam usados de maneira ética e que os impactos sociais sejam positivos e equitativos.

Ramasubramanian e Dutta (2024) propõem o modelo *CODE^SHIFT*, um framework de justiça de dados desenvolvido para promover o impacto coletivo e a transformação social. Baseado na abordagem centrada na cultura (CCA), este modelo visa resistir às ciências sociais hegemônicas dentro da teorização orientada por dados e propõe uma reinterpretação dos dados para desafiar as estruturas de capitalismo e colonialismo que perpetuam práticas de exploração e extração.

## 2.5 DADOS

Na era digital, os dados são considerados o novo petróleo, essencial para o funcionamento eficiente e inovador das organizações. Os dados representam um conjunto de informações estruturadas ou não, que podem ser processadas e analisadas para gerar percepções valiosas (OLIVEIRA; REZENDE; SARAMAGO, 2024). A importância dos dados reside em sua capacidade de transformar informações brutas em conhecimento acionável, essencial para a tomada de

decisões estratégicas e operacionais. Com a crescente digitalização de processos e a expansão da Internet das Coisas (*IoT*), a quantidade de dados gerados aumentou exponencialmente, fazendo com que as organizações precisem desenvolver capacidades robustas para gerenciar e explorar esses dados eficazmente (GANDOMI; HAEGERMANN, 2015).

Segundo Inmon (2005) os dados podem ser classificados em três tipos principais: estruturados, semiestruturados e não estruturados. Dados estruturados são aqueles organizados em formatos predefinidos, como tabelas em bancos de dados relacionais, que facilitam a busca e a análise. Dados semiestruturados, por outro lado, possuem alguma forma de organização, mas não seguem um esquema rígido, como arquivos *XML* e *JSON*.

De acordo com Mello et al., (2024) , esses dados mantêm uma estrutura flexível que facilita a sua adaptação e integração em diferentes sistemas. Finalmente, dados não estruturados não possuem uma estrutura específica, incluindo texto de e-mails, vídeos, imagens e postagens em redes sociais. Marr (2015) afirma que a diversidade de tipos de dados requer abordagens variadas para sua coleta, armazenamento e análise, exigindo ferramentas e técnicas específicas para lidar com cada categoria.

A visão dos dados como um ativo estratégico é altamente reconhecida no contexto organizacional. Segundo Davenport e Prusak (2003), os dados são essenciais para a inovação e a competitividade das empresas, permitindo a personalização de produtos e serviços, a otimização de processos e a antecipação de tendências de mercado. Tratar os dados como ativos permite que as organizações extraiam valor significativo, transformando-os em uma vantagem competitiva sustentável. Além disso, a utilização de dados para a tomada de decisões baseada em evidências possibilita uma gestão mais precisa e eficiente, contribuindo para o sucesso a longo prazo das empresas. Esse enfoque é corroborado por Brynjolfsson e McAfee (2014) que destacam a importância da análise de dados para a definição de estratégias empresariais eficazes.

Nesse contexto, a justiça de dados surge como uma proposta para garantir que esse ativo seja tratado com responsabilidade, para evitar a ampliação das disparidades sociais e econômicas.

## 2.6 COLETA DE DADOS

A coleta de dados é um processo fundamental para qualquer organização, mas, no contexto da justiça de dados, ela deve ser conduzida com transparência, respeito aos direitos individuais e preocupação com a equidade (OLSEN, 2015). Ainda segundo Olsen (2015) às metodologias de coleta de dados podem ser classificadas em quantitativas e qualitativas. As metodologias quantitativas incluem questionários e *surveys*, que permitem a coleta de dados estruturados e numéricos, facilitando a análise estatística. Já as metodologias qualitativas, como entrevistas e grupos focais, são usadas para obter visões mais profundas e compreensivas sobre o comportamento e as percepções dos indivíduos.

### 2.6.1 Fonte dos dados

As fontes de dados são variadas e podem incluir sensores *IoT*, redes sociais, transações comerciais, entre outras. De acordo com Rosa (2020) os sensores *IoT* têm se tornado uma fonte crucial de dados, proporcionando informações em tempo real sobre diversos aspectos de ambientes urbanos, industriais e domésticos.

Redes sociais, no que lhe concerne, oferecem uma vasta quantidade de dados textuais e visuais, que podem ser analisados para entender tendências, comportamentos e sentimentos dos usuários. No entanto, uma abordagem centrada neste cenário exige que essas fontes sejam utilizadas de maneira que não comprometa a privacidade dos indivíduos.

Segundo Costa (2018) as redes sociais são um campo fértil para a coleta de dados devido ao grande volume de interações diárias. Transações comerciais fornecem dados valiosos sobre padrões de consumo e comportamento de compra dos consumidores, sendo essenciais para análises de mercado e estratégias de marketing. Toledo et al., (2024) destaca que esses dados são fundamentais para a criação de perfis detalhados de consumidores e para a personalização de ofertas e campanhas publicitárias.

### 2.6.2 Desafios na coleta de dados

A coleta de dados enfrenta diversos desafios, que incluem problemas relacionados à qualidade dos dados, privacidade, e integração de dados de diferentes fontes. Rêgo (2013) aponta que um dos principais desafios é garantir a qualidade dos dados, que pode ser afetada por erros de medição, inconsistências e dados faltantes.

A qualidade dos dados é um aspecto essencial para garantir que os dados utilizados nas análises e na tomada de decisões sejam precisos e confiáveis. Mahanti (2019) identifica vários critérios para assegurar a qualidade dos dados, incluindo precisão, completude, consistência, atualidade e relevância. A implementação de metodologias para a gestão da qualidade dos dados é fundamental para identificar e corrigir problemas, garantindo que os dados sejam adequados para seus propósitos específicos. Além disso, é ressaltado que a adoção de ferramentas e processos de limpeza, validação e enriquecimento de dados pode melhorar significativamente a qualidade dos dados, aumentando sua utilidade e valor para a organização.

A privacidade dos dados também é uma preocupação crescente, especialmente no contexto de coleta de dados pessoais e sensíveis (MARTINS, 2020). Além disso, a integração de dados provenientes de múltiplas fontes requer técnicas sofisticadas para garantir a compatibilidade e a coerência dos dados (SILVA, 2024). Soluções para esses desafios incluem a implementação de processos robustos de governança de dados, uso de técnicas avançadas de limpeza e validação de dados, e adoção de *frameworks* de segurança da informação.

As considerações éticas e legais são fundamentais na coleta de dados, especialmente no contexto de dados pessoais e sensíveis. Segundo Bioni (2020) a ética na coleta de dados envolve garantir o consentimento informado dos participantes, respeitar a privacidade e garantir a confidencialidade das informações coletadas. No Brasil, a LGPD estabelece diretrizes rigorosas para a coleta, armazenamento e uso de dados pessoais, impondo responsabilidades significativas às organizações que lidam com esses dados (MARTINS, 2020). A conformidade com a LGPD e outras regulamentações é essencial para evitar penalidades legais e garantir a confiança dos usuários e clientes.

## 2.7 ANÁLISE DE DADOS

A análise de dados é um processo crucial que envolve a inspeção, limpeza, transformação e modelagem de dados para descobrir informações úteis, chegar a conclusões e apoiar a tomada de decisões informadas. De acordo com Davenport e Prusak (2003) a análise de dados é fundamental para transformar dados brutos em informações valiosas, permitindo que as organizações entendam padrões, tendências e relações subjacentes. A capacidade de analisar dados de maneira eficaz proporciona uma vantagem competitiva significativa, tornando-se um diferencial crucial para as organizações na era digital.

### 2.7.1 Técnicas e ferramentas de análise

As técnicas de análise de dados variam desde métodos estatísticos tradicionais até abordagens computacionais avançadas. Segundo Montgomery e Runger (2014) técnicas estatísticas como regressão linear, análise de variância (ANOVA) e testes de hipóteses são amplamente utilizadas para explorar e interpretar dados quantitativos. Além disso, métodos como a análise de séries temporais e a modelagem de equações estruturais são empregados para analisar padrões temporal e relações complexas entre variáveis (HAIR et al., 2009).

Em técnicas computacionais, o uso de algoritmos de aprendizado de máquina, como árvores de decisão, redes neurais e máquinas de vetores de suporte, permite a criação de modelos preditivos e a descoberta de padrões ocultos em grandes volumes de dados (OLIVEIRA et al., 2024).

As ferramentas e *softwares* para análise de dados são fundamentais para implementar as técnicas mencionadas. De acordo com Battist e Smolski (2019), ferramentas como R e *Python* são utilizadas por sua flexibilidade e vasta gama de bibliotecas específicas para análise de dados. *Softwares* comerciais como SAS, SPSS e STATA oferecem interfaces amigáveis e funcionalidades avançadas para análise estatística e modelagem preditiva. Além disso, plataformas de big data como *Hadoop* e *Spark* são essenciais para processar e analisar grandes volumes de dados de maneira eficiente (GUAN et al., 2024).

## 2.8 TRANSFORMAÇÃO DIGITAL

A transformação digital consiste no processo de integração de tecnologias digitais em todas as áreas de uma organização, mudando fundamentalmente como ela opera e entrega valor a seus clientes. De acordo com Rogers (2019) a transformação digital não se limita à implementação de novas tecnologias, mas também envolve a mudança de mentalidade e cultura organizacional para aproveitar plenamente as oportunidades oferecidas pelas tecnologias emergentes. Ela envolve a reestruturação dos processos de negócios, a adoção de novas formas de pensar e trabalhar e a criação de novos modelos de negócios mais ágeis e capazes de responder às mudanças nas demandas dos clientes e nas condições do mercado.

A transformação digital afeta profundamente a estrutura, os processos e a cultura organizacional. De acordo com Westerman, Bonnet e McAfee (2014), a digitalização pode levar a uma reorganização completa das operações internas, proporcionando maior eficiência e flexibilidade. As empresas são incentivadas a adotar estruturas organizacionais mais horizontais, que facilitam a colaboração e a inovação. Além disso, a transformação digital promove mudanças culturais significativas, incentivando uma mentalidade de experimentação e aprendizado contínuo.

Albertin e Moura (2021) enfatizam que a transformação digital é impulsionada pela necessidade de inovação constante e pela pressão competitiva, exigindo que as empresas se adaptem rapidamente para permanecerem relevantes.

Essa transformação apresenta desafios e oportunidades para as empresas. Um dos principais desafios é a resistência à mudança, tanto por parte dos funcionários quanto da gerência, o que pode dificultar a adoção de novas tecnologias e processos (KOTTER, 1996). Além disso, a necessidade de investimentos substanciais em tecnologia e infraestrutura pode representar um obstáculo para muitas empresas, especialmente as de menor porte (BHARADWAJ et al., 2013).

No entanto, as oportunidades geradas pela transformação digital são vastas. As empresas podem melhorar significativamente sua eficiência operacional, oferecer melhores experiências aos clientes e inovar em seus produtos e serviços.

