

## ANÁLISE BIBLIOMÉTRICA: REVISÃO DE LITERATURA E PROPOSTA DE FRAMEWORK METODOLÓGICO EM 12 PASSOS

 <https://doi.org/10.56238/arev6n4-146>

Data de submissão: 10/11/2024

Data de publicação: 10/12/2024

### **Marcelo Salvador Celestino**

Doutorando e Mestre em Mídia e Tecnologia (UNESP)  
PPGMiT-DO/FAAC, Universidade Estadual Paulista "Júlio de Mesquita Filho"  
E-mail: [marcelo.salvador@unesp.br](mailto:marcelo.salvador@unesp.br)  
ORCID: <https://orcid.org/0000-0002-5551-8793>

### **Regina Celia Baptista Belluzzo**

Doutora em Ciências da Comunicação (USP)  
Professora voluntária, PPGMiT/FAAC,  
Universidade Estadual Paulista "Júlio de Mesquita Filho"  
E-mail: [rbelluzzo@gmail.com](mailto:rbelluzzo@gmail.com)  
ORCID: <https://orcid.org/0000-0001-9514-2930>

### **João Pedro Albino**

Doutor em Administração de Empresas, Métodos Quantitativos e Informática (USP)  
Professor Associado voluntário, PPGMiT/FAAC,  
Universidade Estadual Paulista "Júlio de Mesquita Filho"  
E-mail: [jp.albino@unesp.br](mailto:jp.albino@unesp.br)  
ORCID: <https://orcid.org/0000-0001-5965-1869>

### **Vânia Cristina Pires Nogueira Valente**

Livre Docente em Representação Gráfica (UNESP), Doutora em Engenharia Civil (USP)  
Professora Associada e Coordenadora do PPGMiT-DO/FAAC,  
Universidade Estadual Paulista "Júlio de Mesquita Filho"  
E-mail: [vania.valente@unesp.br](mailto:vania.valente@unesp.br)  
ORCID: <https://orcid.org/0000-0001-6563-2402>

## RESUMO

A análise bibliométrica é uma metodologia de estudo retrospectivo, quantitativo, estatístico e descritivo da literatura científica, que permite identificar os elementos mais prolíficos e influentes em determinado campo, as interconexões entre autores e a literatura em si, além de identificar temas emergentes de pesquisa. O presente estudo está dividido em duas partes. Primeiramente, apresentou-se uma revisão narrativa da literatura sobre os fundamentos, principais técnicas e desafios que permeiam a condução de uma análise bibliométrica, com o objetivo de elucidar o tema, servindo como ponto de partida para pesquisas dessa natureza. Para ilustrar os exemplos, foi simulada uma análise bibliométrica com o tema Redução de Dose da Radiação Ionizante e Otimização de Protocolos em Tomografia Computadorizada (TC), em caráter experimental, utilizando o pacote *bibliometrix* no *software* RStudio. Em seguida, foi proposto um *framework* metodológico com 12 passos, que têm sido utilizados empiricamente na condução de análises bibliométricas pelos presentes autores. A adição do conceito de simulação e de um ciclo de *feedback* permitiu refinar a metodologia e obter resultados mais confiáveis. Este estudo contribui para a literatura ao propor um roteiro empírico metodológico

orientado para a prática, que promove transparência, reprodutibilidade e rigor metodológico nas análises bibliométricas, promovendo a ciência aberta e colaborativa.

**Palavras-chave:** Análise bibliométrica, Infometria, Mídia e Tecnologia, Metodologia de pesquisa, Simulação.

## 1 INTRODUÇÃO

A análise bibliométrica é uma metodologia utilizada pela informetria para quantificar a produção científica de um determinado campo, por meio de técnicas estatísticas (Ball, 2017). Na perspectiva de Hjørland (2004), os estudos bibliométricos são ferramentas que, por meio do mapeamento dos domínios do conhecimento e das comunidades discursivas, permitem analisar como o conhecimento está estruturado e organizado em diferentes áreas. Essas ferramentas, inseridas no contexto da análise de domínio, têm como principal contribuição a exploração do fluxo de conhecimento entre os pares no universo acadêmico-científico, bem como seu processo de produção e disseminação na sociedade.

A análise bibliométrica tem sido empregada em pesquisas de diversas áreas do conhecimento, com foco nos mais variados temas, como governança e sustentabilidade (N'ze; Tenkoul, 2024), Inteligência Artificial (IA) no reconhecimento da linguagem de sinais (Zhang *et al.*, 2024) e pesquisas globais sobre o câncer (Karger; Kureljusic, 2023). Essa versatilidade nas aplicações evidencia que se trata de uma estratégia de amplo interesse acadêmico e empresarial.

Existem, na literatura, diversos guias e manuais que descrevem as particularidades na condução de uma análise bibliométrica (Aria; Cuccurullo, 2017; Donthu *et al.*, 2021; Uribe; Contreras; Guerrero, 2023; Öztürk; Kocaman; Kanbach, 2024; Passas, 2024), o que pode representar um desafio para pesquisadores ou estudantes que buscam o primeiro contato com essa metodologia. Os objetivos do presente artigo foram: **(i)** apresentar uma revisão narrativa da literatura, explorando os conceitos, e principais desafios de uma análise bibliométrica, com a apresentação de exemplos ilustrados de técnicas selecionadas; e **(ii)** propor um *framework* metodológico baseado em 12 passos, organizados em três módulos, para servir como ponto inicial para estudantes e pesquisadores que desejam compreender os principais fundamentos dessa abordagem de forma objetiva e prática.

## 2 PROCEDIMENTOS METODOLÓGICOS

O estudo adota um *mix* metodológico composto por uma revisão narrativa da literatura (Green; Johnson; Adams, 2006; Ferrari, 2015) e uma abordagem teórico-conceitual exploratória, baseada no modelo interativo de Miles e Huberman (1994) (Período de coleta de dados, redução de dados, exibição de dados, elaboração e verificação de conclusões). A integração de ambas metodologias, somada às vivências dos presentes autores, fundamenta a proposição de um *framework* estruturado para a análise bibliométrica. O material utilizado na revisão narrativa foi selecionado por conveniência, utilizando diversas fontes, como artigos e manuais, sem um recorte temporal definido, mas abrangendo literatura clássica para abordar as leis bibliométricas.

## 2.1 EXEMPLOS ILUSTRADOS

Para ilustrar os conceitos selecionados, foi conduzida uma simulação de análise bibliométrica com o tema Redução de Dose da Radiação Ionizante e Otimização de Protocolos em Tomografia Computadorizada (TC) (tema de interesse de um dos autores). Foi realizada uma busca na base Scopus, incluindo apenas artigos publicados em inglês, no período de 1998 a 2022. A *query* de busca utilizada na base Scopus em *all fields* (todos os campos), é apresentada no **quadro 1**.

**Quadro 1** – *Query* de busca utilizada na base de dados Scopus

(optimization OR "Scan Optimization" OR "Dose, Reduction") AND ("Tomography, Spiral Computed" OR "Computed Tomography, Spiral" OR "Computer-Assisted Tomography, Spiral" OR "Computer Assisted Tomography, Spiral" OR "Spiral Computer-Assisted Tomography" OR "Tomography, Spiral Computer-Assisted" OR "Computerized Tomography, Spiral" OR "Spiral Computerized Tomography" OR "CT Scan, Spiral" OR "Scan, Spiral CT" OR "Scans, Spiral CT" OR "Spiral CT Scan" OR "Spiral CT Scans" OR "Helical CT" OR "CT, Helical" OR "CTs, Helical" OR "Helical CTs" OR "Spiral Computed Tomography" OR "Spiral CT" OR "CT, Spiral" OR "CTs, Spiral" OR "Spiral CTs" OR "Tomography, Helical Computed" OR "CAT Scan, Spiral" OR "CAT Scans, Spiral" OR "Scans, Spiral CAT" OR "Spiral CAT Scans" OR "Helical Computed Tomography" OR "Computed Tomography, Helical" OR "Multislice, Computed Tomography" OR "CT, Angiography") AND NOT ("Magnetic Resonance Imaging" OR MRI OR "Positron Emission Tomography" OR PET OR "Radiation Therapy" OR "radiotherapy" OR "tomotherapy" OR "Ultrasound" OR "Mammography" OR "Fluoroscopy" OR "Radiography" OR "Industrial" OR "Orthodontic" OR "Orthodontics")

**Nota:** Foram utilizados termos MeSH (*Medical Subject Headings*), como *Tomography, Spiral Computed* (Tomografia Computadorizada Espiral), e termos comuns, como *Scan Optimization* (Otimização de Varredura), a fim de ampliar a amostra. **Fonte:** Elaborado pelos autores (2024).

Para a análise dos dados, utilizou-se o pacote gratuito *bibliometrix* (Aria; Cuccurullo, 2017) da linguagem R de programação, no ambiente RStudio (Posit, 2024). Esse *software*, considerado um dos mais completos para mapeamento científico (Terra *et al.*, 2022), organiza fluxos de trabalho para análise bibliométrica descritiva e/ou visual, com integração a métodos estatísticos.

