



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Fabio Lorensi do Canto

**Avaliação de impacto de periódicos Ibero-Americanos com base no índice h5
do Google Scholar Metrics**

Florianópolis
2022

Fabio Lorensi do Canto

**Avaliação de impacto de periódicos Ibero-Americanos com base no índice h5
do Google Scholar Metrics**

Tese submetida ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Doutor em Ciência da Informação.

Área de concentração: Gestão da Informação
Linha pesquisa: Informação, Comunicação Científica e Competência.

Florianópolis

2022

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Canto, Fabio Lorensi

Avaliação de impacto de periódicos Ibero-Americanos com base no índice h5 do Google Scholar Metrics / Fabio Lorensi Canto ; orientador, Adilson Luiz Pinto, coorientador, Thiago Magela Rodrigues Dias , 2022.
190 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós Graduação em Ciência da Informação, Florianópolis, 2022.

Inclui referências.

1. Ciência da Informação. 2. Avaliação de impacto . 3. Periódicos científicos Ibero-Americanos. 4. Índice h5. 5. Google Scholar Metrics. I. Pinto, Adilson Luiz . II. Dias , Thiago Magela Rodrigues. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Ciência da Informação. IV. Título.

Fabio Lorensi do Canto

**Avaliação de impacto de periódicos Ibero-Americanos com base no índice h5 do
Google Scholar Metrics**

O presente trabalho em nível de Doutorado foi avaliado e aprovado, em 22 de julho 2022,
pela banca examinadora composta pelos seguintes membros:

Prof. Douglas Dyllon Jeronimo de Macedo, Dr.
Universidade Federal de Santa Catarina

Alexandre Ribas Semeler, Dr.
Universidade Federal do Rio Grande do Sul

Prof. José Antonio Moreira González, Dr.
Universidad Carlos III de Madrid

Certificamos que esta é a versão original e final do trabalho de conclusão que foi julgado
adequado para obtenção do título de Doutor em Ciência da Informação.

Coordenação do Programa de Pós-Graduação

Prof. Adilson Luiz Pinto, Dr.
Orientador

Florianópolis, 2022.

RESUMO

Proposta de *framework* de avaliação de impacto de periódicos Ibero-Americanos com base no índice h5 do Google Scholar Metrics. O *framework* *Gsm@hdata* é composto por três dimensões: uma teórica, uma tecnológica e uma analítica. A dimensão teórica aplica conceitos de análise de citação no índice h5, a dimensão tecnológica é representada por uma ferramenta bibliométrica de pesquisa, extração e análise de dados do Google Scholar Metrics e a dimensão analítica aplica os conceitos da dimensão teórica na análise dos dados extraídos na dimensão tecnológica. A escolha da região Ibero-Americana foi motivada pela menor visibilidade em bases de dados internacionais. Os sistemas de avaliação do Brasil, Espanha, México e Colômbia definiram o índice h5 como indicador de impacto para periódicos não indexados no *Journal Citation Reports* e na *Scopus*. A validação do *framework* se deu em um universo de 10.553 periódicos de 23 países Ibero-Americanos indexados no diretório Latindex. Por meio da ferramenta bibliométrica *Gsm_hdata* foram executadas rotinas de busca, de análise de citação e de autocitação e de extração de dados do índice h5, inclusive de edições anteriores do Google Scholar Metrics. Foram localizados 4.049 periódicos indexados na edição de 2021 do Google Scholar Metrics, o que corresponde a 38,4% do total e representa mais do que o dobro de visibilidade em comparação a *Scopus*, o *Journal Citation Reports* e o *SciELO*. Os periódicos do Brasil, Espanha, México, Colômbia, Argentina, Chile e Cuba correspondem a 90% do total. Os periódicos de Ciências Sociais, de Artes e Humanidades e de Ciências da Saúde abrangem 72%. O índice h5 é um indicador válido, considerando os níveis de correlação com outros indicadores, e estável, com crescimento médio anual de 11% nos últimos cinco anos. O índice h5 é pouco influenciado por autocitações, com taxa média de 3,6% por periódico. Mas o excesso de autocitações, mais frequentes em periódicos de menor impacto e de áreas especializadas, podem aumentar significativamente o índice h5. Os resultados sugerem que a principal vantagem do índice h5 do Google Scholar Metrics é a cobertura mais ampla de periódicos regionais. Mas os seus escassos recursos limitam o uso em larga escala. O *framework* proposto, especialmente a sua dimensão tecnológica, contorna parte das limitações. Possibilita a análise de impacto de periódicos Ibero-Americanos com base no índice h5 de forma mais confiável e menos suscetível a distorções ocasionadas por excesso de autocitações e eventuais práticas de manipulação de citações. O *framework* poderá ser futuramente aplicado em processos de avaliação de periódicos não indexados em bases de dados, inclusive de outras regiões de baixa visibilidade científica.

Palavras-chave: avaliação de impacto; periódicos Ibero-Americanos; índice h5; Google Scholar Metrics; autocitação de periódico.

ABSTRACT

This study proposes an impact evaluation framework for Ibero-American journals based on the Google Scholar Metrics h5-index. The Gsm@hdata framework is composed of three dimensions, a theoretical, a technological and an analytical dimension. The theoretical dimension applies concepts of citation analysis in the h5-index context, the technological one is represented by a bibliometric tool for searching, extracting and analyzing data from Google Scholar Metrics and the analytical one applies the concepts of the theoretical dimension in the analysis of the extracted data in the technological dimension. The reason for the choice of the Ibero-American region was the lower visibility in international databases. The evaluation systems of Brazil, Spain, Mexico and Colombia defined the h5-index as an impact indicator for journals not-indexed in the Journal Citation Reports and Scopus. The definitive validation of the framework took place in a universe of 10,553 journals from 23 Ibero-American countries indexed in the Latindex directory. Using the bibliometric tool Gsm_hdata, routines of search, citation analysis and self-citation and data extraction from the h5-index were performed, including from previous editions of the GSM. A total of 4,049 journals were found in the 2021 edition of this source, which corresponds to 38.4% and represents more than double the visibility compared to Scopus, Journal Citation Reports and SciELO. Journals from Brazil, Spain, Mexico, Colombia, Argentina, Chile and Cuba correspond to 90%. Journals in the Social Sciences, Arts and Humanities and Health Sciences comprise 72%. The h5-index is a valid indicator, in view of the levels of correlation with other indicators, and stable, with an average annual growth of 11% in the last five years. The h5-index is less influenced by self-citations, with an average rate of 3.6% per journal. But excess self-citations, more frequent in journals with lower average impact and in specialized disciplines, can significantly increase the h5-index. The results suggest that the main advantage of the Google Scholar Metrics h5-index is the broader coverage of regional journals. But its scarce resources limit large-scale use. The proposed framework, especially its technological dimension, overcomes part of the limitations. It makes it possible to analyze the impact of Ibero-American journals based on the h5-index in a more reliable way and less susceptible to distortions caused by excessive self-citations and possible citation manipulation practices. The framework may be applied in the future in evaluation processes of journals not indexed in databases, including those from other regions with low scientific visibility.

Keywords: *impact evaluation; Ibero-American journals; h5-index; Google Scholar Metrics; journal self-citation.*

RESUMEN

Esta tesis presenta a propuesta de un marco de evaluación de impacto para revistas Iberoamericanas basado en el índice h5 de Google Scholar Metrics. El marco Gsm@hdata está compuesto por tres dimensiones: una teórica, una tecnológica y una analítica. La dimensión teórica aplica conceptos de análisis de citas en el índice h5; la tecnológica está representada por una herramienta bibliométrica de búsqueda, extracción y análisis de datos de Google Scholar Metrics y la dimensión analítica aplica los conceptos de la dimensión teórica en el análisis de los datos que fueron extraídos en la dimensión tecnológica. La elección de la región Iberoamericana fue motivada por la menor visibilidad en las bases de datos internacionales. Los sistemas de evaluación de Brasil, España, México y Colombia definieron el índice h5 como indicador de impacto para revistas no indexadas en el Journal Citation Reports y Scopus. La validación definitiva del marco fue realizada en un universo de 10.553 revistas de 23 países Iberoamericanos indexadas en el directorio Latindex. Utilizándose la herramienta bibliométrica Gsm_hdata, se realizaron rutinas de búsqueda, análisis de citas y autocitas y extracción de datos del índice h5, incluso de ediciones anteriores del Google Scholar Metrics. En la edición 2021 del Google Scholar Metrics fueron encontradas 4.049 revistas indexadas, lo que corresponde al 38,4% del total y representa más del doble de visibilidad en comparación con Scopus, Journal Citation Reports y SciELO. Revistas de Brasil, España, México, Colombia, Argentina, Chile y Cuba corresponden al 90% del total. Las revistas de Ciencias Sociales, Artes y Humanidades y Ciencias de la Salud comprenden al 72%. El índice h es un indicador válido, considerándose el índice de correlación con otros indicadores, y estable, con crecimiento en promedio anual de 11% en los últimos cinco años. El índice h5 es poco influenciado por las autocitas, con una tasa media del 3,6% por revista. Las autocitas excesivas, más frecuentes en revistas de menor impacto en promedio y en áreas especializadas, pueden aumentar significativamente el índice h5. Los resultados y la principal ventaja del índice h5 de Google Scholar Metrics es la cobertura más amplia de revistas regionales. Pero sus escasos recursos limitan su uso a gran escala. El marco propuesto, especialmente en su dimensión tecnológica, elude parte de estas limitaciones. Permite un análisis periódico del impacto de las autocitas y los intentos ocasionales de periódicos de manipular números. Futuramente, el marco puede ser aplicado en procesos de evaluación de revistas no indexadas en bases de datos, incluyendo otras regiones de baja visibilidad.

Palabras-chave: *evaluación de impacto; revistas iberoamericanas; índice h5; métricas de Google Académico; auto citación.*

LISTA DE FIGURAS

Figura 1 - Metodologia de cálculo do CiteScore 2019.....	56
Figura 2 - Critérios de avaliação de impacto de periódicos da Espanha.....	78
Figura 3 - Resultado de avaliação de periódico no CRMICYT	79
Figura 4 – Representação do framework <i>Gsm@hdata</i> em diagrama de <i>Venn</i>	94
Figura 5 – Estrutura da ferramenta bibliométrica <i>Gsm_hdata</i>	97
Figura 6 – Busca de títulos no endereço URL do Google Scholar Metrics	98
Figura 7 – Identificação de título equivalente no Google Scholar Metrics.....	99
Figura 8 – Identificação de autocitação no Google Scholar Metrics.....	100
Figura 9 – Campo de ano na URL do Google Scholar Metrics	101
Figura 10 – Subcategorias de assunto de Ciências Sociais no Google Scholar Metrics	105
Figura 11 – Mapa dos países Ibero-Americanos do universo da pesquisa.....	109
Figura 12 – Processo de coleta de dados da pesquisa	112
Figura 13 – Distribuição dos periódicos em países Ibero-Americanos.....	117
Figura 14 – Rede de coocorrência de descritores de assuntos	123
Figura 15 – Matrix de correlação entre indicadores de periódicos Ibero-Americanos.....	126
Figura 16 – Rede de citações de periódico de Ciências Sociais.....	156
Figura 17 – Síntese do <i>framework Gsm@hdata</i>	168

LISTA DE GRÁFICOS

Gráfico 1 – Número de periódicos indexados no SciELO por país (2012-2020).....	32
Gráfico 2 – Distribuição de periódicos por país e por tipo de conteúdo	120
Gráfico 3 – Distribuição de periódicos por país e por natureza da instituição editora	120
Gráfico 4 – Distribuição de periódicos por país e por área do conhecimento	121
Gráfico 5 – Distribuição de periódicos por país e por periodicidade	124
Gráfico 6 – Distribuição de periódicos por país e por idioma	125
Gráfico 7 - Previsão do Fator de Impacto a partir do índice h5	128
Gráfico 8 - Previsão do CiteScore a partir do índice h5	129
Gráfico 9 – Distribuição do índice h5 dos periódicos por país.....	130
Gráfico 10 – Distribuição do índice h5 dos periódicos por área	132
Gráfico 11 – Impacto por país e idioma.....	135
Gráfico 12 – Crescimento do índice h5 por área do conhecimento (2017-2021).....	137
Gráfico 13 – Crescimento do índice h5 por país (2017-2021).....	138
Gráfico 14 – Distribuição das taxas de autocitação de acordo com o índice h5	139
Gráfico 15 - Distribuição das taxas de autocitação por país	140
Gráfico 16 - Distribuição das taxas de autocitação por área	142
Gráfico 17 - Variação entre os índices h5 e h _s 5	143
Gráfico 18 - Taxa de crescimento do índice h5.....	144
Gráfico 19 - Distribuição de citações por ano e por área (2016-2021).....	151
Gráfico 20 - Distribuição de citações em periódicos de Engenharias (2016-2021).....	152
Gráfico 21 - Distribuição de citações em periódicos de Exatas & Naturais (2016-2021)	153
Gráfico 22 - Distribuição de citações em periódicos Multidisciplinares (2016-2021).....	154
Gráfico 23 - Distribuição de citações em periódicos de Ciências Sociais (2016-2021).....	154
Gráfico 24 – Índice h5 de periódicos de Ciências Sociais (2017-2021)	155

LISTA DE QUADROS

Quadro 1 – Sistemas de classificação de periódicos Ibero-Americanos	82
Quadro 2 – Conceitos da dimensão teórica do <i>framework Gsm@hdata</i>	95
Quadro 3 – Formatos de entrada e de saída do <i>Gsm_htline</i>	101
Quadro 4 – Síntese dos procedimentos metodológicos	116

LISTA DE TABELAS

Tabela 1 – Número de documentos indexados por país no Google Scholar	38
Tabela 2 – Categorias e subcategorias de assuntos no Google Scholar Metrics.....	43
Tabela 3 – Síntese dos parâmetros de indicadores de impacto.....	51
Tabela 4 – Identificação do índice h de um conjunto de documentos	58
Tabela 5 – Estrutura do índice h_s5	69
Tabela 6 – Visibilidade de periódicos Ibero-Americanos em bases de dados.....	118
Tabela 7 – Média do índice h5 por país e por área	131
Tabela 8 – Média do índice h5 por país e por natureza da instituição	133
Tabela 9 – Média do índice h5 por país e por periodicidade	134
Tabela 10 – Análise de citação de periódicos de Ciências Agrárias	145
Tabela 11 - Análise de citação de periódicos de Engenharias	145
Tabela 12 - Análise de citação de periódicos de Exatas & Naturais	146
Tabela 13 - Análise de citação de periódicos de Ciências da Saúde	147
Tabela 14 - Análise de citação de periódicos de Artes & Humanidades	148
Tabela 15 - Análise de citação de periódicos de Ciências Sociais.....	149
Tabela 16 - Padrões de citação em periódicos Multidisciplinares	150

LISTA DE ABREVIATURAS E SIGLAS

A&H	Artes & Humanidades
A&HCI	<i>Arts & Humanities Citation Index</i>
AIS	<i>Article Influence Score</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CCA	Ciências Agrárias
CCS	Ciências Sociais
CDS	Ciências da Saúde
CS	<i>CiteScore</i>
E&N	Ciências Exatas & Naturais
ENG	Engenharias
ES	<i>Eigenfactor Score</i>
nES	<i>Normalized Eigenfactor Score</i>
ESCI	<i>Emerging Sources Citation Index</i>
FI	Fator de Impacto
FI5y	Fator de Impacto de cinco anos
FIws	Fator de Impacto sem autocitações
GS	<i>Google Scholar</i>
GSC	<i>Google Scholar Citations</i>
GSM	<i>Google Scholar Metrics</i>
ISI	<i>Institute for Scientific Information</i>
JCR	<i>Journal Citation Reports</i>
JSM	<i>Journal Scholar Metrics</i>
MTD	Multidisciplinar
SCIE	<i>Science Citation Index Expanded</i>
PoP	<i>Publish or Perish</i>
SCR	<i>Self-Citation Rate</i>
SJ&CR	<i>SCImago Journal & Country Rank</i>
SSCI	<i>Social Sciences Citation Index</i>
WoS	<i>Web of Science</i>

SUMÁRIO

1 INTRODUÇÃO	15
1.1 PROBLEMAS DE PESQUISA	19
1.2 OBJETIVOS	19
1.2.1 Objetivo geral	19
1.2.2 Objetivos específicos	19
1.3 JUSTIFICATIVA	19
1.4 ORIGINALIDADE DA PESQUISA	21
2 REVISÃO DE LITERATURA	23
2.1 BASES DE DADOS.....	23
2.1.1 Web of Science	25
2.1.2 Journal Citation Reports	26
2.1.3 Scopus	28
2.1.4 SCImago Journal & Country Rank	30
2.1.5 SciELO	30
2.1.6 Latindex	33
2.1.7 Google Scholar	34
2.1.7.1 Características do Google Scholar	36
2.1.7.2 Tamanho e crescimento do Google Scholar	37
2.1.7.3 Análise de citação no Google Scholar	39
2.1.7.4 Erros de indexação no Google Scholar	40
2.1.8 Google Scholar Citations	40
2.2 GOOGLE SCHOLAR METRICS	42
2.2.1 Tamanho e cobertura do Google Scholar Metrics	44
2.2.2 Vantagens do Google Scholar Metrics	46
2.2.3 Falhas e limitações do Google Scholar Metrics	47
2.3 INDICADORES DE IMPACTO	48
2.3.1 Fator de Impacto	52
2.3.2 Eigenfactor Score	53
2.3.3 Article Influence Score	54
2.3.4 CiteScore	55
2.3.5 SJR	56
2.3.6 SNIP	57
2.3.7 Índice h	58
2.4 CRÍTICAS AOS INDICADORES DE IMPACTO	62
2.5 MANIPULAÇÃO DE CITAÇÕES	64
2.6 AUTOCITAÇÃO DE PERIÓDICO	64
2.6.1 Índice h_s5	68
2.7 PERIÓDICOS IBERO-AMERICANOS.....	70
2.7.1 Sistemas de avaliação	73
2.7.1.1 Qualis Capes	75
2.7.1.2 ARCE/Fecyt.....	77

2.7.1.3 CRMCYT/CONACYT.....	79
2.7.1.4 Pubindex.....	80
2.7.1.5 REVENCYT.....	81
2.7.1.6 Sistemas de outros países Ibero-Americanos.....	81
2.8 AVALIAÇÃO DE IMPACTO NAS HUMANIDADES.....	82
2.9 JOURNAL SCHOLAR METRICS.....	84
2.10 PUBLISH OR PERISH.....	84
3 DEMONSTRAÇÃO DE ORIGINALIDADE.....	86
4 METODOLOGIA.....	91
4.1 CARACTERIZAÇÃO DA PESQUISA.....	91
4.2 PROCEDIMENTOS METODOLÓGICOS.....	92
4.2.1 Framework Gsm@hdata.....	93
4.2.1.1 Dimensão teórica do <i>framework</i>	94
4.2.1.2 Dimensão tecnológica do <i>framework</i>	96
4.2.1.2.1 <i>Gsm_hdata</i>	97
4.2.1.2.2 <i>Gsm_hscite</i>	99
4.2.1.2.3 <i>Gsm_hcitation</i>	100
4.2.1.2.4 <i>Gsm_htline</i>	101
4.2.1.3 Dimensão analítica do <i>framework</i>	102
4.2.1.3.1 Protocolo de análise de citação.....	102
4.2.2 Validação do framework GSM@hdata.....	105
4.2.2.1 Validação parcial interna.....	104
4.2.2.1.1 <i>Teste de precisão 1</i>	105
4.2.2.1.2 <i>Teste de precisão 2</i>	106
4.2.2.1.3 <i>Teste de precisão 3</i>	108
4.2.2.2 Validação parcial externa.....	107
4.2.3 Universo da pesquisa.....	109
4.2.4 Coleta de dados.....	110
4.2.5 Tratamento e análise dos dados.....	112
5 RESULTADOS E ANÁLISE.....	117
5.1 VISIBILIDADE DOS PERIÓDICOS IBERO-AMERICANOS.....	117
5.2 CARACTERÍSTICAS DAS PUBLICAÇÕES.....	119
5.2.1 Tipo de conteúdo.....	119
5.2.2 Natureza da instituição editora.....	120
5.2.3 Aderência temática.....	121
5.2.4 Escopo temático.....	122
5.2.5 Periodicidade.....	123
5.2.6 Idiomas.....	124
5.2.7 Modelo de acesso.....	125
5.3 IMPACTO DOS PERIÓDICOS.....	125
5.3.1 Validade do índice h5.....	125

5.3.2	Previsão do FI a partir do índice h5.....	128
5.3.3	Previsão do CS a partir do índice h5	129
5.3.4	Impacto por país e por área do conhecimento	130
5.3.5	Impacto por país e natureza da instituição editora	133
5.3.6	Impacto por país e periodicidade	134
5.3.7	Impacto por país e por idioma	135
5.3.8	Crescimento anual do índice h5	136
5.4	ANÁLISE DE AUTOCITAÇÃO NO ÍNDICE H5	138
5.4.1	Taxa de autocitação por país	140
5.4.2	Taxa de autocitação por área	141
5.4.3	Variação entre os índices h5 e h _s 5.....	143
5.4.4	Análise de citação de periódicos de Ciências Agrárias.....	144
5.4.5	Análise de citação de periódicos de Engenharias	145
5.4.6	Análise de citação de periódicos de Ciências Exatas & Naturais.....	146
5.4.7	Análise de citação de periódicos de Ciências da Saúde	146
5.4.8	Análise de citação de periódicos de Artes & Humanidades.....	147
5.4.9	Análise de citação de periódicos de Ciências Sociais.....	148
5.4.10	Análise de citação de periódicos Multidisciplinares.....	149
5.5	DISTRIBUIÇÃO DE CITAÇÕES POR ANO	150
5.5.1	Indícios de manipulação de citações	151
5.5.2	Procedimento em caso de indícios de manipulação de citações	156
6	DISCUSSÃO	158
6.1	LIMITAÇÕES DO GOOGLE SCHOLAR METRICS	158
6.2	COBERTURA DO GOOGLE SCHOLAR METRICS.....	159
6.3	CARACTERÍSTICAS DOS PERIÓDICOS IBERO-AMERICANOS	161
6.4	VALIDADE E ESTABILIDADE DO ÍNDICE H5.....	163
6.5	ANÁLISE DE IMPACTO COM BASE NO ÍNDICE H5	164
6.6	INFLUÊNCIA DE AUTOCITAÇÕES NO ÍNDICE H5.....	165
7	SÍNTESE DO <i>FRAMEWORK GSM@HDATA</i>	168
8	CONSIDERAÇÕES FINAIS	170
9	CONCLUSÃO	174
	REFERÊNCIAS.....	175

1 INTRODUÇÃO

Indicadores de impacto baseados em citação são utilizados na avaliação da produção científica, especialmente no contexto de periódicos indexados em bases de dados. A avaliação por indicadores pressupõe que a citação é um parâmetro objetivo de relevância de pesquisa perante a comunidade científica. A quantificação de citações recebidas e a representação por indicadores são utilizadas para medir o nível de impacto de pesquisas nas disciplinas em que estão inseridas ou possuem aderência temática (WALTMAN, 2016; HOU, 2017).

A análise de citação se tornou a técnica mais utilizada para avaliação de publicações científicas. É um elemento de mensuração de qualidade científica que complementa um sistema de avaliação que se inicia antes da publicação, no processo subjetivo de revisão por pares (THELWALL; WILSON, 2014).

Em nível individual, indicadores de impacto são utilizados para a aferição de desempenho de cientistas, pesquisadores e docentes. Entre as atividades em que os indicadores são utilizados se incluem processos seletivos, progressões funcionais, concessões de bolsas e financiamentos, ingresso em programas de pós-graduação, entre outros procedimentos acadêmicos (DING; LIU; KANDONGA, 2020).

Em nível institucional, indicadores são utilizados em sistemas de avaliação de programas de pós-graduação, de departamentos, de laboratórios e de instituições de ensino e de pesquisa (MINGERS; MEYER, 2017). Aplicações mais amplas incluem análises geográficas, ou seja, descrições do impacto científico de cidades, estados, países e até continentes (YANG; WANG, 2015).

Uma das razões para a popularização de indicadores é que o crescimento exponencial das atividades científicas (PRICE, 1976) torna complexo analisar individualmente toda a produção científica publicada anualmente. Critérios objetivos visam a garantir imparcialidade e minimizar a influência de valores, de opiniões, de preferências e de outros fatores subjetivos no processo avaliativo (ALPERIN; ROZEMBLUM, 2017).

As bases de dados bibliométricas facilitam processos avaliativos em larga escala, especialmente os realizados por agências de fomento, entidades responsáveis por políticas nacionais de avaliação em ciência e tecnologia (VASEN; VILCHIS, 2017). Os indicadores mais utilizados atualmente são o Fator de Impacto (FI) do Journal Citation Reports (JCR), o CiteScore (CS) da Scopus e o SJR do

Scimago Journal & Country Rank (SJ&CR) (WALTMAN, 2016; VERA-BACETA; THELWALL; KOUSHA, 2019).

Os indicadores do JCR e da Scopus funcionam em sistemas fechados, calculados a partir de dados de citação restritos aos periódicos indexados. Essas publicações são consideradas as de maior qualidade e prestígio em suas respectivas disciplinas (MARTÍN-MARTÍN; ORDUNA-MALEA; LÓPEZ-CÓZAR, 2018). Por outro lado, a maioria das publicações indexadas nessas fontes são de origem anglo-saxônica, de idioma inglês e com aderência temática às áreas de ciências exatas e naturais, engenharias e medicina (WALTMAN, 2016).