### 2.8.1 Tecnologias emergentes

O conceito de *big data* refere-se ao processamento e análise de conjuntos de dados que são grandes demais para serem gerenciados por ferramentas tradicionais de banco de dados. Segundo Mayer-Schönberger e Cukier (2013), *big data* é caracterizado pelos “cinco Vs”: volume, variedade, velocidade, veracidade e valor. A capacidade de analisar grandes volumes de dados em tempo real permite às organizações tomar decisões mais rápidas e precisas. No entanto, o conceito também apresenta desafios significativos, como a necessidade de infraestrutura de TI robusta, habilidades especializadas em análise de dados e preocupações com a privacidade e segurança dos dados.

Além disso, dados são agora considerados um ativo estratégico, crucial para a tomada de decisões informadas e a criação de valor. Oliveira, Rezende e Saramago (2024) enfatizam que a gestão eficaz de dados envolve não apenas a coleta e o armazenamento, mas também a análise preditiva e a aplicação de *big data* e inteligência artificial (IA) para gerar percepções acionáveis.

A aplicação de IA e aprendizado de máquina (ML) na análise de dados revoluciona como as informações são processadas e interpretadas. Segundo Goodfellow, Bengio e Courville (2016), a IA e o ML permitem a criação de modelos que podem aprender e fazer previsões baseadas em dados. Algoritmos de ML, como redes neurais profundas, árvores de decisão e *k-means*, são usados para identificar padrões complexos e realizar tarefas preditivas com alta precisão, o papel dos dados na transformação digital é particularmente significativo.

Contudo, a aplicação dessas tecnologias precisa ser orientada pela justiça de dados, garantindo que seu uso respeite seus princípios. A análise de grandes volumes de dados permite decisões mais rápidas e precisas, mas deve ser acompanhada de cuidados com os seus processos.



### 3 METODOLOGIA

A metodologia adotada neste estudo combina uma análise bibliométrica com o uso de ferramentas especializadas para mapear e compreender a produção científica sobre o tema. Esta abordagem permite identificar os principais autores, palavras-chave, redes de coautoria e tendências emergentes no campo. Utilizando dados extraídos da base Web of Science e ferramentas como o VOSviewer, foram analisados aspectos como coocorrência de termos, clusters temáticos e conexões entre os estudos.

Tal metodologia se justifica pela sua capacidade de oferecer uma representação, por meio de metadados, sobre o estado atual da pesquisa, contribuindo para a compreensão das dimensões, impactos e perspectivas relacionados à justiça de dados no cenário global.

#### 3.1 TIPO DE PESQUISA

Esta pesquisa foi conduzida utilizando uma abordagem bibliométrica, reconhecida como um método robusto para avaliar e mapear a produção científica em um campo de estudo. A bibliometria é uma técnica quantitativa que analisa a distribuição de publicações, citações e padrões de colaboração científica, oferecendo percepções sobre a dinâmica de um campo de pesquisa, incluindo seus autores mais citados, redes de coautoria e palavras-chave recorrentes (DONTHU et al., 2021).

A análise bibliométrica permite não apenas identificar as principais contribuições teóricas e empíricas de uma área, mas também avaliar a evolução histórica do tema, as lacunas na literatura e as oportunidades para novas investigações. Este método é utilizado em áreas como gestão, ciências sociais e tecnologias emergentes, aplicado aqui para explorar o tema “Data Justice” no contexto mundial. Estudos de referência, como Small (1973) destaca a relevância da cocitação para revelar agrupamentos intelectuais, enquanto Garfield (1972) estabelece os fundamentos do crescimento exponencial da literatura científica, ambos fornecendo a base teórica para este trabalho.

Um dos principais recursos da análise bibliométrica é a criação de mapas de rede, que ilustram as conexões entre autores e os termos mais frequentemente usados nos artigos. Ferramentas como o VOSviewer são amplamente utilizadas para gerar esses mapas, facilitando a identificação de “clusters” temáticos e de autoria (ECK et al., 2010). Esses clusters mostram áreas de maior colaboração entre pesquisadores e indicam linhas de pesquisa que estão sendo mais exploradas em conjunto.

### 3.2 COLETA DE DADOS

A coleta de dados foi realizada em três etapas principais: planejamento, seleção e extração de dados.

Na fase de planejamento, definiu-se que o escopo da pesquisa, incluindo a base de dados a ser utilizada e os critérios de inclusão e exclusão de artigos. A busca foi limitada a artigos publicados entre 2013 e 2024.

Foram incluídos os estudos que:

- Abordam diretamente o conceito de data justice ou seus equivalentes (justiça de dados, governança de dados).
- Foram publicados entre 2013 e 2024, assegurando a relevância atual do tema.

Foram excluídos os estudos que:

- Não abordam diretamente as questões de justiça social na coleta, uso ou distribuição de dados.
- Não estão disponíveis em texto completo.
- Estão em línguas diferentes de português, inglês ou espanhol.

A base de dados *Web of Science* foi selecionada para esta pesquisa por sua abrangência e relevância no campo acadêmico, especialmente por incluir periódicos de alto impacto e ampla cobertura nas ciências sociais, humanas e tecnológicas. Essa plataforma oferece recursos avançados para análise bibliométrica, permitindo identificar tendências, mapear redes de colaboração e avaliar a produtividade científica em diversas áreas. Além disso, a *Web of Science* conta com filtros e critérios de indexação rigorosos, o que assegura a qualidade e a relevância dos

estudos sobre justiça de dados, facilitando uma análise aprofundada e confiável do tema.

Para conduzir a busca, foi utilizada a seguinte *query*: ***Refine results for "data justice" (Topic) and Article (Document types)***. O processo resultou em 124 artigos que foram analisados.

A extração de dados foi realizada em formato *BibTeX*, utilizada em contextos acadêmicos para gerenciar referências e citações. No processo, foram coletadas informações essenciais, como: autor(es), ano de publicação, periódico, país e palavras-chave. Esse procedimento garantiu a padronização e a consistência na coleta de informações essenciais para a análise subsequente.

### 3.3 ANÁLISE DE DADOS

A primeira parte da análise de dados nesta pesquisa foi conduzida por meio de uma abordagem detalhada e científica, com foco em três pilares principais: análise bibliométrica, identificação de *clusters* de autoria e agrupamentos temáticos. Cada uma dessas técnicas contribuiu para fornecer uma compreensão abrangente da produção acadêmica sobre *data justice*, permitindo identificar padrões, lacunas e tendências emergentes.

Durante a análise bibliométrica foi utilizado o *software* VOSviewer, permitindo a criação de mapas de rede que ilustram as relações entre autores, países e periódicos. Foram identificados *clusters* de autoria, bem como *clusters* temáticos que refletem as principais áreas de foco no campo de *data justice*. Essa análise foi estruturada em quatro dimensões:

- A análise de coautoria visa identificar os principais pesquisadores que colaboram em estudos sobre *data justice*. Por meio dessa análise, foi possível visualizar os *clusters* de coautoria, ou seja, grupos de autores que frequentemente publicam juntos. A identificação desses *clusters* permitiu compreender as redes de colaboração científica e identificar os pesquisadores centrais no desenvolvimento do tema.
- A análise de autoria focou-se na identificação dos autores mais produtivos e influentes no campo de *data justice*. Isso ajudou a

mapear quais autores têm se destacado e contribuído para o avanço da discussão sobre justiça de dados.

- Foram analisadas as palavras-chave mais frequentemente associadas aos estudos sobre *data justice*, utilizando uma técnica de co-ocorrência de termos. Esta análise permitiu identificar os principais tópicos e subtemas discutidos na literatura, bem como visualizar os *clusters* temáticos formados em torno de termos centrais.
- A análise de co-citação examinou como os artigos e autores são citados em conjunto, revelando as relações entre as pesquisas e os estudos que têm servido de base teórica para o desenvolvimento do conceito de *data justice*. Essa análise permitiu identificar os principais marcos teóricos e estudos de referência que moldam o campo. Através de *clusters* de cocitação, foi possível observar a formação de "escolas de pensamento" e as abordagens dominantes sobre o tema.

A segunda parte da análise de dados foi conduzida utilizando o *software* R e o pacote *bibliometrix*, reconhecido por sua eficiência e flexibilidade na exploração de grandes conjuntos de dados bibliográficos. Esta etapa teve como objetivo aplicar as Leis de Bradford (1934) e Lotka (1926) para compreender padrões de dispersão de publicações científicas em periódicos e a produtividade de autores, respectivamente.

Os dados bibliográficos utilizados na análise foram obtidos em formato “.bib”, exportados da base de dados *Web Of Science*. Esse formato foi escolhido por sua compatibilidade com ferramentas de análise bibliométrica e por conter informações completas sobre artigos, autores e fontes (periódicos).

Inicialmente, os dados foram estruturados em um formato adequado para análise, utilizando uma ferramenta de conversão específica disponível no pacote *bibliometrix*. Esse processo garantiu que os metadados, como nomes de periódicos, nomes de autores, títulos de artigos e referências citadas, fossem devidamente organizados. Após a preparação dos dados, foi realizada uma verificação de consistência para assegurar que o número de periódicos, autores e citações fosse adequado à aplicação das leis bibliométricas.

A partir disso, a Lei de Bradford foi utilizada para identificar a dispersão da literatura científica entre os periódicos analisados. Segundo Bradford (1934), uma pequena proporção de periódicos, conhecidos como "núcleo", é responsável por uma parcela significativa das publicações, enquanto as demais contribuições estão distribuídas entre uma maior quantidade de periódicos menos produtivos.

Com base nos dados processados, os periódicos foram classificados de acordo com o número de artigos publicados. Posteriormente, os periódicos foram segmentados em três zonas, conforme preconiza a Lei de Bradford: o núcleo, que concentra a maior parte dos artigos, e duas zonas subsequentes, que representam contribuições menos significativas. Essa segmentação foi visualizada por meio de um gráfico logarítmico, no qual os periódicos foram dispostos no eixo horizontal em ordem de produtividade, e o número de artigos publicados foi representado no eixo vertical. O núcleo foi destacado no gráfico como uma área sombreada, facilitando a identificação dos periódicos mais relevantes (BRADFORD, 1934).

Ademais, A Lei de Lotka foi aplicada para analisar a distribuição de produtividade dos autores no conjunto de dados. Lotka (1926) descreve que poucos autores são altamente produtivos, enquanto a maioria contribui com apenas um ou dois trabalhos. Para verificar essa distribuição, foi calculada a produtividade de cada autor, ou seja, o número de artigos publicados por eles.

Os dados empíricos foram ajustados a um modelo teórico proposto por Lotka, no qual a produtividade dos autores segue uma relação inversamente proporcional ao quadrado do número de artigos publicados (LOTKA, 1926).

### 3.4 CRONOGRAMA DE EXECUÇÃO

Para garantir que todas as etapas da pesquisa sejam concluídas dentro do prazo de 3 meses, o cronograma foi dividido em semanas, detalhando as atividades principais de cada etapa, como pode ser analisado no Quadro 1.