## 2.2 BASES DOS *FRAMEWORK* PROPOSTO

Para a elaboração do *framework* metodológico, foram consultados guias e análises bibliométricas conduzidas na prática, a fim de compreender as diferentes formas com que os autores aplicam essa metodologia (Pizzani; Da Silva; Hossne, 2010; Vasconcelos, 2014; Aria; Cuccurullo, 2017; Donthu *et al.*, 2021; Karger; Kureljusic, 2023; Öztürk; Kocaman; Kanbach, 2024). Diferentemente dos estudos utilizados como referência, foi adicionada uma etapa de simulação com ciclo de *feedback*, que corresponde à realização de testes desde a *query* de busca até a aplicação de diversas técnicas bibliométricas, como a co-ocorrência de palavras-chave, a fim de explorar a integridade dos metadados, a aderência ao problema de pesquisa e, conforme sugerido por Donthu *et al.* (2021), selecionar as técnicas a serem aplicadas com antecedência. Um dos benefícios da simulação é a possibilidade de realizar testes sem custos adicionais ou prejuízos, permitindo identificar falhas precocemente e refinar o processo original (Celestino; Valente, 2021). No presente estudo, essa

abordagem crítica possibilitou uma análise mais consciente e contextualizada dos dados bibliométricos, contribuindo para uma compreensão mais aprofundada e reflexiva sobre como refinar o processo metodológico e apresentar resultados mais confiáveis e consistentes.

### 3 REVISÃO DE LITERATURA

#### 3.1 O QUE É UMA ANÁLISE BIBLIOMÉTRICA?

A origem do termo bibliometria é controversa na literatura, envolvendo Paul Otlet e Alan Pritchard. Para Momesso e Noronha (2017), é mais provável que Paul Otlet seja considerado o autor do termo, criado no livro *Traité de Documentation* (Otlet, 1934)<sup>1</sup>, sob uma perspectiva ampla e complexa de documentação. Entretanto, o termo foi formalizado por Alan Pritchard (1969)<sup>2</sup>, que descreveu bibliometria como a aplicação de métodos estatísticos à análise de publicações. Para Alvarenga (1998), a bibliometria oferece *insights* sobre como os saberes são construídos e legitimados, permitindo uma compreensão mais crítica das práticas científicas contemporâneas.

Nesse contexto, a análise bibliométrica é entendida como uma metodologia quantitativa que examina a produção científica de forma retrospectiva, utilizando métricas que avaliam a performance e a influência da literatura. Além disso, utiliza-se de técnicas que permitem identificar padrões, redes de colaboração e explorar as inter-relações que se formam entre os elementos envolvidos na produção científica (Donthu *et al.*, 2021; Passas, 2024). Ao quantificar aspectos da comunicação científica, a análise bibliométrica fornece um panorama sobre a disseminação e a influência do conhecimento produzido. Os dados podem orientar decisões estratégicas sobre investimentos em pesquisa, educação e políticas públicas de saúde, considerando as relações formadas em redes de pesquisa integradas (Ninkov; Frank; Maggio, 2022; Byl *et al.* 2024).

Öztürk, Kocaman e Kanbach (2024) destacam a confusão comum entre os termos *pesquisa* e *análise* bibliométrica. Eles alertam para a variação na abordagem e na qualidade dos estudos que utilizam a análise bibliométrica, muitos dos quais são incorretamente considerados como pesquisa bibliométrica, quando, na verdade, são revisões narrativas da literatura. Além disso, observam que, em alguns casos, as técnicas de análise bibliométrica são aplicadas superficialmente, sem uma avaliação profunda dos resultados ou discussão sobre sua relevância para o campo de pesquisa específico.

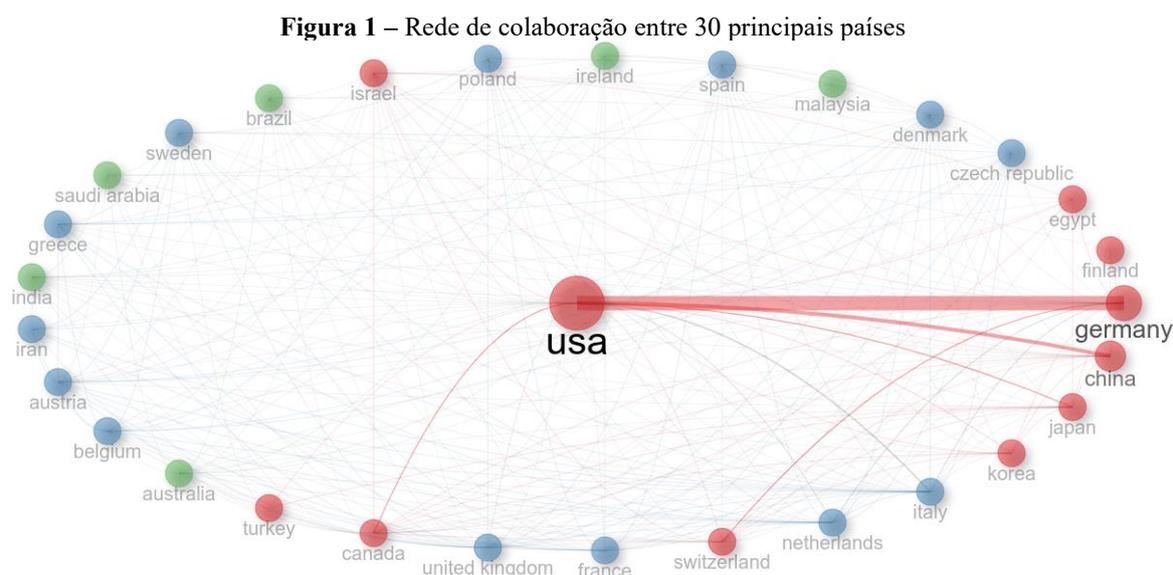
---

<sup>1</sup> OTLET, P. *Traité de documentation: le livre sur le livre: théorie et pratique*. Bruxelas: Mundaneum, 1934. Disponível em: [http://lib.ugent.be/fulltxt/handle/1854/5612/Traite\\_de\\_documentation\\_ocr.pdf](http://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf). Acesso em: 11 abr. 2016.

<sup>2</sup> PRITCHARD, A. Statistical bibliography or bibliometrics? *Journal of Documentation*, Leeds, v. 25, n. 4, p. 348-349, 1969.

### 3.2 O CONCEITO DE NÓ

Em uma rede, o nó representa a unidade fundamental que se conecta a outras, formando uma estrutura complexa. Os nós que se formam em uma rede podem representar qualquer entidade, desde pessoas em redes sociais, autores, instituições, até genes em redes biológicas. Imagine uma rede como uma teia: os nós são os pontos de encontro dos fios, e as linhas que os conectam representam as relações entre eles. Ao estudar as propriedades desses nós e suas interações, é possível compreender a estrutura e o funcionamento de diversas redes, como redes sociais e tecnológicas. Conceitos como *degree* (grau) (número de conexões), *centrality* (centralidade) (importância na rede) e *clusters* (agrupamentos) (grupos de nós densamente conectados) são essenciais para analisar essas estruturas complexas (Newman, 2010). A **figura 1** ilustra uma rede de colaboração formada entre países, na produção científica da análise bibliométrica simulada.



**Nota:** Rede de colaboração entre 30 países, formada nas pesquisas sobre Redução de Dose da Radiação Ionizante e Otimização de Protocolos em TC. Cada círculo (nó) representa um país participante da pesquisa, enquanto as linhas que os conectam indicam as colaborações entre eles. Identificaram-se *clusters* de acordo com as cores azul, verde e vermelho, sendo esse último o mais destacado, evidenciando uma forte colaboração entre os Estados Unidos e a Alemanha. **Fonte:** Dados da pesquisa analisados com o *bibliometrix* (Aria; Cuccurullo, 2017).

Para analisar redes visuais e suas comunidades, utilizam-se algoritmos como *Walktrap* e *Louvain*. O *Walktrap* simula conexões curtas entre nós para identificar comunidades, criando hierarquias nas divisões da rede. Por outro lado, o *Louvain* otimiza a modularidade, reagrupando nós iterativamente até alcançar a melhor divisão (Smith *et al.*, 2020; Mukerjee, 2021). Enquanto o *Walktrap* destaca pequenas comunidades, o *Louvain* é mais eficiente em grandes redes, focando na divisão geral ideal.

### 3.3 MÉTRICAS E INDICADORES

Geisler (2000, p. 48) define o termo métricas como um sistema de medidas que inclui o item objeto da medida, a unidade de medida e o valor da unidade. Além disso, classifica as métricas como objetivas ou subjetivas. A avaliação pelos pares é citada por esse autor como uma métrica subjetiva, enquanto a contagem de patentes é considerada uma métrica objetiva. Ainda de acordo com a definição desse autor, as métricas podem assumir vários formatos, como uma medida única, uma razão (entre duas medidas), um índice ou uma medida integrada que combine várias métricas, inclusive com atributos diferentes, objetivos e subjetivos.

As métricas são essenciais para avaliar o impacto e a relevância de pesquisas publicadas. Entretanto, a avaliação de pesquisas não deve ser reducionista; a análise deve considerar múltiplas métricas (ou indicadores), apoiadas pela opinião de especialistas, a fim de obter *insights* e perspectivas críticas da produção acadêmica (Elsevier, 2009).

Enquanto o conceito de métrica se refere a medidas quantitativas utilizadas para avaliar e analisar a produção científica, os indicadores são considerados medidas específicas que representam e quantificam características mais particulares dentro da bibliometria (Glänzel, 2003). O **quadro 2** apresenta uma síntese de exemplos de métricas avaliadas em uma análise bibliométrica.