Por consequência, o alcance geográfico e temático do JCR e da Scopus é limitado. Parte significativa das publicações de países emergentes, de idiomas regionais e das áreas de Ciências Humanas, Sociais e de Artes não são indexadas (MINGERS; MEYER, 2017).

Não seria adequado, portanto, definir os indicadores dessas duas bases de dados como critérios únicos ou preponderantes em sistemas de avaliação multidisciplinar de países de baixa visibilidade (NEDERHOF, 2006; WALTMAN, 2016).

Nos últimos anos, o Google Scholar Metrics (GSM) passou a ser considerado uma fonte alternativa de análise de impacto. A gratuidade de acesso, a interface amigável e o funcionamento simplificado são alguns dos seus aspectos favoráveis. A sua política de indexação reflete o conteúdo acadêmico da web, minimizando o viés comercial. O GSM inclui periódicos de países emergentes, de idiomas e de disciplinas que usualmente têm menor visibilidade em bases de dados internacionais (LÓPEZ-CÓZAR; MARTÍN-MARTÍN, 2019; MARTÍN-MARTÍN; AYLLÓN; ORDUÑA-MALEA, 2014).

O GSM mede o impacto de publicações por meio do índice h5, o indicador de Hirsch (2005), com janela temporal de cinco anos. Esse indicador é calculado com dados de citação de fontes web, previamente mapeadas pelo Google Scholar (GS) (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012, 2013).

No cálculo do índice h5 são incluídos dados de citação de outros tipos de documentos, tais como teses e dissertações, periódicos não indexados e relatórios técnicos. Essa sistemática torna a análise de impacto de citação mais abrangente, não restrita ao contexto de periódicos de alto impacto ou prestígio (KOUSHA; THELWALL, 2007).

Uma das principais vantagens do GSM é possibilitar a análise de impacto de periódicos não avaliados em outras fontes. Essa característica tem aumentado a relevância do GSM no cenário da avaliação científica. Pesquisas vêm sendo realizadas com intuito de compreender o funcionamento, os critérios de indexação e o mecanismo de registro de citações dessa fonte bibliométrica (JACSÓ, 2012a; HARZING, 2013; LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2013; ORDUÑA-MALEA; AYTAC; TRAN, 2019).

Mas o GSM também possui disfunções, as quais colocam em dúvida a sua adequação com fonte oficial de avaliação científica (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012). A falta de transparência da política de indexação (HARZING, 2014), a inexistência de recursos de análise e de extração de dados e a falta de confiabilidade são alguns dos problemas apontados.

A indexação automatizada possui falhas. Entre as falhas reportadas estão a duplicação de entradas por título, a não indexação de publicações, os erros em contagens de citações e a falta de controle e padronização de registros (JACSÓ, 2012a; LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2013).

A não divulgação de uma lista geral das publicações indexadas e a inexistência de recurso de busca por ISSN e por país de origem dificultam análises em larga escala a partir do buscador padrão do GSM (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012, 2013).

A impossibilidade de exclusão de autocitações de periódicos e a possibilidade de manipulação do índice h5 também são inconsistências do sistema (MARTIN-MARTIN et al. 2014), principalmente se o indicador for utilizado como parâmetro de desempenho em processos acadêmicos (LÓPEZ-CÓZAR; ROBINSON-GARCÍA; TORRES-SALINAS, 2014).

É justamente o que vêm ocorrendo nos últimos anos. O índice h5 vem sendo incluído como critério de classificação de periódicos em sistemas de alguns países. Na Espanha, desde 2013, o índice h5 do GSM é um dos critérios de pontuação de impacto e de visibilidade de revistas acadêmicas no sistema de aferição do selo de qualidade FECYT (DE FILIPPO; ALEIXANDRE-BENAVENT; SANZ-CASADO, 2020).

O Brasil, a Colômbia e o México, países com maior produção de periódicos da América Latina, reformularam suas políticas de avaliação, definindo o índice h5 entre os indicadores bibliométricos utilizados como parâmetros de classificação

(ALPERIN; ROZEMBLUM, 2017; COSTA; CANTO; PINTO, 2020).

Na Colômbia e no México, o índice h5 foi definido como um indicador alternativo das políticas nacionais de avaliação de periódicos, respectivamente, o Publindex/Colciencias e o CRMICYT/CONACYT. Nos dois sistemas, o índice h5 se destina à avaliação de periódicos não indexados no JCR e na Scopus, que são classificados nos níveis inferiores (PALACIOS-GÓMEZ, 2017; VASEN; VILCHIS, 2017).

O Brasil ia além, tendo estabelecido o índice h5 como indicador exclusivo ou subsidiário ao CS e ao FI na nova metodologia do sistema Qualis Periódicos (BRASIL, 2021). Em algumas áreas de avaliação, o índice h5 do GSM poderia vir a ser o principal indicador para classificação de periódicos no quadriênio 2017-2020. Mas a judicialização da questão e outros problemas complexos tornaram incerto o uso do GSM no país.

Essa discussão, portanto, parece ser mais pertinente em países de menor visibilidade em bases de dados internacionais. A menor visibilidade é mais evidente em países de idioma não inglês, com preponderância de periódicos das ciências humanas e sociais e que divulgam mais pesquisas de interesse local e regional (ALPERIN; ROZEMBLUM, 2017; VASEN; VILCHIS, 2017).

Resultados preliminares já publicados indicam que o índice h5 é o indicador de impacto mais abrangente, tanto para periódicos do Brasil (COSTA; CANTO; PINTO, 2020), quanto dos demais países da América-Latina e do Caribe (CANTO et al. 2022).

Indicaram, ainda, a utilidade do índice h5 para análise de impacto em estudos geográficos de larga escala, tendo em vista ser um indicador mais representativo de cenários regionais (CANTO et al. 2022).

Ampliada a abrangência geográfica, o universo definitivo desta pesquisa é o conjunto de periódicos Ibero-Americanos indexados no GSM. A lista geral de periódicos é obtida no Latindex, considerando a ampla cobertura da região.

As abordagens iniciais foram limitadas às características das publicações e aos níveis de impacto por país e por área do conhecimento.

A análise definitiva compreende também a validade e a estabilidade do índice h5 e um modelo para estimar outros indicadores a partir do seu resultado. Um enfoque é dado ainda para a questão da autocitação, incluindo o cálculo do índice h5 excluindo autocitações e a mensuração de taxas de autocitação por periódico,

por país e por área.

A viabilização deste estudo foi possível a partir do estabelecimento de uma perspectiva tecnológica. Foi desenvolvida uma ferramenta bibliométrica de busca, extração e análise de dados de citação do índice h5 do GSM.

O estabelecimento de rotinas computacionais de extração e de gerenciamento de dados *web* permitiu superar as limitações técnicas do GSM e investigar essa fonte a partir de uma perspectiva longitudinal (HARZING, 2013) e de larga escala.

1.1 PROBLEMAS DE PESQUISA

No contexto de definição do índice h5 como indicador de impacto alternativo de periódicos, diante da ausência de recursos e de incertezas quanto a sua validade e confiabilidade, é que se insere a seguinte pergunta de pesquisa: como utilizar o índice h5 do GSM para avaliação de impacto de periódicos Ibero-Americanos?

1.2 OBJETIVOS

1.2.1 Objetivo geral

Propor um *framework* para avaliação de impacto de periódicos Ibero-Americanos com base no índice h5 do GSM.

1.2.2 Objetivos específicos

a) Desenvolver uma ferramenta bibliométrica de pesquisa, de extração e de análise de dados do índice h5 do GSM.

b) Identificar as características editoriais, os níveis de impacto e as taxas de autocitação e de crescimento do índice h5 dos periódicos Ibero-Americanos indexados no GSM.

c) Testar a validade e a estabilidade do índice h5 na avaliação de impacto de periódicos Ibero-Americanos.

d) Definir critérios para identificação de indícios de manipulação de citações no índice h5 do GSM.

1.3 JUSTIFICATIVA

A justificativa para a definição do tema de pesquisa encontra fundamento em dois aspectos, um aspecto de caráter geral e outro de caráter específico.

Sob o aspecto geral, o GSM vem sendo objeto de investigação no campo dos estudos métricos da informação desde o seu lançamento. Este interesse se deve à possibilidade de sua utilização como fonte bibliométrica alternativa, especialmente para análise de periódicos não indexados em outras bases de dados.

Embora alguns dos problemas das primeiras versões tenham sido corrigidos, outros ainda persistem no GSM. Logo, é necessário o seu monitoramento a cada atualização para atestar a sua validade e estabilidade na avaliação de periódicos acadêmicos e científicos.

Sob o aspecto específico, se justifica a escolha do tema em razão da nova metodologia do Qualis/CAPES do quadriênio 2017-2020, que define o uso do índice h5 do GSM, de forma exclusiva ou subsidiária ao FI e ao CS, para avaliação de periódicos.

É provável que o índice h5 do GSM seja o indicador mais utilizado para avaliação de impacto no sistema Qualis, considerando a cobertura mais abrangente do GSM em comparação com a Scopus e com o JCR.

Se por um lado a escolha do GSM pode ser considerada positiva, também é motivo de atenção. É preciso identificar violações de padrões internacionais de boas práticas editoriais, sobretudo aqueles relacionados à citação e à autocitação. O GSM é suscetível à manipulação de dados e não conta com um sistema de monitoramento como o de outras bases de dados.

Assim, o desenvolvimento de ferramentas e de técnicas de pesquisa, de extração e de análise de dados de citação com base no índice h5 adquire a relevância acadêmica necessária em estudos de doutorado.

Por fim, a ampliação do universo da pesquisa encontra justificativa na premissa de que os modelos do JCR e da Scopus não possibilitam uma avaliação abrangente dos periódicos Ibero-Americanos.

Assim, o uso do índice h5 para classificação de periódicos deve ser expandir para outros países da região, seguindo os modelos da Espanha, do Brasil, da Colômbia e do México.

O estudo proposto pode contribuir para essa discussão, considerando-se a escassez de dados quantitativos e de resultados de análises em larga escala, tanto no cenário da comunicação científica quanto no do impacto da ciência regional.

1.4 ORIGINALIDADE DA PESQUISA

Uma das condições para validação de teses de doutorado é a originalidade da pesquisa (VAGARINHO, 2019).

A originalidade na ciência

implica desvendamento de dúvidas, apresentação de perspectivas à abordagem de problemas, revelação de resoluções, explicitação de correlações importantes, enfim, descrição de novidades que enriqueçam o conhecimento sobre um assunto. É, portanto, a condição nuclear pela qual os limites da ciência se expandem, devendo ser distinguida como característica de mérito para a publicação (BICAS, 2008, p. 473).

A originalidade, portanto, significa a investigação de algo novo, que nunca foi estudado anteriormente. Pode englobar tanto a forma (procedimentos metodológicos) quando o conteúdo (universo e objeto de pesquisa), de modo que os resultados representem um acréscimo de conhecimento em relação ao tema pesquisado (BICAS, 2008; VAGARINHO, 2019).

Com a finalidade de certificar a originalidade desta tese, foi realizada uma busca sistematizada de literatura seguida de uma síntese dos estudos com relação temática e metodológica com a presente pesquisa.

As fontes foram escolhidas de acordo com os tipos de documentos indexados e a abrangência geográfica do conteúdo. Para a busca de artigos foram selecionadas a Web of Science (WoS) e a Scopus, duas bases de dados referenciais de escopo multidisciplinar amplamente utilizadas em revisões de literatura.

Para a busca de teses e dissertações foram selecionadas a *ProQuest Dissertations and Theses*, de abrangência internacional, e a Biblioteca Digital e Teses e Dissertações (BDTD), de abrangência nacional.

Como resultado, não foram localizados estudos, modelos, metodologias ou ferramentas que, de simultaneamente, 1) tratem da avaliação de impacto de periódicos a partir do índice h5 obtido pelo GSM; 2) utilizem ferramentas, métodos ou técnicas, próprias ou já existentes, para análise em larga escala de dados de citação do GSM, incluindo a identificação de autocitações e cálculo do índice h5 com exclusão de autocitações; 3) proponham modelo, metodologia, sistema ou *framework* para análise de periódicos Ibero-Americanos a partir do índice h5 do GSM.

Diante disso, a presente pesquisa se reveste de originalidade. A descrição dos procedimentos de busca sistemática, de inclusão e exclusão de registros e de

análise de documentos selecionados são apresentados detalhadamente na seção 3.

2 REVISÃO DE LITERATURA

2.1 BASES DE DADOS

As bases de dados se tornaram fontes de informação indispensáveis para o desenvolvimento científico e tecnológico. Foram sendo estruturadas nas últimas décadas, no processo de migração do formato impresso para o eletrônico de documentos (FLADUNG, 2007).

Esse novo formato ampliou as possibilidades de armazenamento, de recuperação e de acesso à informação especializada. As bases de dados se tornaram as principais fontes de conteúdo acadêmico e científico em praticamente todas as áreas do conhecimento humano (GU; BLACKMORE, 2017; GUSEMBAUER, 2019).

Bases de dados internacionais atestam a qualidade do conteúdo disponibilizado, considerando a fixação de rígidas políticas de inclusão e exclusão de publicações. São avaliados critérios de qualidade, regularidade e impacto das publicações indexadas. Entre os critérios principais se incluem a composição internacional do corpo editorial, revisão por pares cega ou duplo cega, políticas de preservação digital e de direitos autorais, periodicidade e regularidade, baixa endogenia, observância de boas práticas acadêmicas, indicadores de impacto, entre outros (DE WINTER; ZADPOOR; DODOU, 2014; MONTANARI; PACKER, 2014).

A indexação em bases de dados garante mais visibilidade para as publicações. Essa maior visibilidade amplia o prestígio, que normalmente se traduz em uma melhora da qualidade dos trabalhos recebidos para avaliação por pares.

Com melhores trabalhos publicados, usualmente há um aumento no número de citações recebidas. O aumento do impacto pode proporcionar a indexação em uma base de dados de maior alcance, ampliando ainda mais a visibilidade da publicação (GONZALEZ-PARDO; REPISO; ARROYAVE-CABRERA, 2020).

As estratégias comerciais adotadas pelas editoras também contribuíram para a expansão das bases de dados internacionais. Editores multinacionais criaram amplos recursos de conteúdo científico relevante e atualizado, que se tornaram de aquisição quase obrigatória para universidades, laboratórios e centros de pesquisa (FLADUNG, 2007; GUEDON, 2010; LARIVIÈRE; HAUSTEIN; MONGEON, 2015).

As bases de dados podem conter diversos tipos de documentos, tais como livros, teses e dissertações, relatórios técnicos, patentes, imagens e vídeos.

Entretanto, as bases de dados de periódicos se consolidaram como as fontes mais utilizadas nas atividades de ciência, tecnologia e inovação. São utilizadas para o registro, a recuperação e a análise da produção científica em níveis geográfico (regional, nacional e internacional) e temático (especializada ou multidisciplinar) (LARIVIÈRE; SUGIMOTO, 2018).

Os recursos de pesquisa, de análise e de extração de dados das bases de dados facilitam as atividades de cientistas, pesquisadores, acadêmicos e profissionais relacionadas à gestão de informação especializada. As principais bases com cobertura internacional e de escopo multidisciplinar são a WoS, a Scopus e o Google Scholar (GS) (WALTMAN, 2016).

Essas bases de dados possuem recursos de análise de citação, cujos dados são utilizados em estudos métricos e na avaliação de impacto de publicações. Os dados de citação são processados e representados na forma de indicadores de impacto, que são atualizados e divulgados anualmente (GUSEMBAUER, 2019).

A avaliação de desempenho científico por meio de indicadores pressupõe que a citação é um indicativo de relevância na ciência. Assim, as publicações mais citadas seriam as mais relevantes em seu campo científico (NEDERHOF, 2006; WALTMAN, 2016). Esse modelo baseado em análise de citação se expandiu na ciência internacional, sendo incorporado em sistemas de avaliação de diversos níveis.

Os indicadores podem ser apresentados em recursos nas próprias bases ou em fontes bibliométricas vinculadas a elas, tais como o JCR, SJ&CR e o GSM (ORDUNA-MALEA; AYTAC; TRAN, 2019; MARTÍN-MARTÍN et al., 2018).

A WoS e a Scopus são bases de dados mantidas por empresas de conteúdo científico, acessíveis total ou parcialmente por assinatura, que têm por finalidade a recuperação de dados referenciais da produção científica para a realização de revisões de literatura e de estudos bibliométricos. As principais funcionalidades são a coleta e a análise de dados de citação, que possibilitam a geração de indicadores de impacto (VANCLAY, 2012; FERRER-SAPENA et al., 2016; WALTMAN, 2016; WALTMAN; TRAAG, 2017; LARIVIÈRE; SUGIMOTO, 2018).

O GS é uma base de dados disponível gratuitamente na forma de um buscador web. Ao contrário das bases de coleção restrita, o GS recupera documentos disponíveis em fontes acadêmicas. O GS também registra dados de citação dos documentos indexados, utilizando-os para a geração de indicadores de

impacto em suas duas extensões voltadas para análise bibliométrica, o GSC e o GSM.

2.1.1 Web of Science

A WoS teve a sua origem na década de 1950, a partir do trabalho de Eugene Garfield, pesquisador da Universidade da Pensilvânia, estado da Filadélfia, Estados Unidos. Aliando os princípios da biblioteconomia e da documentação com técnicas pioneiras da computação, iniciou pesquisas sobre indexação, cabeçalhos de assuntos, citação e referência e outros temas relacionados à documentação com o auxílio de máquinas (GARFIELD, 1952).

Na época, Garfield (1952) já mencionava as vantagens de se realizar o tratamento técnico e temático da informação científica por meio de métodos não convencionais (impressos), mais sim em dispositivos como microfimes, cartões IBM e disquetes. Em 1960, fundou o *Institute for Scientific Information* (ISI), instituição precursora no tratamento automatizado de informação científica.

No âmbito do ISI foram desenvolvidos o *Science Citation Index* (SCI), base de dados de citação, e o FI. A técnica utilizada nesses dois recursos era a de cruzar listas de referências de artigos de periódicos, criando uma base de dados de citação. Isso permitiu calcular a frequência de citações não apenas dos artigos, mas de autores e periódicos, sendo estes na época limitados a aproximadamente 4000 títulos (GARFIELD, 1976).

Foram acrescentadas duas coleções temáticas, o *Social Sciences Citation Index* (SSCI) e o *Arts & Humanities Citation Index* (A&HCI) NEDERHOF (2006). O reconhecimento da relevância da citação no processo de elaboração de trabalhos científicos contribuiu para impulsionar o reconhecimento internacional do ISI (KOUSHA; THELWALL, 2008).

O sucesso do projeto aliado ao crescimento do mercado de informação bibliométrica levou o ISI a ser adquirido em 1992 pela Thompson Scientific, a então divisão científica da gigante editorial Thompson. Em 2016, toda a divisão de propriedade intelectual e de ciência da Thompson passou para o controle da Clarivate Analytics, incluído a WoS e o JCR.

Atualmente, a WoS é formada por uma coleção principal (*WoS Core Collection*) composta por sete coleções: SCIE (1945-presente); SSCI (1956-); A&HCI (1975-); *Emerging Sources Citation Index* (ESCI) (2015-); *Book Citation Index*; *Current*

Chemical Reactions; Index Chemicus.

A WoS inclui outras coleções específicas, como a *Derwent Innovations Index* (1963-presente) de análise de patentes, a *Korean Journal Database* (1980-presente) de periódicos coreanos, a *Russian Science Citation Index* (2005-) de periódicos russos, a *SciELO Citation Index* (2002-) de periódicos de acesso aberto de países ibero-americanos e africanos, a *Conference Proceedings Citation Index - Science* (1990-) e a *Conference Proceedings Citation Index - Social Science & Humanities* (1990-), ambas de trabalhos de eventos (CLARIVATE ANALYTICS, 2022).

As vantagens dessa fonte são a padronização de metadados, que possibilita a pesquisa em diversos campos, a possibilidade de extração de dados com registros completos, incluindo referências. Também se destaca a ampla cobertura temporal, com documentos a partir da década de 1940 (AMARA; LANDRY, 2012; WALTMAN, 2016; HOU, 2017; MARTÍN-MARTÍN *et al.*, 2018; ORDUÑA-MALEA; AYTAC; TRAN 2019).

Além disso, a classificação das publicações em aproximadamente 250 categorias de assuntos facilita a sistematização do conhecimento. Pode-se realizar análises comparativas entre publicações em uma mesma categoria. Também possibilita o estudo dos padrões dos documentos publicados em diferentes categorias (GU; BLACKMORE, 2017; MARTÍN-MARTÍN *et al.*, 2018).

Por outro lado, a WoS sempre foi considerada essencialmente voltada para a ciência produzida por países desenvolvidos, especialmente da Europa e América do Norte. Publicações de outras regiões, como América Latina, Ásia e África são pouco representadas, assim como os temas de interesse predominante dessas regiões (ALPERIN; ROZEMBLUM, 2017).

Com as reiteradas críticas, nos últimos anos a empresa iniciou um processo de ampliação da cobertura da ciência regional, com destaque para a criação do ESCI em 2015 (MACHIN-MASTROMATTEO, 2019).

Outra funcionalidade da WoS é a coleta das citações recebidas pelas publicações indexadas. Os dados são tratados e disponibilizados em duas bases de dados bibliométricas: os dados brutos no InCites e os dados estruturados na forma de indicadores no JCR.

2.1.2 Journal Citation Reports

O JCR é uma base de dados que tem por finalidade mensurar o impacto dos

periódicos indexados na WoS. Pode ser considerada uma extensão voltada especificamente para análises bibliométricas. É utilizada principalmente para análise de impacto em variados contextos, como geográficos e temáticas (GU; BLACKMORE, 2017).

Até 2020, o JCR somente incluía o SCIE e no SSCI, os dois índices principais da WoS. A partir de 2021, foram incluídas o A&HCI e o ESCI, representando a continuidade da ampliação de cobertura tocada pela Clarivate. Mas o cálculo do FI se manteve restrito aos dois primeiros índices. As duas coleções inseridas são avaliadas por meio dos demais indicadores (CANTO et al., 2022).

O JCR possui ampla cobertura temporal, com registros a partir de 1997 para em alguns casos. O escopo temático é inclinado para as ciências exatas, tecnológicas e médicas. Com a ampliação, o total de periódicos passou de aproximadamente 12 mil para 20.994 em 2021 (CLARIVATE, 2022).

Essa ampliação teve como alvos os periódicos regionais e diminuiu a vantagem quantitativa da Scopus (CANTO et al. 2022).

Mesmo assim, o JCR ainda pode ser considerado uma fonte com viés científico. As limitações geográfica e temática sempre foram consideradas as suas principais desvantagens (WALTMAN, 2016). Os periódicos dos Estados Unidos e do Reino Unido representam aproximadamente 50% de toda coleção e os das ciências duras mais de 60%. O SCIE tem mais do que o dobro de periódicos que o SSCI e o A&HCI.

Os pontos fortes do JCR são os recursos, que permitem análises comparativas, extração da dados e relatório, visualização por gráficos e conferência de dados usados nos cálculos de indicadores (LEYDESDORFF; BORNMANN, 2016).

O principal indicador é o FI, que possui versões de dois, de cinco anos e uma sem autocitações. Outros indicadores são o Eigenfactor Score (ES), o Article Influence Score (AIS) e o Jornal Citation Impact (JCI) (KRAMPL, 2019).

As coleções são classificadas em quartis (Q1, Q2, Q3 e Q4) de acordo com o impacto nas categorias de assuntos. Assim, periódicos podem obter quartis diferentes quando classificados e mais de assunto (CAMPANARIO, 2018).

O JCR possui um sistema de monitoramento de padrões de citação e de autocitação, visando identificar manipulações (FRANDSEN, 2007; TAŞKIN et al., 2021). Com base nesse monitoramento, publicações podem ser suspensas se

apresentarem mudanças atípicas nos padrões de citação (HUANG; CATHY LIN, 2012; LIU; FANG, 2020).

Isso reforça o padrão de qualidade das publicações indexadas, mantendo o JCR como destaque no contexto de análise de citação. Os seus maiores concorrentes são a Scopus e o SJ&CR.

2.1.3 Scopus

A Scopus¹ é uma base de dados referencial de escopo multidisciplinar lançada em novembro de 2004 pelo grupo Elsevier. Tornou-se a primeira concorrente da WoS no segmento. Com funcionamento semelhante, se destina à recuperação de documentos científicos e à análise bibliométrica. Também se estrutura a partir de uma coleção restrita, com periódicos selecionados de acordo com critérios de qualidade pré-estabelecidos (WALTMAN, 2016; MARTIN-MARTÍN et al., 2017; ORDUÑA-MALEA; AYTAC, TRAN, 2019).

Até 2020, a Scopus indexava 25.100 publicações, sendo 23.452 periódicos com revisão por pares, 9.8 milhões de trabalhos apresentados em eventos, 852 séries de livros e 294 relatórios comerciais. Abrangia aproximadamente 77 milhões de registros e 1.7 bilhão de referências citadas desde 1970 (ELSEVIER, 2020).