QUADRO 1 – CRONOGRAMA DE EXECUÇÃO

| SEMANA | ATIVIDADES   |
|--------|--|
| 1      | Revisão da literatura e desenvolvimento do referencial teórico |

|    |   |
|----|---|
|    | Planejamento da estratégia de busca (definição da query e bases de dados)<br>Revisão inicial da literatura e refinamento dos critérios de inclusão/exclusão |
| 2  | Execução da busca de artigos na base Web of Science<br>Identificação de artigos relevantes e remoção de duplicatas  |
| 3  | Início da triagem dos artigos (leitura de títulos e resumos)<br>Continuação da triagem para exclusão de artigos irrelevantes                                |
| 4  | Leitura completa dos artigos selecionados<br>Avaliação detalhada da adequação dos artigos aos critérios de inclusão   |
| 5  | Extração de dados dos estudos selecionados<br>Organização dos dados coletados em categorias específicas   |
| 6  | Início da análise bibliométrica (uso do software VOSviewer)<br>Criação de mapas de rede para identificação de clusters de autoria e temáticos               |
| 7  | Interpretação dos resultados da análise bibliométrica<br>Identificação de padrões, lacunas e tendências na pesquisa sobre justiça de dados                  |
| 8  | Elaboração de gráficos e tabelas para visualização dos resultados<br>Discussão dos achados com base na literatura existente                                 |
| 9  | Redação do relatório final da revisão sistemática e análise bibliométrica   |
| 10 | Inclusão dos resultados, discussões e conclusões  |
| 11 | Revisão e ajustes no relatório final<br>Garantia da clareza, coerência e correção dos dados apresentados  |
| 12 | Término e submissão do TCC  |

FONTE: A autora (2024).

## 4 RESULTADOS

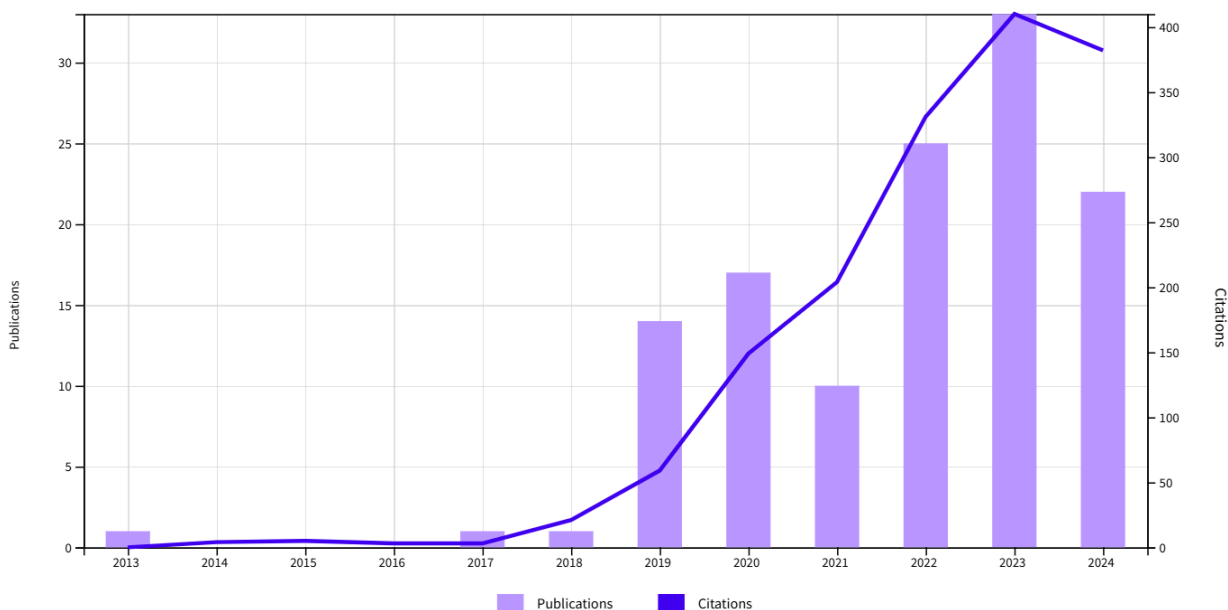
Os resultados deste estudo oferecem uma visão detalhada sobre a produção científica relacionada ao conceito de *data justice*, destacando padrões de publicação, tendências acadêmicas e contribuições significativas ao campo. Por meio da análise bibliométrica, foi possível identificar os principais pesquisadores, periódicos de maior impacto, redes de colaboração entre autores e temas centrais que estruturam as discussões na área. A aplicação das leis de Lotka e Bradford permitiu avaliar a distribuição da produtividade científica e a concentração do conhecimento em fontes específicas, elucidando aspectos fundamentais da organização e disseminação da pesquisa. Essas descobertas são essenciais para compreender as direções e lacunas existentes no debate sobre justiça de dados em nível global.

### 4.1 ANÁLISE GERAL DAS PUBLICAÇÕES

Primeiramente, é importante identificar a distribuição das publicações ao longo do tempo. No período analisado, 2013 a 2024, foi possível identificar 124 artigos publicados sobre justiça de dados (Gráfico 1) que somaram 1.571 citações, das quais 1.416 são citações sem autocitações. Isso indica que a produção científica analisada possui reconhecimento significativo fora do círculo dos próprios autores.

A média de citações por item é de 12,67, o que sugere que cada publicação possui, em média, uma visibilidade considerável no campo acadêmico. Por fim, o ano de 2023 foi o mais produtivo nos últimos 10 anos, com 33 artigos publicados.

GRÁFICO 1 – DISTRIBUIÇÃO TEMPORAL DAS PUBLICAÇÕES



FONTE: *Web Of Science (2024)*.

Os aspectos gerais dos registros analisados no Quadro 2 mostram que as 124 publicações foram escritas por 252 autores, com uma média de coautoria de aproximadamente 2 por artigo. Esses autores são filiados a 86 revistas diferentes, resultando em uma média de 3 artigos por revista, indicando uma distribuição considerável da produção entre diferentes veículos científicos.

QUADRO 2 – CARACTERÍSTICAS GERAIS DAS PUBLICAÇÕES

| <b>Web of Science</b> | <b>Visão geral dos resultados</b> |
|-----------------------|-----------------------------------|
| Publicações           | 124                               |
| Autores               | 252                               |
| Palavras-chave        | 618                               |
| Revistas              | 86                                |

FONTE: *Web Of Science (2024)*.

As publicações selecionadas apresentam um índice *h-index* igual a 18, o que significa que 18 artigos foram citados pelo menos 18 vezes, conforme a definição de *h-index* apresentada pela *Web of Science (2019)*, baseada em uma lista de publicações classificadas em ordem decrescente pela contagem de citações. Esse índice reflete a relevância e o impacto das publicações no campo, cobrindo o

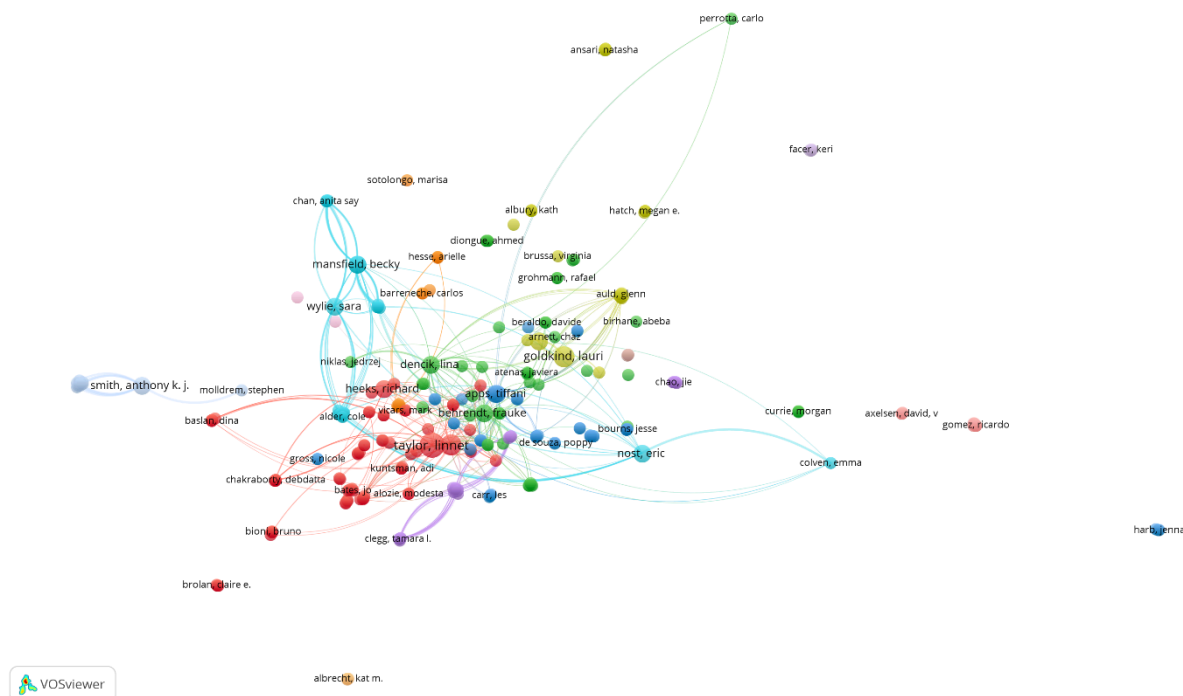


período de 2013 a 2024, sendo que 2013 foi o primeiro ano de publicação sobre o tópico especificado.

## 4.2 ANÁLISE DOS PRINCIPAIS AUTORES

A visualização da rede dos principais autores citados no campo da justiça de dados e campos relacionados oferece uma análise aprofundada das interconexões e da influência de cada pesquisador no desenvolvimento e na consolidação desse campo de estudo. Segundo a Figura 1, essa rede permite identificar tanto os autores mais centrais, cujas obras são amplamente referenciadas e servem de base teórica, quanto os agrupamentos temáticos, que indicam diferentes abordagens e áreas de especialização. Com base nessa análise, é possível entender a dinâmica colaborativa e as principais linhas de pesquisa, além de observar a presença de temas emergentes que ainda estão nos estágios iniciais de desenvolvimento.

FIGURA 1 – ANÁLISE RELACIONAL DOS PRINCIPAIS AUTORES



FONTE: A autora (2024).

Na rede, Linnet Taylor se destaca por sua posição central e proeminente, refletindo sua importância como uma das autoras mais citadas da área. A

centralidade de Taylor indica que suas publicações são usadas como referência teórica para estudos subsequentes, indicando que seu trabalho aborda conceitos fundamentais ou estruturas metodológicas para o campo da justiça de dados. Ao ocupar um ponto de convergência, Taylor está se tornando um ponto de referência para pesquisas que exploram tópicos como ética, governança de dados e os impactos sociais do uso massivo de dados, sendo uma das autoras que compõem a base conceitual do campo.