**Quadro 2** – Exemplos de métricas

Métrica	Descrição	Ref.
Citações ponderadas	Dá mais peso a citações de artigos/autores influentes, similar ao algoritmo <i>PageRank</i> .	[4]
Colaboração e coautoria	Análise da colaboração entre pesquisadores e instituições, que pode influenciar a visibilidade e o impacto das publicações.	[5]
Fator de impacto	Mede a frequência com que os artigos de uma revista são citados, refletindo sua qualidade e relevância.	[4]
<i>g-index</i> (índice g)	Índice que amplifica as citações dos artigos mais citados para destacar distinções entre pesquisadores.	[3]
<i>h-index</i> (índice h)	Métrica que fornece uma visão equilibrada da produtividade e impacto do autor. Ex.: um <i>h-index</i> de 10 indica 10 publicações citadas ao menos 10 vezes.	[2]
<i>m-index</i> (índice m)	Métrica complementar ao <i>h-index</i> , que caracteriza a influência do pesquisador ao longo do tempo.	[2]
Total de citações	Contagem de quantas vezes os trabalhos de um autor ou instituição foram citados por outros pesquisadores.	[1]
Total de publicações	Total de artigos acadêmicos publicados por um autor, grupo ou instituição.	[1]

**Nota:** [1] = Donthu *et al.* (2021); [2] = Hirsch (2005); [3] = Egghe (2006); [4] = Van Noorden (2010); [5] = Elsevier (2009).

**Fonte:** Elaborado pelos autores (2024).

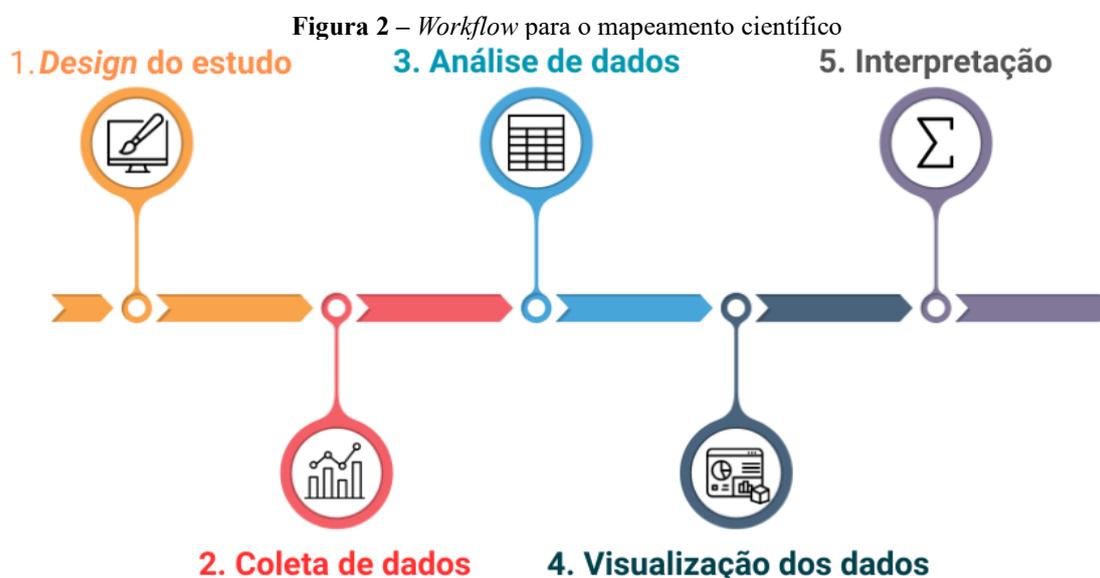
O número de artigos publicados em periódicos é considerado um fator importante, já que os periódicos mais produtivos sobre um determinado tema podem ser considerados os mais relevantes dentro do contexto investigado, com base na Lei de Bradford (Alabi, 1979). Porém, Anthony van Raan,

citado por Van Noorden (2010, p. 864-865, tradução nossa), afirma que "se há uma coisa com a qual todo bibliometrista concorda, é que nunca se deve usar o fator de impacto do periódico para avaliar o desempenho de pesquisa de um artigo ou de um indivíduo — isso é um pecado mortal."

### 3.4 ETAPAS NA CONDUÇÃO DE UMA ANÁLISE BIBLIOMÉTRICA

De modo geral, a análise bibliométrica é dividida em: **(i)** análise descritiva; **(ii)** análise de performance; e **(iii)** mapeamento científico (Donthu *et al.*, 2021). A análise de performance objetiva a classificação de acordo com a quantificação das métricas, enquanto o mapeamento científico se concentra na visualização das relações e interconexões que se formam entre a produção e a interpretação, devendo seguir um caminho capaz de associar-se às lacunas na literatura, produzindo considerações relevantes para a área (Öztürk; Kocaman; Kanbach, 2024).

No livro *Introducción a la bibliometría práctica* (Introdução à bibliometria prática, tradução nossa) (Uribe; Contreras; Guerrero, 2023), é proposto um processo estruturado para a análise bibliométrica que parte da definição do objetivo da pesquisa até a apresentação dos resultados. Já Aria e Cucurullo (2017) apresentam um *workflow* (fluxograma) baseado em Zupic e Čater (2015), que consiste em cinco etapas principais, como pode ser visto na **figura 2**.



**Nota:** *Design* (Desenho) do estudo se refere à estrutura metodológica utilizada na pesquisa. **Fonte:** Elaborado pelos autores (2024), com base em Aria e Cucurullo (2017) e Zupic e Čater (2015).

Cada uma das etapas pode compreender subetapas, como no passo 3 (Análise de dados), em que são realizadas a escolha do *software*, o tratamento dos dados ou a definição de subgrupos a serem analisados. Aria e Cucurullo (2017) consideram o mapeamento científico de forma holística, como

um processo contínuo da análise bibliométrica, utilizado para entender a evolução do conhecimento. Para o presente artigo, adota-se a perspectiva de Donthu *et al.* (2021), em que o mapeamento científico é uma das etapas de avaliação presentes na análise bibliométrica, o que permite uma apresentação sequencial e mais estruturada dos resultados.

### 3.5 APRESENTAÇÃO DOS RESULTADOS

Uma forma de iniciar a apresentação dos resultados é por meio dos dados gerais da amostra, que incluem o número total de documentos, autores, dados de colaborações, entre outros (Volpe *et al.* 2023), seguidos da análise de performance dos autores, com as revistas mais relevantes e uma apresentação do contexto histórico, mencionando o primeiro trabalho publicado da amostra (Montazeri *et al.*, 2023). Sequencialmente, apresentam-se as técnicas de mapeamento científico, como mencionado anteriormente. A **tabela 1** resume os dados e compara os *datasets* (conjuntos de dados) 1 (3331 artigos) e 2 (2252 artigos) utilizados na simulação de exemplo.

**Tabela 1** – Comparativo de dados gerais das amostras

<b>Informações Principais sobre os Dados</b>	<b>Dataset 1</b>	<b>Dataset 2</b>
<i>Timespan</i> (Período de tempo)	1998:2022	1998:2022
<i>Sources</i> ( <i>Journals, Books, etc.</i> ) (Fontes - periódicos, livros, etc.)	761	496
<i>Documents</i> (Documentos - Artigos)	3331	2252
<i>Annual growth rate %</i> (Taxa de crescimento anual %)	11,04	24,68
<i>Document average age</i> (Idade média dos documentos)	8,9	8,23
<i>Average citations per doc</i> (Média de citações por documento)	27,67	25,55
<i>References</i> (Referências)	80633	55909
<i>Keywords plus</i> (Palavras-chave adicionais)	12342	10189
<i>Author's keywords</i> (Palavras-chave do autor)	5272	4566
<i>Authors</i> (Autores)	13122	9616
<i>Authors of single-authored docs</i> (Autores de documentos com um único autor)	73	29
<i>Single-authored docs</i> (Documentos com um único autor)	79	34
<i>Co-authors per doc</i> (Coautores por documento)	6,54	6,77
<i>International co-authorships %</i> (Coautorias internacionais %)	20,23	21,05

**Fonte:** Dados da pesquisa analisados com o *bibliometrix* (Aria; Cuccurullo, 2017).

O panorama apresentado na **tabela 1** fornece uma visão inicial das características gerais de cada amostra, permitindo iniciar o relato dos achados com uma análise descritiva do estudo (Aria; Cuccurullo, 2017). Uma breve comparação dos dados revela algumas variações discretas, como a *idade média dos documentos*, mas uma variação maior que o dobro na *taxa de crescimento anual*.

A descrição dos resultados dependerá do objetivo do estudo e das informações que se pretende evidenciar. Por uma questão de estrutura e organização textual, sugere-se que as inferências e considerações dos autores sobre esses resultados sejam apresentadas na seção de discussão, seguindo

a mesma ordem em que apareceram nos resultados (Montazeri *et al.*, 2023; European Radiology, 2024).

### 3.6 ANÁLISE DE PERFORMANCE E TÉCNICAS DE MAPEAMENTO CIENTÍFICO

O **quadro 3** apresenta duas categorias e as principais técnicas utilizadas em uma análise bibliométrica (Donthu *et al.*, 2021), possibilitando ao pesquisador realizar uma análise abrangente e contextualizada do *corpus* da literatura.