As publicações são classificadas em 27 grandes áreas do conhecimento e 309 categorias de assuntos. A distribuição do conteúdo entre as áreas é de 30,4% de CDS (Medicina, Enfermagem e Odontologia), 28% de Ciências Físicas (Química, Física, ENG), 26,2% de CCS (Psicologia, Economia, Administração e Artes e Humanidades) e 15,4% de Ciências da Vida (Neurociência, Farmacologia e Biologia) (ELSEVIER, 2020).

O principal indicador de impacto calculado é o CS. A partir dele, as publicações são classificadas em quartis (Q1, Q2, Q3 e Q4) de acordo com as categorias de assuntos. Os demais indicadores são o SJR, o SNIP e o índice h (esse em nível de publicação, autor e conjunto de documentos). Outras métricas em nível de publicação são o total de citações, total de documentos e porcentagem de documentos citados, todos com o período de quatro anos do CS (WALTERS, 2017; JAMES et al., 2019).

A Scopus padroniza metadados dos campos “tipo de documento”, “resumo”,

¹ <https://www.scopus.com>

“palavras-chave”, “referencias citadas”, “afiliação”, “autor” e “ORCID”. Permite a exportação de dados para gerenciadores bibliográficos, tais como o Mendeley, RefWorks e EndNote, e possui uma ferramenta de visualização (ELSEVIER, 2020).

A Elsevier divulga dados quantitativos e estatísticos, alterações em políticas de indexação e de permanência. Os critérios básicos de indexação de periódicos são ISSN ativo, revisão por pares, conteúdo legível internacionalmente (escrita em romano e títulos e resumos em inglês), políticas de ética em pesquisa e responsabilização por má conduta (ELSEVIER, 2020).

Outros aspectos são política editorial, diversidade geográfica do comitê editorial e dos autores, editor permanente, regularidade e pontualidade, adequação entre escopo e conteúdo publicado, artigos com citação em publicações indexadas na própria base, versão em inglês do sítio entre outros (ELSEVIER, 2020).

Também são aplicados critérios de permanência. Os indicadores individuais são comparados com as médias das áreas do conhecimento em que estão classificadas. Os principais parâmetros são o impacto (citações e autocitações), a produtividade (número de artigos publicados) e o uso (acesso ao texto completo e leitura de resumo) (ELSEVIER, 2020).

As autocitações não podem ter uma taxa igual ou superior a duas vezes a média da área. A taxa média de citação, o CS, o número de artigos publicados, de visualização de resumos e de acessos aos textos completos devem ser superiores a metade da média da área (ELSEVIER, 2020).

A política de indexação da Scopus era considerada mais abrangente que a da WoS, contendo um volume maior de publicações regionais, das áreas de ciências humanas e sociais e de idiomas não inglês (CHINCHILLA-RODRÍGUEZ et al., 2016). Até 2015, foi estimado que 97% dos periódicos indexados na WoS estavam incluídos na Scopus, sendo aquela considerada praticamente um subconjunto desta (WALTMAN, 2016). Essa diferença de cobertura foi diminuindo nos últimos anos, especialmente com a criação do ECSI e a sua recente inclusão no JCR (CANTO et al., 2022).

Uma vantagem da Scopus é o acesso aos dados bibliométricos. É possível extrair os dados de toda coleção no período de dez anos (2011-2020) em um arquivo único, incluindo indicadores e características editoriais. Esse mesmo nível de acesso não é possível nas fontes da Clarivate, que limitam o número de registros na extração.

Análises com base na coleção da Scopus também podem ser obtidas por meio de uma outra fonte bibliométrica: o SJ&CR.

2.1.4 SCImago Journal & Country Rank

O SJ&CR é uma base de dados bibliométrica desenvolvida em 2007 pelo grupo de pesquisa espanhol SCImago. As análises são realizadas a partir da coleção e dos dados de citação da Scopus. Assim, as mesmas características de cobertura se aplicam a ambas (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2013).

O principal indicador do SJ&CR é o SJR, que corresponde à média de citações recebidas por artigos do periódico em um período de três anos, atribuídos pesos distintos às citações, variáveis de acordo com o impacto do periódico citante. As autocitações são limitadas a um terço do total de citações recebidas. Com isso, o SJR pode ser considerado um indicador de prestígio ou de influência, e não apenas de impacto (GONZÁLEZ-PEREIRA; GUERRERO-BOTE; MOYA-ANEGÓN, 2010).

O SJ&CR disponibiliza outros indicadores, tais como o índice h, a média de citações por artigo em dois anos de análise (similar ao FI), o total de documentos e o número total de citações, entre outras métricas.

Não tendo coleção própria, o SJ&CR se destaca por seus recursos de análise, de visualização e de extração de dados. Permite comparações entre periódicos com base no escopo geográfico (país, região e continente). Possui, ainda, uma ferramenta de visualização de dados (VizTools) e filtros de identificação de publicações de acesso aberto e de solapamento entre outras fontes (SciELO e WoS) (SCIMAGO, 2020).

Como nem todos esses recursos estão disponíveis na Scopus, o SJ&CR se tornou uma fonte amplamente utilizada em pesquisas bibliométricas, especialmente as de larga escala devido à facilidade de extração de dados (MACHIN-MASTROMATTEO, 2019).

2.1.5 SciELO

O SciELO é uma base de dados de periódicos eletrônicos criada no Brasil em 1998. Foi desenvolvida por meio de uma parceria entre a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e o Centro Latino-Americano e do Caribe de Informação em CDS (BIREME/OPAS/OMS).

Na época de criação, o programa tinha dois objetivos gerais. Primeiro, o de

desenvolver *expertise* para migração de periódicos brasileiros para o formato eletrônico, dentro de padrões internacionais de qualidade. Segundo, o de aumentar a visibilidade, o uso e o impacto das publicações, auxiliando o desenvolvimento da pesquisa nacional (PACKER; MENEGHINI, 2014).

O êxito do projeto levou a sua expansão, primeiro para o Chile, depois para outros países da América-Latina. Foi criada a rede SciELO, um programa de cooperação internacional baseada em uma infraestrutura comum de periódicos de acesso aberto (BUSTOS-GONZÁLEZ; PALMA; 2014).

Desde 2013, a rede SciELO conta com coleções de 15 países Ibero-Americanos e mais a da África do Sul. Os princípios gerais, as metodologias e as tecnologias do programa são comuns a todos. Cumpre a cada país a gestão de suas coleções, tanto de recursos, quanto de questões editoriais e de indexação (PACKER; COP; SANTOS, 2014).

O estabelecimento dos requisitos técnicos e editoriais da rede contribuiu para que o alcance de padrões internacionais de comunicação científica. Também possibilitou o aumento de visibilidade mesmo para publicações com escopo regional e de idioma não inglês. A maioria dos periódicos mantiveram os idiomas regionais de publicação, predominantemente o português e o espanhol.

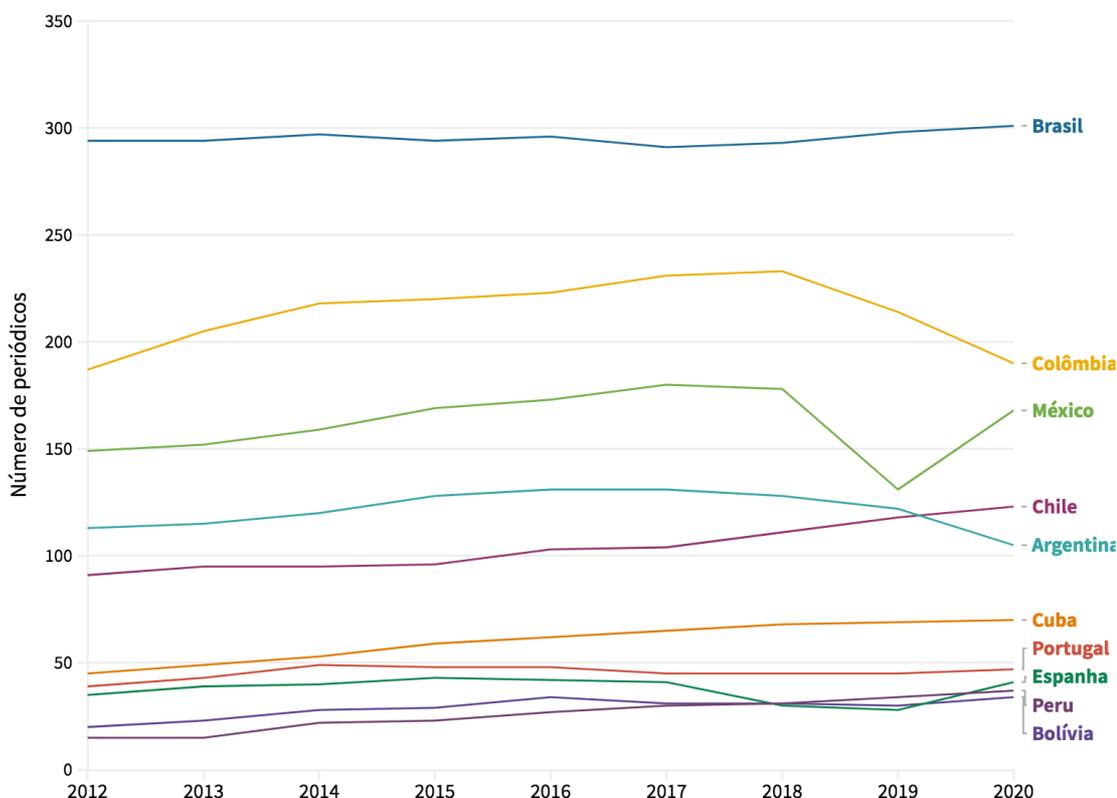
Um fato relevante para a expansão internacional do projeto foi a criação do SciELO Citation Index, uma coleção integrada à WoS (VÉLEZ-CUARTAS; LUCIO-ARIAS; LEYDESDORFF, 2016). Isso permitiu o uso dos recursos disponíveis na WoS a partir da coleção do SciELO (MINNITI; SANTORO; BELLI, 2018). Esta incorporação acirrou a competição por uma melhor cobertura regional entre a WoS e a Scopus (VÉLEZ-CUARTAS; LUCIO-ARIAS; LEYDESDORFF, 2016).

Em razão desse viés regional, há quem defenda que a indexação no SciELO deveria ser mais valorizada no sistema Qualis (CUETO, 2019), ao invés de priorizar bases internacionais. Essa valorização ocorre em alguns países, nos quais a indexação no SciELO é um critério de qualidade e visibilidade em sistemas de avaliação, tais como o da Argentina, do México (VASEN; VILCHIS, 2017) e do Chile.

Pode ser afirmar que um empecilho para a definição do SciELO como critério de visibilidade em sistemas de avaliação é o lento ritmo de crescimento de sua coleção.

O Gráfico 1 mostra o que número de periódicos no SciELO nas dez maiores coleções por país se mantém praticamente estável no período 2012-2020.

Gráfico 1 – Número de periódicos indexados no SciELO por país (2012-2020)



Fonte: SciELO (2022).

O total de periódicos da rede no ano de 2012 era de 1.051; alcançou o número máximo em 2017 com 1.308; e encerrou o período com 1.215 periódicos indexadas. Esses números representam um crescimento anual médio próximo de 1% (PACKER et al. 2019), ritmo que provavelmente não acompanha a evolução das publicações dos países participantes.

Além disso, nenhum país foi incorporado a rede SciELO nesse período (PACKER; COP; SANTOS, 2014), o que indica uma estagnação na expansão geográfica.

Para uma visão mais precisa, seria necessário comparar os dados de ingresso e de exclusão de publicações. O SciELO possui uma taxa não desprezível de desindexação, o que pode limitar a sua expansão quantitativa.

Mas de forma geral, os dados sugerem que o projeto não consegue acompanhar o mesmo ritmo de crescimento (exponencial) das atividades científicas (MERTON, 1976). Até porque os últimos dez anos foram os de maior crescimento da produção e da indexação de periódicos Latino-Americanos em bases de dados internacionais (COLLAZO-REYES, 2014; VÉLEZ-CUARTAS; LUCIO-ARIAS;

LEYDESDORFF, 2016; CRESPO-GASCÓN; TORTOSA; GUERRERO-CASADO, 2019; IBÁÑEZ-MARTI, 2018).

Outrossim, não é razoável afirmar que o SciELO é o parâmetro principal de qualidade de periódicos dos países participantes, haja vista que muitas publicações não indexadas foram incluídas na WoS e na Scopus nos últimos anos (CANTO et al., 2022).

De toda forma, o SciELO se tornou um modelo internacional de acesso aberto à informação científica, em um cenário dominado por editores comerciais e por bases de dados por assinatura.

Manter toda a infraestrutura do projeto requer fluxo de recursos, que nem sempre estão disponíveis para patrocinar uma expansão no mesmo ritmo do crescimento da ciência.

2.1.6 Latindex

O Latindex foi criado em 1995 na Universidade Nacional Autônoma do México. Dois anos depois se tornou uma rede de cooperação regional baseada em um diretório de periódicos da América Latina e no Caribe. Em 1999, o Latindex ampliou a abrangência geográfica, passando a incluir publicações da Espanha e de Portugal. Também passaram a ser indexadas publicações de outros países com conteúdo dedicado à região (LATINDEX, 2022).

O ingresso é autorizado somente para periódicos com conteúdo acadêmico, sendo vedada publicações empresariais de divulgação de produtos e serviços. Os demais requisitos técnicos são mínimos: estar em vigência, com um número publicado e ISSN ativo ou em trâmite (LATINDEX, 2022).

O sistema de avaliação de períodos é composto por 38 características divididas em cinco conjuntos: básicas, de apresentação, de gestão e política editorial, de conteúdo e características das revistas *online* (LATINDEX, 2022).

As sete características do conjunto das básicas são obrigatórias. Depois disso, cada característica cumprida dos demais conjuntos valem um ponto. A partir do alcance de 30 pontos, o periódico ingressa no Catálogo 2.0, ou seja, a coleção dos periódicos de maior qualidade editorial (LATINDEX, 2022).

Assim como o SciELO, o Latindex ajudou a consolidar a ciência latino-americana por meio da construção de uma infraestrutura básica de editoração de periódicos (ALPERIN; ROZEMBLUM, 2017). Também possibilitou a centralização de

informações e o estabelecimento de padrões editoriais a partir de critérios internacionais de qualidade (GONZALEZ-PARDO; REPISO; ARROYAVE-CABRERA, 2020).

Considerando essa relevância no cenário Ibero-Americano, o Latindex é utilizado com frequência em estudos sobre a ciência da região (GONZALEZ-PARDO; REPISO; ARROYAVE-CABRERA, 2020; ALPERIN; ROZEMBLUM, 2017; CRESPO-GASCÓN; TORTOSA; GUERRERO-CASADO, 2019; CAMACHO et al., 2019).

O Latindex também passou a ser fixado como parâmetro de qualidade e de visibilidade em sistemas de classificação de periódicos de alguns países, tais como a Argentina, o Brasil, a Costa Rica e o Chile (VASEN; VILCHIS, 2017).

2.1.7 Google Scholar

Lançado em 1998, o Google logo se tornou o principal sistema de busca da *Internet*. Atualmente é uma espécie de porta de entrada para a rede mundial de computadores. Por meio do monitoramento de padrões de uso, a Google detectou que parte significativa das buscas tinha por finalidade a recuperação de informação acadêmica (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009). A empresa observou que o padrão de busca desse tipo de usuário era distinto, o que despertou o interesse da equipe de desenvolvimento da empresa (ORDUNA-MALEA; AYTAC; TRAN, 2019).

O Google empreendeu na filtragem de conteúdo para ingressar de forma mais incisiva no mercado acadêmico (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009). Um projeto pioneiro se iniciou em 2000 por meio de cooperação com a *CrossRef* e mais 45 editores. O projeto conjunto teve por objetivo o desenvolvimento de um sistema coletor de citações de documentos acadêmicos em formato eletrônico (JACSÓ, 2010).

Os resultados desse projeto culminaram no desenvolvimento de sistema de indexação de documentos acadêmicos e científicos visando a pesquisa e a recuperação de dados referenciais (título, autor, resumo, palavras-chave, etc.). Esse sistema foi lançado em 2004 sob a denominação de Google Scholar.

Rapidamente o GS se tornou referência de acesso aos textos completos de artigos, teses, dissertações, relatórios, anais de eventos e outros documentos com

estrutura e origem acadêmica (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009; MARTÍN-MARTÍN *et al.*, 2018).

Diferentemente de bases de dados de conteúdo restrito, o GS realiza a indexação automatizada de documentos em fontes acadêmicas (HARZING, 2013). As fontes são domínios especializados, tais como sítios de universidades e instituições de pesquisa, repositórios, portais de periódicos, bases de dados abertas, catálogos de bibliotecas e *blogs* de pesquisadores (JACSÓ, 2011; GUSENBAUER, 2019).

O processo de indexação do GS ocorre em três etapas: a compilação de fontes, a localização de documentos e indexação. Para que documentos sejam alcançados pelo robô indexador (*Googlebot*) são necessários alguns requisitos. Visando facilitar a estruturação de sítios acadêmicos, esses requisitos estão disponíveis no espaço de ajuda do sistema (LÓPEZ-CÓZAR; ORDUÑA-MALEA; MARTIN-MARTIN, 2019).

O acesso deve ser autorizado, sem bloqueios de segurança contra raspagem de conteúdo, e os metadados e os arquivos dos documentos devem estar em um local de fácil identificação. Quando o acesso é permitido, a indexação alcança a lista de referências dos documentos. A coleta dos dados referenciais aumentou a funcionalidade do GS, pois possibilitou a realização de análise de citação, atividade antes restrita a poucas bases de dados (LÓPEZ-CÓZAR; ORDUÑA-MALEA; MARTIN-MARTIN, 2019).

O sistema de busca é baseado no tradicional buscador Google, acessível por qualquer pessoa conectada à Internet, independente de assinatura, identificação ou instalação adicional. Essa característica diferenciou o GS dos demais serviços dessa natureza, que normalmente são de acesso restrito.

Enquanto bases de dados fechadas acirravam a competição por receitas decorrentes do emergente mercado acadêmico, o GS se tornou alternativa, especialmente para usuários, bibliotecas ou instituições que não podem arcar com os custos de assinatura (JACSÓ, 2010).

O aumento do uso do GS despertou interesse da comunidade acadêmica, dando origem a uma vasta literatura (KOUSHA; THELWALL, 2008; HARZING, 2014; JACSÓ, 2011; GUSENBAUER, 2019). Sob uma perspectiva histórica, três fases podem ser observadas em pesquisas sobre o GS: a primeira de curiosidade e ceticismo com a ferramenta; a segunda de investigação sistemática e crítica em

relação às limitações; e a terceira e atual fase de otimismo com o aprimoramento do sistema e o potencial de uso (MARTIN-MARTIN et. al., 2017).

2.1.7.1 Características do Google Scholar

A facilidade de uso é considerada um dos pontos fortes do GS (COTHRAN, 2011; PITOL; DE GROOTE, 2014). A pesquisa de documentos pode ser realizada de forma anônima, sem necessidade de *login*. Há opção de busca avançada, incluindo operadores booleanos. A busca pode ser feita em campos básicos documento (autor, título, fonte, data ou texto).

É possível filtrar e classificar os resultados por data ou relevância, mantendo registros em todos ou em idiomas específicos. Há opções para inclusão de patentes e citações nos resultados. Por outro lado, não estão disponíveis recursos de pesquisa presentes em outras bases de dados, o que pode afetar a precisão em uso avançado (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2013).

Os documentos indexados pelo GS estão nos principais formatos textuais utilizados em âmbito acadêmico web, tais como o PDF, o html, o ppt, o word e o *postscript* (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009). Em registros que há acesso ao texto completo, o formato de visualização é semelhante. Mas nesse caso o link remete à fonte do documento, estando o acesso sujeito às restrições (TORRES-SALINAS, RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009).

O GS permite a visualização dos registros das citações coletadas. São mostrados os dados do documento citante, sem *link* para o documento citado (TORRES-SALINAS, RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009).

Uma das principais disfunções apontadas nas primeiras versões do GS era a apresentação de registros duplicados, na proporção em que trabalhos iguais eram indexados em várias fontes. Esse problema gerava distorções, tanto em âmbito da recuperação quanto no da análise citação. As citações eram divididas entre as versões de um mesmo documento (PITOL; DE GROOTE, 2014).

Esse problema foi corrigido a partir de um mecanismo de unificação de documentos e de citações. Foi inserido um comando de visualização de todas as versões, mantendo o somatório das citações no registro principal (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009; PITOL; DE GROOTE, 2014).

Para uso de alguns recursos do GS é necessária a identificação de usuário. São utilizados os dados de acesso dos demais serviços do Google. Os recursos

restritos ficam na “Minha biblioteca”, espaço de personalização de configurações. É possível configurar o envio de alertas, enviar resultados para listas, exportar registros e outros serviços mínimos desse tipo.

2.1.7.2 Tamanho e crescimento do Google Scholar

O GS possui ampla cobertura, resultado de sua política de indexação sem os critérios de qualidade usuais em outras bases (COTHRAN, 2011). Além de detalhes sobre o processo de indexação, pouco se sabe sobre o tamanho e o crescimento dessa fonte (HARZING, 2014; GUSENBAUER, 2019; LÓPEZ-CÓZAR; ORDUÑA-MALEA; MARTÍN-MARTÍN, 2019).

A falta de transparência foi criticada, especialmente nos primeiros anos do GS. Nos anos seguintes foram observados alguns ajustes. O primeiro foi a ampliação dos idiomas, que era restrito ao inglês. A recuperação de documentos em português, por exemplo, teve início somente em janeiro de 2006. Atualmente, o sistema abrange múltiplos idiomas, inclusive de alfabeto não romano (LÓPEZ-CÓZAR; ORDUÑA-MALEA; MARTÍN-MARTÍN, 2019).

Uma segunda evolução no GS ocorreu no acesso aos metadados, que era bloqueado por alguns editores comerciais. Isso resultava em lacunas temáticas. A *American Chemistry Society*, por exemplo, não dava acesso, prejudicando a recuperação da área de química. Isso também ocorria com Elsevier, a Science e a Nature.

Acordos com as editoras resolveram a questão (TORRES-SALINAS, RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009) e atualmente o GS abrange os principais editores de todas as áreas do conhecimento (HARZING, 2014).

O acesso à texto completo ainda depende da política dos editores. Nos casos em que o usuário não possui acesso, o GS apresenta apenas dados referenciais. O acesso ao texto completo vai depender de permissão, como as fornecidas por meio de VPN's de redes universitárias (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009).

Considerando-se que o GS reflete parte expressiva do conteúdo acadêmico da Internet, é impossível calcular com precisão o número total de documentos indexados. Estimativas apontaram algo em torno de 170 milhões de registros em 2013 (ORDUNA-MALEA; LÓPEZ-CÓZAR, 2015), 389 milhões de registros em 2018

(GUSENBAUER, 2019) e 400 milhões em 2019 (ORDUNA-MALEA; AYTAC, TRAN, 2019).

A taxa de crescimento também é desconhecida. Uma estimativa é que esteja na faixa de 30 milhões de novos registros indexados anualmente. Orduña-Malea e López-Cózar (2014) calcularam a quantidade de documentos indexados no GS de acordo com os países de origem das fontes do conteúdo, conforme mostra a Tabela 1.

Tabela 1 – Número de documentos indexados por país no Google Scholar

País	Agosto 2010	Novembro 2012	Março 2013	Abril 2020*
China	7.520.000	18.900.000	30.700.00	52.400.000
EUA	7.873.000	14.531.000	16.019.000	N/A
Japão	1.873.000	10.100.000	10.400.000	22.100.000
França	2.820.000	4.430.000	4.210.000	7.150.000
Espanha	907.000	3.140.000	2.990.000	4.410.000
Brasil	1.440.000	2.170.000	2.320.000	4.100.000
Coréia do Sul	481.000	1.630.000	1.720.000	4.480.000
Reino Unido	430.000	4.230.000	1.200.000	6.620.000
Alemanha	684.000	1.060.000	1.030.000	2.000.000
Itália	308.000	723.000	798.000	1.870.000
Holanda	219.000	717.000	774.000	1.440.000

Fonte: Orduña-Malea; López-Cózar (2014). * Dados obtidos a partir da mesma técnica utilizada pelos autores.

Os dados foram atualizados em 2020 a partir da mesma metodologia (ORDUÑA-MALEA; LÓPEZ-CÓZAR, 2014), visando a ampliar a análise temporal.

O Brasil registrou 90% de crescimento no número de documento em sete anos, o que corresponde a um crescimento anual entre 10% e 15%. Esses dados definem o GS como a maior fonte acadêmica da atualidade.

A importância de se estimar esses números reside não apenas em conhecer o próprio GS, mas de forma reflexa conhecer o tamanho e o crescimento de toda informação acadêmica disponível eletronicamente (GUSENBAUER, 2019).

Na medida em que a quantidade de documentos, a variedade de idiomas e de disciplinas abrangidas são aspectos positivos do GS, algumas falhas no controle dos metadados são evidentes.

Há falta de padronização dos registros bibliográficos, ocorrência de erros de indexação e de apresentação de dados. Não é possível identificar quais fontes são revisadas por pares, o que compromete a qualidade e a confiabilidade do conteúdo.

Em razão disso, o uso irrestrito do GS para fins científicos pode ser desaconselhado. Seria mais indicado o seu uso como uma fonte acadêmica complementar.