Além de Taylor, autores como Richard Heeks e Lauri Goldkind também se destacam por sua centralidade e múltiplas conexões na rede. Esses autores ocupam posições estratégicas, ligando diferentes grupos temáticos e contribuindo para a interdisciplinaridade do campo. A presença de Heeks e Goldkind como pontos de interseção indica que suas pesquisas são aplicáveis a vários subcampos da justiça de dados, demonstrando um diálogo constante entre disciplinas e abordagens teóricas. Ambos desempenham um papel importante na integração de diversos temas, que vão desde a governança de dados até as implicações éticas e sociais da aplicação de dados.

A estrutura da rede também destaca a formação de grupos distintos, representando grupos de autores com interesses temáticos semelhantes. Um desses grupos, onde Jo Bates e Tamara I. Clegg, parece concentrar sua pesquisa em temas relacionados à justiça social e à ética. Esse grupo demonstra um interesse comum em entender como o uso de dados afeta direitos, privacidade e questões éticas, destacando a importância da justiça de dados no contexto social e seus efeitos sobre indivíduos e comunidades. A proximidade entre esses autores sugere colaborações frequentes e uma convergência de ideias sobre as implicações éticas do uso de dados.
















Outro grupo significativo inclui autores como Richard Heeks e Lina Dencik, cuja produção acadêmica parece se concentrar em temas de governança e regulamentação de dados. Esse grupo explora estruturas de controle e políticas de dados, refletindo um debate sobre a necessidade de regulamentações para garantir o uso ético e responsável dos dados nas sociedades contemporâneas. A concentração de autores nesse agrupamento mostra como as discussões sobre governança são centrais no campo da justiça de dados, principalmente no que diz respeito ao monitoramento e à transparência nas práticas de coleta e uso de

informações pessoais.

Também é possível identificar um grupo de autores, liderado por Carlo Perrotta, cuja pesquisa se concentra no impacto e no uso de dados em cidades inteligentes e na aplicação de tecnologias em ambientes urbanos. Esse grupo representa uma linha de pesquisa que visa a entender como a coleta e a análise de dados em contextos urbanos influenciam a formulação de políticas públicas e a governança urbana. A presença desse grupo na rede sugere uma expansão do campo da justiça de dados em áreas que exploram a interseção entre dados, tecnologia e desenvolvimento urbano, refletindo a diversidade e a complexidade das abordagens presentes no campo.

Como visto no Quadro 3, as redes de autores também destacam a presença de colaborações frequentes entre pesquisadores, com conexões significativas entre aqueles que compartilham interesses temáticos. Essa proximidade entre autores sugere uma comunidade acadêmica ativa e interconectada, em que a produção colaborativa contribui para o fortalecimento das bases teóricas e práticas do campo. As colaborações interdisciplinares são evidentes, especialmente entre autores cujas pesquisas abordam diferentes dimensões da justiça de dados, promovendo um diálogo entre áreas diversas e fomentando o desenvolvimento de uma compreensão mais abrangente das implicações sociais, éticas e políticas do uso de dados.

#### QUADRO 3 – RESUMO GERAL DA REDE DOS PRINCIPAIS AUTORES

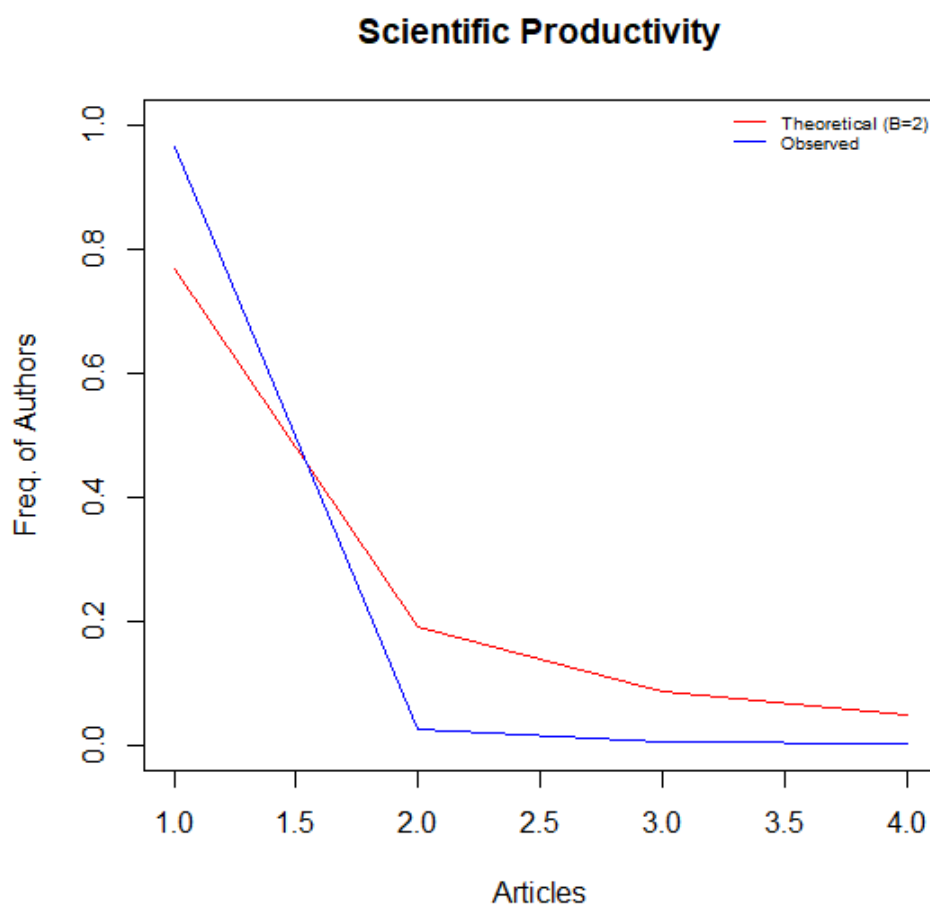
| Posição | Author                 | Cluster (Cor)   | Citações | Força Total de Conexão |
|---------|------------------------|---|----------|------------------------|
| 1       | Mansfield, Becky       |    | 69       | 1190                   |
| 2       | Murphy, Michelle       |    | 69       | 1190                   |
| 3       | Behrendt, Frauke       |    | 41       | 1152                   |
| 4       | Masiero, Silvia        |    | 57       | 1110                   |
| 5       | Smith, Anthony K. J.   |    | 56       | 1012                   |
| 6       | Dencik, Lina           |    | 23       | 968                    |
| 7       | Nost, Eric             |    | 19       | 939                    |
| 8       | Taylor, Linnet         |    | 367      | 921                    |
| 9       | Wylie, Sara            |    | 26       | 892                    |
| 10      | Broadly, Timothy R. R. |    | 11       | 809                    |
| 11      | Cook, Teddy            |   | 11       | 809                    |
| 12      | Davis, Mark D. M.      |  | 11       | 809                    |
| 13      | Duck-Chong, Elizabeth  |  | 11       | 809                    |
| 14      | Ellard, Jeanne         |  | 11       | 809                    |
| 15      | Holt, Martin           |  | 11       | 809                    |

FONTE: A autora (2024).

#### 4.2.1 Análise da produtividade científica por autor segundo a Lei de Lotka

O Gráfico 2, baseado na Lei de Lotka, ilustra a distribuição de produtividade científica dos autores em termos de número de artigos publicados, comparando a frequência observada com a frequência teórica esperada com base em um modelo teórico de produtividade. Na linha azul, observa-se a distribuição real dos dados coletados, enquanto a linha vermelha representa o modelo teórico, que, neste caso, assume um parâmetro  $B = 2$ . Esse tipo de análise é essencial para entender como a produção científica se distribui entre os autores, permitindo avaliar se os dados empíricos seguem ou divergem de um padrão esperado de produtividade.

GRÁFICO 2 – VISUALIZAÇÃO DA DISTRIBUIÇÃO DE PRODUTIVIDADE CIENTÍFICA ENTRE OS AUTORES



FONTE: A autora (2024).

No eixo horizontal (x), o gráfico exibe o número de artigos publicados por autor, enquanto o eixo vertical (y) mostra a frequência relativa dos autores que publicaram uma determinada quantidade de artigos. Assim, pode-se observar que a maioria dos autores publicou apenas um artigo, conforme indicado pela alta frequência de autores na categoria de um único artigo. Essa alta concentração de produtividade em um único artigo está representada pela interseção inicial e alta de ambas as linhas (azul e vermelha) no ponto correspondente a “1” no eixo dos artigos.

À medida que o número de artigos aumenta, a frequência de autores diminui acentuadamente. Esse padrão é evidenciado pela curva descendente das duas linhas. Tanto a linha observada quanto a teórica indicam que, para um número crescente de artigos (de 2 a 4, neste caso), a proporção de autores diminui

rapidamente. No entanto, a linha azul, representando a frequência observada, cai mais abruptamente em comparação com a linha teórica, indicando que, na realidade, o número de autores com múltiplas publicações é ainda menor do que o previsto pelo modelo teórico com  $B = 2$ .

Enquanto o modelo teórico assume uma distribuição específica onde alguns autores teriam uma produtividade relativamente mais alta, os dados observados indicam haver uma menor proporção de autores que publicam mais de um artigo. Este padrão é comum em várias áreas do conhecimento, onde um pequeno grupo de autores é responsável por uma maior quantidade de publicações, enquanto a maioria dos autores contribui com um único artigo. No entanto, neste caso específico, o número de autores com múltiplas publicações é ainda inferior do que o esperado teoricamente, o que pode ser indicativo do fenômeno "*one timers*".

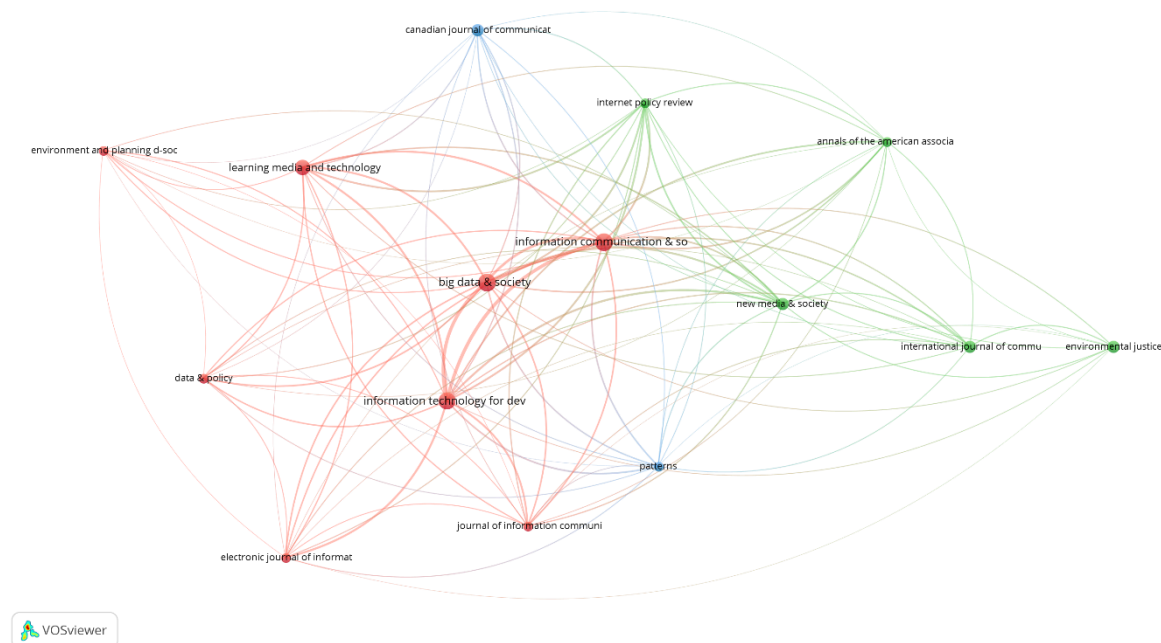
Muitas vezes, os "*one-timers*" aparecem em trabalhos colaborativos, desempenhando um papel específico em projetos maiores. Mesmo que publiquem apenas uma vez, sua participação pode ser crucial para a qualidade e impacto do trabalho.