**Quadro 3** – Exemplos de técnicas empregadas na análise bibliométrica.

Categoria	Técnicas	Descrição	Ref.
Análise de Performance	Análise de métricas de produção	Avaliação de contribuições de autores, instituições, países e periódicos, incluindo publicações (TP), e métricas de impacto (TC, TCm, <i>h-index</i> etc.), conforme ilustrado no <b>quadro 2</b> .	[1]
	Análise de Citação	Avalia a influência de publicações com base no número de citações que recebem.	[2], [3]
Mapeamento Científico	Análise de Cocitação	Mede a similaridade entre documentos, autores ou periódicos com base na frequência com que são citados juntos.	[1], [2], [3]
	Acoplamento Bibliográfico	Examina a relação entre documentos que citam as mesmas referências.	[1], [3]
	Análise de co-ocorrência de palavras	Identifica conexões entre conceitos que ocorrem juntos em títulos, palavras-chave ou resumos.	[1], [2]
	Análise de Coautoria	Avalia a colaboração entre autores com base em coautorias em publicações.	[2]
	Mapeamento de Redes	Visualiza relações entre documentos, autores e instituições usando <i>software</i> de mapeamento.	[2]
	Análise de Tendências	Examina a evolução de tópicos e temas ao longo do tempo para identificar áreas emergentes.	[2]

**Nota:** [1] = Donthu *et al.* (2021); [2] = Aria e Cuccurullo (2017); [3] = Zupic e Čater (2015). Ref. = Referências; TC = Total de citações; TCm = Média de citações por publicação; TP = Total de publicações. **Fonte:** Elaborado pelos autores (2024).

A **figura 3** apresenta uma rede de co-ocorrência baseada em 100 palavras-chave gerais, construída com o algoritmo *Louvain*. Na imagem, são identificados três *clusters* formados em torno de tópicos centrais, que estabelecem conexões entre si, criando uma rede ampla, na qual determinados termos aparecem frequentemente juntos em um conjunto de publicações, sugerindo uma relação semântica ou temática entre eles.



### 3.7.2 Lei de Bradford

Descrita por Samuel C. Bradford em 1934, a Lei de Bradford estabelece que a maioria das citações em um campo de estudo provém de um número relativamente pequeno de periódicos principais. Bradford observou que, ao ordenar os periódicos por número de citações recebidas, os periódicos mais citados (Zona 1 ou *Core Source*) (Fontes Principais) respondiam por uma grande proporção das citações totais. Essa proporção decresceu gradativamente à medida que se avançava para os periódicos menos citados (Zona 2 e Zona 3) (Alabi, 1979).

Em uma análise bibliométrica sobre modelos de inovação, Guimarães, Moreira e Bezerra (2021) identificaram uma aderência da amostra de 919 periódicos à Lei de Bradford, apresentando uma distribuição coerente de acordo com essa lei (Zona 1 = 8,27% periódicos,  $n = 33,16\%$  artigos ( $\frac{1}{3}$  dos artigos da amostra; Zona 2 = 36,56% periódicos,  $n = 33,88\%$  artigos; e Zona 3 = 55,17% periódicos,  $n = 32,96\%$  artigos). A distribuição de artigos está equilibrada, com um pequeno destaque para a Zona 2. Os periódicos da Zona 1, apesar de apresentarem a menor parcela da amostra, são responsáveis por praticamente um terço da produção, enquanto a Zona 3, que contém a maioria dos periódicos, produziu uma quantidade ligeiramente menor de artigos. Isso significa que, embora a Zona 3 compreenda a maior parte dos periódicos, a Zona 1 é proporcionalmente mais eficiente em termos de produção científica, concentrando uma produção significativa em um número reduzido de periódicos.

Na análise bibliométrica de Navarro-Ballester *et al.* (2023), sobre as publicações da Covid-19 nos jornais de radiologia entre 2020 e 2021, o *European Radiology* apareceu como o periódico mais produtivo ( $n = 167$  artigos) e o segundo em número de artigos cocitados ( $n = 32, 4,95\%$ ) em um *cluster* temático. No estudo de Volpe *et al.* (2023), o *European Radiology* apareceu em segundo lugar em produtividade ( $n = 255, 4,95\%$ ). Já na análise bibliométrica sobre a IA para detecção do câncer de Karger e Kureljusic (2023), este periódico apareceu como o oitavo mais produtivo ( $n = 69$ ). Isso demonstra que a produtividade ou relevância de um periódico varia de acordo com o tema e escopo investigados.

### 3.7.3 Lei de Zipf

A Lei de Zipf, publicada em 1949, descreve uma relação inversa entre a frequência de ocorrência de palavras e sua posição no *ranking* (classificação) de frequência. Como descrito no trecho original, as frequências "decrecem de acordo com a seguinte série harmônica simples: 1, 1/2, 1/3, ... 1/n, desde que cada frequência,  $f$ , quando multiplicada por seu *rank*,  $r$ , resultará em uma constante" (Zipf, 1949, p.35). Isso significa que, em um *corpus* linguístico, palavras de maior frequência

aparecem mais frequentemente, enquanto aquelas de menor frequência seguem essa distribuição regular, indicando uma estrutura previsível mesmo em grandes volumes de dados. Essa lei pode ser aplicada em diferentes contextos como a compreensão de dados, Processamento de Linguagem Natural (PLN), identificação de palavras-chave em texto, e estratégias de SEO (*Search Engine Optimization*) (Otimização de Motores de Busca).

Cassettari *et al.* (2015), ao analisar os *rankings* de frequência das palavras, identificaram padrões distintos de uso: algumas palavras são muito comuns, como artigos e preposições, que possuem um papel estrutural no texto, enquanto outras de cunho técnico, como semântica, possuem frequências menores. Isso não só confirmou a validade da Lei de Zipf em ambos os contextos, mas também destacou diferenças significativas na escolha vocabular entre escrita e fala. Embora a Lei de Zipf seja de grande valia para complementação de uma análise bibliométrica, na busca por tendências baseadas nas palavras-chave, é necessário considerar que existem limitações na padronização dessas palavras, conforme será descrito na próxima seção.

### 3.8 DESAFIOS E LIMITAÇÕES

A condução de uma análise bibliométrica eficaz envolve diversos desafios, desde a seleção das bases de dados mais adequadas até a interpretação coerente dos resultados. A necessidade de habilidades técnicas específicas, a capacidade de integrar diferentes métodos de análise e a importância da colaboração interdisciplinar são fatores cruciais para garantir a qualidade e a confiabilidade dos resultados (Da Silva; Hayashi C.; Hayashi M., 2011).

Uma limitação a ser considerada nessa metodologia é a falta de homogeneidade nas palavras-chave, o que pode resultar em uma análise incompleta ou distorcida na identificação em técnicas como a tendência temática (Geisler, 2002). Isso é ilustrado por Cassettari *et al.* (2015, p. 161-162) em um estudo comparativo da Lei de Zipf entre textos e discursos orais:

Todas estas palavras mencionadas aparecem e/ou tem sentido direto com o título do texto apresentado. Se considerar as palavras "livro" e "livros" como uma só, elas somam 41 ocorrências, elevando-as à oitava posição e, no mesmo caso, com as palavras "citação" e "citações", juntas somariam 39 ocorrências, passando para a nona posição (Cassettari *et al.* 2015, p. 161-162).

No estudo mencionado de Cassettari *et al.* (2015), percebe-se, portanto, que o agrupamento de palavras com o mesmo sentido (diferenciadas por singular e plural) modifica a posição dos termos em relação à frequência de ocorrência em um texto ou discurso. Dessa forma, embora a semelhança entre as palavras seja evidente para o pesquisador, ela pode complexificar o processo e levar a uma análise fragmentada dos dados, potencialmente distorcendo os resultados finais de uma análise de palavras-

chave. Ignorar a presença de alguns termos que refletem o mesmo significado ou interpretar um termo com suposições prévias pode resultar em uma análise enviesada. Um caminho para mitigar isso é complementar a análise de palavras-chave dos autores com uma análise de palavras-chave dos títulos, ampliando a capacidade de inferências e interpretações dos dados e contribuindo para a compreensão de tendências emergentes de pesquisa (Chen *et al.*, 2024).

Outro desafio diz respeito à qualidade dos dados. Para ilustrar isso, utiliza-se como exemplo a pesquisa simulada na base Scopus, mencionada na metodologia. O resultado inicial retornou 5184 artigos, reduzidos para 3331 (*dataset 1*) após a aplicação de filtros e a remoção de sete duplicatas. Após isso, foi realizada uma análise exploratória dos dados, identificando falhas na completude de metadados, como palavras-chave. Optou-se por excluir intencionalmente artigos com metadados incompletos (1079), para ilustrar o impacto dessa exclusão na análise, resultando no *dataset final* (*dataset 2*), constituído por 2252 artigos.

A análise dos *datasets* revelou discrepâncias na completude dos metadados, impactando diretamente a qualidade dos resultados. No *dataset 1*, o campo *author's keywords* apresentou uma ausência de 21,22%, classificado pelo *bibliometrix* como pobre. Esta falha compromete a precisão de uma análise dependente de palavras-chave. Apesar disso, o *dataset 1* poderia ser considerado adequado exclusivamente para analisar a performance, já que as métricas relacionadas ao desempenho estavam completas. Após a exclusão de 1079 artigos, o *dataset 2* apresentou uma melhora nos metadados que estava com baixa completude, e foram então classificados pelo *software* como excelentes. No entanto, ainda houve falhas no número de referências citadas, o que poderia comprometer uma análise de citações, além do fato de ter sido introduzido um viés na seleção. Essas discrepâncias evidenciam a importância de uma escolha criteriosa das métricas e técnicas em uma análise bibliométrica. Esses resultados corroboram as críticas identificadas na literatura sobre as limitações da análise bibliométrica, devido à sua dependência de dados precisos, manipulação de citações e limitações na captura de nuances qualitativas do conhecimento. Para superar essas limitações, pesquisadores têm recomendado a adoção de métodos híbridos que integrem análises qualitativas e quantitativas, o aprimoramento na precisão dos metadados e o uso de abordagens inovadoras, como a webometria e a altimetria (Glänzel, 2003; Aria; Cuccurullo, 2017; Ball, 2017; Ninkov; Frank; Maggio, 2022). Além disso, faz-se necessário considerar as possibilidades de preenchimento desses dados, como será explicado a seguir no *framework* proposto.