2.1.7.3 Análise de citação no Google Scholar

Outro recurso de destaque no GS é a coleta de dados de citação. Essa funcionalidade confere maior utilidade ao sistema, tendo em vista que a análise de citação é um dos pilares da avaliação científica e dos estudos métricos (MARTÍN-MARTÍN *et al.*, 2018).

A coleta de dados ocorre nas referências dos documentos. A partir disso, algumas pesquisas buscaram identificar as fontes das citações, isto é, quais são os tipos de documentos citantes. Os resultados foram comparados com os padrões de outras bases de dados (KOUSHA; THELWALL, 2007, 2008; MARTÍN-MARTÍN *et al.*, 2018).

Kousha e Thelwall (2007) analisaram citações em 1650 artigos de 108 periódicos de acesso aberto publicados em 2001 em várias disciplinas, comparando o registro pelo GS e pela WoS. Constataram que o GS coleta significativamente mais citações do que a WoS nas áreas de Ciência da Computação e de Ciências Sociais, especialmente as oriundas de anais de eventos.

Em estudo posterior, Kousha e Thelwall (2008) analisaram citações registradas no GS em relação a 39 periódicos de acesso aberto e que não foram registradas por outras bases de dados. Constataram que de um total de 5.589 citações, 35% eram provenientes de artigos de periódicos, 25% de anais, 22% de trabalhos depositados em repositórios e 8% de teses.

Martín-Martín *et al.* (2018) analisaram 2.448.055 citações feitas a 2.299 artigos em inglês de 252 categorias de assuntos publicados em 2006. Constataram que o GS registrou o maior percentual de citações em todas as áreas (entre 93%-96%), seguido da Scopus (35%-77%) e da WoS (27%-73%). Os menores percentuais das duas últimas fontes são de áreas de Ciências Humanas e Sociais.

O GS registrou quase 95% das citações registradas pela WoS e 92% das registradas pela Scopus. A maioria das citações registradas apenas pelo GS não

eram de periódicos, mas de teses, livros, trabalhos de anais e materiais não publicados. Além disso, o GS é a base que mais coleta citações de documentos em idioma que não o inglês.

2.1.7.4 Erros de indexação no Google Scholar

Nos primeiros anos foram observados problemas de padronização de entradas por autor no GS. Por vezes o sistema indexava termos do corpo do texto, como “*introducción*”; “*estado de la cuestión*”, “*índice*” e “*contenido*” registrado em documentos em espanhol (TORRES-SALINAS; RUIZ-PÉREZ; LÓPEZ-CÓZAR, 2009) e “*Payment Options*”, “*Please Login*”, “*Strategic Plan*”, “*I Background*”, “*II Objectives*”, “*Forgot Password*”, “*I Introduction*” e “*Subscribe*” em inglês (JACSÓ, 2010).

Como consequência, os respectivos não eram recuperados por autor e as citações não eram computadas, prejudicando o resultado de indicadores de impacto (JACSÓ, 2010).

Embora o processamento automatizado de grandes quantidades de dados implique, naturalmente, em um percentual de erros, Jacsó (2010) entende que os mecanismos de correção do Google eram ineficientes, sem o tratamento adequado de metadados.

Parte dos problemas parecem ter sido sanados nas edições posteriores do GS (JACSÓ, 2012b). Mas os erros do sistema de indexação automática ainda são a principal razão da falta de credibilidade, não apenas do GS, mas também de suas duas extensões alimentadas pelos mesmos dados, o GSC e o GSM.

2.1.8 Google Scholar Citations

O Google Scholar Citations (GSC)² é uma base de dados bibliométrica lançada em novembro de 2011 visando a identificar e monitorar citações recebidas em trabalhos acadêmicos. É uma ferramenta vinculada ao GS que consiste na criação de um perfil pessoal do autor, com as áreas de pesquisa de seu interesse (ORTEGA, 2015; TEIXEIRA DA SILVA; DOBRÁNSZKI, 2018).

O GSC rastreia os documentos publicados e as citações recebidas pelo autor. É apresentado um painel de resumo, contendo indicadores básicos de impacto. São

²<https://scholar.google.com/citations>

calculados o total de citações recebidas e a distribuição anual. Os indicadores de impacto disponíveis são o índice h e o índice i10, que representa o número de documentos com no mínimo 10 citações. É possível inserir coautores do titular do perfil e inserir ou excluir documentos manualmente, corrigindo eventuais erros da indexação automática.

Os mesmos aspectos positivos atribuídos ao GS podem ser atribuídos ao GSC. Gratuidade, facilidade de uso, indexação automática de documentos e cálculo de indicadores de impacto são as principais vantagens mencionadas (AHMAD et al. 2021; TEIXEIRA DA SILVA, 2018).

Mas o GSC pode apresentar as mesmas inconsistências do GS, especialmente erros de contagem de citações e de atribuição de autoria. Teixeira da Silva e Dobránszki (2018) relatam que a ferramenta deve ser monitorada constantemente pelo seu titular, visando à identificação e à correção de erros. Como os perfis são públicos, o monitoramento e a correção de erros evitam suspeitas sobre o seu próprio titular.

Foram identificados, ainda, problemas de confiabilidade no GSC. Um deles é a possibilidade de “enganar” o sistema de indexação, por meio da criação de documentos e citações falsas (JACSÓ, 2011).

Utilizando-se tais técnicas é possível aumentar-se o número de citações recebidas e, conseqüentemente, os indicadores de impacto de autores, periódicos e até instituições. Jacsó (2011) alerta para a possibilidade de alterar documentos com “palavras brancas”, ou seja, por meio de textos invisíveis para humanos, mas legíveis pelo robô indexador.

O GSC também reflete os problemas de indexação de autor em contagens de citações. Teixeira da Silva (2018) constatou que o GS, ao atribuir publicações ou citações indevidas, pode aumentar artificialmente o índice h do titular do perfil.

Em uma pesquisa envolvendo o monitorando de um autor altamente produtivo durante o período de quatro meses, foram indexadas 325 citações indevidas, ocasionando um aumento do índice i10 de 282 para 290 documentos (TEIXEIRA DA SILVA, 2018).

Em estudo com 50 autores altamente citados na Ciência da Informação, foram identificadas inconsistências em mais da metade dos dados apresentados no GS. As implicações práticas do estudo são que autores devem evitar práticas antiéticas; que editores devem adotar técnicas de monitoramento; que instituições

devem tomar precauções ao usar o índice h em processos de avaliação e que o Google deve melhorar o funcionamento de seus sistemas (AHMAD et al. 2021).

Esse estudo traz à tona ainda novas formas de manipulação do índice h e do índice i10 do GS por meio de falsas alegações de autoria (AHMAD et al. 2021).

2.2 GOOGLE SCHOLAR METRICS

Em abril 2012 foi lançado o *Google Scholar Metrics for Publications*³, extensão voltada para mensuração de impacto de publicações a partir dos dados do GS. Esse lançamento, que já estava sendo previsto desde a criação do GSC, reposicionou a empresa no mercado acadêmico, aumentando a concorrência com os demais serviços de avaliação da produção científica (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2013; JACSÓ, 2012a).

O GSM mantém a mesma política de acesso do GS. O reconhecimento da marca aliado à desnecessidade de instalação de software ou *plug-in* são fatores que favorecem a sua utilização. A interface e funcionalidades seguem a tendência minimalista do GS e os recursos oferecidos são limitados em comparação com outras bases de dados (JACSÓ, 2012a).

O indicador de impacto é o índice h5, ou seja, o índice h de Hirsch (2005) com um período de cinco anos, fixado conforme a data de atualização do sistema. Na primeira versão, o período de análise foi de 2007 a 2011, com citações coletadas até data de lançamento, em abril de 2012 (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012).

Em novembro de 2012 o sistema recebeu uma primeira atualização, incluindo as citações desse período (ORDUNA-MALEA; LÓPEZ-CÓZAR, 2013). Desde então, as atualizações anuais passaram a ser no mês de julho. O período atual engloba trabalhos publicados entre 2016 e 2020 e citações coletadas até julho de 2021.

Além do índice h5, o sistema calcula a mediana h5, uma métrica que corresponde à mediana do número de citações recebidas pelos artigos do índice h5 (h5-core). A função da mediana parece ser a de desempatar publicações com o mesmo índice h5 (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012; JACSÓ, 2012a).

Há três formas de explorar as publicações indexadas: visualizar listas por idioma, visualizar listas por categorias (grandes áreas) e subcategorias de assunto (áreas de pesquisa) e pesquisar por título ou palavra do título da publicação.

³https://scholar.google.com/citations?view_op=top_venues&hl=en

É possível listar as publicações de maior impacto por doze idiomas: inglês, chinês, português, espanhol, alemão, russo, francês, japonês, coreano, polonês, ucraniano e indonésio. As listas contêm as 100 publicações de maior impacto por idiomas (ORDUNA-MALEA; LÓPEZ-CÓZAR, 2014).

As publicações de maior impacto podem ser listadas em oito categorias (grandes áreas) e 298 subcategorias de assuntos (áreas de pesquisa), mas nestes casos, somente aquelas em inglês. Mas as listas por categorias por assuntos são de apenas 20 publicações.

Algumas subcategorias podem estar presentes em mais de uma categoria de assunto. Biblioteconomia e Ciência da Informação (*Library and Information Science*), por exemplo, está inserida nas categorias de Engenharia e Ciência da Computação (*Engineering & Computer Science*) e de Sociais Aplicadas (*Social Applied*). As vinte publicações desta subcategoria, portanto, são replicadas nas duas categorias mencionadas (LÓPEZ-CÓZAR; ORDUÑA-MALEA; MARTIN-MARTIN, 2019).

A Tabela 2 mostra as categorias e o respectivo número de subcategorias de assuntos no GSM.

Tabela 2 – Categorias e subcategorias de assuntos no Google Scholar Metrics

Categoria	Nº subcategorias
<i>Business, Economics & Management</i>	16
<i>Chemical & Materials Science</i>	17
<i>Engineering & Computer Science</i>	56
<i>Health & Medical Science</i>	69
<i>Humanities, Literature & Arts</i>	26
<i>Life Sciences & Earth Sciences</i>	39
<i>Physics & Mathematics</i>	23
<i>Social Sciences</i>	52
Total	298

Fonte: GSM (2021).

A busca por publicações pode ser efetuada por título ou termo do título, digitado em campo único, sem opção de busca avançada. Os resultados são limitados a 20 registros (CABEZAS-CLAVIJO; LÓPEZ-CÓZAR, 2013). A falta de um recurso de busca por ISSN é uma falha evidente, considerando as inúmeras variações possíveis para o título de uma publicação (JACSÓ, 2012a).

O próprio o menu “ajuda” reconhece que foram detectadas 959 variações para título o periódico “*Proceedings of the National Academy of Sciences (PNAS)*”. A

recomendação é que diante que uma busca sem resultados, sejam feitas buscas por abreviaturas ou variações do título (GSM, 2020).

Essa recomendação indica que o sistema não é pensado para uso em larga escala, ou seja, a partir de uma lista com grande número de publicações. Casos de homonímia ou similaridade entre termos do título e de suas variações certamente afetariam a precisão dos resultados (REGENBRECHT; LANGLOTZ, 2015).

A criação de um filtro de busca por ISSN, que é básico em outras bases de dados, seria uma melhoria expressiva no GSM.

Para cada publicação são identificados os trabalhos citantes dos artigos do índice h5. Os registros das citações são apresentados no mesmo formato tradicional do GS, contendo as informações básicas de autor, título do trabalho e da publicação e o ano (GSM, 2020).

A inclusão das publicações no GSM é automática. Os únicos critérios são que a publicação tenha pelo menos 100 documentos publicados no período e que tenha recebido pelo menos uma citação, ou seja, com índice h5 > 0 (GSM, 2020).

Além de periódicos, o GSM indexa anais de eventos de Ciência da Computação e de Engenharia Elétrica, bem com repositórios de pré-prints (LÓPEZ-CÓZAR; ORDUÑA-MALEA; MARTIN-MARTIN, 2019).

Recomendações técnicas são dispostas em um guia para desenvolvedores. Aspectos técnicos podem favorecer o trabalho do *Googlebot*, como a edição dos metadados e o uso das versões atualizadas dos softwares de repositórios (*Eprints*, *Digital Commons* e *DSpace*) e de editoração recomendados (*Atypon*, *Highwire*, *Ingenta*, *Silverchair* e *OJS*). Não é recomendada a adoção de *Dublin Core*, pois esse padrão dificulta a indexação automatizada (GSM, 2020).

Conforme se observa, o GSM possui funcionamento distinto das demais bases de dados científicas. A indexação é automática, sem controle de qualidade do conteúdo ou filtragem de publicações revisadas por pares.

A política de indexação proporciona uma coleção abrangente. Mas passados 10 anos de seu lançamento, os dados sobre o tamanho e a cobertura do GSM não são conhecidos (GU; BLACKMORE, 2017).

2.2.1 Tamanho e cobertura do Google Scholar Metrics

O tamanho e a cobertura do GSM, ou seja, o número de publicações e as suas características ainda não são conhecidos. Isso porque a Google não divulga

dados quantitativos ou a lista geral de publicações (GU; BLACKMORE, 2017). Essa ausência de informação é considerada falta de transparência, característica indesejável quando se trata de informação especializada (JACSÓ; 2012a, 2012b).

Extraindo-se as listas de publicações por categorias e subcategorias de assuntos e por idiomas, é possível identificar aproximadamente 5.565 publicações. Esse conjunto é limitado às publicações de maior impacto, sendo que os rankings por assunto abrangem apenas publicações de língua inglesa (GS, 2020).

Algumas pesquisas de cobertura restaram inconclusivas ou apresentaram apenas estimativas. López-Cózar e Cabezas-Clavijo (2013) inicialmente estimaram em torno de 40 mil o total de publicações indexadas. Mas esse número parece ser significativamente inferior ao número real, principalmente se forem consideradas publicações nos vários idiomas além do inglês indexados pelo sistema.

Visando obter um número mais preciso, Gu e Blackmore (2017) constataram que de 41.787 periódicos científicos revisados por pares e de todas as áreas extraídos do UlrichsWeb, somente 10.354 (20.8%) estavam indexados no GSM.

O método utilizado por Gu e Blackmore (2017), contudo, exigia a correspondência exata entre as grafias do título nas duas fontes. Logo, a falta de padronização do campo de título pode ter contribuído para um resultado significativamente inferior ao real. Esse estudo evidencia a necessidade de realizar buscas por variações do título, visando diminuir a ocorrência de erros de correspondência (REGENBRECHT; LANGLOTZ, 2015).

Outro estudo de cobertura teve como resultado a criação do Journal Scholar Metrics (JSM), base de dados⁴ com aproximadamente 9.196 publicações de artes, de humanidades e de Ciências Sociais indexadas na edição de GSM. Deste total, aproximadamente 27% são dos Estados Unidos, 21% do Reino Unido, 9% da Espanha, 5% do Brasil, 4% da Holanda, 3% França, 2% Alemanha e 29% de outros países (MARTÍN-MARTÍN; ORDUNA-MALEA; LÓPEZ-CÓZAR, 2017).

O JSM provavelmente é a maior fonte secundária de dados do GSM. Embora nunca tenha sido atualizada, é útil por manter os endereços URL das publicações, que podem ser objeto de estudos históricos e de cobertura do GSM.

Alguns estudos tiveram enfoque geográfico. A cobertura e o índice h5 das publicações da Espanha vêm sendo monitoradas anualmente. De 2.668 publicações

⁴<http://www.journal-scholar-metrics.infoec3.es/layout.php?id=home>

desse país listadas no UlrichsWeb, foram encontradas 1.352 (48,7%) na edição de 2018 (LÓPEZ-CÓZAR; MARTÍN-MARTÍN, 2018) e 1.448 (54,3%), na edição de 2019 do GSM. Esse volume era quase o triplo dos periódicos espanhóis indexados na SJ&CR (571) e mais de dez vezes no JCR (126) na época de realização do estudo (LÓPEZ-CÓZAR; MARTÍN-MARTÍN, 2019).

As áreas com maior cobertura de publicações espanholas são as Ciências Sociais (48,7%), as Artes e Humanidades (29%), as Ciências da Saúde (11,2%) as Ciências Naturais e as Engenharias (11,1%). Foram detectados ainda 95 periódicos com entradas duplicadas, o que corresponde a 6,5% de duplicação (LÓPEZ-CÓZAR; MARTÍN-MARTÍN, 2019).

Estudos com delimitação temática também foram conduzidos por López-Cózar e Caballero (2013), que constataram que de 663 periódicos da área de Comunicação de diversos países, 277 foram localizadas no GSM, número que representa 41,7% do total.

Costa, Canto e Pinto (2020), identificaram que de 46 periódicos brasileiros de Ciência da Informação listados na BRAPCI, 21 deles estavam indexados no GSM, o que representa 45% do total.

López-Cózar e Martín-Martín (2019) acreditam que o principal critério que impede uma maior cobertura dos GSM em relação os periódicos espanhóis é a necessidade de produção inferior a 100 artigos. Acreditam que se a produtividade mínima fosse reduzida a 50 artigos, haveria um aumento significativo do número de periódicos indexados.

Essa é a mesma constatação de Costa, Canto e Pinto (2020) em relação aos periódicos brasileiros de Ciência da Informação, pois notaram uma relação entre não indexação e menor produtividade.

A cobertura mais ampla, inclusive nas áreas de Ciências Humanas e Sociais, de publicações de idiomas não inglês e de países emergentes dá relevância ao GSM no cenário de avaliação da ciência. O sistema preenche uma lacuna deixada pela JCR e pela Scopus na geração de indicadores de impacto para publicações de menor prestígio (BORNMANN et al., 2016).

2.2.2 Vantagens do Google Scholar Metrics

Desde o seu lançamento, o GSM vem sendo considerado como uma fonte alternativa para mensuração do impacto de publicações em nível nacional (ZHANG;

LUN, 2019) e internacional (JACSÓ, 2012a; LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2013).

A primeira vantagem do GSM é a sua gratuidade, que é um diferencial diante de uma gama de serviços acessíveis mediante pagamento. Em um momento que o acesso aberto desponta como novo paradigma, a existência de uma fonte gratuita de avaliação de impacto se torna uma aliada desse movimento (KOUSHA; THELWALL, 2008).

Outra vantagem é simplicidade, especialmente para usuários que não visam a um uso avançado. O reconhecimento da marca também contribuiu para o bom posicionamento do GSM (JACSÓ, 2012a; ORDUÑA-MALEA; LÓPEZ-CÓZAR, 2014), já que o GS é considerado a fonte de informação acadêmica mais utilizada (GUSENBAUER, 2019).

A adoção do índice h como indicador oficial do GSM é bem vista. É um indicador consolidado, com forma de cálculo apta para representar a relação entre produtividade e impacto. O índice h é considerado superior em alguns aspectos em relação aos demais indicadores (FI e CS). É estável ano a ano, não estando sujeito a alterações abruptas decorrentes de muitas citações a um único artigo (ORDUNA-MALEA; LÓPEZ-CÓZAR, 2014; BARNES, 2017).

Essas vantagens - gratuidade, facilidade de uso e reconhecimento da marca e do indicador - pesam mais em relação ao usuário em geral. Já o uso para fins de avaliação e de estudos métricos requerem um olhar específico, que podem relevar algumas falhas e limitações do sistema.

2.2.3 Falhas e limitações do Google Scholar Metrics

A não divulgação do número de periódicos indexados ou de uma lista geral (*master list*) é vista como falta de transparência (JACSÓ, 2012a; LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012). Periódicos que atendem aos requisitos técnicos e de impacto podem não ser indexados sem razão aparente em alguma edição (desindexação pontual) (COSTA; CANTO; PINTO, 2020).

Falhas incluem ainda a falta de padronização de entradas por título, o que pode resultar em duplicação de registros (AYLLÓN; LÓPEZ-CÓZAR; ORDUNA-MALEA, 2016).

O GSM pode apresentar erros em contagens de citações (JACSÓ, 2012a) e é suscetível de manipulação do índice h5 por meio de citações falsas (LÓPEZ-

CÓZAR; ROBINSON-GARCÍA; TORRES-SALINAS, 2014). Muitas das falhas têm origem na coleta de citações na web, que pode sofrer com a instabilidade característica de fontes eletrônicas.

A limitação de 20 periódicos nas listas de resultados por categorias de assuntos e busca por título dificulta a recuperação de títulos em conjunto.

Não há divulgação dos critérios de classificação de assuntos, tampouco se essa decisão é humana ou automatizada. A impossibilidade de cálculo do índice h5 excluindo-se as autocitações é uma limitação evidenciada (CABEZAS-CLAVIJO; LÓPEZ-CÓZAR, 2013).

A cobertura temporal limitada impede análises mais amplas e dificultam a mensuração da produção científica em nível institucional (ORDUNA-MALEA; AYTAC; TRAN, 2019).

A inexistência de uma ferramenta nativa de extração de dados é outra falha relevante. Análises métricas utilizando dados do GSM podem consumir muito tempo, inviabilizando a realização de estudos (ORDUNA-MALEA; AYTAC; TRAN, 2019). Por estes motivos, o *Publish or Perish* (PoP) é a ferramenta mais utilizada para realização de análises de citação com dados do GS (JACSÓ, 2009).

Diante dessas questões, não há unanimidade com relação à utilização do GSM como fonte confiável para análise de impacto em processos oficiais (CABEZAS-CLAVIJO; LÓPEZ-CÓZAR, 2013). Em alguns casos, é vista somente como fonte complementar, visando à coleta de indicadores para análise de periódicos não indexados em bases tradicionais.

Mas conforme conclui Jacsó (2012b), nenhuma base de dados é perfeita e capaz de representar fielmente a qualidade e o impacto da ciência. Somente uma combinação de vários indicadores bibliométricos com o processo de revisão por pares pode refletir a natureza multidimensional do real impacto científico de periódicos, pesquisadores e grupos de pesquisa.

2.3 INDICADORES DE IMPACTO

Indicadores de impacto podem ser definidos como valores atribuídos a uma unidade de análise com a finalidade de quantificar o seu impacto científico em um determinado contexto. Os indicadores complementam o processo de avaliação da produção científica, que se inicia na etapa subjetiva de revisão por pares (WALTMAN, 2016).

Os indicadores baseados em análise de citação servem como instrumentos de mensuração do impacto de pesquisas após a publicação. São amplamente utilizados como critérios objetivos em atividades avaliativas (THELWALL; WILSON, 2014; WALTMAN, 2016).

Os primeiros indicadores foram criados para mensurar a quantidade de citações recebidas por periódicos científicos (GARFIELD, 1972). Atualmente são utilizados do mesmo modo para análise de outros elementos da produção científica, tais como autores, grupos de pesquisa, laboratórios e departamentos, programas de pós-graduação, regiões e países. O crescimento do número de indicadores e de suas aplicações foi favorecido pelo desenvolvimento dos índices de citação em bases de dados (YANG; WANG, 2015), além do avanço das técnicas de análise de redes sociais (JAMES *et al.*, 2019).

As fórmulas dos indicadores incluem as principais variáveis da produção científica, tais como produtividade (número de trabalhos publicados) e impacto (número de citações recebidas), analisadas por determinado período (entre dois e cinco anos). Nesse sentido, a fórmula mais usada se baseia na média de citações recebidas por artigo de uma publicação (WALTMAN, 2016; WALTERS, 2017).

Alguns indicadores podem também considerar a influência da fonte do documento citante e a aderência temática entre as fontes dos documentos citante e citado (WALTMAN, 2016; HOU, 2017; JAMES *et al.*, 2019). Com formulações matemáticas mais complexas, as chamadas métricas de influência utilizam algoritmos recursivos (WALTERS, 2017).

Devido à utilidade dos indicadores para as atividades de avaliação, há vasta literatura sobre o tema. São pontuados os aspectos positivos e negativos, comparativamente entre vários indicadores criados nos últimos anos. Os aspectos pontuados são a forma de cálculo, o período de análise, a abrangência, tanto no que se refere ao número de publicações quanto no das tipologias documentais, bem como atualização, transparência e acesso aos dados (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2013; SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015; WALTMAN, 2016; HOU, 2017; WALTERS, 2017; JAMES *et al.*, 2019).

Alguns indicadores são normalizados, ou seja, incluem uma variável relacionada ao padrão de citação de área de pesquisa. A normalização permite que publicações de disciplinas diferentes possam ser comparadas. Indicadores podem ainda ser dependentes ou independentes ao tamanho da amostra, isto é, terem seu

resultado influenciado pelo padrão de produtividade (WALTERS, 2017).

Os principais indicadores bibliométricos são o FI, o Eigenfactor Score (ES) e o Article Influence Score (AIS) calculados pelo JCR; o CS, o SJR e o SNIP calculados pela Scopus e pelo SJ&CR; o índice h, calculado em várias fontes; e o índice e a mediana h5 calculado pelo GSM (WALTMAN, 2016; MARTÍN-MARTÍN *et al.*, 2018).

A Tabela 3 apresenta uma síntese dos parâmetros dos indicadores de impacto mais conhecidos e utilizados atualmente.

Considerando os diferentes parâmetros (BORNMANN *et al.*, 2012), estudos investigaram a correlação entre resultados de indicadores de conjuntos de publicações (ELKINS *et al.*, 2010; KIM; HUH; CHU, 2014; SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015; OKAGBUE; TEIXEIRA DA SILVA, 2020). Os resultados identificaram níveis relevantes de correlação positiva entre os indicadores disponíveis, o que indica estabilidade e validade convergente (SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015).