No entanto, este padrão também pode refletir características do campo de pesquisa, como barreiras para publicações repetidas ou um perfil de autores que participam de forma esporádica, contribuindo apenas uma vez. Alternativamente, essa baixa produtividade em múltiplas publicações também pode sugerir que o campo é altamente segmentado ou competitivo, com a maioria dos autores publicando apenas uma vez e um número muito reduzido de pesquisadores alcançando uma alta produtividade.

### 4.3 ANÁLISE DAS PRINCIPAIS REVISTAS

A Figura 2 indica uma divisão dos periódicos em grupos, organizados de acordo com suas coocorrências em citações dos artigos analisados. Cada *cluster* representa um grupo de periódicos com interesses temáticos semelhantes, refletidos próximo de suas interconexões no mapa de acoplamento bibliográfico. Essa disposição ajuda a identificar áreas de interesse comum entre as publicações, facilitando a compreensão dos principais eixos temáticos que envolvem o conceito de *data justice*.

FIGURA 2 – VISUALIZAÇÃO DA REDE DE COCITAÇÃO DAS REVISTAS



FONTE: A autora (2024).

Em primeiro lugar, no *cluster* vermelho central, os periódicos “*Big Data & Society*” e “*Information Communication & Society*” se destacam como nós centrais, cada um com mais de 350 citações. Esse grupo sugere uma forte conexão com temas relacionados à governança de dados, à ética digital e às implicações sociais da *big data*. A centralidade desses periódicos indica que as discussões sobre justiça de dados estão intimamente ligadas a questões de tecnologia e comunicação, refletindo o foco de muitos estudos sobre a compreensão do impacto social e ético do uso de *big data*. Assim, essas publicações se consolidam como referências importantes para a pesquisa crítica em ciência de dados e justiça de dados.

Por outro lado, o grupo azul é representado por periódicos como o “*Journal of Information Policy*” e o “*Internet Policy Review*”, que se destacam por seu foco em políticas públicas e regulamentação da informação. Esse agrupamento destaca o crescente interesse acadêmico nas interseções entre justiça de dados e governança política, sugerindo que a regulamentação e a política pública se tornaram áreas centrais nas discussões sobre o uso ético e responsável dos dados.

Outro agrupamento significativo é o *cluster* verde, que inclui publicações como

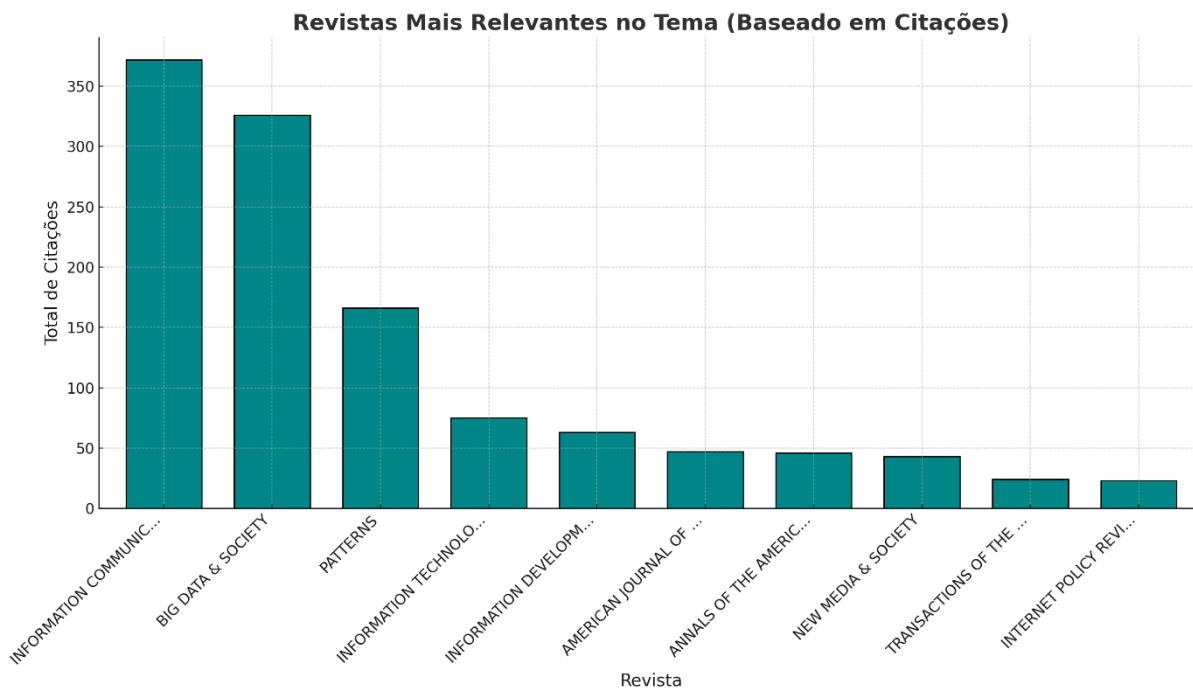
“*Environmental Justice*” e “*Public Health*”, com foco em temas de saúde pública e justiça ambiental. A presença desses periódicos reflete a preocupação do campo de justiça de dados com o uso de dados para regular e aprimorar melhorias em áreas sensíveis, como saúde e meio ambiente. Esse grupo enfatiza a expansão do conceito de justiça de dados para abranger dimensões que transcendem a tecnologia, incluindo preocupações éticas sobre o impacto dos dados na saúde e no meio ambiente.

Por fim, o grupo amarelo, composto por periódicos como “*Learning Media and Technology*” e “*Educational Technology & Society*”, concentra-se na aplicação de tecnologia e *big data* na educação. A inclusão dessas publicações sugere que o campo da justiça de dados é relevante no contexto educacional, com atenção voltada para a privacidade dos dados educacionais e o uso ético da análise na aprendizagem. Esse grupo, portanto, sinaliza uma crescente conscientização sobre as implicações do uso de dados na educação, enfatizando a proteção e o uso justo dos dados relativos a alunos e educadores.

Assim, a análise dos grupos de periódicos nos permite identificar uma diversificação temática no campo da justiça de dados, abrangendo aspectos regulatórios e éticos e preocupações com áreas específicas, como saúde, meio ambiente e educação.

O Gráfico 3 mostra que determinadas revistas têm um impacto significativamente maior em termos de números de citações, com “*Information Communication & Society*” e “*Big Data & Society*” liderando com mais de 350 citações cada. Essa alta contagem de citações ressalta a posição dessas revistas como fontes confiáveis e altamente respeitadas para pesquisas sobre justiça de dados e estudos críticos em ciência de dados.





FONTE: A autora (2024).

A análise de citações sugere que o campo é altamente interconectado, com algumas revistas atuando como “hubs” que centralizam a literatura. A alta concentração de citações em determinadas revistas indica que os pesquisadores tendem a recorrer a essas fontes para apoiar as discussões sobre os impactos éticos e sociais da tecnologia e dos dados.

Revistas como “*Patterns*” e “*Information Technology for Development*” aparecem em posições intermediárias no gráfico, com cerca de 150 a 200 citações. Essas publicações se destacam por sua abordagem multidisciplinar, abrangendo tópicos de tecnologia, ciência de dados e desenvolvimento social.

A presença de periódicos multidisciplinares nessa classificação indica que o campo da justiça de dados não se limita a uma única disciplina, mas está integrado a vários domínios, incluindo desenvolvimento sustentável, governança de dados e ciência da informação.

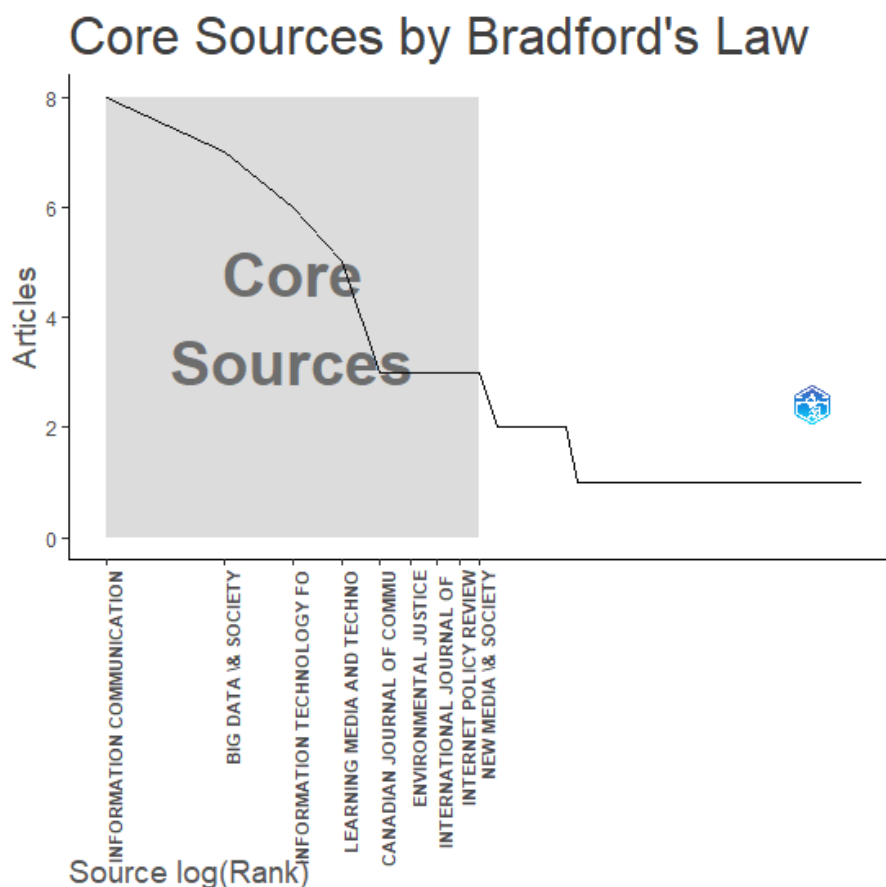
Além das principais revistas, o gráfico mostra uma presença equilibrada de várias outras revistas, cada uma contribuindo com um número substancial de citações. Essa distribuição equilibrada sugere que o campo da justiça de dados é altamente colaborativo, com várias revistas contribuindo para a construção do conhecimento.