#### 4 *FRAMEWORK* METODOLÓGICO PARA A CONDUÇÃO DE UMA ANÁLISE BIBLIOMÉTRICA

Na sequência, com o aporte dos conceitos e definições apresentados, são descritos os passos utilizados para o desenvolvimento de uma análise bibliométrica. Esse *framework* metodológico foi construído com base na literatura e, principalmente, nas percepções e experiências práticas dos autores. Sua estrutura é composta por 12 passos (**Quadro 4**) distribuídos em três módulos, os quais:

- I. Procedimentos iniciais e compreensão dos dados:** Inclui o delineamento do estudo, definição de objetivos, seleção das bases de dados, criação de *queries* de busca e escolha do *software* e das bases de dados a serem utilizadas (passos 1 a 4).
- II. Simulação:** Envolve testes exploratórios das *queries* nas bases de dados, importação e análise preliminar dos dados, criação de cenários e perguntas hipotéticas, e identificação de falhas nos metadados, garantindo que os dados atendam aos objetivos da pesquisa (passos 5 a 7).
- III. Condução da análise bibliométrica e documentação:** Refere-se à etapa principal do estudo, onde são realizados o refinamento da amostra, a seleção de métricas e técnicas com base na qualidade dos dados, a visualização das informações na interface do *software*, além da análise e descrição dos resultados (passos 8 a 12).

**Quadro 4** – *Framework* metodológico com 12 passos para a condução da análise bibliométrica

Módulo	Passo	Descrição e fundamentos
<b>I</b> Procedimentos iniciais e compreensão dos dados	1 - Definição dos objetivos e do escopo	Estabelecer claramente os objetivos da pesquisa, elaborar as questões-problema que nortearão a pesquisa e delimitar o escopo do estudo. Com base em Aria e Cuccurullo (2017), tem-se que as questões gerais de um estudo podem ser divididas em: <b>(i)</b> identificação da base do conhecimento, por meio do mapeamento intelectual (métricas de performance, frequência de co-citação); <b>(ii)</b> exame da fronteira da pesquisa, para identificar as tendências emergentes de pesquisa (evolução temática, mapa temático, TC); e <b>(iii)</b> produção de uma estrutura de rede (análise de coautoria, co-citação). Exemplos de perguntas são: <b>(1)</b> Quais os autores, instituições e países mais produtivos? <b>(2)</b> Qual o tópico mais pesquisado? (Öztürk; Kocaman; Kanbach, 2024).
	2 - Criação da(s) <i>query(ies)</i> de busca	Criar a(s) <i>query(ies)</i> de busca utilizando descritores consolidados ou termos de tesouros, como <i>MeSH</i> , <i>Engineering main heading</i> , <i>Engineering uncontrolled terms</i> , <i>EMTREE medical terms</i> , com o objetivo de minimizar vieses. No entanto, a inclusão de termos livres pode ampliar o escopo da amostra, pois nem todos os artigos publicados utilizam descritores consolidados ou termos controlados de tesouros. Essa discussão deve ser realizada antes das pesquisas. Pode-se aplicar operadores booleanos (ex.: "AND NOT" ("E NÃO")) para excluir resultados irrelevantes, garantindo a precisão e especificidade na recuperação dos dados. Karger e Kureljusic (2023) elaboraram uma <i>string</i> de busca sobre o diagnóstico do câncer por IA, composta por duas partes: uma voltada aos termos técnicos, como (" <i>artificial intelligence</i> " AND " <i>machine intelligence</i> "), e outra relacionada ao domínio de aplicação, como (" <i>cancer detect</i> " AND " <i>cancer diagnos*</i> "). Essa abordagem pode ser adaptada para a construção de <i>queries</i> focadas em diferentes perspectivas, como categorias temáticas distintas ("sustentabilidade" AND "economia circular"), contextos geográficos variados ("políticas públicas" AND "América Latina"), interações

		entre conceitos (ética <i>AND</i> inteligência artificial) ou até tipos de estudos e metodologias ("análise bibliométrica" <i>AND</i> educação). A colaboração de um especialista em biblioteconomia e ciência da informação é fortemente recomendada (Da Silva; Hayashi C.; Hayashi M., 2011). Para elaborar uma <i>query</i> de busca, consulte Bramer <i>et al.</i> (2018).
	3 - Escolha da ferramenta de análise	Selecionar um <i>software</i> adequado para o processamento de dados e para a análise bibliométrica. Essa etapa deve ser realizada antes de se iniciar as buscas, a fim de evitar a coleta de dados de bases incompatíveis ou que apresentem limitações de análise pelo <i>software</i> , como a junção de dados de diferentes bases. Exemplos de ferramentas incluem VOSviewer, Gephi, CiteSpace e <i>bibliometrix</i> (Aria; Cuccurullo, 2017). A escolha depende desde a preferência do pesquisador, com base em suas habilidades e competências (Da Silva; Hayashi C.; Hayashi M., 2011) e da compatibilidade dos tipos de dados até as aplicações mais adequadas (por exemplo, VOSviewer para visualização de redes e CiteSpace para análise de citações). O <i>bibliometrix</i> (Aria; Cuccurullo, 2017) pode ser utilizado para a condução de uma análise bibliométrica completa; contudo, pode-se utilizar mais de uma ferramenta de análise, dependendo da complexidade do estudo, como descrito por Karger e Kureljusic (2023).
	4 - Seleção e testes de bases de dados	Identificar e testar bases de dados adequadas para a coleta de artigos, assegurando que as fontes sejam confiáveis e abrangentes. Adaptar a <i>query</i> de busca para diferentes bases (Bramer <i>et al.</i> , 2018) e considerar se a junção de dados de múltiplas bases é necessária. Sobre a junção de bases de dados, Donthu <i>et al.</i> (2021) e Öztürk, Kocaman e Kanbach (2024) sugerem trabalhar com apenas uma base, enquanto Echchakoui (2020) sugere (no contexto geral) que sejam utilizadas Scopus e WOS. Na prática, identificaram-se limitações severas na junção das bases Scopus, WOS, PubMed, Cochrane e Dimensions, utilizando a ferramenta <i>bibliometrix</i> . A junção dessas bases resultou em problemas na completude de metadados e criação de linhas aleatórias no conjunto de dados resultante, impossibilitando a visualização gráfica dos dados, sem o tratamento adequado. Isso ocorre porque cada base exporta os dados com uma <i>tag</i> (rótulo ou título da coluna de dados) diferente, e a base adequada para o uso no <i>bibliometrix</i> são as <i>tags</i> da Clarivate. Para conhecer as características de diferentes bases de dados, consulte Martinez <i>et al.</i> (2023). Uma alternativa para juntar os dados da Scopus e WOS é apresentada por Lim, Kumar e Donthu (2024).
II Simulação	5 - Busca preliminar, pré-triagem e exportação dos dados	Realizar as buscas nas bases de dados escolhidas, conforme o escopo e os filtros previamente definidos, exportando os dados da base selecionada. Os arquivos de dados devem ser exportados contendo todas as informações possíveis. Além disso, é recomendado que sejam gerados em arquivos com diferentes extensões, como CSV, BIB, RIS e XLSX, pois podem ser úteis em eventuais análises complementares por outros <i>softwares</i> . Donthu <i>et al.</i> (2021) sugerem que sejam avaliados, no mínimo, 500 estudos da área de interesse. Entretanto, na prática isso varia, como relatado na discussão do artigo. É necessário considerar que diferentes tipos de documentos (artigos, capítulos etc.) apresentam diferentes peso de citação (Wallin, 2005). Neste caso, sugere-se: (i) seleção de diferentes tipos de documentos para estudos com o objetivo de compreender a dispersão geral da produção; e (ii) seleção apenas de artigos para um estudo de tendência, com implicações práticas e atuais.
	6 - Importação, tratamento, análise exploratória e validação dos dados	Importar, processar e limpar os dados coletados, removendo duplicatas e inconsistências para garantir a integridade da análise. Realizar novas análises exploratórias para assegurar que os dados são adequados. Criar perguntas e simular cenários diversos, aplicando técnicas de análise de performance e mapeamento científico, verificando a completude dos metadados e se a amostra é suficiente para responder às perguntas de pesquisa. Para uma análise histórica, é necessário investigar desde a primeira publicação da amostra, sem recortes temporais. Para identificar tendências recentes, pode-se focar no período de maior aumento nas publicações. Uma análise bibliométrica completa requer