Mas os níveis de correlação variam de acordo com a amostra e o período analisado (OKAGBUE; TEIXEIRA DA SILVA, 2020). Em algumas comparações, os coeficientes de correlação são muito fortes, como no caso de indicadores baseados na média de citações por artigo, tais como o FI, o FI5 e CS (OKAGBUE *et al.*, 2019; KIM; HUH; CHU, 2014; SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015). Níveis muito fortes no teste de Spearman também foram detectados entre SJR e FI (GONZÁLEZ-PEREIRA *et al.*, 2010), e entre SJR e AIS em um amplo conjunto de periódicos indexados simultaneamente na WoS e na Scopus (SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015).

As análises com outros indicadores mostram resultados diversos, mas ainda em níveis estatisticamente relevantes. Essas variações são normais, considerando que alterar os parâmetros utilizados em cada indicador tende a produzir uma pontuação diferente, mantendo um padrão mínimo de concordância (ELKINS *et al.*, 2010; SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015; OKAGBUE *et al.*, 2019).

O índice h possui maior variação nos níveis de correlação com outros indicadores. Em uma amostra de quase 10 mil periódicos indexados simultaneamente na WoS e na Scopus, foram identificados os seguintes coeficientes de correlação entre o índice h calculado pela Scopus e os demais indicadores: FI=0,76, ES=0,87, AIS=0,70, SJR=0,73, SNIP=0,60.

Tabela 3 – Síntese dos parâmetros de indicadores de impacto

Indicador	Fonte	Período das citações	Período dos documentos	Normalizado ?	Dependente de tamanho?	Inclui autocitações?	Documentos citantes	Documentos citados
AIS	JCR	1 ano	5 anos	Sim	Não	Não	Publicações indexadas	Artigos, revisões
ES	JCR	1 ano	5 anos	Sim	Sim	Não	Publicações indexadas	Artigos, revisões
nES	JCR	1 ano	5 anos	Sim	Sim	Não	Publicações indexadas	Artigos, revisões
FI	JCR	1 ano	2 anos	Não	Não	Sim	Publicações indexadas	Artigos, revisões
Flws	JCR	1 ano	2 anos	Não	Não	Não	Publicações indexadas	Artigos, revisões
FI5	JCR	1 ano	5 anos	Não	Não	Sim	Publicações indexadas	Artigos, revisões
JCI	JCR	3 anos	3 anos	Sim	Não	Sim	Publicações indexadas	Artigos, revisões
CS	Scopus	4 anos	4 anos	Não	Não	Sim	Artigos, revisões, anais, capítulos de livros	Artigos, revisões, trabalhos de eventos, capítulos de livros
Índice h	Scopus, SJR, WoS	Todo	Todo	Não	Sim	Sim	Artigos, revisões, anais, capítulos de livros	Artigos, revisões, trabalhos de eventos, capítulos de livros
SJR	Scopus, SJR	1 ano	3 anos	Sim	Não	Sim, limitado a 33%	Artigos, revisões, anais, capítulos de livros	Artigos, revisões, trabalhos de eventos, séries de livros
SNIP	CWTS, Scopus, SJR	1 ano	3 anos	Sim	Não	Sim	Artigos, revisões, anais, capítulos de livros	Articles, reviews, conference papers, book series
Índice h5	GSM, GSC	5 anos	5 anos	Não	Sim	Sim	Periódicos, anais, livros, relatórios, teses, dissertações	Artigos, trabalhos de eventos de Engenharia e Computação
Mediana h5	GSM	5 anos	5 anos	Não	Sim	Sim	Periódicos, anais, livros, relatórios, teses, dissertações	Artigos, trabalhos de eventos de Engenharia e Computação

Fontes: Waltman (2016), Walters (2017), James et al. (2019), JCR (2022), Elsevier (2022), SJ&CR (2022).

Mesmo com variações, todos os coeficientes resultaram em níveis moderado ou forte, o que indica que o índice h é um parâmetro válido de análise de impacto. Essa validade é reforçada quando o índice h é utilizado para ranquear ou classificar publicações, resultados estes que foram obtidos por meio do teste de Spearman, que calcula a correlação em função das posições ocupadas pelas variáveis (SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015).

Além da análise de correlação, outra técnica estatística utilizada em estudos métricos a partir de dados de citação é a regressão. Diferentes modelos de regressão podem ser utilizados a partir de dados de citação, cada qual adequado ao tipo dos dados, bem como aos objetivos da análise. Um dos modelos é o da regressão linear simples (OKAGBUE et al., 2019; THELWALL; WILSON, 2014).

Uma das aplicações da regressão linear simples é a de avaliar periódicos a partir de indicadores diferentes. O modelo de regressão permite estimar os valores de uma variável (explicada) a partir de outra variável (explicativa) (CHEIN, 2019).

A ideia básica da regressão linear é estabelecer uma reta que melhor descreve a relação entre essas duas variáveis (CHEIN, 2019). Nesse caso, as duas variações são os valores dos indicadores para um mesmo conjunto de periódicos. A partir do modelo é possível prever o valor que um periódico teria em um determinado indicador a partir do valor de outro indicador que este possua.

Pelo que foi divulgado até o momento, será utilizado um modelo de regressão linear simples para comparar periódicos a partir de indicadores diferentes no sistema de avaliação Qualis (BRASIL, 2021).

2.3.1 Fator de Impacto

A história do FI se confunde com a própria história dos indicadores baseados em análise de citação. Foi o pioneiro e é o mais conhecido indicador voltado para avaliação de periódicos a partir de uma base de dados (JAMES *et al.*, 2019). Foi desenvolvido na Década de 1960 por Garfield no âmbito do ISI, com a função de representar matematicamente a relação entre o número de artigo publicados e o número de citações recebidas em uma lista de aproximadamente 4 mil periódicos selecionados na época (GARFIELD, 1972).

O FI é o indicador mais utilizado para avaliação da ciência mundial. Em sua fórmula tradicional de dois anos, representa o número de citações recebidas pelos

artigos publicados pelo periódico no ano de análise, dividido pelo número de artigos publicados nos dois anos anteriores (WALTMAN, 2016).

O FI atual é referente ao ano de 2020 e pode ser expresso pela fórmula:

$$JIF = \frac{\text{Citações recebidas em 2019 por artigos publicados em 2018 e 2019}}{\text{Número de artigos publicados em 2018 e 2019}}$$

Na fórmula do FI, o numerador (documentos citantes) é composto pelas citações recebidas pelo periódico como um todo, considerando todos os tipos de documento. Já o denominador (documentos citáveis) é composto por apenas três tipos: artigos e revisões, principalmente, e trabalhos de eventos (JAMES *et al.*, 2019).

O FI é um indicador não normalizado, então a sua fórmula se aplica de forma idêntica diferentes áreas de pesquisa. Assim, não é recomendado comparar publicações de áreas diversas a partir dele.

Por outro lado, é um indicador independente de tamanho, ou seja, as variações dos níveis de produtividade dos periódicos (número de documentos publicados) são ocultada por sua fórmula, baseada na média de citações por documento (WALTERS, 2017).

O FI tradicional de dois anos possui duas outras versões. Uma varia o período de análise, que passa a ser de cinco anos (*5 Year Impact Factor*). A outra versão é calculada a partir da exclusão de autocitações (FIws) (*Impact Factor Whithout Self Citites*) (WALTERS, 2017).

O período mais longo é mais indicado para algumas disciplinas. Já a versão sem autocitações visa a uma análise de impacto mais precisa (HUANG; CATHY LIN, 2012; CHORUS; WALTMAN, 2016).

2.3.2 Eigenfactor Score

O Eigenfactor Score (ES) foi desenvolvido em 2007 visando a mensuração do desempenho de periódicos de economia, mas logo passou a ser utilizado em outras áreas. Seu funcionamento é baseado no algoritmo *PageRank*, desenvolvido pela Google. Utiliza a teoria de análise de redes sociais para a quantificação de relações entre publicações (elementos) e citações (ligações) (WALTMAN; VAN ECK, 2010).

O ES tem uma fonte própria⁵, mas também foi incorporado pelo JCR. O seu cálculo é semelhante ao do FI de cinco anos, excluindo as autocitações. Todavia, são atribuídos pesos distintos às citações, cuja variação corresponde ao valor do FI do periódico citante. Quanto mais elevado for o nível de impacto de um periódico, maior será o peso atribuído à citação efetuada por seus documentos (KRAMPL, 2019).

O ES é uma medida do desempenho total de um periódico e não do desempenho médio por artigo, como é o FI. Periódicos com mais influência em um campo científico tendem a ter um resultado mais elevado a partir desse indicador (WALTMAN; VAN ECK, 2010).

Embora tente representar mais precisamente a qualidade de um periódico no contexto de sua disciplina científica, o ES tem seu cálculo realizado por meio de um parâmetro algorítmico pré-definido, o que torna complexa a sua reprodução por um usuário externo e sem acesso ao código e aos dados de citação.

Além disso, é um indicador insensível ao de baixo impacto, evidenciando o desempenho dos periódicos altamente influentes.

O ES possui uma versão normalizada (nES). Essa normalização tem a finalidade de adequar a fórmula de acordo com o número de periódicos indexados no JCR em cada ano (WALTERS, 2017). Um periódico médio vai ter nES=1. Um periódico com nES=5 será considerado cinco vezes mais influente que o padrão médio da base de dados.

A versão normalizada do indicador facilita a compreensão do resultado, já que em periódicos de baixo impacto o ES costuma apresentar resultados numéricos ínfimos (por ex: 0.00017).

2.3.3 Article Influence Score

O *Article Influence Score* (AIS) é um indicador que mede a influência média dos artigos de um periódico nos primeiros cinco anos após a publicação. É calculado pela fórmula: $(0,01 \times \text{resultado do ES}) \div X$, sendo X o número de artigos publicados pelo periódico nos cinco anos, dividido pelo número de todos os artigos que receberam citações na base de dados (JCR, 2022).

⁵<http://www.eigenfactor.org/>

Da mesma forma que o ES, um padrão médio será implicará em um AIS=1. Uma pontuação diferente significa que cada artigo do periódico tem uma influência superior ou inferior à da média (WALTERS, 2017).

2.3.4 CiteScore

O CS é o principal indicador de impacto calculado a partir dos dados de citação da base de dados Scopus. Foi lançado em dezembro de 2016 com a mesma forma de cálculo do FI, mas com um período de análise de três anos para os trabalhos publicados. O CS de 2018, por exemplo, foi calculado a partir do total de citações recebidas em 2018, dividido pela soma das citações recebidas em 2015, 2016 e 2017 (JAMES *et al.*, 2019; OKAGBUE; TEIXEIRA DA SILVA, 2020).

Nessa primeira versão, o CS considerava citações coletadas a partir de qualquer publicação indexada na Scopus, o que inclui periódicos, anais, séries de livros e publicações comerciais. Além disso, documentos de todas as seções da publicação eram inicialmente inseridos do rol de documentos citantes e citados, o que incluía material editorial, cartas, notas e outras seções que não são inseridas na lista de itens citáveis do FI (JAMES *et al.*, 2019).

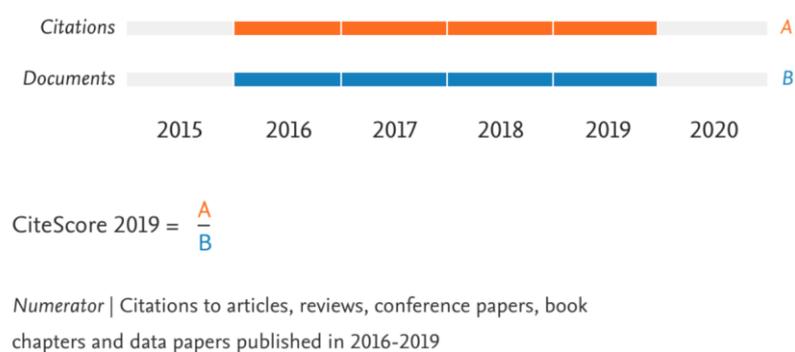
Em junho de 2020, o cálculo do CS sofreu modificações, que foram aplicadas para o índice de 2019. A primeira foi no período de análise, que passou de três para quatro anos. Além disso, passaram a ser consideradas todas as citações recebidas no período, ao contrário da metodologia anterior, que só considerava as citações recebidas no ano do indicador.

Uma segunda modificação foi a inclusão somente de documentos revisados por pares, tais como artigos científicos, artigos de revisão e trabalhos de eventos. Cartas, material editorial, notícias e notas, que via de regra não passam por processos de avaliação por pares, passaram a ser desconsideradas. Essa medida tornou a análise mais robusta, considerando que esses materiais não se tratam, efetivamente, de resultados de pesquisa e não devem representar o seu impacto (ELSEVIER, 2020).

Uma última modificação foi realizada na formatação do número final do indicador, que passou a ser apresentado com apenas duas casas decimais, em vez de três casas, como era anteriormente (ELSEVIER, 2020).

A Figura 1 mostra a metodologia de cálculo do CS de 2019.

Figura 1 - Metodologia de cálculo do CiteScore 2019
New CiteScore methodology: CiteScore 2019



Fonte: Elsevier (2020).

O lançamento do CS havia sido visto com ceticismo por alguns pesquisadores, que questionaram a necessidade de outra métrica de citação, a despeito da existência de outras com a mesma finalidade (JAMES *et al.*, 2019).

As mudanças realizadas diferenciam ainda mais o CS do FI, atendendo a algumas das reivindicações da comunidade científica. O aumento da janela beneficiou mais disciplinas, principalmente as de ritmo mais lento de obsolescência e baixos níveis de impacto (WALTMAN, 2016).

2.3.5 SJR

O SJR é o indicador de impacto calculado pelo SJ&CR a partir dos dados de citação das publicações indexadas na Scopus. Foi lançado em 2007 e recebeu uma atualização em sua formulação em 2012, configuração que ainda permanece atualmente. Adota um período de três anos para os trabalhos publicados e de um ano para coleta das citações recebidas. Assim como outros indicadores da Scopus, se baseia no algoritmo *PageRank*, desenvolvido pelo Google (FALAGAS *et al.*, 2008; GONZALEZ-PEREIRA; GUERRERO-BOTE; MOYA-ANEGON, 2010; JAMES *et al.*, 2019).

O SJR é um indicador semelhante ao AIS, pois também considera pesos distintos para as citações recebidas, de acordo com o impacto do periódico citante. O SJR, contudo, possui uma definição matemática ainda mais complexa, pois o peso de uma citação depende não apenas do impacto da publicação, mas também de uma medida da proximidade temática entre as publicações citantes e citadas.

Assim, uma citação de um periódico citante que é tematicamente mais próxima ao periódico citado recebe um peso superior do que um periódico citante de escopo temático mais distante. Baseia-se, portanto, nas medidas de proximidade da teoria de grafos (WALTMAN, 2016; JAMES *et al.*, 2019).

O SJR tenta traduzir com maior precisão o real impacto das publicações, abrangendo duas variáveis relevantes não consideradas no FI. Uma das vantagens dessa é a possibilidade de comparar diferentes campos de assunto, considerando que o seu resultado é normalizado. Por outro lado, seu cálculo algorítmico dificulta a conferência de resultados (JAMES *et al.*, 2019).

2.3.6 SNIP

O *Source Normalized Impact Per Paper* (SNIP) é um indicador proposto pelo *Centre for Science and Technology Studies* do *Leiden University* (CWTS). É calculado pelo número de citações recebidas no ano base para documentos publicados nos últimos três anos, dividido pelo número total de documentos no mesmo período. A principal característica no SNIP é a utilização de uma abordagem de normalização visando corrigir diferenças nos padrões de citação entre campos científicos (MOED, 2010).

Mas esse processo de normalização não exige a definição prévia dos parâmetros, pois são definidos a partir dos próprios dados de citação. No cálculo do SNIP, a quantidade de referências dos documentos é utilizada como parâmetro de tamanho das áreas (WALTMAN *et al.*, 2013; MOED, 2010).

Basicamente, o cálculo do SNIP considera que se um periódico é citado majoritariamente por publicações com amplas listas de referências, isso sugere que se encontra em um campo com alta densidade de citações (WALTMAN *et al.*, 2013).

Por outro lado, se as citações têm origem em publicações com listas mais restritas de referências, o periódico tem aderência a um campo de baixa densidade. Assim, esse parâmetro (densidade de citações) é utilizado para definir o número médio de citações esperado para cada publicação (WALTMAN *et al.*, 2013).

O SNIP é calculado a partir das publicações indexadas na Scopus, estando disponível também no SJ&CR.

2.3.7 Índice h

O índice h foi criado por Jorge G. Hirsch em 2005 com o objetivo de avaliar a performance científica de pesquisadores a partir da relação entre produtividade e impacto de publicações. Representa o maior número h de trabalhos publicados que tenham recebido pelo menos o mesmo número h de citações.

O índice de Hirsch significa, portanto, que se determinado autor possuir 15 documentos publicados, tendo 10 destes recebidos 10 ou mais citações cada, o índice h será 10, conforme mostra a Tabela 4.

Tabela 4 – Identificação do índice h de um conjunto de documentos

# documento	# citações	
1	30	
2	25	
3	22	
4	22	
5	21	
6	19	
7	18	
8	15	
9	13	
10	10	Índice h
11	9	
12	7	
13	4	
14	0	
15	0	

Fonte: Elaborado pelo autor.

O índice h foi proposto como um novo parâmetro de comparação entre o desempenho científico de pesquisadores competindo por um mesmo recurso de financiamento ou colocação profissional (HIRSCH, 2005). O trabalho de Hirsch (2005), que é físico e nunca havia trabalhado com bibliometria anteriormente, teve um forte impacto nos estudos métricos, recebendo milhares de citações (BORNMANN, 2014).

A função original do índice h, de avaliar cientistas, pesquisadores e acadêmicos, foi ampliada. Atualmente é utilizado para avaliação de desempenho de grupos de pesquisa, periódicos, laboratórios e departamentos, instituições e até de países (BORNMANN et al., 2011; BARNES, 2017).

Uma das primeiras vantagens do índice h é a facilidade de ser calculado. Não exige formulações matemáticas complexas e resulta em um único número inteiro. De forma manual, o índice h de um autor pode ser verificado rapidamente em uma planilha. Para amplas amostras é possível utilizar a WoS, a Scopus ou o SJ&CR,

que possuem esse indicador entre os recursos de análise de impacto (BARNES, 2017).

Outro mérito seu é englobar quantidade e qualidade, duas relevantes abordagens de avaliação, em apenas um número. O valor h evidencia o número de artigos publicados que têm qualidade proporcional, medida por meio do número de citações recebidas. Estabelece um parâmetro médio de desempenho, as vezes considerado mais justo que outras medidas (BARNES, 2017).

O índice h é um indicador estável, que não sofre variações bruscas em razão de trabalhos altamente citados (*outliers*), bem como de trabalhos pouco citados ou não citados. Não é tão sensível a erros pontuais na lista de publicações, tais como citações contadas para mais ou para menos (BORNMANN et al., 2011).

Do mesmo modo, não é tão influenciado pelas autocitações quanto os indicadores baseados na média de citações por artigo. Isso porque as citações (incluindo autocitações) que excedam o valor h são desconsideradas (HIRSCH, 2005; WALTMAN, 2016).

Além das vantagens inerentes à forma de cálculo, uma crescente insatisfação com o FI pode ter favorecido a boa recepção do índice h . O uso do FI não apenas para periódicos, mas também indiretamente para avaliar a qualidade da produção de autores não é bem-visto por parte da comunidade científica (BORNMANN; DANIEL, 2009).

Assim, o índice h se tornou uma alternativa viável no cenário de avaliação da ciência (BARNES, 2017).

Esses aspectos em um ambiente favorável a um novo modelo de avaliação tornou o índice h um parâmetro universal. Passou a ser adotado por gestores de ensino e pesquisa em processos nas quais é necessário comparar desempenho acadêmico. Em alguns casos, instituições fixam um valor de índice h mínimo para candidatos a cargos, promoções ou financiamentos universitários (BARNES, 2017).

Mas o índice h também possui limitações. Primeiramente, a sua fórmula de cálculo considera apenas a parcela de artigos altamente citados, ignorando artigos pouco citados e não citados e não indica o número total de citações recebidas.

Além disso, não há variações de peso para trabalhos publicados por um único autor e por trabalhos em coautoria. Um artigo de autoria única tem peso subestimado, enquanto um trabalho com vários coautores acaba sendo superestimado para cada um deles (BARNES, 2017).

Outra limitação é que o índice h só é adequado em análises comparativas em uma mesma área. Sendo um indicador não normalizado (WALTERS, 2017), as variações nos níveis de produtividade e de impacto entre as diversas áreas de pesquisa podem influenciar no seu resultado.

O mais indicado é a análise comparativa em nível de subárea (BARNES, 2017). É um indicador dependente do tamanho da amostra analisada (WALTERS, 2017). Assim, periódicos mais produtivos tendem a ter o índice h mais elevado que periódicos menos produtivos (CANTO et al., 2022). Mas a maioria dessas limitações parece não ser específica no índice h , muitas delas ocorrendo simultaneamente em outros indicadores (BARNES; 2017).

Por outro lado, há uma característica que parece ser específica da fórmula de Hirsch: a tendência em favorecer pesquisadores com carreiras mais longas, mesmo os de menor impacto, em detrimento de pesquisadores brilhantes em início de carreira. Isso porque o valor do índice h nunca vai ultrapassar o número máximo de trabalhos publicados (BORNMANN, 2014).

Assim, o critério quantidade resta evidenciado em relação ao critério qualidade (impacto). Assim, em uma análise limitada ao critério objetivo (somente o valor) as diferenças entre os dois pesquisadores do exemplo não podem ser detectadas sem uma análise curricular.

O índice h não está vinculado a nenhuma base de dados específica. Dependendo da fonte, os resultados h podem ser diferentes. Algumas das bases que calculam o índice h são a WoS, a Scopus, o SJR, o GSC, o GSM e o ResearchGate (TEIXEIRA DA SILVA; DOBRÁNSZKI, 2018).

Desde o seu lançamento, o índice h vem sendo objeto de ampla investigação no contexto dos estudos métricos. Destacam-se as propostas de modificação em sua forma de cálculo, visando suprimir supostas distorções de sua fórmula original.

Em meta-análise relativas aos primeiros cinco anos de lançamento do índice h , Bornmann et al. (2011) identificaram 37 propostas de variações de cálculo a partir da fórmula original de Hirsch (2005). Essas propostas foram classificadas em cinco grupos, de acordo com a natureza da vantagem apresentada com a inovação métrica.

Um primeiro conjunto de propostas visa diminuir as diferenças entre campos de pesquisa, compensando as variações nos índices médios de citação no cálculo do índice h . Um segundo conjunto objetiva eliminar a influência das autocitações no

cálculo do índice (BORNMANN et al., 2011). Destaca-se, nesse conjunto, a proposta do índice h_s , definido pelo maior número de artigos que receberam h citações independentes, ou seja, excluindo as autocitações (SCHREIBER, 2009).

O terceiro grupo, definido pelo índice h_m , é obtido contando as citações com peso inversamente proporcional ao número de coautores, ou seja, atribuindo peso maior a trabalhos com único autor, seguindo de trabalhos com dois autores, três e assim por diante (BORNMANN et al., 2011).

O quarto grupo de variações do índice h leva em conta o tempo de carreira do pesquisador ou data das publicações avaliadas. Essa questão já havia sido inclusive levantada pelo próprio Hirsch (2005) em seu artigo inaugural, considerando que o favorecimento de carreiras antigas é a principal limitação do índice h .

Como solução parcial para esse problema, foi proposta uma relação h/y , na qual o quociente y (*year*) representa o número de anos passados desde a publicação do primeiro trabalho do autor analisado.

Se por um lado essa proposta corrige a comparação entre pesquisadores de diferentes tempos de carreira, não leva em conta a data do trabalho, pois mesmo trabalhos antigos continuam sendo computados indefinidamente, inclusive após a aposentadoria ou morte do autor (BORNMANN et al., 2011)

Por fim, o quinto grupo de propostas se refere a variações de atribuição de peso maior para citações recebidas pelos trabalhos altamente citados entre aqueles que compõem o índice h .

Entre essas propostas se destaca o índice g , definido como o maior número g de trabalhos que somados receberam g^2 ou mais citações. Com essa fórmula, o índice g será sempre igual ou maior que o índice h , levando mais em conta o volume de citações recebidas pelos artigos (BORNMANN et al., 2011)

As simulações feitas por Bornmann et al. (2011) indicam que há forte correlação entre os resultados obtidos pelo índice h e os obtidos com as 37 variações propostas na literatura.

Assim, não haveria justificativa para proliferação de outros índices baseados na relação entre produtividade e impacto, pois ocasionaria redundância e não melhorias efetivas em processos avaliativos.

Em estudo similar, Ding, Liu e Kadonga (2020) analisaram as limitações do índice h e de mais cinco indicadores de desempenho científico similares, quais sejam, o índice g , o índice AR, o índice p , o indicador de impacto integrado (I3) e o

traço acadêmico. Os aspectos analisados foram a consistência, o grau de discriminação e a relação estatística entre os valores dos indicadores e os números de publicações e citações.

Constataram que todos os indicadores são altamente consistentes e mais suscetíveis ao número de publicações do que a frequência de citações e que o índice h apresenta o menor grau de discriminação, seguido pelo índice g , pelo I_3 , pelo índice AR , pelo índice p e pelo traço acadêmico.