Esse equilíbrio também sugere que os pesquisadores estão consultando uma variedade de fontes para apoiar suas pesquisas, indicando uma diversidade saudável de perspectivas e abordagens metodológicas dentro do campo.

#### 4.3.1 Fontes de informação em zonas de produtividade segundo a Lei de Bradford

O gráfico apresentado baseia-se na Lei de Bradford, que categoriza as fontes de informação em zonas de produtividade, destacando as fontes nucleares, ou seja, aquelas que concentram a maior parte das publicações relevantes em um campo específico de estudo. No caso da análise realizada, o Gráfico 4 evidencia as principais fontes relacionadas ao tema da justiça de dados (*data justice*), classificando-as conforme sua frequência de publicação.

GRÁFICO 4 – FONTES DE INFORMAÇÃO EM ZONAS DE PRODUTIVIDADE



FONTE: A autora (2024).

Na região destacada como "*Core Sources*" (Fontes Nucleares), encontram-se

periódicos que possuem maior número de artigos publicados sobre o tema, indicando sua relevância e centralidade para os pesquisadores da área. Dentre as fontes identificadas, destacam-se periódicos como *Information Communication, Big Data & Society*, e *Information Technology & Society*, que desempenham um papel central no desenvolvimento e disseminação do conhecimento sobre justiça de dados. Essas fontes nucleares são fundamentais para entender o estado da arte e as principais discussões teóricas e empíricas no campo.

À medida que avançamos no eixo horizontal do gráfico, observa-se uma diminuição na frequência de artigos publicados por outras fontes. Esse comportamento está alinhado ao que a Lei de Bradford prediz: uma concentração significativa de publicações em poucos periódicos nucleares, enquanto as publicações restantes estão distribuídas em uma ampla gama de fontes periféricas, com menor frequência de artigos.

Este padrão reforça a importância de identificar e priorizar as fontes nucleares durante revisões bibliográficas e análises bibliométricas, uma vez que elas representam os canais mais produtivos e influentes no campo de pesquisa. No contexto da justiça de dados, o mapeamento dessas fontes não apenas auxilia na consolidação do conhecimento existente, mas também contribui para a identificação de lacunas e oportunidades para pesquisas futuras. Além disso, a identificação de fontes periféricas, ainda que menos produtivas em termos de volume, pode oferecer perspectivas complementares ou inovadoras sobre o tema, ampliando o escopo da discussão acadêmica.

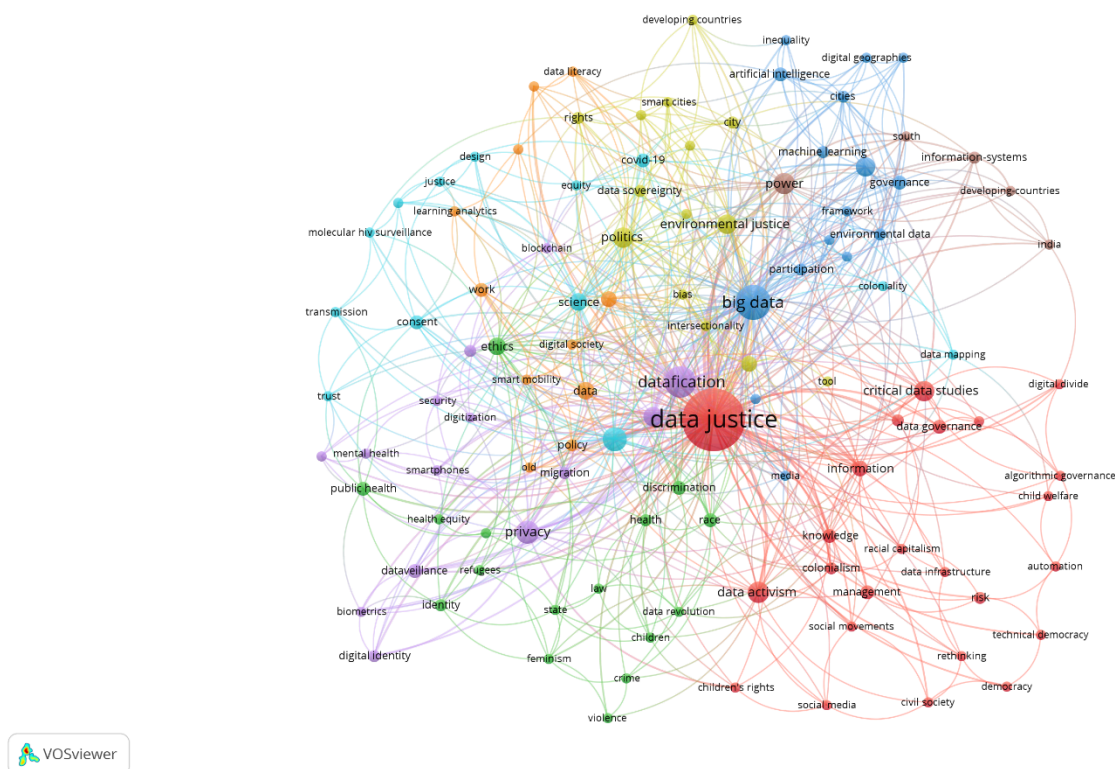
Portanto, o gráfico evidencia a estrutura da literatura científica sobre justiça de dados e destaca a relevância de fontes nucleares como pilares para a compreensão das dinâmicas teóricas e práticas que permeiam este campo emergente.

#### 4.4 ANÁLISE DE COCORRÊNCIA DE PALAVRAS-CHAVE

A Figura 3 representa a rede de coocorrência de 618 palavras-chave extraídas dos artigos, com um limite mínimo de ocorrência para uma palavra-chave definida como 2. Desse total, 113 palavras-chave atingem o limite estabelecido, indicando sua relevância em discussões acadêmicas. Foi usado o método de contagem total, o que significa que cada link de coocorrência tem o mesmo peso na

configuração do mapa.

FIGURA 3 – VISUALIZAÇÃO DA REDE DE COCORRÊNCIAS DE PALAVRAS-CHAVE (ALL KEYWORDS)



FONTE: A autora (2024).
















A figura tem vários grupos distintos (8 no total), sendo que o maior contém 24 itens e o menor, 5 itens. Isso sugere a presença de subtemas bem estabelecidos, bem como de áreas mais específicas e emergentes. Os grupos que tratam de ética, privacidade e consentimento formam conexões estreitas, sugerindo que a ética e a privacidade são preocupações centrais ao se discutir a justiça de dados.

O Quadro 4 apresenta uma análise de palavras-chave relacionadas ao tema “*data justice*” agrupadas por *clusters* e com indicadores de ocorrências e força total de conexão. A palavra-chave “*data justice*” ocupa a primeira posição, destacando-se com o maior número de ocorrências (74) e força de conexão (245), indicando sua centralidade no contexto analisado. Outros termos relevantes incluem “*Big Data*” e “*Datafication*”, que aparecem com frequências menores, mas ainda assim significativas, evidenciando a relação entre esses conceitos no debate acadêmico. *Clusters* diferenciados por cores indicam agrupamentos temáticos, como justiça

social, vigilância, privacidade, e tecnologia, refletindo a interdisciplinaridade do tema. A presença de conceitos como “*Environmental Justice*” e “*Social Justice*” sugere uma conexão entre a justiça de dados e questões socioambientais mais amplas.

Além disso, outro grupo forte está relacionado à governança de dados e aos estudos críticos de dados, abrangendo temas como algoritmos, infraestrutura de dados e o impacto do colonialismo digital, indicando uma análise crítica da estrutura e do controle de dados.

QUADRO 4 – RESUMO GERAL DE CO-OCORRÊNCIAS DE PALAVRAS-CHAVE (ALL KEYWORDS)

| Posição | Palavra-chave         | Cluster (Cor)2  | Ocorrências | Força Total de Conexão |
|---------|-----------------------|---|-------------|------------------------|
| 1       | Data Justice          |      | 74          | 245                    |
| 2       | Big Data              |      | 24          | 117                    |
| 3       | Datafication          |      | 19          | 84                     |
| 4       | Surveillance          |     | 11          | 51                     |
| 5       | Privacy               |    | 10          | 42                     |
| 6       | Power                 |    | 9           | 46                     |
| 7       | Data Activism         |    | 9           | 27                     |
| 8       | Politics              |    | 8           | 40                     |
| 9       | Technology            |    | 7           | 31                     |
| 10      | Environmental Justice |   | 7           | 24                     |
| 11      | Information           |  | 5           | 28                     |
| 12      | Algorithms            |  | 6           | 27                     |
| 13      | Governance            |  | 4           | 19                     |
| 14      | Social Justice        |  | 5           | 25                     |
| 15      | Critical Data Studies |  | 5           | 22                     |

FONTE: A autora (2024).

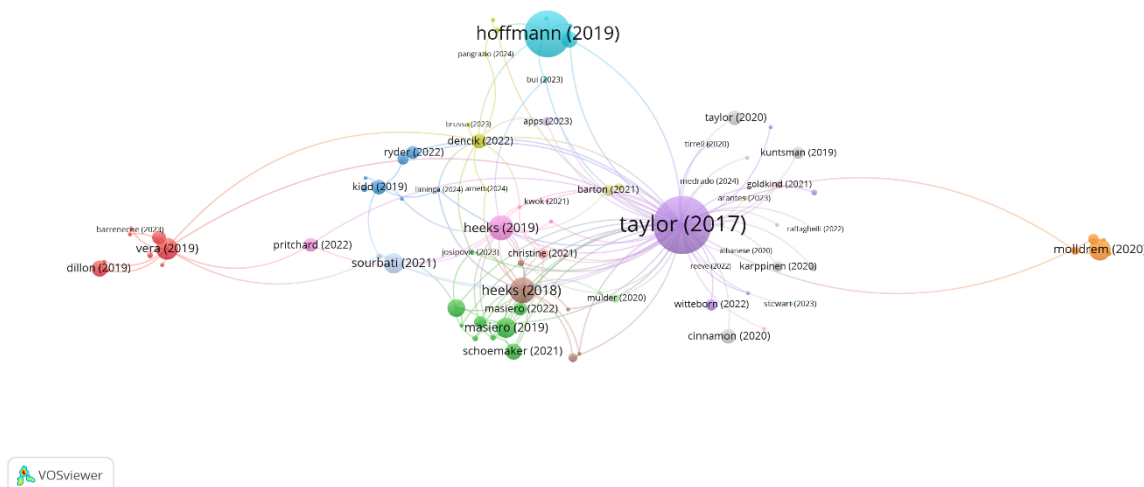
O termo “justiça de dados” ocupa uma posição central e é o maior nó do gráfico, indicando sua relevância e conexão com várias outras palavras-chave. Esse

termo vincula temas como governança de dados, privacidade, direitos e discriminação, mostrando que as discussões sobre justiça de dados são amplas e envolvem diversas perspectivas.