		<p>duas abordagens: <b>(i)</b> geral, sem recorte temporal, e <b>(ii)</b> focada no período mais atual, considerando o início do aumento da produção de forma exponencial, identificado na visualização preliminar dos dados. Contudo, podem ser realizadas subanálises com períodos específicos. Em relação a falha na completude de metadados, existe a possibilidade de preencher os dados faltantes manualmente; contudo, isso pode ser um grande desafio e tomar tempo, já que pode haver ambiguidade nos dados ou, ainda, eles não serem localizados. Essa é uma escolha dos autores, que devem considerar o tempo de resposta necessário para os resultados do estudo, visto que a análise bibliométrica é uma metodologia com perfil ágil. A automação do tratamento de dados pode ser feita via APIs (<i>Application Programming Interfaces</i>) (Interfaces de Programação de Aplicações) das bases ou <i>scripts</i> personalizados, sendo sugerido consultar sempre o manual do <i>software</i> escolhido para verificar as possibilidades.</p>
	7 - Revisão da(s) <i>query(ies)</i> de busca (opcional)	<p>Caso sejam encontradas inconsistências ou problemas com a qualidade dos dados, de modo que possam comprometer a resposta ao problema de pesquisa, deve-se ajustar a <i>query</i> de busca e repetir o processo a partir do passo 5. Esse ciclo de <i>feedback</i> garantirá que os dados atendam aos objetivos da pesquisa. Se os problemas persistirem, pode ser necessário revisar os objetivos do estudo e reiniciar o processo desde o passo 1. Caso não seja necessário revisar a <i>query</i>, prossiga para o passo 8.</p>
III Condução da análise bibliométrica e documentação	8 - Seleção dos artigos por dupla leitura ou duplo cego	<p>Criar um roteiro para refinar ainda mais a seleção dos artigos com base na leitura dos resumos, com perguntas como: <b>1)</b> O <i>design</i> do estudo é adequado para responder ao problema de pesquisa proposto? <b>2)</b> Os objetivos do estudo foram declarados? <b>3)</b> O estudo adere ao problema de pesquisa [declarar o problema]? Para maior rigor, é sugerido que os resumos dos artigos sejam: <b>(i)</b> selecionados por especialista da área com dupla leitura dos resumos; ou <b>(ii)</b> selecionados por dois autores independentes (duplo-cego), com apoio de um especialista, para resolver eventuais divergências na seleção. Esse procedimento confere maior rigor metodológico ao estudo, considerando que, mesmo com esforços para uma seleção criteriosa dos artigos por meio de uma <i>query</i> bem elaborada, podem ser recuperados textos relacionados, mas que não aderem diretamente ao problema de pesquisa. Após a seleção dos artigos, deve-se refazer o passo 6 (e o 7, caso necessário) antes de prosseguir para o passo 9. Isso se deve ao fato de que, dependendo da quantidade de artigos excluídos, duas situações podem ocorrer: <b>(i)</b> a amostra pode se tornar pequena demais, comprometendo sua representatividade; <b>(ii)</b> a exclusão de artigos pode impactar as métricas da amostra, exigindo uma nova análise da qualidade dos dados. Sugere-se que a exclusão seja feita diretamente em uma cópia do arquivo (planilha), a ser importada posteriormente para a análise.</p>
	9 - Seleção de métricas e técnicas	<p>Escolher as métricas e técnicas que serão aplicadas, de forma apropriada para a análise bibliométrica, levando em conta as características dos metadados disponíveis, e os resultados identificados no passo 6. Nem sempre será necessária a condução de uma análise bibliométrica completa, sendo sugerido focar na aplicação das técnicas que responderão aos problemas de pesquisa, conforme exemplos e referências dos <b>quadro 2 e 3, e tabela 1</b>. Adicionalmente recomenda-se consultar: Byl <i>et al.</i> (2024) e Rahman <i>et al.</i> (2024).</p>
	10 - Importação dos dados, condução da análise bibliométrica e descrição dos resultados	<p>Importar os dados resultantes do passo 8 para o <i>software</i> escolhido e conduzir a análise bibliométrica (Zupic; Čater, 2015; Aria; Cuccurullo, 2017; Donthu <i>et al.</i>, 2021; Öztürk; Kocaman; Kanbach, 2024). Após a análise dos dados, recomenda-se utilizar como base o guia <i>BIBLIO</i> para estruturar o artigo de análise bibliométrica (Montazeri <i>et al.</i>, 2023). Uma sequência para a condução e descrição dos resultados, baseada em Donthu <i>et al.</i> (2021), envolve três etapas: <b>(i) Análise descritiva:</b> apresentar os dados gerais da amostra em tabelas e descrevê-los (<b>tabela 1</b>). <b>(ii) Análise de performance:</b> aplicar métricas como TP, TC e <i>h-index</i>, e técnicas como a Lei de Bradford e a Lei de Lotka. Pode-se apresentar os resultados na seguinte ordem: artigo mais antigo e artigos mais</p>

		<p>citados da amostra; autores, periódicos, instituições e países mais produtivos e mais citados (influentes). <b>(iii) Mapeamento científico:</b> A análise de palavras-chave frequentes, <i>trend topics</i>, mapeamento temático e mapas de colaborações internacionais pode ser suficiente para entender o estado atual da pesquisa e as tendências emergentes. Essa abordagem pode ser complementada por técnicas como acoplamento bibliográfico, co-ocorrência de palavras-chave e análise fatorial, conforme os objetivos. A apresentação deve ser descritiva e focada nos resultados, sem inferências dos autores. Aria e Cuccurullo (2017) sugerem usar visualizações como redes, mapas e diagramas para facilitar a interpretação e as conexões entre dados. Um <i>framework</i> (quadro) integrativo (Öztürk; Kocaman; Kanbach, 2024) pode organizar categorias de métricas, destacando, por exemplo, os três periódicos mais produtivos, os mais citados, redes de colaboração, dentre outras. Essa estratégia garante informações essenciais sobre desempenho científico, mesmo quando o foco do estudo não seja a análise de performance. Volpe <i>et al.</i> (2023) apresentam os resultados correlacionando colaborações e impactos, exemplificando a flexibilidade na organização dos resultados.</p>
	<p>11 - Análise, interpretação dos dados e discussão</p>	<p>Realizar a discussão do estudo com ênfase nos achados bibliométricos, traçando inferências e relações entre as métricas e respondendo às perguntas iniciais de pesquisa. Recomenda-se que a discussão siga a ordem em que os dados foram apresentados. Sugestão de passos para a discussão dos achados, com base Montazeri <i>et al.</i> (2023) e European Radiology (2024) e perspectivas deste autores: inicie com uma breve retomada e contextualização do objetivo do estudo, seguida da discussão das principais métricas e achados bibliométricos, e tópicos de tendência. Realize uma análise crítica, confrontando os resultados com a literatura, e finalize abordando implicações do estudo (acadêmicas, sociais, clínicas etc.), potencialidades, limitações e perspectivas futuras, seguido da seção de conclusão. Recomenda-se a leitura de Öztürk, Kocaman e Kanbach (2024) para ampliar o escopo da discussão. Embora o foco seja a análise quantitativa, uma análise qualitativa de estudos selecionados, como os mais citados ou os principais em determinado <i>cluster</i>, pode complementar as perspectivas da pesquisa.</p>
	<p>12 - Elaboração do material suplementar e manual de reprodutibilidade (opcional)</p>	<p>As análises bibliométricas são estudos que frequentemente geram um grande volume de dados devido à complexidade dos procedimentos e às diversas possibilidades de análise. O uso de material suplementar deve ser encarado como um complemento à pesquisa, evitando que o texto principal dependa dele para fazer sentido ao leitor. O texto principal deve ser autossuficiente e claro, enquanto o material suplementar pode conter informações adicionais, como: descrições detalhadas dos procedimentos de busca, um glossário com os termos utilizados, além de tabelas e figuras complementares que não sejam essenciais para a compreensão do estudo. No material suplementar, também é possível detalhar todos os procedimentos metodológicos, incluindo os <i>scripts</i> utilizados nos <i>softwares</i>, o que facilita a reprodutibilidade ou adaptações por outros pesquisadores. No entanto, este passo pode ser dispensado caso o <i>template</i> (modelo) do artigo permita uma descrição suficientemente detalhada para a reprodução da pesquisa diretamente no texto principal, ou se as normas do periódico escolhido não suportarem a inclusão desse tipo de material. Para hospedar e disponibilizar o material suplementar, sugere-se o uso de repositórios como o Figshare (<a href="https://figshare.com">https://figshare.com</a>), que permite o armazenamento e o compartilhamento de dados de forma prática e acessível.</p>

**Nota:** TC = Total de citações; TP = Total de publicações. **Fonte:** Elaborado pelos autores (2024), com base na revisão de literatura apresentada e percepções práticas da aplicação do *framework* em outras pesquisas em andamento.

## 5 DISCUSSÃO

A análise bibliométrica tem se consolidado como uma ferramenta indispensável para compreender o panorama quantitativo da produção científica em diferentes áreas do conhecimento.

Este artigo apresentou uma revisão dos principais conceitos, abordagens e desafios relacionados à aplicação dessa metodologia. Além disso, propôs um *framework* metodológico em 12 passos, distribuídos em três módulos, com o objetivo de nortear pesquisadores em suas investigações. A partir dessa proposta, foram identificados elementos cruciais que impactam a eficácia e a qualidade das análises, bem como as decisões metodológicas que as acompanham.

### 5.1 ABRANGÊNCIA DA LITERATURA E BASES DE DADOS

De acordo com Donthu *et al.* (2021), a análise bibliométrica é recomendada para campos de pesquisa com 500 ou mais publicações, garantindo uma base robusta para métricas e mapeamentos. Para os autores, em áreas com poucos artigos (50 a 300), sua aplicação é desnecessária, pois resultados significativos dificilmente seriam obtidos, sendo mais apropriadas revisões sistemáticas ou tradicionais. Entretanto, na literatura, foram identificados estudos que abrangeram tamanhos de amostras variáveis, como 6450 estudos (Karger; Kureljusic, 2023), 31169 (Ruiz-Fresneda; Morales-Álvarez, 2024) e 164 artigos (Ínan, 2023).