Observa-se que tão logo o trabalho que deu origem ao índice h foi publicado por Hirsch (2005), surgiram inúmeras propostas de variações, tanto relativas à novas fórmulas de cálculo quanto de novas aplicações.

Entre essas propostas, destaca-se a de Braun, Glänzel e Schubert (2006) no sentido da utilização do índice h para avaliação do impacto de periódicos. Nessa proposta, diferentemente da fórmula original, o índice não é calculado com base em todo o período de existência da publicação, mas fixando-se um limite de tempo.

2.4 CRÍTICAS AOS INDICADORES DE IMPACTO

Em paralelo à ampliação do número de indicadores (WALTERS, 2017) e de suas fontes de cálculo (GUSENBAUER, 2019), houve um aumento na utilização de critérios objetivos em procedimentos de avaliação. Mas esse fenômeno, descrito como obsessão pelo “fator de impacto” (CHORUS; WALTMAN, 2016), não ocorreu de forma unânime.

Críticas ao uso indiscriminado de indicadores como critérios preponderantes de avaliação são observadas no cenário da comunicação científica (GUEDÓN, 2010; WALTMAN, 2016).

As críticas atingem primeiramente os editores científicos e seus mantenedores, que se utilizam de estratégias gerais de negócios em uma área de profundo interesse humano e social, como é ciência (MACROBERTS; MACROBERTS, 1989; GUEDÓN, 2010; LARIVIÈRE; SUGIMOTO, 2018). Abrangem todo o sistema consolidado em torno do conceito de avaliação de impacto por meio de análise de citações (GUEDÓN, 2010).

Uma divergência é relacionada ao uso da citação como elemento central de representação de impacto, de qualidade e de relevância científica (MACROBERTS; MACROBERTS, 1989). Essa divergência é histórica e teve início simultaneamente às primeiras propostas de uso da análise de citação (GARFIELD, 1972).

A discussão inicial é centrada no comportamento de citação (KAPLAN, 1965), isto é, os motivos que levam autores a citar determinados documentos. Ocorre que esses motivos nem sempre são estritamente científicos e, em alguns casos, a citação não é cientificamente necessária (CRONIN, 1984; MACROBERTS; MACROBERTS, 1989).

Citações tendenciosas, citar colegas de instituição, omissão de autores clássicos são alguns dos comportamentos reprováveis na prática de citação e que influenciam os resultados dos indicadores (MACROBERTS; MACROBERTS, 1989; HENEBERG, 2016).

Citações contrárias também são computadas em indicadores de impacto, sendo atribuído o mesmo peso as citações com teor positivo e negativo. Essa crítica, no entanto, não foi acompanhada de evidências empíricas que indicassem se tratar de um problema relevante. Até porque citações negativas representam apenas uma pequena porção do total de citações (XU; DING; LIN, 2022).

Alguns estudos sugerem, pelo contrário, que as citações negativas não interferem significativamente; que os artigos com mais citações negativas são justamente os de maior impacto; e que a exclusão de citações negativas não diminuiriam o nível geral de impacto de um conjunto de documentos (CATALINI; LACETERA; OETTL, 2015; XU; DING; LIN, 2022).

Outras questões se mostraram mais preocupantes (CHORUS; WALTMAN, 2016). A busca pelo aumento de indicadores pode levar alguns autores a praticar condutas reprováveis. Citações cruzadas, empilhamento de citações e inserção de citações falsas são algumas das práticas questionáveis que surgiram como efeito colateral da corrida pelo impacto (VAN NOORDEN, 2013; HENEBERG, 2016; LÓPEZ-CÓZAR; ROBINSON-GARCÍA; TORRES-SALINAS, 2014).

Como resposta, as bases de dados têm incrementado a vigilância e o monitorando dos padrões de citação. A Clarivate, por exemplo, tem utilizado inteligência artificial para identificar a manipulação de dados de citação no JCR (VAN NOORDEN, 2013).

Parte das alegações não se referem estritamente a função da citação no processo de elaboração de documentos. Aduzem que nem sempre a citação é realizada pelos autores dentro das normas tácitas historicamente estabelecidas pela ciência (KAPLAN, 1965; CRONIN, 1984; MACROBERTS; MACROBERTS, 1989).

Parte dos problemas mencionados não são restritos à análise de citação, mas incluem também má conduta na escrita científica. As más práticas podem ocorrer em outros contextos, tais como plágio, manipulação ou fabricação de dados de pesquisa, conflito de interesse e falsas atribuições de autoria (FONG; WILHITE, 2017).

As críticas direcionadas a análise de citação deveriam motivar a melhoria de todos os processos de avaliação da ciência, não se restringindo apenas à análise de citação e ao modelo de avaliação por indicadores de impacto.

Mas, de fato, alguns dos problemas são específicos de análise de citação, especialmente o da manipulação de citações.

2.5 MANIPULAÇÃO DE CITAÇÕES

A mencionada “obsessão pelo impacto” (CHORUS; WALTMAN, 2016) pode levar editores a utilizar estratégias para elevar o número de citações recebidas pelo periódico que editam. Mas nem sempre essas estratégias são orientadas pelos padrões internacionais de boas práticas acadêmicas.

As estratégias de aumento de impacto podem ir além de melhorias no processo de revisão por pares e de ampliação de visibilidade, entre outras medidas de incremento de qualidade editorial. Em sentido oposto, podem envolver práticas inadequadas, entre as quais a manipulação de citações (FONG; WILHITE, 2017).

A manipulação de citações pode ocorrer de diferentes formas. Pode haver, por exemplo, vínculo entre o periódico citante e citado, a mesma editora ou instituição (*publisher self-citation*) (ZHOU, 2021).

Algumas formas envolvem a participação de mais de uma publicação, como os cartéis de citação (*citation cartels*) (FONG; WILHITE, 2017) e o empilhamento de citações (*citation-stacking*) (VAN NOORDEN, 2013; FONG; WILHITE, 2017).

Existem processo mais sérios, como a criação de citações falsas (LÓPEZ-CÓZAR; ROBINSON-GARCÍA; TORRES-SALINAS, 2014).

Mas a tipo mais comum de manipulação é por meio de autocitação, que é considerada a forma mais fácil de aumentar indicadores de impacto (TAŞKIN et al., 2021).

2.6 AUTOCITAÇÃO DE PERIÓDICO

A autocitação (*self-citation*) é definida basicamente como a prática de um autor citar um trabalho anterior de sua própria autoria ou coautoria (IOANNIDIS, 2015; HENEBERG, 2016; WALTMAN, 2016; SZOMSZOR et al., 2020; TAŞKIN et al., 2021). Não é, *a priori*, considerada uma prática inadequada. É autorizada e muitas vezes necessária para facilitar o fluxo e a difusão de conhecimento entre trabalhos científicos (YU; YU; WANG, 2014; FLATT; BLASIMME; VAYENA, 2017; SIMÕES; CRESPO, 2020).

A autocitação cumpre a função de divulgação de pesquisas anteriores do próprio autor, podendo ser considerada uma forma promoção de visibilidade e reputação acadêmica (GONZÁLEZ-SALA; OSCA-LLUCH; HABA-OSCA, 2019).

A autocitação pode ser configurada de outras formas, tais como em nível de colaboração (citar colegas de pesquisa), nível de periódico (citar um artigo do mesmo periódico) (ROUSSEAU, 1999; HARTLEY, 2009), em nível de instituições (citar uma publicação mesma instituição/editora) (ZHOU, 2021), em nível de país (citar uma publicação do mesmo país) e assim por diante (HUANG; CATHY LIN, 2012; FRANDSEN, 2007; WALTMAN, 2016).

Não há unanimidade acerca da necessidade de exclusão de autocitações no cálculo de indicadores de impacto (CAMPANARIO, 2011; HUANG; CATHY LIN, 2012).

Há consenso de que em níveis *micro* (autor, periódico) e *meso* (instituição) a influência da autocitação é significativa na análise de impacto (YU; YU; WANG, 2014), podendo ser necessária a exclusão.

Já em nível macro (país), a influência não é significativa, sendo desnecessária a exclusão (WALTMAN, 2016). Mas é unânime a opinião de que não se pode simplesmente ignorar a existência de autocitações na comunicação científica (FRANDSEN, 2007).

Diferentemente da autocitação de autor (*author self-citation*), na qual há inequívoca intenção de se autocitar (seja ela justificável ou não), a autocitação de periódico (*journal self-citation*) (HARTLEY, 2009) pode ocorrer de forma involuntária.

Não é raro que diversas pesquisas sobre um mesmo assunto sejam publicadas em um mesmo periódico. Assim, a citação de parte dessas pesquisas em um novo trabalho da mesma publicação é cientificamente válida (GAMI et al., 2004; FRANDSEN, 2007; YU; YU; WANG, 2014; GAZNI; DIDEGAH, 2021).

O problema tem início quando a autocitação é desnecessária e não há relação temática ou metodológica entre o trabalho citante e o citado. Autocitações desnecessárias ou excessivas são indicadores de anormalidade que podem afetar a credibilidade da pesquisa e da própria publicação.

Em excesso, autocitações de periódico podem interferir na avaliação de impacto, aumentando artificialmente indicadores baseados em citação (BARTNECK; KOKKELMANS, 2011; FLATT; BLASIMME; VAYENA, 2017).

Em 2007 o FI passou a ter uma versão com exclusão de autocitações. A partir de 2009, o JCR passou a identificar e monitorar as autocitações das publicações indexadas. Diversas publicações sofreram sanções de base de dados em razão de irregularidades envolvendo autocitações no FI (HUANG; CATHY LIN, 2012; LIU; FANG, 2020).

Paralelamente, novos indicadores de impacto foram criados com base na exclusão de autocitações, tais como o ES e o AIS (BERGSTROM; WEST; WISEMAN, 2008), ou na limitação de autocitações, como o SJR (GONZÁLEZ-PEREIRA et al., 2010). Esses indicadores partem do princípio de que as autocitações interferem no impacto ou a na influência da publicação em seu campo científico.

A diferenciação entre citações e autocitações facilitaram estudos em larga escala, especialmente os de identificação de padrões ou de taxas de autocitação.

Taxas de autocitações de periódicos são definidas de duas formas. A primeira forma (*self-citing*) é definida pelo percentual de autocitações entre o número total de referências de um periódico. A segunda (*self-cited*) é pelo percentual de autocitações no número total de citações recebidas por um periódico (ROUSSEAU, 1999; FRANDSEN, 2007).

Taxas de autocitação (*self-citation rates*) são estabelecidas em diferentes níveis. Podem ser calculadas por periódico, por disciplina ou área de pesquisa, por editora ou instituição, por idiomas e até por países (TAŞKIN et al., 2021).

As taxas de autocitação são maiores logo após a publicação do trabalho, diminuindo sua ocorrência com o passar dos anos. Tendem a ser superiores em periódicos de menor impacto (HENENBERG, 2016), de idiomas regionais (TAŞKIN et al., 2021), de assuntos especializados (LIVAS; DELLI, 2018; SANFILIPPO et al., 2021) e de áreas de menor visibilidade (ROUSSEAU, 1999; FRANDSEN, 2007).

Por outro lado, as taxas médias de autocitação parecem não apresentar diferenças relevantes entre categorias de assunto, mantendo-se em percentuais próximos em estudos de diferentes disciplinas (TAŞKIN et al., 2021).

Em estudo com 5.876 periódicos indexados no JCR de 2002, foi constatado que 82% das publicações possuíam taxas inferiores a 20%, com média geral de 5.87% e mediana de 9%. Esses padrões do JCR parecem não ter variado nos anos seguintes.

Um novo estudo realizado em 2018 com 11.866 periódicos indexados no JCR constatou que aproximadamente 10% possuíam 25% ou mais de autocitação e 5% não possuem nenhuma autocitação. As taxas médias de autocitação, entre 5% e 10%, aumentam na medida em que decresce o quartil das publicações, ou seja, são maiores em publicações de menor impacto (TAŞKIN et al., 2021).

Em estudo histórico com aproximadamente 24 mil periódicos indexados na WoS (GAZNI; DIDEGAH, 2021) constataram que, embora o volume de autocitação tenha aumentado, a média geral da taxa de dois anos caiu de 21% em 1975 para 12% em 2018. Por consequência, a influência das autocitações no resultado do fator de impacto também diminuiu de 13% em 1997 para 9% em 2017.

As taxas de autocitação de 35 periódicos de medicina intensiva variam de 0 a 35.4%, com mediana de 8.8% (SANFILIPPO et al., 2021). As taxas de 85 periódicos de odontologia caíram de 13.725% em 2014 para 10.667% em 2016 (LIVAS; DELLI, 2018). Em ambos os estudos, os periódicos de assuntos específicos em cada disciplina registraram taxas superiores aos de escopo amplo (LIVAS; DELLI, 2018; SANFILIPPO et al., 2021).

Alguns estudos visaram identificar padrões excessivos de autocitação de periódico. Usualmente, taxas superiores a 20% ou 25% são consideradas excessivas, independente a área do conhecimento. Taxas ainda mais extremas podem ser indícios de má prática (TAŞKIN et al., 2021).

Mas taxas de autocitação acima da média da disciplina, por exemplo, não são necessariamente decorrentes de manipulação (YU et al. 2014). Para uma análise mais precisa é recomendado realizar análises em larga escala (FRANDSEN, 2007).

Más práticas de autocitação de periódicos podem ser voluntárias (antecipadas) ou induzidas coercitivamente (CHORUS; WALTMAN, 2016).

No primeiro caso, autores inserem voluntariamente autocitações imaginando que isso pode facilitar o aceite do manuscrito em determinado periódico. No

segundo caso, editores ou revisores exigem ou recomendam fortemente ao autor que cite um ou mais trabalhos do próprio periódico (CHORUS; WALTMAN, 2016; WILHITE; FONG, 2012).

Nessas duas formas, voluntária ou coerciva, a autocitação de periódico tem como consequência direta o aumento artificial (anormal) de indicadores de impacto baseados na frequência de citações recebidas (IOANNIDIS, 2015).

Alguns indícios de más práticas de autocitação já foram identificados. O mais evidente é o aumento brusco de um indicador de impacto de um ano a outro. A concentração de autocitações em um ano específico do período analisado também pode indicar a ocorrência de manipulação (YU et al. 2014; CHORUS; WALTMAN, 2016).

A identificação de alguma forma de vínculo entre os periódicos citante e citado também pode ser considerado má prática, caso o volume das citações seja superior ao das demais citações (IOANNIDIS, 2015).

2.6.1 Índice h_{s5}

Diferentemente do que ocorre com o FI, estudos de autocitação de periódico a partir de dados do índice h são raras. Uma possível razão para essa lacuna é que apenas os artigos mais citados são considerados no cálculo desse indicador (BARNES, 2017). O resultado h desconsidera tanto os artigos que receberam menos citações que o valor h quanto as citações que ultrapassem o valor h nos artigos considerados (BORNMANN, 2014; TEIXEIRA DA SILVA; DOBRÁNSZKI, 2018).

Estima-se que apenas 5% do total de artigos publicados por um autor ou um periódico integram o resultado h . Assim, em tese, as autocitações poderiam não interferir no índice h na mesma intensidade que interfere nos indicadores baseados na média de citações, tais como o FI e o CS (HIRSCH, 2005; WALTMAN, 2016; BARNES, 2017).

Mesmo assim, uma versão do índice h calculado sem autocitações, denominado índice h_{s5} , foi proposta por Schreiber (2009). Essa proposta foi baseada na percepção de que, ao contrário do que imaginou Hirsch (2005), as autocitações podem exercer influência no resultado, especialmente se confrontado com taxas extremas de autocitação (VÍIU, 2016).

O conceito do índice h_s5 pode ser representado pela Tabela 5, que utiliza os mesmos dados da Tabela 4, referente ao índice h tradicional.

Tabela 5 – Estrutura do índice h_s5

	# D	# C	# A	C-A	
	1	30	8	22	
	2	25	6	19	
	3	22	5	17	
	4	22	6	16	
	5	21	7	14	
	6	19	6	13	
	7	18	5	13	
	8	15	7	8	índice h_s
	9	13	5	8	
índice h	10	10	3	7	
	11	9	2	7	
	12	7	3	4	
	13	4	1	3	
	14	0	0	0	
	15	0	0	0	

Onde:
D – Documentos
C – Citações
A - Autocitações

Fonte: Elaborado pelo autor.

Mas a proposta de Schreiber (2009), assim como a maioria dos estudos de autocitação com base no índice h , se referem apenas à autocitação de autor (VÍIU, 2016; TEIXEIRA DA SILVA; DOBRÁNSZKI, 2018).

A influência das autocitações de periódico no índice h , por outro lado, ainda não são bem conhecidas. No caso do GSM, a inexistência de recursos de análise e extração de dados dificultam a realização de estudos mais amplos (CABEZAS-CLAVIJO; LÓPEZ-CÓZAR, 2013). A falta de padronização de títulos (HARZING, 2014) também pode interferir na análise de correspondência dos títulos das fontes citantes e citadas.

Diante disso, a falta de parâmetros de autocitação pode ser um entrave no uso do índice $h5$ em sistemas de avaliação. Periódicos com excesso de autocitações podem obter vantagem na definição de níveis de impacto (rankings, quartis ou estratos) em detrimento de publicações com taxas de autocitação em padrões normais.

2.7 PERIÓDICOS IBERO-AMERICANOS

A ideia de que os países anglo-saxões e os norte europeus ocupam uma posição hegemônica na ciência ocidental possui certo consenso. Como contraponto, alguns estudos tentaram compreender o papel e as características de regiões tidas como periféricas no cenário da produção e da comunicação científica.

Entre essas regiões se inclui a Ibero-Americana, ora estudada como um todo (AGUADO-LOPEZ et al., 2014), ora fragmentada entre as três sub-regiões, a América-Latina (CAMACHO et al., 2019), o Caribe e a península Ibérica (MORENO-GIL et al., 2020).

Os principais aspectos da ciência Ibero-Americana investigados são as características da produção científica, especialmente aspectos temáticos, idiomáticos (MORENO-GIL et al., 2020), a visibilidade regional e internacional (MACHIN-MASTROMATTEO et al., 2017), a predominância do modelo de acesso aberto (MINNITI et al., 2018) e a qualidade e impacto, representados em diferentes indicadores de citação (CHINCHILLA-RODRÍGUEZ et al., 2015, 2016; ROMERO-TORRES et al., 2013; VÉLEZ-CUARTAS; LUCIO-ARIAS; LEYDESDORFF, 2016).

Iniciativas como Latindex e SciELO contribuíram para melhorar o controle de qualidade editorial, favorecendo o aumento da visibilidade regional (CAMACHO et al., 2019). O lançamento do SciELO Citation Index na WoS é visto como um reflexo do aumento de qualidade da ciência da região (MINNITI et al., 2018).

Além disso, o número de periódicos Ibero-Americanos indexados em bases de dados internacionais aumentou substancialmente nos últimos anos (AGUADO-LOPEZ et al., 2014; MACHIN-MASTROMATTEO, 2019).

Se em 2005 havia 86 periódicos Ibero-Americanos indexados na WoS e 388 na Scopus, em 2013 já eram 373 e 1.044, respectivamente (AGUADO-LOPEZ et al., 2014). Nos anos seguintes, a cobertura da Scopus teve poucas variações. A da WoS, por outro lado, registrou um aumento substancial e atualmente é superior à de sua concorrente. Esse crescimento se deve especialmente em razão da criação do ESCI em 2015, coleção que possibilitou a entrada de centenas de publicações regionais na WoS (MACHIN-MASTROMATTEO, 2019).

Além disso, na última atualização de julho de 2021, os periódicos indexados no ESCI e o AHCI foram incluídos no JCR. Essa mudança favoreceu a análise bibliométrica dos periódicos da região, embora o cálculo do FI e de suas duas variações (FIws, FI5y) ainda tenham permanecido restritos ao SCIE e ao SSCI

(CANTO et al. 2022).

Não obstante esse crescimento de visibilidade, algumas barreiras ainda precisam ser superadas. A endogenia é uma característica da região. A prevalência de autores do mesmo país do periódico reduz a internacionalização e a variação dos temas de pesquisa (COLLAZO-REYES; LUNA-MORALES; LUNA-MORALES, 2017).

O Brasil é o país de maior endogenia, com 88% dos artigos em periódicos locais publicados por brasileiros (COLLAZO-REYES, 2014). O alto percentual de artigos publicados em idiomas regionais (português ou espanhol) reduz a visibilidade e, conseqüentemente, o impacto das publicações (CHINCHILLA-RODRÍGUEZ et al., 2015).

As preferências por temas de pesquisa locais e regionais e a baixa qualidade de algumas publicações também são mencionadas como limitadores de um melhor desempenho da região (ROMERO-TORRES et al., 2013).

Os periódicos de menor impacto são aqueles com maior proporção de autocitações e a maioria dos artigos publicados em idiomas regionais. Em contrapartida, os periódicos de maior impacto são aqueles com maior percentual de artigos de autores de fora da região (COLLAZO-REYES et al., 2008, 2017) e os artigos em inglês são os mais citados (CHINCHILLA-RODRÍGUEZ et al., 2015, 2016; COLLAZO-REYES, 2014; COLLAZO-REYES; LUNA-MORALES; LUNA-MORALES, 2017).

Outro padrão da região é a baixa presença de editoras comerciais. Universidades, associações e institutos de pesquisa representam 90% dos periódicos Latino-Americanos no SciELO (VÉLEZ-CUARTAS; LUCIO-ARIAS; LEYDESDORFF, 2016).

Essa é a mesma proporção de periódicos brasileiros indexados no WoS e Scopus que são mantidos por universidades e associações (RODRIGUES; ABADAL, 2014). Embora não se possa dizer que todos os periódicos acadêmicos tenham baixa qualidade, deficiências no financiamento do ensino superior da região podem dificultar o complexo processo de editoração científica.

Além disso, a menor produtividade dos periódicos acadêmicos também reflete na visibilidade internacional, pois WoS e Scopus priorizam as publicações mais produtivas (REPISO; ORDUÑA-MALEA; AGUADED, 2019).

As editoras comerciais tendem a ter estratégias de internacionalização mais eficazes, o que pode aumentar a visibilidade e o impacto das publicações

(COLLAZO-REYES; LUNA-MORALES; LUNA-MORALES, 2017).

Aproximadamente 90% das publicações da SSCI, SCIE e AHCI da WoS são de editoras comerciais, com a maioria dos periódicos publicados por universidades classificadas nos quartis inferiores do JCR (REPISO; ORDUÑA-MALEA; AGUADED, 2019).

Há correlação entre indicadores socioeconômicos dos países da região (PIB, população, extensão territorial, número de pesquisadores) e os indicadores de produção científica da WoS (IBÁÑEZ-MARTI, 2018) e Scopus (COLLAZO-REYES, 2014).

Nesse sentido, a Espanha, o Brasil, o México, a Colômbia, o Chile, a Argentina, Cuba e a Venezuela correspondem a mais de 90% dos indicadores de produção científica da região Ibero-Americana.

A produção científica brasileira é mais da metade do total da América-Latina e quase metade do número de periódicos indexados no WoS (COLLAZO-REYES, 2014; CRESPO-GASCÓN; TORTOSA; GUERRERO-CASADO, 2019; IBÁÑEZ-MARTI, 2018). Os periódicos do país também apresentam a maior produtividade da América Latina, com uma média de aproximadamente 100 artigos por volume (COLLAZO-REYES, 2014). As ciências agrárias e biológicas e as ciências da saúde são as áreas de pesquisa brasileira com maior indexação e impacto internacional (RODRIGUES; ABADAL, 2014).

O México tem o segundo maior volume de produção e número de periódicos indexados em bases de dados internacionais (COLLAZO-REYES, 2014; CRESPO-GASCÓN; TORTOSA; GUERRERO-CASADO, 2019; IBÁÑEZ-MARTI, 2018; VÉLEZ-CUARTAS; LUCIO-ARIAS; LEYDESDORFF, 2016).

As áreas de maior produtividade são ciências da saúde, agrárias e biológicas, física e astronomia e engenharias, áreas que também registram maior colaboração internacional (LANCHO-BARRANTES; CANTÚ-ORTIZ, 2019).

Argentina e Chile têm indicadores significativos de produção científica em bases de dados internacionais e a Colômbia registrou o maior aumento no número de periódicos indexados na WoS nos últimos anos (COLLAZO-REYES, 2014; MACHIN-MASTROMATTEO, 2019).

Venezuela e Cuba completam o grupo de países mais produtivos, embora no caso venezuelano o número de periódicos indexados e os indicadores de produção científica tenham apresentado queda (MACHIN-MASTROMATTEO, 2019).

Os demais países da AL&C, especialmente os da América Central e do Caribe, têm baixos níveis de atividades científicas (MACHIN-MASTROMATTEO, 2019; MACHIN-MASTROMATTEO et al., 2017) e recebem poucas menções em estudos sobre a comunicação acadêmica da região (COLLAZO-REYES, 2014; MINNITI et al., 2018).

A Espanha, por outro lado, tem uma alta visibilidade internacional. Os periódicos espanhóis são quase a metade de todos os periódicos Ibero-Americanos indexados na Scopus. Também são em número significativamente superior na WoS (AGUADO-LOPEZ et al., 2014; NAVAS-FERNÁNDEZ; ABADAL; RODRIGUES, 2018).

Esse desempenho pode ser atribuído em parte à política científica do país, que definiu a internacionalização como prioridade. Aspectos como colaboração com autores estrangeiros, composição internacional de corpo editorial e publicação de trabalhos no idioma inglês foram incentivados (NAVAS-FERNÁNDEZ; ABADAL; RODRIGUES, 2018).