#### 4.5 ANÁLISE DOS PRINCIPAIS ARTIGOS

A análise do gráfico de vinculação bibliográfica entre documentos fornece uma compreensão aprofundada das interconexões e influências no campo da justiça de dados e assuntos relacionados. Esse tipo de visualização revela como os artigos estão vinculados em termos de citações, destacando aqueles que desempenham um papel central e influente na literatura. Por meio do mapeamento de nós e conexões, é possível identificar os artigos mais citados, os agrupamentos temáticos e a evolução do campo ao longo do tempo.

FIGURA 4 – REDE DOS PRINCIPAIS ARTIGOS BASEADO EM CITAÇÕES



VOSviewer

FONTE: A autora (2024).

Consoante a Figura 4, o artigo de Taylor (2017) ocupa uma posição de destaque no gráfico, atuando como o nó central e mais proeminente. Essa centralidade reflete o quão citado é o artigo no campo. O trabalho de Taylor (2017) fornece uma base teórica sólida para o campo da justiça de dados, estabelecendo diretrizes conceituais ou metodológicas que serviram de base para vários estudos subsequentes.

O gráfico também revela a existência de vários *clusters*, cada um representando um grupo de estudos interconectados por suas citações, sugerindo subtemas específicos no campo da justiça de dados. O *cluster* azul, onde se destacam artigos como Hoffmann (2019) e Ryder (2022), parece concentrar discussões sobre governança de dados e ética. A organização desse *cluster* reflete o interesse em questões de controle e regulamentação de dados, tópicos cruciais para entender o impacto social e ético do uso excessivo de dados. Por outro lado, o *cluster* vermelho, que inclui Vera (2019) e Dillon (2019), destaca a justiça social e os impactos ambientais, revelando que a justiça de dados também está ligada a preocupações socioambientais. Esse conjunto indica que a análise de dados é relevante não apenas no contexto digital, mas também para questões de sustentabilidade e responsabilidade social.

Além dos grupos centrais, podemos observar no gráfico a presença de artigos que atuam como “artigos de ligação” e desempenham um papel estratégico ao conectar diferentes subtemas. Artigos como Heeks (2018, 2019) e Masiero (2019, 2022) aparecem em posições estratégicas, estabelecendo conexões entre vários *clusters*. A função desses artigos como pontes sugere que eles ampliam o escopo dos conceitos centrais, aplicando e adaptando estruturas teóricas fundamentais em novos contextos e estudos de caso. Essa característica destaca a interdisciplinaridade do campo da justiça de dados, em que estudos baseados em um núcleo conceitual comum exploram diferentes abordagens e áreas de aplicação, contribuindo para a expansão e a diversificação do campo.

É interessante notar a presença de artigos mais recentes, como Barreneche (2023) e Bui (2023), que, embora ainda não estejam totalmente integrados ao núcleo central, já mostram uma conexão com os artigos mais citados, em especial com Taylor (2017).

De outro modo, alguns artigos mostram menos conectividade no gráfico, como Molldrem (2020), que aparece relativamente isolado dos *clusters* centrais. Esse isolamento pode indicar que o trabalho aborda um tópico emergente ou especializado, que ainda não foi totalmente integrado ao corpo principal de conhecimento sobre justiça de dados. A relativa desconexão deste artigo também pode indicar uma especialização metodológica ou temática, tornando-o potencialmente relevante para estudos futuros que busquem expandir os limites do

campo com novas perspectivas.

A presente análise evidencia ainda uma progressão cronológica no desenvolvimento do campo, desde publicações iniciais, como Taylor (2017), que estabeleceram as bases teóricas, até estudos mais recentes, que exploram aplicações e temas emergentes.

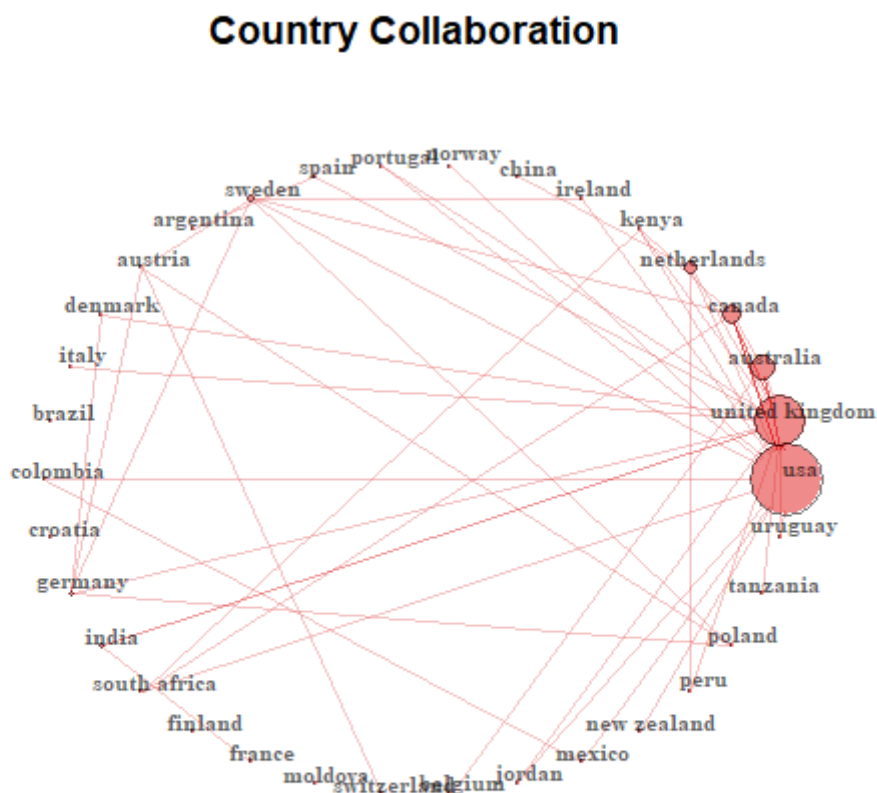
Esse progresso cronológico reflete a maturação da área de justiça de dados, que evoluiu de uma fase inicial de construção conceitual para uma fase de aplicação e diversificação temática. A presença de documentos publicados entre 2017 e 2024 sugere que o campo está crescendo em complexidade, abrangendo novas áreas e respondendo a desafios contemporâneos.

#### 4.6 ANÁLISE DA REDE DE CONTRIBUIÇÃO POR PAÍS

A Figura 5 representa a rede de colaboração internacional entre diferentes países no campo de estudo em questão. Cada nó do gráfico corresponde a um país, enquanto as linhas que conectam os nós, indicam relacionamentos de coautoria entre pesquisadores de diferentes nações. O tamanho dos nós reflete a intensidade ou a frequência dessas colaborações; os países com um número maior de parcerias aparecem os maiores nós, demonstrando sua centralidade e influência na rede de colaboração científica global.



FIGURA 5 – REDE DE CONTRIBUIÇÃO POR PAÍSES



FONTE: A autora (2024).

Países como os Estados Unidos, o Reino Unido e a Austrália se destacam por seu tamanho proeminente no gráfico, indicando serem os principais centros de colaboração internacional nessa área de pesquisa. A posição central e o tamanho desses nós, revelam que esses países desempenham um papel fundamental na condução e na disseminação do conhecimento, estabelecendo conexões com uma ampla gama de parceiros internacionais. Esse padrão sugere que os pesquisadores sediados nesses países não apenas produzem um volume significativo de trabalho, mas também participam ativamente de parcerias internacionais, ampliando o alcance e a influência de suas pesquisas.

A conexão entre os Estados Unidos, o Reino Unido e o Canadá, destacada pela proximidade e pela espessura das linhas, mostra uma forte rede de colaboração entre esses países, que frequentemente compartilham interesses de pesquisa e abordagens metodológicas comuns. A frequência de coautoria entre esses países indica uma interdependência científica, em que ideias e descobertas são desenvolvidas em conjunto, promovendo um ambiente de inovação e troca de

conhecimento benéfico para o campo na totalidade.

Além dos principais centros de colaboração, o gráfico também revela a participação de países de outras regiões, como Brasil, Índia, África do Sul e China, que, embora representados por nós menores, mostram um compromisso crescente com a pesquisa internacional. Essas colaborações podem indicar uma expansão geográfica do campo, em que o conhecimento e a inovação são distribuídos além dos centros de pesquisa tradicionais, permitindo que os países emergentes contribuam com novas perspectivas e contextos. A presença desses países na rede de colaboração sugere uma diversificação no campo de estudo, com possíveis contribuições regionais que podem enriquecer as discussões globais sobre o assunto.

Entretanto, o gráfico também sugere um certo desequilíbrio na colaboração científica internacional, com uma concentração de parcerias em países de língua inglesa. Essa centralização pode refletir não apenas a disponibilidade de recursos e infraestrutura para pesquisa em países como os Estados Unidos e o Reino Unido, mas também questões como a predominância do idioma inglês na ciência e a prevalência de redes acadêmicas consolidadas nesses países.

Por fim, essa centralização pode limitar a diversidade de perspectivas, ressaltando a importância de promover colaborações com uma variedade maior de nações, especialmente aquelas que ainda têm um perfil mais baixo na produção científica global.

## 5 CONSIDERAÇÕES FINAIS

Este estudo conduziu uma análise bibliométrica das publicações globais, abrangendo o período de 2013 a 2024, sobre o conceito de justiça de dados. Por meio do uso de ferramentas como o VOSviewer e dados extraídos do banco de dados *Web of Science*, foram identificados os principais autores, redes de coautoria, palavras-chave e grupos temáticos relacionados ao tema. Essa abordagem permitiu mapear o estado da produção científica sobre justiça de dados, apresentando as principais linhas de discussão e os pontos centrais que estruturam esse campo de estudo.

A análise de coautores, cocitações e agrupamentos temáticos revelou fortes vínculos entre autores e instituições, bem como o surgimento de subtemas como vigilância, governança de dados, privacidade e poder. Essa interpretação forneceu uma visão geral da evolução do campo, destacando não apenas os avanços teóricos, mas também os limites da produção científica, como a predominância de estudos em regiões específicas e a falta de pesquisas aplicadas em contextos sub-representados.

Apesar da robustez metodológica da análise bibliométrica, são reconhecidas limitações intrínsecas, como o viés introduzido pela base de dados utilizada, que prioriza a produção científica em inglês e de regiões com maior apoio financeiro à pesquisa. Assim, os resultados aqui apresentados devem ser entendidos como um corte transversal da literatura científica que, embora relevante, não abrange todas as publicações sobre o assunto.

Apesar disso, o trabalho alcançou seus objetivos ao identificar as características centrais do campo de estudos sobre justiça de dados e responder à pergunta de pesquisa: como se organiza a literatura científica de justiça de dados no mundo?

O mapeamento das contribuições acadêmicas permitiu compreender não apenas os principais autores e instituições envolvidas, mas também as lacunas existentes na literatura, especialmente no que diz respeito à aplicação prática dos princípios de justiça de dados em contextos regionais e em setores como educação, saúde e governança pública. A análise bibliométrica demonstrou como os debates sobre justiça de dados vêm se consolidando, destacando a importância de integrar

perspectivas éticas e sociais na governança de tecnologias emergentes.