Dois pontos merecem destaque. O primeiro é que, para temas emergentes e recentes, talvez a amostra de estudos não seja ampla o suficiente, mas o estudo bibliométrico pode contribuir para identificar as bases preliminares da literatura. A repetição periódica da análise (por exemplo, a cada um ou dois anos) pode proporcionar um acompanhamento atualizado das pesquisas, mantendo a ciência atualizada sobre determinado tema. O segundo ponto refere-se ao fato de que, embora seja utilizada uma amostra de milhares de documentos, pode haver falta de homogeneidade no peso das citações (Wallin, 2005), especialmente se abranger tipos de documentos diferentes (como artigos e capítulos de livro). Além disso, resultados potencialmente distorcidos podem surgir caso os artigos não sejam triados adequadamente quanto à aderência ao tema e a análise seja conduzida sobre uma amostra gerada apenas com base em filtros automáticos. Para mitigar esse problema, o *framework* metodológico proposto sugere a inclusão de um passo (8) para triagem manual, com dupla checagem do texto, ou, alternativamente, por dois revisores independentes, com o auxílio de um especialista, garantindo que a análise seja conduzida sobre uma amostra consistente, removendo artigos impertinentes para o estudo.

Em relação à seleção das bases de dados, observa-se uma dependência excessiva da ferramenta a ser utilizada. A fusão dos conjuntos de dados pode esbarrar em aspectos práticos, devido à divergência na sua estrutura, de acordo com cada base de dados. Na literatura, foram identificados estudos baseados principalmente na Scopus e na Web of Science (WOS), embora outras bases como Medline, Philosophers Index e Lilacs também tenham sido utilizadas (Pizzani *et al.*, 2010). Donthu *et*

al. (2021) e Öztürk, Kocaman e Kanbach (2024) sugerem o uso de apenas uma base de dados. Por outro lado, Echchakoui (2020) aponta que a fusão da Scopus e WOS pode trazer mais consistência para uma pesquisa. Nos presentes testes práticos, a junção das bases Scopus, WOS, Dimensions, PubMed e Cochrane, suportadas pelo *bibliometrix* resultou em problemas nos conjuntos de dados, como a inserção de linhas em branco, que impossibilitaram a importação do arquivo na interface gráfica pela função *biblioshiny()*, sem que fossem deletadas diretamente na planilha. Resultou ainda em métricas que, embora completas inicialmente, tornaram-se incompletas, o que possivelmente ocorreu devido às diferentes nomenclaturas das *tags* de cada base de dados.

Lim, Kumar e Donthu (2024) sugerem uma alternativa para combinar as bases Scopus e WOS. Atualmente, o *bibliometrix* apresenta uma função para mesclar conjuntos de dados na interface gráfica. Recomenda-se sempre confirmar a real necessidade de juntar as bases, pois a maioria dos periódicos indexados na WOS também está indexada na Scopus. Além disso, reforça-se a necessidade de atenção à qualidade final dos metadados.

Em um guia de 24 passos para a condução de uma revisão sistemática, Muka *et al.* (2019, p. 52, tradução nossa) afirmam que "uma busca bibliográfica deve incluir pelo menos quatro bases de dados online: Embase, MEDLINE, Web of Science e Google Scholar". Embora o interesse da revisão sistemática seja avaliar a qualidade das evidências, isso levanta uma questão sobre a abrangência e representatividade da amostra. Logo tem-se a reflexão de que trabalhar exclusivamente com bases como WOS e Scopus, embora populares, pode limitar a representatividade da amostra, já que estudos potenciais para a análise podem estar contidos em outras bases.

## 5.2 ESTRUTURA DA ANÁLISE E MÉTRICAS

As análises bibliométricas são tradicionalmente estruturadas em três frentes principais: descritiva, de performance e de mapeamento científico (Donthu *et al.* 2021). Esse equilíbrio metodológico permite não apenas a interpretação quantitativa, mas também a contextualização qualitativa indireta das inter-relações identificadas (Geisler, 2000; Glänzel, 2003). Apesar de sua natureza quantitativa, a integração com uma breve síntese ou análise qualitativa—como, por exemplo, dos estudos mais citados ou mais relevantes dentro de um determinado *cluster*—pode enriquecer a interpretação dos resultados. Essa abordagem amplia a relevância das contribuições acadêmicas, proporcionando uma visão mais abrangente das tendências, lacunas e potenciais aplicações práticas dos achados bibliométricos. Essa integração é reafirmada pelo estudo de De Medeiros Filho e Russo (2018), sobre marcas como um indicador no contexto das empresas, que mesclou uma revisão

sistemática com análise bibliométrica, demonstrando como as metodologias podem se complementar para oferecer *insights* mais amplos.

### 5.3 IMPACTO DA QUALIDADE DOS METADADOS

Wallin (2005) destaca que a ausência de metadados, como endereços completos de autores em 17% dos registros da SCIE (*Science Citation Index Expanded*, Índice Expandido de Citação Científica), compromete análises bibliométricas, especialmente em níveis intermediários, dificultando a identificação de afiliações institucionais e a avaliação do impacto científico. Isso tende a se agravar com a falta de padronização dos nomes das instituições. Por outro lado, ao tentar realizar o preenchimento manual, o pesquisador pode deparar-se com a afiliação múltipla e simultânea de um autor, que pode comprometer a precisão da análise bibliométrica ao distorcer métricas como produtividade, colaboração e impacto acadêmico.

O exemplo da exclusão de artigos com falhas em palavras-chave nos *datasets* analisados (**Tabela 1**) ilustra como a ausência de dados pode comprometer análises dependentes de co-ocorrência de termos, reduzindo a quantidade de informações disponíveis. Entretanto, ressalta-se que palavras-chave adicionais não podem ser corrigidas manualmente, pois são geradas automaticamente por algoritmo a partir dos títulos das referências citadas por um artigo, o que limita a intervenção do pesquisador e pode introduzir vieses nos resultados da análise.

Vale ressaltar que a exclusão de artigos com falhas nos metadados deve ser conduzida com máxima cautela, pois pode impactar e alterar significativamente os resultados das métricas, além de introduzir viés, enfraquecendo as bases metodológicas da pesquisa. Assim, qualquer decisão deliberada de trabalhar com limitações na completude dos metadados ou de excluir estudos deve ser discutida com um especialista da área e devidamente reportada na seção de metodologia ou limitações do estudo, alertando os leitores sobre os potenciais riscos na interpretação dos resultados.

Espera-se que uma análise bibliométrica produza resultados de forma ágil, considerando a inserção frequente de novas publicações na literatura, o que pode impactar os resultados sobre temas emergentes. Isso sugere que, em temas de inovação, pode ocorrer a rápida obsolescência da análise bibliométrica, o que torna necessário realizar novas pesquisas periodicamente sobre um tema, para que as decisões sejam fundamentadas em dados atualizados e relevantes.

### 5.4 POTENCIALIDADES, LIMITAÇÕES E PERSPECTIVAS FUTURAS

O *framework* metodológico apresentado demonstra que a flexibilidade na escolha de métricas e técnicas é fundamental para adaptar a análise aos objetivos específicos da pesquisa. Além disso, ele

pode ser ajustado para a condução de pesquisas em qualquer área do conhecimento. Testes simulados e ciclos de *feedback*, intencionalmente inseridos, contribuem para ajustar o foco da investigação, garantindo que a amostra seja representativa e adequada às questões-problema, além de proporcionar resultados mais transparentes e confiáveis.

A principal limitação é que utilizou-se apenas o *bibliometrix* para conduzir as análises, sendo sugerido que outras ferramentas sejam agregadas para ampliar a perspectiva do *framework* e explorar diferentes aplicações de técnicas aos dados. Como continuidade, sugere-se a adoção de abordagens multimodais, combinando elementos quantitativos e qualitativos para alcançar uma compreensão mais abrangente e detalhada da produção científica. Também se propõe a aplicação do *framework* em diferentes bases de dados e contextos temáticos, a fim de refinar e expandir o modelo apresentado. Por fim, ressalta-se a necessidade do desenvolvimento de estratégias para manter o rigor na completude dos metadados e na integração entre diferentes bases de dados, possibilitando uma análise verdadeiramente ampla e representativa da literatura.

## 6 CONCLUSÃO

O presente estudo apresentou uma revisão narrativa sobre a metodologia de análise bibliométrica, destacando sua importância como ferramenta de avaliação quantitativa e das conexões que se formam na produção científica, e propôs um *framework* metodológico em 12 passos, a ser utilizado por pesquisadores e estudantes interessados no tema. A análise bibliométrica, apesar de suas limitações, permanece uma metodologia importante para mapear a produtividade, a influência e as comunidades que se formam em torno da produção acadêmico-científica, bem como identificar temas emergentes de pesquisa. A aplicação do *framework* proposto pode contribuir para práticas metodológicas mais rigorosas e confiáveis. Como continuidade, sugere-se explorar a integração de novas métricas, abordagens híbridas (quantitativas e qualitativas) e técnicas avançadas de análise, ampliando a aplicabilidade e a profundidade das análises bibliométricas.

## AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior -Brasil (CAPES) - Código de Financiamento 001.

## REFERÊNCIAS

- ABBAS, A. F. et al. Bibliometrix analysis of information sharing in social media. *Cogent Business & Management*, London, v. 9, n. 1, p. 2016556, 2022.
- ALABI, G. Bradford's law and its application. *International Library Review*, [s.l.], v. 11, n. 1, p. 151-158, jan. 1979.
- ALVARENGA, L. Bibliometria e arqueologia do saber de Michel Foucault: traços de identidade teórico-metodológica. *Ciência da Informação*, Brasília, v. 27, n. 3, set. 1998.
- ARIA, M.; CUCCURULLO, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, [s.l.], v. 11, n. 4, p. 959-975, nov. 2017.
- BALL, Rafael. *An Introduction to Bibliometrics*. Amsterdam: Chandos Publishing, 2017.
- BRAMER, W. M. et al. A systematic approach to searching: an efficient and complete method to develop literature searches. *Journal of the Medical Library Association: JMLA*, Chicago, v. 106, n. 4, p. 531, 2018.
- BYL, L. et al. Measuring Research Output through Bibliometrics. 2016. Disponível em: <https://bit.ly/3YbJujs>. Acesso em: 16 out. 2024.
- CASSETTARI, R. R. B. et al. Comparação da Lei de Zipf em conteúdos textuais e discursos orais. *El profesional de la información*, [s.l.], v. 24, n. 2, p. 157-167, 2015.
- CHEN, G. et al. Comparing semantic representation methods for keyword analysis in bibliometric research. *Journal of Informetrics*, [s.l.], v. 18, n. 3, p. 101529, 2024.
- DA SILVA, M. R.; HAYASHI, C. R.; HAYASHI, M. C. P. I. Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo. *InCID: revista de ciência da informação e documentação*, Ribeirão Preto, v. 2, n. 1, p. 110-129, 9 jun. 2011.
- DONTHU, N. et al. How to conduct a bibliometric analysis: an overview and guidelines. *Journal of Business Research*, [s.l.], v. 133, p. 285-296, set. 2021.
- ECHCHAKOUI, S. Why and how to merge Scopus and Web of Science during bibliometric analysis: the case of sales force literature from 1912 to 2019. *Journal of Marketing Analytics*, [s.l.], v. 8, p. 165-184, 2020.
- EGGHE, L. Theory and practise of the g-index. *Scientometrics*, [s.l.], v. 69, p. 131-152, 20 jun. 2006.
- ELSEVIER. *Research Metrics Guidebook*. 2009. Disponível em: <https://bit.ly/4b9PfnH>. Acesso em: 07 ago. 2024.
- EUROPEAN RADIOLOGY. *European Radiology Manuscript requirements for submissions*. (Ed. 2). 2024. Disponível em: <https://bit.ly/3CNn54V>. Acesso em: 24 nov. 2024.

- FERRARI, R. Writing narrative style literature reviews. *Medical writing*, [s.l.], v. 24, n. 4, p. 230-235, 2015.
- GEISLER, E. *The metrics of science and technology: evaluation and measurement of research, development, and innovation*. Westport, CT: Quorum Books. 2000.
- GLÄNZEL, W. *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. 2003. Disponível em: <https://bit.ly/3yr5OwZ>. Acesso em: 14 ago. 2024.
- GREEN, B. N.; JOHNSON, C. D.; ADAMS, A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *Journal of chiropractic medicine, Lombard*, v. 5, n. 3, p. 101-117, 2006.
- GUIMARÃES, A. J. R.; MOREIRA, P. S. C.; BEZERRA, C. A. Modelos de inovação: análise bibliométrica da produção científica. *Brazilian Journal of Information Science*, Marília, n. 15, p. 6, 4 maio 2021.
- HIRSCH, J. E. An index to quantify an individual's scientific research output. *arXiv:physics/0508025 [physics.soc-ph]*, [s.l.], v. 102, n. 46, p. 16569-16572, 29 set. 2005.
- HJØRLAND, B. Domain analysis: A socio-cognitive orientation for information science research. *Bulletin of the American Society for Information Science and Technology*, Silver Spring, v. 30, n. 3, p. 17-21, 31 fev.-mar. 2004.
- İNAN, Ü. S. E. Evaluation of digital marketing from a bibliometric analysis perspective. *Socialis Series in Social Science*, Jaipur, v. 4, p. 45-58, 2023.
- KARGER, E.; KURELJUSIC, M. Artificial intelligence for cancer detection—a bibliometric analysis and avenues for future research. *Current Oncology*, Basel, v. 30, n. 2, p. 1626-1647, 2023.
- LIM, W. M. ; KUMAR, S.; DONTU, N. How to combine and clean bibliometric data and use bibliometric tools synergistically: guidelines using metaverse research. *Journal of Business Research*, [s.l.], v. 182, p. 114760, 2024.
- LOTKA, A. J. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, New York, v. 16, n. 12, p. 317-323, 19 jun. 1926.
- MARTINEZ, E. C. et al. Ten steps to conduct a systematic review. *Cureus*, Palo Alto, v. 15, n. 12, 2023.
- MEDEIROS FILHO, A. R.; RUSSO, S. L. Marcas como um indicador: revisão sistemática e análise bibliométrica da literatura. *Biblios*, Pittsburgh, n. 71, p. 50-67, 2018.
- MILES, M. B.; HUBERMAN, A. M. *Qualitative data analysis: an expanded sourcebook*. 2nd ed. Thousand Oaks: SAGE Publications, 1994.
- MOMESSO, A. C.; NORONHA, D. P. Bibliométrie ou Bibliometrics: o que há por trás de um termo?. *Perspectivas em ciência da informação*, v. 22, p. 118-124, 2017.

MONTAZERI et al. Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. *Systematic Reviews*, [s.l.], v. 12, n. 1, 15 dez. 2023.

MUKA, T. et al. A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. *European journal of epidemiology*, v. 35, p. 49-60, 2020.

MUKERJEE, S. A systematic comparison of community detection algorithms for measuring selective exposure in co-exposure networks. *Scientific Reports*, London, v. 11, n. 1, p. 15218, 2021.

NAVARRO-BALLESTER, A. et al. Publications on COVID-19 in radiology journals in 2020 and 2021: bibliometric citation and co-citation network analysis. *European Radiology*, Berlin, v. 33, n. 5, p. 3103-3114, maio 2023.

NEWMAN, Mark E. J. *Networks: an introduction*. Oxford; New York: Oxford University Press, 2010.

NINKOV, A.; FRANK, J. R.; MAGGIO, L. A. *Bibliometrics: Methods for studying academic publishing. Perspectives on medical education*, London, v. 11, n. 3, p. 173-176, 2022.

N'ZE, A. A. P.; TENKOUL, A. Global Research on Good Governance and Sustainability: a bibliometric analysis. *International Journal of Sustainability in Economic, Social, and Cultural Context*, [s.l.], v. 20, n. 2, p. 47-76, 20 set. 2024.

ÖZTÜRK, O.; KOCAMAN, R.; KANBACH, D. K. How to design bibliometric research: an overview and a framework proposal. *Review of managerial science*, [s.l.], p. 1-29, 6 mar. 2024.

PASSAS, I. *Bibliometric Analysis: The Main Steps*. Encyclopedia, Basel, v. 4, n. 2, p. 1014–1025, 2024.

PIZZANI, L.; DA SILVA, R. C.; HOSSNE, W. S. Análise bibliométrica dos 40 anos da produção científica no Brasil e no mundo. *Revista Bioethikos*, São Paulo, v. 4, n. 4, p. 453-460, 2010.

POSIT. RSTUDIO IDE The most trusted IDE for open source data science. 2024. Disponível em: <https://posit.co/products/open-source/rstudio/>. Acesso em: 06 out. 2024.

RAHMAN, A. et al. A comprehensive bibliometric and content analysis of artificial intelligence in language learning: tracing between the years 2017 and 2023. *Artificial Intelligence Review*, [s.l.], v. 57, n. 4, p. 107, 2024.

RUIZ-FRESNEDA, M. A.; GIJÓN, A.; MORALES-ÁLVAREZ, P. Bibliometric analysis of the global scientific production on machine learning applied to different cancer types. *Environmental Science and Pollution Research*, [s.l.], v. 30, n. 42, p. 96125-96137, set. 2023.

SMITH, N. R. et al. A guide for choosing community detection algorithms in social network studies: the question alignment approach. *American journal of preventive medicine*, Philadelphia, v. 59, n. 4, p. 597-605, 2020.

TAGUE-SUTCLIFFE, J. An introduction to informetrics. *Information processing & management*, [s.l.], v. 28, n. 1, p. 1-3, 1992.

TERRA et al. Análise bibliométrica com o software bibliometrix. In: XLII ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 2022, Foz do Iguaçu, PR. Anais [online]. Foz do Iguaçu: Enegp, 2022. Disponível em: <https://bit.ly/4cyl1Ko>. Acesso em: 20 ago. 2024.

URIBE, F. F. M; GUERRERO, J. C. O.; CONTRERAS, F. C. Introducción a la bibliometría práctica. Asociacion de Bibliotecólogos del Peru. 2023. Disponível em: <https://bit.ly/3WWNncN>. Acesso em: 10 ago. 2024.

VAN NOORDEN, R. A profusion of measures. Nature, London, v. 465, n. 7300, p. 864-867, jun. 2010.

VASCONCELOS, Y. L. Estudos bibliométricos: procedimentos metodológicos e contribuições. Revista de Ciências Jurídicas e Empresariais, [s.l.], v. 15, n. 2, p. 211-220, set. 2014.

VOLPE, S. et al. Quo vadis Radiomics? Bibliometric analysis of 10-year Radiomics journey. European Radiology, Vienna, v. 33, n. 10, p. 6736-6745, 2023.

WALLIN, J. A. Bibliometric methods: pitfalls and possibilities. Basic & clinical pharmacology & toxicology, Oxford, v. 97, n. 5, p. 261-275, 2005.

ZHANG, Y. et al. Artificial intelligence in sign language recognition: a comprehensive bibliometric and visual analysis. Computers and Electrical Engineering, [s.l.], v. 120, p. 109854, 2024.

ZIPF, G. K. On the Economy of Words. In: ZIPF, G. K. Human Behavior and the principle of least effort. An Introduction to Human Ecology. Massachusetts: Addison-Wesley Press, INC, 1949. p. 19-55. Disponível em: <https://bit.ly/3OsYbKO>. Acesso em: 24 nov. 2024.

ZUPIC, I.; ČATER, T. Bibliometric methods in management and organization. Organizational research methods, [s.l.], v. 18, n. 3, p. 429-472, 2015.