Assim, os periódicos espanhóis foram se desenvolvendo a partir de padrões internacionais de visibilidade e de impacto, especialmente os exigidos pelo JCR e pela Scopus. Indicadores de impacto foram definidos como critérios em sistemas pioneiros de avaliação (IN-RECJ), mesmo de Ciências Humanas e Sociais (IN-RECS e IN-RECH) (NAVAS-FERNÁNDEZ; ABADAL; RODRIGUES, 2018; DE FILIPPO; ALEIXANDRE-BENAVENT; SANZ-CASADO, 2020).

Além disso, algumas características dos periódicos espanhóis diferem dos demais da região. O país possui menor proporção de periódicos acadêmicos e maior de editores comerciais em comparação com o Brasil, por exemplo (RODRIGUES; ABADAL, 2014).

Também possui mais periódicos de acesso pago ou restrito do que as países latino-americanos, que tem o acesso aberto como característica marcante (NAVAS-FERNÁNDEZ; ABADAL; RODRIGUES, 2018; MINNITI et al., 2018).

2.7.1 Sistemas de avaliação

Assim como em outras atividades de interesse público, a ciência está sujeita à procedimentos de avaliação de resultados (MINGERS; MEYER, 2017).

Sistemas nacionais de avaliação científica usualmente são conduzidos por agências ou conselhos de fomento. Processos avaliativos têm a finalidade de

orientar a distribuição de recursos e identificar pontos positivos e negativos em política de gestão científica. Os países definem critérios conforme o estágio de desenvolvimento científico e os objetivos almeçados (VASEN; VILCHIS, 2017).

Entre as dimensões abrangidas em sistemas de avaliação está a da qualidade de periódicos. Uma das vantagens da avaliação de periódicos é a possibilidade de avaliação reflexa de outros elementos, tais como artigos, autores e instituições. É um modelo que não tem a finalidade de avaliar somente periódicos, mas também a de definir padrões de qualidade proporcionais de acordo com os níveis de gestão editorial, de revisão por pares, de visibilidade e de impacto dos periódicos avaliados (GREGORIO-CHAVIANO, 2018).

É possível quantificar o desempenho de um autor, de um laboratório, de um programa de pós-graduação ou de uma universidade por meio do nível de qualidade dos periódicos utilizados para divulgação das pesquisas que conduzem. Quanto mais elevado for o nível do periódico, a tendência é que maior será o rigor científico necessário para que manuscritos submetidos sejam aceitos para publicação (GREGORIO-CHAVIANO, 2018).

São variados os critérios utilizados para avaliação de qualidade de periódicos. Parte deles são intrínsecos à própria publicação, como os relacionados às políticas de gestão e composição do conselho editorial, modelo de revisão por pares, critérios éticos e de direitos autorais, periodicidade e regularidade, estrutura tecnológica, de preservação digital e interoperabilidade (GREGORIO-CHAVIANO, 2018).

Outros critérios são extrínsecos, tais como a indexação em bases de dados (visibilidade) e o nível de impacto (indicadores, quartis). Em alguns casos, a indexação implica automaticamente no preenchimento de um conjunto de critérios intrínsecos, considerando que estes são requisitos obrigatórios para inclusão em determinadas bases de dados (COSTA; CANTO; PINTO, 2020; ALPERIN; ROZEMBLUM, 2017).

No cenário Latino-Americano, a indexação no SciELO e o cumprimento de características do Latindex são utilizadas como padrão de qualidade editorial em alguns sistemas de avaliação (MINNITI et al., 2018; CAMACHO et al., 2019).

Mas nos últimos anos, a excelência dos periódicos da região passou a ser mais atrelada ao impacto, ou seja, relacionada aos resultados de indicadores bibliométricos ou de posições ocupadas em *rankings* de bases de dados, especialmente o JCR e a Scopus (VASEN; VILCHIS, 2017).

A tendência de avaliação por meio de indicadores na região pode ser vista nos sistemas CRMICYT do México, Publindex da Colômbia e REVENCYT da Venezuela. Esses países reformularam suas políticas de avaliação, enfatizando os critérios de visibilidade e de impacto internacional como padrão de excelência de periódicos nacionais (PALACIOS, 2017; VASEN; VILCHIS, 2017).

Com isso, os sistemas dos três países foram ao encontro dos sistemas da Espanha (Fecyt) (DE FILIPPO; ALEIXANDRE-BENAVENT; SANZ-CASADO, 2020) e do Brasil (Qualis) (COSTA; CANTO; PINTO, 2020), que já estipulavam a visibilidade e o impacto internacional como parâmetros de excelência.

Países de menor visibilidade, por outro lado, optaram por manter sistemas baseados em critérios editoriais e de indexação regional, especialmente com base no SciELO, no Latindex e no Redalyc (GREGORIO-CHAVIANO, 2018).

2.7.1.1 Qualis Capes

O Qualis Periódicos é o sistema de classificação de publicações periódicas conduzido pela Capes com a finalidade de subsidiar a avaliação da produção científica brasileira. Constitui-se na atribuição de um conceito para a definição da qualidade de periódicos que publicaram trabalhos científicos de pesquisadores vinculados aos programas de pós-graduação do país. O Qualis utiliza uma plataforma *online* de coleta e visualização de dados denominada Sucupira⁶ (CAPES, 2020).

Os periódicos são avaliados por coordenações de 49 áreas do conhecimento. As áreas são agrupadas em dois níveis, sendo o primeiro nível formado por três colégios (Colégio de Ciências da Vida, Colégio de Humanidades e Colégio de Ciências Exatas, Tecnológicas e Multidisciplinar) e o segundo nível formado por nove grandes áreas (Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Humanas, Ciências Sociais Aplicadas, Linguística, Letras e Artes, Ciências Exatas e da Terra e Multidisciplinar) (CAPES, 2018).

Na grande área Ciências Sociais Aplicadas, por exemplo, a área de avaliação 31 – Comunicação e Informação abrange as áreas básicas de Comunicação, Ciência da Informação e Museologia. As atividades de avaliação de cada área são registradas em relatórios de áreas (COSTA; CANTO; PINTO, 2020).

⁶<https://sucupira.capes.gov.br/>

A avaliação ainda vigente é referente ao período 2013-2016. Os periódicos foram enquadrados em estratos indicativos do nível de qualidade ou impacto científico: A1 e A2, considerados de excelência internacional; B1 e B2 de excelência nacional; B3, B4, B5 e ainda o estrato C, para periódicos não científicos. Cada periódico pode ser avaliado em mais de uma área, considerando a ampla aderência temática (COSTA; CANTO; PINTO, 2020).

Os critérios básicos de classificação eram comuns, mas a definição das publicações nos estratos A1 e A2 variavam de acordo com o desenvolvimento da área, o nível de competitividade científica e número de periódicos existentes.

Em algumas áreas, a indexação em bases de dados, especialmente WoS e Scopus, era requisito suficiente para classificação no estrato A1. Em outras áreas era exigido um nível de impacto mínimo para classificação nos estratos de excelência internacional (JUKEMURA; DINIZ, 2015).

Em 2019 a Capes propôs uma nova metodologia para o Qualis. A estratificação de periódicos manteve os dois grupos (A e B), mas cada um com quatro níveis: A1, A2, A3, A4, B1, B2, B3, B4. O estrato C continuou reservado para periódicos não científicos, e o estrato NP, para publicações não periódicas. A classificação seria em apenas uma área, a chamada área mãe, ou seja, aquela com a maior quantidade de trabalhos publicados pelo periódico no quadriênio (COSTA; CANTO; PINTO, 2020).

A partir dessa metodologia, em meados de 2019 foi divulgada uma lista com os resultados preliminar da avaliação dos periódicos referentes ao período 2017/2019 (avaliação de meio termo)⁷.

Após um período de suspensão da avaliação em razão da Pandemia de COVID-19, bem como atrasos no cronograma da avaliação decorrentes de outras questões, os fundamentos da nova metodologia foram definidos por meio da Portaria nº 145, de 10 de setembro de 2021 (BRASIL, 2021).

Embora a Portaria tenha sido revogada dois meses após a sua edição, algumas definições sintetizaram a forma de escolha geral dos indicadores da nova metodologia.

Primeiramente a Seção III, que mencionava os indicadores utilizados para avaliação de impacto dos periódicos:

⁷Disponível em: <https://www.ufrgs.br/ppgge/ppgge/wp-content/uploads/2019/12/QUALIS-NOVO-1.pdf>

Art. 12. A estratificação far-se-á por meio de utilização de um dos seguintes agrupamentos de indicadores:

I - QR1: uso do CiteScore, obtido a partir da base de dados Scopus, e da base de dados Journal of Citation Report [sic] (JCR), ou, subsidiariamente, dos índices h5 ou h10 gerados por meio da ferramenta de busca Google Metrics;

II - QR2: uso exclusivo dos índices h5 ou h10 gerados por meio da ferramenta de busca Google Metrics;

Para compatibilizar a análise de periódicos a partir de indicadores diferentes, estava prevista a utilização de um modelo de regressão linear (BRASIL, 2021). A regressão é técnica estatística permite estimar o valor aproximado do indicador para uma publicação com base no valor de outro indicador (CHEIN, 2019).

Outra questão positiva da portaria foi estabelecer a necessidade de observância de boas práticas de pesquisa. Entre elas, definir o excesso de autocitações e a citações cruzadas como más práticas (BRASIL, 2021).

Por outro lado, era impossível não ler as imprecisões da Portaria. A primeira foi a definição do JCR como critério, como um todo, sem deixar claro quais dos seus indicadores poderiam ser utilizados.

A orientação para a obtenção do índice h10 a partir do Google [Scholar] Metrics também causou estranheza, já que o GSM não calcula esse indicador.

Esses e outros pontos da normativa da Capes, talvez redigidas por pessoas estranhas ao campo dos estudos métricos, podem ter contribuído para a súbita revogação da portaria.

É provável ainda que a existência de uma Ação Civil Pública questionando a legalidade dos novos critérios do Qualis, especialmente no que tange a aplicação retroativa, seja mais um motivo para os atrasos na conclusão da avaliação.

Por essas e outras questões, as definições descritas na *natimorta* portaria não estão garantidas para esse quadriênio, especialmente as que se referem a utilização do índice h5.

Mas para fins deste tudo, não se poderia deixar de mencionar a tentativa de inclusão do indicador do GSM no maior sistema de avaliação de periódicos região Ibero-Americana,

2.7.1.2 ARCE/Fecyt

O ARCE é o sistema de classificação de periódicos acadêmicos e científicos espanhóis promovido pela *Fundación Española para la Ciencia y Tecnología* (Fecyt).

Até a última convocatória realizada em 2020, 518 periódicos já haviam alcançado o Selo de Qualidade Fecyt.

As duas dimensões de avaliação são o impacto e a visibilidade. A dimensão de impacto representa 80% da pontuação e é dividida em três elementos: citações, índice h e quartis. As fontes de impacto consideradas são a WoS (SSCI, SCIE, AHCI e ESCI), o JCR, a Scopus, o SciELO, o GSM e o ICDS (MIAR).

Conforme mostra a Figura 2, são atribuídos percentuais de ponderação para as fontes de avaliação de impacto, que são categorizadas em quatro níveis.

Figura 2 - Critérios de avaliação de impacto de periódicos da Espanha

CRITERIO	PONDERACIÓN
Citas WoS Citas SCOPUS Citas ESCI Citas Scielo	60% de la puntuación total
Cuartil JCR Cuartil SJR	10% de la puntuación total
h-Index WoS h-Index SJR h5-Index Google S.Metrics	10% de la puntuación total
Índice Compuesto de Difusión Secundaria de MIAR (ICDS)	20% de la puntuación total

Fonte: Fecyt (2021).

A pontuação final vai de 0 a 100 e os periódicos que obtêm o selo são classificados em quartis (C1, C2, C3 e C4) de acordo com 13 categorias temáticas (Arqueologia, Antropologia e Etnologia, Belas Artes, Ciências da Educação, Ciências Jurídicas, Ciências Políticas e Sociologia, Comunicação, Informação e Documentação Científica, Economia, Filosofia, Geografia, História, Linguística, Literatura, Psicologia).

Um dos aspectos positivos do sistema espanhol é considerar diversas fontes para análise de impacto e de visibilidade das publicações, incluindo regionais (SciELO) e alternativas (GSM e MIAR). Por outro lado, a pontuação mais elevada é reservada para as bases de dados comerciais (Scopus, WoS/JCR), o que pode dificultar a obtenção do selo por periódicos não indexados ou sem acesso aos quartis superiores.

Considerando que mais de 90% das publicações avaliadas são das áreas de ciências humanas e sociais (DE FILIPPO; ALEIXANDRE-BENAVENT; SANZ-CASADO, 2020), poderia haver um maior equilíbrio entre as fontes.

Também não há previsão de controle ou de limite para autocitações. Como a pontuação é aferida por contagem de citações, o excesso de autocitações pode

ocasionar distorções. Logo, a ausência de controle pode motivar a adoção de estratégias de aumento de autocitações.

2.7.1.3 CRMICYT/CONACYT

O CRMICYT é o sistema de classificação de periódicos mexicanos do *Consejo Nacional de Ciencia y Tecnología* - CONACYT. O sistema do país é baseado em seis dimensões: política e gestão editorial, qualidade de conteúdo, nível de citação, regularidade, acessibilidade e visibilidade internacional (VASEN; VILCHIS, 2017; MÉXICO, 2020).

A classificação ocorre em oito níveis, sendo os quatro superiores (Q1, Q2, Q3 e Q4) reservados a periódicos indexados na WoS e na Scopus. Os indicadores adotados nesses níveis são o FI, o nES e o total de citações (JCR), e o SNIP, o SJR e o índice h, obtidos por meio do SJ&CR (VASEN; VILCHIS, 2017).

Já nos estratos inferiores, destinados aos periódicos não indexados nas duas bases de dados, a classificação é definida como RCI (periódicos de abrangência internacional), RCN (periódicos de abrangência nacional), REC (periódicos em consolidação) e RED (periódicos em desenvolvimento) (MÉXICO, 2021).

A Figura 3 mostra o resultado da avaliação de um periódico mexicano de Ciências Sociais a partir dos indicadores mencionados.

Figura 3 - Resultado de avaliação de periódico no CRMICYT

Resultado de evaluación

Revista: Investigación Bibliotecológica: Archivonomía, Bibliotecología e Información
 Área: V. Ciencias Sociales
 ISSN: 0187-358X
 Clasificación de la Revista: Q4

Se considera:			
Journal Citation Report (JCR)		Scopus / Scimago Journal & Country Rank	
Journal Impact Factor:	0.176	Source Normalized Impact per Paper:	0.293
Eigenfactor Score:	0.00006	Scimago Journal Rank:	0.141
Total Cites:	103	H Index:	7

México (2020).

Na análise de periódicos não indexados na WoS e na Scopus, é utilizada a contagem de citações por meio de uma pesquisa no GS. A forma de coleta de dados consiste em duas buscas por nome da publicação, com resultados limitados ao período de três anos da respectiva avaliação. Na primeira busca se inclui e na

segundo se excluí as citações dos resultados. Ao final, se subtrai os números de registros encontrados, cuja diferença é o número de citações recebidas (MÉXICO, 2021).

Essa metodologia não parece ser a mais adequada. Primeiramente por depender de um procedimento manual de verificação, algo questionável atualmente, dada a variedade de fontes de indicadores automaticamente calculados. Além disso, a instabilidade do GS nas contagens de citações *web* pode levar a resultados contraditórios, conforme relatado em estudos anteriores (JACSÓ, 2012).

Na ponderação do critério de visibilidade internacional é avaliada a indexação no SciELO, o cumprimento dos critérios do Latindex e o índice h5 do GS, obtido por meio do PoP (MÉXICO, 2021). Assim como no procedimento de contagem de citações diretamente no GS, o uso do PoP para cálculo do índice h5 pode acarretar distorções.

Seria mais recomendado o uso do GSM para obtenção do índice h5, nos moldes dos sistemas do Brasil e da Espanha. O resultado do indicador e os dados de citação são mantidos por um ano, o que confere mais confiabilidade que os métodos manuais referidos.

2.7.1.4 Publindex

O Publindex é o sistema de avaliação de periódicos mantido pelo *Colciencias*, agência de fomento à ciência, tecnologia e inovação da Colômbia. Em 2017 o Publindex atualizou a sua política de classificação, incluindo indicadores de impacto como critérios essenciais de classificação de periódicos colombianos em quatro estratos (A1, A2, B e C) (VASEN; VILCHIS, 2017).

Os estratos A1 e A2 são reservados para os periódicos classificados nos quartis Q1 e Q2 da Scopus e do JCR, respectivamente. Já os estratos B e C são atribuídos aos periódicos classificados no quartis Q3 e Q4. Para periódicos não indexados é utilizado o índice h5 para classificação nos dois estratos mais baixos (COLOMBIA, 2018).

A inserção de indicadores no sistema colombiano representa uma transição de um modelo centrado em aspectos editoriais para um centrado na valorização do impacto. A definição do índice h5 foi uma solução para incorporação do critério de impacto mesmo no caso de periódicos não indexados no JCR e na Scopus,

conforme nota de justificativa apresentada no *Modelo de Clasificación de Revistas do Publindex*:

La incorporación del índice H, permite ampliar significativamente el universo de fuentes académicas de citas para las revistas científicas colombianas, y contribuye a su vez a reducir las limitaciones de los índices citacionales clásicos (JCR y SJR) relacionadas con el predominio del idioma inglés o el abordar objetos de investigación de relevancia local o regional. Por tanto, el índice H se utiliza como una ruta alterna para medir el impacto las revistas nacionales y clasificarlas (COLOMBIA, 2018).

Da mesma forma que o México, o modelo colombiano não definiu o GSM como fonte de índice h5 das publicações, optando pela sua obtenção por meio do PoP. Logo, se aplicam as mesmas preocupações com o uso dessa ferramenta em detrimento obtenção do índice h5 em sua fonte bibliométrica oficial, o GSM.

2.7.1.5 REVENCYT

O REVENCYT é o sistema de classificação de revistas venezuelanas de Ciência e Tecnologia. Foi criado em 1990 para certificar a qualidade editorial das revistas nacionais. A partir de 2015 foi definida uma nova metodologia, que ainda é aplicada atualmente.

Consiste em uma avaliação de cinco fases: 1) ingresso, com análise de critérios básicos de composição da revista; 2) cumprimento de aspectos editoriais, que são quantificados em um sistema pontuação; 3) certificação, para revistas que atingirem uma pontuação mínima e esteja indexada em alguma base de dados; 4) visibilidade regional, considerando a indexação no Redalyc, no SciELO ou no Latindex; 5) visibilidade internacional, considerando a WoS, a Scopus ou o DOAJ.

2.7.1.6 Sistemas de outros países Ibero-Americanos

Alguns países dão menor ênfase aos indicadores de impacto em seus sistemas de avaliação, priorizando os aspectos de qualidade editorial e de visibilidade.

A visibilidade regional assume papel central nesses sistemas, representada especialmente por meio da indexação em bases de dados como o SciELO, o Latindex e o Redalyc.

Entre os países Ibero-Americanos com sistemas nesses moldes se incluem a Argentina, o Chile, a Costa Rica e o Equador.

O Quadro 1 mostra uma síntese dos critérios de visibilidade e impacto dos periódicos Ibero-Americanos.

Quadro 1 – Sistemas de classificação de periódicos Ibero-Americanos

País	Sistema	Aspecto	
		Visibilidade	Impacto
Argentina	Caicyt	SciELO, Latindex	Não avalia
Brasil	Qualis		CS, FI, índices h5 e h10
Chile	Conicyt	SciELO, Latindex	Não avalia
Colômbia	Publindex		Quartis (JCR e Scopus), índice h5 (PoP).
Costa Rica	UCRindex	Latindex	Não avalia
Equador	CACES	Sim, não definido	Não avalia
Espanha	FECYT	WoS, Scopus, GSM	Total de citações, Quartis JCR, índice h5, índice h5 (GSM)
México	CRMCYT	índice h5 (PoP).	FI, nES, total de citações (JCR) SJR: Total de documentos (3 anos), SJR e índice h.
Venezuela	REVENCYT	Latindex, Redalyc, SciELO, WoS, Scopus	Não avalia

Fonte: Elaborado pelo autor.

Questão provavelmente afeta a todos sistemas de avaliação é a necessidade de distinção dos critérios de acordo com os padrões de comunicação de diferentes disciplinas.

Uma das preocupações recai sobre a avaliação das humanidades, grande área do conhecimento que possui características peculiares. Nesse sentido, as características das disciplinas com aderência às humanidades deveriam se refletir nas metodologias de avaliação de periódicos.

2.8 AVALIAÇÃO DE IMPACTO NAS HUMANIDADES

Os parâmetros das bases de dados bibliométricas são mais ajustados para periódicos com elevados níveis de produtividade, com a assimilação e a reprodução de conhecimentos em períodos curtos (alta obsolescência) (WALTMAN, 2017, WALTERS, 2017).

Mas esse padrão varia substancialmente de acordo com a área e a disciplina. Entre as ciências duras (*hard sciences*) as diferenças estão presentes, mas são sutis. Entre as ciências duras e as humanidades (ciências humanas, sociais e artes) as diferenças são marcantes (NEDERHOF, 2006).

Disciplinas como o direito, a linguística, a literatura, a sociologia, a administração pública e a ciência política, entre outras, têm maior orientação temática para questões de interesse local, nacional ou regional (NEDERHOF, 2006; HUANG; CHANG, 2009).

Essa menor internacionalização temática também é idiomática. As humanidades têm maior propensão de uso de idiomas regionais, em detrimento do inglês. Pesquisas das humanidades não utilizam o periódico como principal meio de comunicação. Livros, teses e dissertações são utilizados com mais frequência (GIMÉNEZ-TOLEDO, 2018; HARZING, 2014).

A preferência por essas tipologias documentais se reflete em um tempo mais lento de assimilação de novos conhecimentos (baixa obsolescência). Isso exige um período maior na análise de impacto de citação, entre cinco e dez anos.

Os pesquisadores de humanidades também têm maior propensão à publicação de documentos como autor único, em comparação com uma pesquisa mais colaborativa das ciências duras. Isso tem consequências em contagens de citações em nível de autor (índice h, por exemplo), já que a citação de um documento em coautoria dá crédito a todos os coautores individualmente (BARNES, 2017).

Por fim, os trabalhos das humanidades são direcionados para um público não acadêmico com mais frequência que as demais áreas (NEDERHOF, 2006). Há diferenças, inclusive, nos padrões de escrita de documentos de acordo com o perfil do leitor a que se destinam.

O tratamento diferenciado das humanidades era evidente na política do ISI, mantida pela Thompson, de não incluir os periódicos do A&HCI no JCR. A inclusão da coleção de humanidades na base de dados bibliométrica só ocorreu em 2021, sob uma nítida política de expansão conduzida pela atual controladora, a Clarivate Analytics (CANTO et al., 2022).

Algumas alternativas são apontadas para a avaliação de pesquisas nas humanidades. Parte delas no sentido de substituir a análise de impacto baseada em citações por outros elementos (BARBOSA, 2020).

As propostas mais amplas apontam para um modelo multidimensional que inclua uma análise de impacto não científico (social, econômico), métricas alternativas (redes sociais), colaboração entre outros fatores (BARBOSA, 2020).

No contexto da avaliação de impacto tradicional, isto é, baseada em análise

de citação, as propostas de avaliação das humanidades são direcionadas para o uso de outros indicadores que não os do JCR e da Scopus/SJ&CR.

Muitos autores defendem, nesse sentido, o uso do GS e de suas extensões (GSM e GSC) para avaliação das áreas de artes, humanidades e ciências sociais (BORNMANN et al., 2016; MARTÍN-MARTÍN; ORDUNA-MALEA; LÓPEZ-CÓZAR, 2017).

2.9 JOURNAL SCHOLAR METRICS

Uma proposta interessante para avaliação das humanidades é o JSM, base de dados bibliométrica alimentada com dados de citação do GSM. São mais de 9.000 periódicos agrupados em duas grandes áreas, Artes & Humanidades e Ciências Sociais, e em nove e 13 categorias de assuntos, respectivamente.

Os indicadores do JSM são o índice h5 e a mediana h5 extraídos do GSM. A partir desses indicadores foram calculados o total de citações recebidas, o índice h5 e a mediana h5 com exclusão de autocitações e, ainda, a taxa de autocitação por periódico (MARTÍN-MARTÍN; ORDUNA-MALEA; LÓPEZ-CÓZAR, 2017).

Não foram realizadas análises mais profundas a partir dos dados do JSM, como, por exemplo, identificação de taxas de autocitações por país ou por disciplina.

Além disso, as técnicas ou ferramentas utilizadas para a extração e análises de dados não foram divulgadas, o que impede a reprodutibilidade e provavelmente, impediu a própria atualização dessa base de dados.

Até porque o JSM foi planejado para receber atualizações anuais, mas acabou não sendo atualizado, permanecendo até hoje com seus dados relativos à edição de 2015 do GSM.