Além disso, a análise revelou a predominância de temas interdisciplinares, como vigilância algorítmica, transformação digital e governança, reforçando a relevância do conceito de justiça de dados nas discussões contemporâneas. No entanto, tópicos como o impacto da justiça de dados em regiões sub-representadas, aplicações práticas em políticas públicas e a interseção com outras dimensões da justiça social continuam sendo lacunas importantes a serem exploradas em pesquisas futuras.

No que se refere às recomendações, sugere-se que futuras pesquisas explorem mais profundamente as relações entre justiça de dados e outras dimensões da justiça, como justiça ambiental, educacional e interseccional. Além disso, seria pertinente investigar as diferenças nas abordagens de justiça de dados em contextos geográficos e culturais diversos, com atenção especial às regiões sub-representadas na literatura atual. A integração de análises empíricas que investigam a implementação de princípios de justiça de dados em políticas públicas e práticas organizacionais também é recomendada, com vistas a ampliar a aplicabilidade prática dos resultados acadêmicos.

## REFERÊNCIAS

ALBERTIN, A. L.; MOURA, R. D. Transformação digital: gerando valor para o “novo futuro”. **GV-Executivo**, v. 20, n. 1, p. 26 – 29, 2021.

ALHASSAN, I.; SAMMON, D.; DALY, M. Data governance activities: an analysis of the literature. **Journal of Decision Systems**, v. 25, n. sup1, p. 64 – 75, 2016.

BATTIST, I. D. E.; SMOLSKI, F. M. da S. **Software R: Análise estatística de dados utilizando um programa livre**. Bagé, RS: Editora Faith, 2019.

BHARADWAJ, A. et al. Digital business strategy: toward a next generation of insights. **MIS Quarterly**, p. 471 – 482, 2013.

BIONI, B. R. **Proteção de Dados Pessoais: A Função e os Limites do Consentimento**. Rio de Janeiro: Renovar, 2020.

BRADFORD, Samuel C. Sources of information on specific subjects. **Engineering**, v. 137, p. 85-86, 1934.

BRYNJOLFSSON, E.; MCAFEE, A. **The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies**. [S.I.]: WW Norton & Company, 2014.

CARMO, Hermano; FERREIRA, Manuela. Metodologia da Investigação–Guia para Auto-aprendizagem (2ª edição). **Lisboa: Universidade Aberta**, v. 15, p. 16, 2008.

COSTA, B. R. L. Bola de neve virtual: o uso das redes sociais virtuais no processo de coleta de dados de uma pesquisa científica. **Revista interdisciplinar de gestão social**, v. 7, n. 1, 2018.

CRAWFORD, K.; CALO, R. There is a Blind Spot in AI Research. **Nature**, 2016.

CRESWELL, John W.; POT, Cheryl N. **Qualitative inquiry and research design: Choosing among five approaches**. Sage publications, 2016.

DAVENPORT, T. H.; HARRIS, J. G. **Competing on Analytics: The New Science of Winning**. [S.I.]: Harvard Business Review Press, 2017.

DAVENPORT, T. H.; PRUSAK, L. **Conhecimento Empresarial: Como as Organizações Gerenciam o seu Capital Intelectual**. 13ª. ed. Rio de Janeiro: Elsevier, 2003.

DENCIC, L.; HINTZ, A.; CABLE, J. Data Justice: Understanding Datafication in Relation to Social Justice. Cardiff University, 2016.

DENCIC, L. et al. Exploring Data Justice: Conceptions, Applications and Directions. **Information, Communication & Society**, v. 22, n. 7, p. 873 – 881, 2019.

DONTHU, N. et al. How to conduct a bibliometric analysis: An overview and guidelines. v. 133, p. 285 – 296, 9 2021. Acesso em: 2023-12-04T00:00:00+00:00.

ECK, N. J. V. et al. A Comparison of Two Techniques for Bibliometric Mapping: Multidimensional Scaling and VOS. **Journal of the American Society for Information Science and Technology**, v. 61, p. 2405 – 2416, 2010.

ECONOMIST, T. The world's most valuable resource is no longer oil, but data. 2017.

EUBANKS, V. **Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor**. New York: St. Martin's Press, 2018.

FLENSBURG, Sofie; LOMBORG, Stine. Datafication research: Mapping the field for a future agenda. **New Media & Society**, v. 25, n. 6, p. 1451-1469, 2023.

FLICK, Uwe. **Introdução à metodologia de pesquisa: um guia para iniciantes**. Penso Editora, 2012.

GANDOMI, A.; HAEGERMANN, M. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, v. 35, n. 2, p. 137 – 144, 2015.

GANGADHARAN, S. P. Decentering Technology in Discourse on Data Justice. **Information, Communication & Society**, v. 23, n. 6, p. 873 – 881, 2020.

GARFIELD, E. Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. **Science**, v. 178, n. 4060, p. 471 – 479, 1972.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.I.]: MIT Press, 2016.

GUAN, S. et al. Hadoop-based secure storage solution for big data in cloud computing environment. **Digital Communications and Networks**, v. 10, n. 1, p. 227 – 236, 2024.

HAIR, J. F. et al. **Multivariate Data Analysis**. 7th. ed. [S.I.]: Pearson, 2009.

HEEKS, R.; RENKEN, J. Data justice for development: What would it mean? **Information Development**, v. 34, n. 1, p. 90 – 102, 2018.

HILL, Michael; VARONE, Frédéric. **The public policy process**. Routledge, 2016.

INMON, W. H. **Building the Data Warehouse**. 4. ed. Indianapolis: Wiley Publishing, Inc., 2005.

JANSSEN, Marijn et al. Data governance: Organizing data for trustworthy Artificial

Intelligence. **Government information quarterly**, v. 37, n. 3, p. 101493, 2020.

KHATRI, Vijay; BROWN, Carol V. Designing data governance. **Communications of the ACM**, v. 53, n. 1, p. 148-152, 2010.

KOTTER, J. P. **Leadership Change**. Boston, MA, USA: Harvard Business School Press, 1996.

LOSHIN, David. **Master data management**. Morgan Kaufmann, 2010.

LOTKA, Alfred J. The frequency distribution of scientific productivity. **Journal of the Washington academy of sciences**, v. 16, n. 12, p. 317-323, 1926.

MAHANTI, R. **Data Quality: Dimensions, Measurement, Strategy, Management, and Governance**. [S.I.]: Quality Press, 2019.

MARR, B. **Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance**. [S.I.]: John Wiley & Sons, 2015.

MARTINS, A. D. F. LGPD: A Formal Concept Analysis and Its Evaluation. In: **Anais do XXXV Simpósio Brasileiro de Bancos de Dados**. [S.I.]: SBC, 2020. p. 259 – 264.

MASIERO, Silvia; DAS, Soumyo. Datafying anti-poverty programmes: Implications for data justice. **Information, Communication & Society**, v. 22, n. 7, p. 916-933, 2019.

MASIERO, Silvia; BUDDHA, Chakradhar. **Data Justice in Digital Social Welfare: A Study of the Rythu Bharosa Scheme**. 2021.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data: A Revolution That Will Transform How We Live, Work, and Think**. [S.I.]: Houghton Mifflin Harcourt, 2013.

MEJIAS, U. A.; COULDRY, N. Datafication. **Internet Policy Review**, v. 8, n. 4, p. 1 – 10, 2019.

MELLO, R. D. S. et al. **Dados Semi-Estruturados**. 2024.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied Statistics and Probability for Engineers**. 6th. ed. [S.I.]: Wiley, 2014.

OLIVEIRA, J. A. D.; REZENDE, C.; SARAMAGO, G. Análise de Dados em Pesquisas Qualitativas: Algumas Orientações, Perspectivas e Softwares. **Cadernos da FUCAMP**, v. 29, 2024.

OLSEN, W. **Coleta de Dados: Debates e Métodos Fundamentais em Pesquisa Social**. [S.I.]: Penso Editora, 2015.

O'NEIL, C. **Weapons of Math Destruction**: How Big Data Increases Inequality and Threatens Democracy. New York: Crown, 2016.

OTTO, B.; WEBER, K. Data Governance. In: **Daten-und Informationsqualität**: Auf dem weg zur information excellence. [S.l.: s.n.], 2011. p. 277 – 295.

PATTON, Michael Quinn. **Qualitative research & evaluation methods: Integrating theory and practice**. Sage publications, 2014.

RAMASUBRAMANIAN, S.; DUTTA, M. J. The CODE ^ SHIFT Model: A Data Justice Framework for Collective Impact and Social Transformation. **Human Communication Research**, v. 50, n. 2, p. 173 – 183, 2024.

RÊGO, B. L. **Gestão e Governança de Dados**: Promovendo Dados como Ativo de Valor nas Empresas. [S.l.]: Brasport, 2013.

ROGERS, D. L. **Transformação Digital**: Repensando Seu Negócio para a Era Digital. [S.l.]: Autêntica Business, 2019.

ROSA, A. Internet of Things (IoT). 2020.

SILVA, M. Aplicações de Inteligência Artificial e Machine Learning no Brasil. **Revista de Tecnologia e Sociedade**, v. 12, n. 2, p. 45 – 60, 2024.

SMALL, H. Co-citation in the scientific literature: A new measure of the relationship between two documents. **JASIS**, v. 24, n. 4, p. 265 – 269, 1973. Disponível em: <http://dx.doi.org/10.1002/asi.4630240406>.

SOUTHERTON, C. Datafication. Springer International Publishing, p. 358 – 361, 2022.

TAYLOR, L. What is data justice? The case for connecting digital rights and freedoms globally. **Big Data & Society**, v. 4, n. 2, p. 2053951717736335 –, 2017.

TOLEDO, N. M. de et al. Bafafá: Análise de sentimentos de comentários em redes sociais como ferramenta de suporte a decisão (SAD). 2024.

UNION, E. General Data Protection Regulation. **Official Journal of the European Union**, 2016.

VERA, Lourdes A. et al. When data justice and environmental justice meet: Formulating a response to extractive logic through environmental data justice. **Information, Communication & Society**, v. 22, n. 7, p. 1012-1028, 2019.

WANG, Richard Y.; STRONG, Diane M. Beyond accuracy: What data quality means to data consumers. **Journal of management information systems**, v. 12, n. 4, p. 5-33, 1996.



WESTERMAN, G.; BONNET, D.; MCAFEE, A. *Leading Digital: Turning Technology into Business Transformation*. Harvard Business Press, 2014.

WU, Chengke et al. Critical review of data-driven decision-making in bridge operation and maintenance. **Structure and infrastructure engineering**, v. 18, n. 1, p. 47-70, 2021.

YIN, Robert K. **Estudo de Caso-: Planejamento e métodos**. Bookman editora, 2015.

ZUBOFF, S. Big Other: Surveillance Capitalism and the Prospects of an Information Civilization. **Journal of Information Technology**, v. 30, n. 1, p. 75 – 89, 2015.

ZUBOFF, S. **The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power**. New York: PublicAffairs, 2019.