2.10 PUBLISH OR PERISH

O Publish or Perish⁸ é uma ferramenta bibliométrica gratuita desenvolvida por Harzing (2007). Se presta à extração e à análise de dados de citação do GS, bem como de outras bases de dados científicas. O PoP tem sido utilizado em estudos métricos envolvendo dados do GS, tendo em vista a ausência de recursos de extração e análise próprios nesse sistema (HARZING; VAN DE WAL, 2008; JACSÓ,

⁸<https://harzing.com/resources/publish-or-perish>

2009; AMARA, LANDRY, 2012; MINGERS; MEYER, 2017; MARTÍN-MARTÍN *et al.*, 2018).

Atualmente, o PoP está em sua versão de nº 8 e pode ser instalado nos sistemas operacionais Windows, Mac e Linux. A ferramenta extrai dados do GS, o Microsoft Academic, o CrossRef, o PubMed, o WoS e Scopus. É possível realizar-se buscas por título do periódico, ISSN, título do artigo e nome de autor. Pode-se delimitar também o período da publicação (JACSÓ, 2009).

Os campos extraídos pelo sistema são ordem, número de citações recebidas, autoria, título, ano, fonte, domínio, link. O sistema calcula automaticamente nove métricas: total de artigos e total de citações; número médio de citações por artigo; número de citações por autor; número de artigos por autor e citações por ano; índice h, índice g, o índice h contemporâneo, e três variações do índice h contemporâneo (JACSÓ, 2009).

É possível efetuar correções nos dados, desmarcando registros considerados errados ou duplicados. Por fim, há a possibilidade de cópia e extração dos dados nos formatos RIS, Bibtex, CSV e Excel.

Por outro lado, o PoP não identifica ou exclui autocitações nas métricas calculadas. Também é limitado a requisições individuais de busca, ou seja, não faz busca por listas de publicações. Isso dificulta seu uso em estudos em larga escala. Usualmente, são necessárias etapas e limpeza de dados para uma aferição mais confiável dos indicadores gerados por essa ferramenta (PAVAS; ARZOLA DE LA PEÑA, 2020).

O México e a Colômbia já definiram o PoP como ferramenta de obtenção do índice h5 de periódicos avaliados em seus sistemas de classificação (ALPERIN; ROZEMBLUM, 2017). O Brasil pretende utilizá-lo apenas nos casos de periódicos não indexados no GSM (COSTA; CANTO; PINTO, 2020).

Recomenda-se, portanto, o monitoramento do desempenho do PoP, pois quando utilizado a partir dos dados do GS, tende a replicar os erros já existentes ou ainda ampliá-los (JACSÓ, 2009).

3 DEMONSTRAÇÃO DE ORIGINALIDADE

A demonstração de originalidade da pesquisa foi considerada para o caso da não identificação de estudo que atendesse, simultaneamente, os seguintes critérios de similaridade temática e metodológica:

- a) tratar de avaliação de impacto de periódicos a partir do índice h5 do GSM;
- b) propor modelo, metodologia, técnica, procedimento ou *framework* para análise de impacto de periódicos regionais;
- c) utilizar ferramentas, métodos ou técnicas, próprias ou já utilizadas, para análise de dados de citação do índice h5 do GSM, inclusive com a identificação de autocitações e o cálculo do índice h5 com exclusão de autocitações.

Para busca de artigos científicos relacionados ao tema de pesquisa foram selecionadas a WoS e a Scopus, bases de dados referenciais de escopo multidisciplinar amplamente utilizadas em revisões de literatura.

Foi definida a seguinte estratégia de busca:

((("google scholar metrics") OR (h5-index OR "h5 index")) AND (evaluation OR assessment OR ranking OR rating OR classification OR quality) AND (journa* OR publicatio* OR periodica* OR magazin*)) OR ((h5-index OR "h5 index") AND (bibliometric OR "citation analysis") AND (tool OR software OR system OR database OR engine))

A estratégia foi aplicada nos campos título, resumo e palavras-chave dos documentos. Foram considerados documentos publicados de 2012, ano de lançamento do GSM, até 2020, ano de qualificação do projeto de tese e, conseqüentemente, de início formal de realização da pesquisa doutoral.

Na WoS e na Scopus foram incluídos artigos, artigos de revisão, artigos de conferências e capítulos de livros. Cartas, materiais editoriais e outros tipos de documentos foram desconsiderados.

Inicialmente foram localizados 46 documentos na WoS e 61 na Scopus, totalizando 107 registros. Os registros foram exportados para o gerenciador bibliográfico EndNote Web. Foram eliminadas 41 duplicatas, resultando em um conjunto de 64 documentos únicos. Foram também excluídos dois artigos vinculados a esta tese publicados preliminarmente (COSTA; CANTO; PINTO, 2020; PINTO et al; 2020).

Foi realizada a leitura dos títulos e dos resumos dos documentos e aplicados os seguintes critérios de exclusão: 1) estudos não relacionados ao tema de pesquisa; 2) estudos que incluam o índice h5 ou o GSM de forma superficial ou restritos à análise de impacto em nível de autor a partir do índice h.

Foram excluídos 22 documentos com base no critério de exclusão 1 e 25 com base no critério de exclusão 2.

Foi então realizada a leitura do texto completo dos 17 documentos restantes com ênfase em três aspectos: objetivos, metodologia e resultados.

Foi constatado que estudos pioneiros sobre o GSM investigaram o seu funcionamento, usabilidade, cobertura, recursos, falhas e limitações, tanto do próprio GSM, como do índice h5 na condição de principal indicador (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012; JACSÓ, 2012; PABLOS-COELLO; MARTÍN; LÓPEZ, 2013; MARTÍN-MARTÍN; AYLLÓN; ORDUÑA-MALEA, 2014; ORDUNA-MALEA; LOPEZ-COZAR, 2014).

Parte desses estudos compararam os resultados do índice h5 com os obtidos a partir de indicadores de outras fontes, especialmente WoS/JCR, Scopus, SJ&CR (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012; MINASNY et al., 2013; RODRÍGUEZ-MORALES, 2014; WALTERS, 2016; WALTERS, 2017; CASTELLÓ-COGOLLOS et al., 2018).

Outro conjunto de estudos se ocuparam de questões relacionadas ao tamanho e à cobertura do GSM, comparando seu mecanismo de indexação automatizada com o de outras bases de dados, tanto de forma geral (GU; BLACKMORE, 2017) como delimitando por área do conhecimento (DELGADO; REPISO, 2013) ou país (RODRÍGUEZ-MORALES; OCHOA-OROZCO; TRISTÁN, 2014; ZARIFMAHMOUDI; JAMALI; SADEGHI, 2015; PABLOS-COELLO; MARTÍN; LÓPEZ, 2013).

De forma geral, os estudos fixaram os aspectos pertinentes e as lacunas a serem exploradas sobre o tema. Foram baseados no índice h5 a partir do resultado apresentado na edição corrente do GSM, sem análises longitudinais e raramente em larga escala. São pontados os aspectos positivos no uso de GSM como fonte de avaliação. Também são mencionados os problemas e as limitações ocasionadas pela simplicidade do sistema (LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012; JACSÓ, 2012; MARTÍN-MARTÍN; AYLLÓN; ORDUÑA-MALEA, 2014; ORDUNA-MALEA; LOPEZ-COZAR, 2014).

Logo, a maioria dos trabalhos relacionados (CASTELLÓ-COGOLLOS et al., 2018; GU; BLACKMORE, 2017; JACSÓ, 2012; LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012; MARTÍN-MARTÍN; AYLLÓN; ORDUÑA-MALEA, 2014; MINASNY et al., 2013; ORDUNA-MALEA; LOPEZ-COZAR, 2014; PABLOS-COELLO; MARTÍN; LÓPEZ, 2013; RODRÍGUEZ-MORALES, 2014; RODRÍGUEZ-MORALES; OCHOA-OROZCO; TRISTÁN, 2014; WALTERS, 2016; WALTERS, 2017; ZARIFMAHMOUDI; JAMALI; SADEGHI, 2015) atendem ao critério “a” de similaridade.

Mas como não propõem modelos, metodologias ou frameworks, tampouco dão tratamento às autocitações no índice h5, não alcançam os critérios “b” e “c” de similaridade, não atingindo a originalidade desta tese.

Três estudos, entretanto, exigiram uma análise mais minuciosa (LOPEZ-COZAR; CABEZAS-CLAVIJO, 2013; REGENBRECHT; LANGLOTZ, 2015; CAMACHO et al. 2019), pois foram considerados relacionados.

Entre as diversas pesquisas conduzidas pelos autores sobre este tema, uma em especial tem como objetivo verificar a possibilidade de uso do GSM como fonte de análise de impacto alternativa aos tradicionais JCR e SJ&CR (LOPEZ-COZAR; CABEZAS-CLAVIJO, 2013). É realizada uma comparação entre as três fontes, abrangendo os aspectos de cobertura, políticas de indexação, busca e visualização, indicadores bibliométricos, opções de análise de resultados, custo econômico e diferenças na classificação dos periódicos. O estudo forneceu um amplo diagnóstico de funcionamento GSM, evidenciando as suas particularidades, seus méritos e deficiências.

Entretanto, foi um estudo eminentemente analítico e não propositivo. Logo, não atendeu aos critérios “b” e “c” de similaridade.

Regenbrecht e Langlotz (2015) apresentam um sistema de classificação de anais de eventos e periódicos da área de Interação Humana-Computador e disciplinas correlatas baseado no resultado do índice h5 do GSM. A razão alegada foi a inadequação do FI e de outros sistemas para avaliação de impacto desse escopo temático. O sistema denominado “CHINZ ranking” prevê a classificação de publicações em quatro estratos (A*, A, B e C) de acordo com o valor do índice h5 (30+, 20-29, 10-19, 0-9). O sistema foi testado a partir de uma lista de 272 publicações (*master list*) obtida na base de dados *ArnetMinder*. A busca de publicações foi realizada manualmente no GSM, com tentativas de variações do título e abreviaturas. Como resultado, foram localizadas 120 publicações (44,1%),

que foram classificadas nos quatro estratos de acordo com o índice h5 identificado.

Em síntese, o estudo analisado propõe um modelo de classificação de publicações inteiramente baseado no índice h5. Atende aos critérios “a” e “c” de similaridade, embora tenha como objeto uma disciplina específica. Por outro lado, a inexistência de uma técnica de busca e de extração automatizada dificulta a aplicação desse modelo em larga escala. Além disso, qualquer análise de impacto que ignore a existência de autocitações está sujeita a resultados contraditórios. Diante disso, o estudo de Regenbrecht e Langlotz (2015) não atende ao critério “c” de similaridade, não ofuscando a originalidade dessa tese.

Camacho et al. (2019) propõem um modelo de análise de impacto de periódicos Latino-Americanos indexados no Latindex a partir do índice h5. Desenvolveram uma forma de cálculo para estimar o índice h na Scopus a partir de periódicos que só possuam o índice h5 do GSM. Esse estudo é convergente com o desenvolvido nesta tese. Além das duas fontes principais em comum (Latindex e GSM) a delimitação geográfica é bastante próxima.

Camacho et al. (2019) também apresentam um modelo para estimar o valor de outros indicadores a partir do índice h5. Por outro lado, a técnica de busca de periódicos no GSM não é revelada, a qual se supõe que tenha sido feito manualmente. O modelo também não faz menção à autocitação, questão afeta em sistemas de avaliação com base no índice h5.

Logo, a pesquisa alcança os critérios “a” e “c” de similaridade, o que é insuficiente para atingir a originalidade desta tese.

Para busca de teses e dissertações foram selecionadas a BDTD, de abrangência nacional e a *ProQuest*, de abrangência internacional.

Considerando as diferentes configurações entre os motores de busca das fontes de teses, a estratégia de busca aplicada na WoS e na Scopus não obteve resultados satisfatórios, considerando o conceito de precisão na recuperação da informação.

Foi então definida uma estratégia mais abrangente, representada pela busca do termo “*Google Scholar Metrics*” no resumo dos documentos.

Como resultado, foi encontrada um registro na Proquest e seis na BDTD, sendo um duplicado. Da leitura dos títulos e dos resumos dos documentos, foi possível constatar as pesquisas conduzidas não tinham o GSM ou o índice h5 como objetos principais de investigação. Apenas os relacionaram como parâmetros de

análise de impacto, dentre outras fontes e indicadores bibliométricos existentes.

Diante disso, nenhuma das teses e dissertações localizadas alcança os três requisitos de similaridade definidos, não obstante a originalidade desta pesquisa.

4 METODOLOGIA

4.1 CARACTERIZAÇÃO DA PESQUISA

Pesquisas descritivas se destinam à resolução de problemas a partir da análise de fenômenos e de suas relações com o ambiente na qual estão inseridos. Visam obter a maior precisão possível, evitando interferências tanto do pesquisador, quanto as técnicas e dos instrumentos utilizados na pesquisa. Se baseiam em levantamentos das propriedades, das características ou dos perfis da população, do processo, do objeto, do fato ou do fenômeno investigado (MICHEL, 2009; HERNANDEZ-SAMPIERI, FERNANDEZ-COLLADO; BAPTISTA-LUCIO, 2013).

Além da descrição do objeto principal, as pesquisas de alcance descritivo também se baseiam na identificação de possíveis variáveis as quais o objeto possa se relacionar. Podem incluir medições e coletas de dados, do objeto principal e de suas variáveis (MICHEL, 2009; HERNANDEZ-SAMPIERI, FERNANDEZ-COLLADO; BAPTISTA-LUCIO, 2013).

A presente pesquisa, portanto, se enquadra no conceito de pesquisa de alcance descritivo. O tema de pesquisa é a utilização de indicadores de impacto na avaliação de atividades científicas. Delimitado, o tema atinge a avaliação de periódicos Ibero-Americanos a partir do índice h5 do GSM, tanto em sistemas de classificação de periódicos como em estudos métricos em larga escala.

São realizados levantamentos de dados do objeto de estudo (índice h5) da população (periódicos Ibero-Americanos), e das variáveis as quais eles se relacionam (características editoriais, países, áreas, dados de citação, entre outros).

Pesquisas descritivas, em muitos casos, são classificadas também como de alcance exploratório. Essa definição pode ocorrer quando o objetivo é examinar um tema ou problema de pesquisa pouco explorado, sobre o qual ainda pairam muitas dúvidas. Esse cenário de incerteza exige a realização de etapas prévias de exploração e de levantamento de dados. São identificados os conceitos sob os quais o objeto de estudo se fundamenta, visando melhor conhecê-lo e descrevê-lo (HERNANDEZ-SAMPIERI, FERNANDEZ-COLLADO; BAPTISTA-LUCIO, 2013).

Mas esse não é o cenário observado na presente pesquisa. Desde o seu lançamento, o GSM vêm sendo amplamente estudado e avaliado. Seu funcionamento, seus recursos e suas limitações já bem são conhecidos. Essa

pesquisa já parte de um conhecimento de base consolidado. Logo, se caracteriza como uma pesquisa descritiva, mas não exploratória.

Já no que tange ao método de abordagem, a pesquisa se classifica como qualiquantitativa, também chamada de método misto. Como o nome indica, pesquisas de método misto combinam as abordagens qualitativa e quantitativa em uma única investigação científica. Visam a resolução do problema de pesquisa de forma mais abrangente do que seria possível com a utilização de uma abordagem única (CRESWELL, 2007).

A escolha do método misto tem sido observada em pesquisas de diversas áreas do conhecimento, especialmente nas ciências humanas e sociais. É um tipo de abordagem que viabiliza a coleta e a análise de dados qualitativos e quantitativos de maneira integrada e sistemática (CRESWELL, 2007).

Em síntese, o método misto pode utilizar dados de origem numérica ou estatística, extraído de bases de dados ou outras fontes quantitativas. Mas ao mesmo passo, pode utilizar dados verbais, textuais, visuais, simbólicos ou de outros tipos caracterizados como qualitativos (HERNANDEZ SAMPIERI; FERNANDEZ COLLADO; BAPTISTA LUCIO, 2013).

No contexto do método misto, a abordagem quantitativa é destinada à análise e à apresentação de dados sobre as características e o impacto dos periódicos analisados. São utilizadas técnicas estatísticas para tratamento de dados numéricos e análise de variáveis. São ainda utilizados recursos de visualização de dados para facilitar a compreensão dos resultados.

Já a abordagem qualitativa terá como enfoque o alcance do objetivo geral de pesquisa referente à proposta de *framework* de utilização do índice h5 do GSM. Além do aspecto quantitativo, a abordagem também será subsidiada por uma análise qualitativa, relativo a esquemas, critérios de classificação e categorização que transcenderão as formulações matemáticas típicas de abordagens estritamente quantitativas.

4.2 PROCEDIMENTOS METODOLÓGICOS

Os procedimentos metodológicos são explicados em quatro etapas: apresentação do *framework*, delimitação do universo da pesquisa, exposição do procedimento de coleta de dados e descrição das técnicas de tratamento e de análise dos dados.

4.2.1 Framework *Gsm@hdata*

Essa pesquisa é orientada por conceitos de análise de citação, em especial os relacionados ao uso de indicadores para avaliação de impacto de periódicos. Abrange também o desenvolvimento de recursos que permitam a aplicação prática desses conceitos diante da ausência de recursos suficientes para o alcance dos objetivos propostos.

Essa estrutura de procedimentos de pesquisa, que alia aspectos teóricos e aplicações práticas, pode ser caracterizada como um *framework*.

Um *framework* é considerado uma estrutura menos desenvolvida que a de uma teoria. É formado por sentenças que relacionam conceitos abstratos à dados empíricos. *Frameworks* conceituais são desenvolvidos para explicar ou descrever fenômenos que ocorrem sob condições semelhantes (LESHEM; TRAFFORD, 2007).

Considerando a ampla aplicação, não é raro o uso de *frameworks* na elaboração de teses de doutorado na área de Ciências Sociais (LESHEM; TRAFFORD, 2007). São utilizados inclusive em estudos métricos envolvendo o uso de indicadores de impacto de citação (MOED, 2015). Além disso, um *framework* abrange todo o sistema de avaliação de qualidade de pesquisas de ensino superior do Reino Unido, o *Research Excellence Framework – REF* (STOCKHAMMER; DAMMERER; KAPUR, 2021).

Se constata que *frameworks* são utilizados em contextos análogos ao dessa pesquisa.

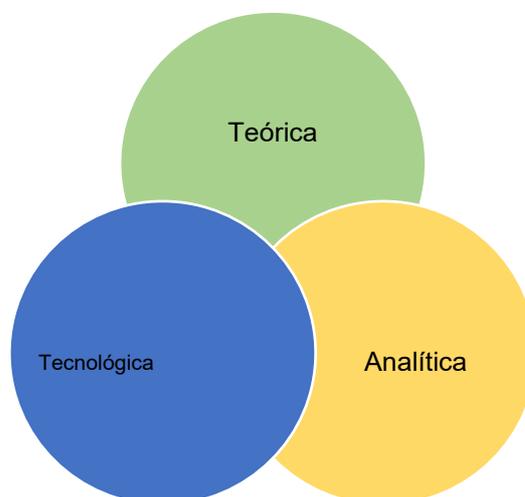
Diante disso, é proposto o framework *Gsm@hdata* para sistematização de procedimentos de avaliação de impacto de periódicos a partir do índice h5 do GSM. O *framework* é composto de três dimensões, sendo uma dimensão teórica, uma dimensão tecnológica e uma dimensão analítica, conforme mostra a Figura 4.

A dimensão teórica é sedimentada na aplicação de conceitos de análise de citação, especialmente na sua aplicação na mensuração de impacto de periódicos científicos. A dimensão tecnológica é representada por um *software* bibliométrico desenvolvido especificamente para extração e análise de dados do GSM.

Por fim, a dimensão analítica do *framework* se ocupa da interpretação dos dados extraídos na dimensão tecnológica com base nos conceitos introduzidos pela dimensão teórica. Essa dimensão tem as finalidades de definir critérios e de orientar procedimentos de tomada de decisão em sistemas de avaliação de periódicos ou

ainda em estudos métricos que compreendam a avaliação de impacto científico em larga escala.

Figura 4 – Representação do *framework* *Gsm@hdata* em diagrama de Venn



Fonte: Elaborado pelo autor.

Cada uma das três dimensões do *framework* é explicada em detalhes a seguir.

4.2.1.1 Dimensão teórica do framework

A dimensão teórica do *framework* *Gsm@hdata* se refere à aplicação de conceitos e fundamentos próprios de análise de citação (*citation analysis*). Possui enfoque na caracterização da citação como elemento de representação do impacto científico por meio de indicadores (*citation-based indicators*) (WALTMAN, 2016).

Essa dimensão teórica inclui ainda variáveis frequentemente utilizadas em estudos de bibliometria e cientometria envolvendo avaliação de periódicos. Entre essas variáveis se destacam a de impacto (frequência de citações recebidas; autocitações), a de produtividade (frequência de trabalhos publicados), a de visibilidade (presença em bases de dados) e a de aderência ou escopo temático (classificação em áreas do conhecimento ou temas de pesquisa).

A dimensão teórica engloba o conhecimento de conceitos básicos de estatística descritiva, usuais em análises de dados em estudos métricos. Aplica os conceitos básicos de análise de redes sociais à fenômenos da comunicação científica, tais como as relações entre temas de pesquisa e entre fontes de citações recebidas por periódicos.

O Quadro 2 relaciona os principais conceitos abrangidos pela dimensão teórica do *framework* proposto nessa pesquisa.

Quadro 2 – Conceitos da dimensão teórica do *framework* *Gsm@hdata*

Conceito original	Conceito traduzido	Definição	Fontes
<i>h-index</i>	Índice h	Número <i>h</i> de documentos publicados que tenham recebido no mínimo o número <i>h</i> de citações	Hirsch (2005)
<i>h5-index</i>	Índice h5	Número <i>h</i> de documentos de uma publicação que tenham recebido no mínimo o número <i>h</i> de citações, no período de cinco anos	GSM; Braun, Glänzel e Schubert (2006)
<i>h_s-index</i>	Índice h _s	Número <i>h</i> de documentos publicados que tenham recebido no mínimo o número <i>h</i> de citações, excluídas as autocitações	Schreiber (2009)
<i>Journal self-citation</i>	Autocitação de periódico	Identidade de fonte entre os documentos citante (principal) e citado (presente na lista de referência)	Chorus e Waltman (2016)
<i>Self-cited rate</i>	Taxa de autocitação	Percentual de autocitação entre o total de citações recebidas	Rousseau (2009)
<i>Pearson coefficient</i>	Coeficiente de <i>Pearson</i>	Medida de correlação entre dois conjuntos de variáveis de acordo com a os valores por elas assumidos	Salvador-Oliván e Agustín-Lacruz (2015)
<i>Spearman coefficient</i>	Coeficiente de <i>Spearman</i>	Medida de correlação entre dois conjuntos de variáveis de acordo com a posição por elas ocupadas	Elkins et al. (2010); Salvador-Oliván e Agustín-Lacruz (2015)
<i>Strength of the relationship between variables</i>	Interpretação de coeficientes de correlação	Definição dos níveis de forçados coeficientes de correlação	Akoglu (2018)
<i>Linear regression</i>	Regressão linear	O modelo para estimar os valores de uma variável (explicada) a partir de outra (explicativa)	Thelwall e Wilson (2014)
<i>Outlier</i>	Ponto fora da curva	Valor que difere drasticamente do padrão de distribuição de um conjunto de variáveis	Szomszor et al. (2020)
<i>Citation network analysis</i>	Análise de rede de citação	Análise de relações entre documentos citantes e citados em uma rede de citações	Moed (2010)

Fonte: Elaborado pelo autor.

Parte desses conceitos orientaram a construção de uma ferramenta bibliométrica, o instrumento que caracteriza a dimensão tecnológica do *framework*.

4.2.1.2 Dimensão tecnológica do *framework*

A dimensão tecnológica do *framework* do *Gsm@hdata* visa enfrentar os problemas computacionais causados pela ausência de recursos de busca, de extração e de análise de dados do GSM. A dimensão tecnológica é representada pela ferramenta bibliométrica *Gsm_hdata*.

O *Gsm_hdata* foi idealizado e projetado por este pesquisador e codificado pelo Prof. Dr. Marcos Talau⁹, da Universidade Tecnológica Federal do Paraná (UTPFR), por meio de um projeto informal de colaboração técnica e científica.

O código fonte do *Gsm_hdata* foi escrito em linguagem Python 3.0, para ser executado a partir do sistema operacional *Linux Ubuntu* nativo. Pode também ser executado por meio de aplicações do tipo *Virtual Machine*, aplicação que roda um sistema Linux de forma virtual em outros sistemas operacionais (MacOS; Windows).

O *Gsm_hdata* realiza rotinas computacionais a partir de dados web disponíveis em endereços URL do GSM.

Utiliza a extensão *GeckoDriver* para operação da biblioteca *Selenium Webdriver*. Essa extensão possibilita a simulação de requisições realizadas por humanos a partir de um navegador *Firefox*. As requisições são temporizadas em períodos aleatórios entre 30 e 60 segundos, aproximadamente. Essas configurações têm o objetivo de ocultar a origem automática das requisições e evitar bloqueios de acesso.

As requisições podem ser feitas a partir de qualquer tipo de conexão de Internet. Para coleta dos dados de pesquisa, entretanto, o acesso foi realizado por meio da rede VPN UFSC. Esse caminho de rede indica a origem acadêmica da conexão e contribui para justificar o grande volume de requisições em longos períodos.

O *Gsm_hdata* é composto por quatro recursos com funcionalidades distintas, mas complementares. São eles o *Gsm_hdata*, o *Gsm_hscite*, o *Gsm_hcitation* e o *Gsm_htline*.

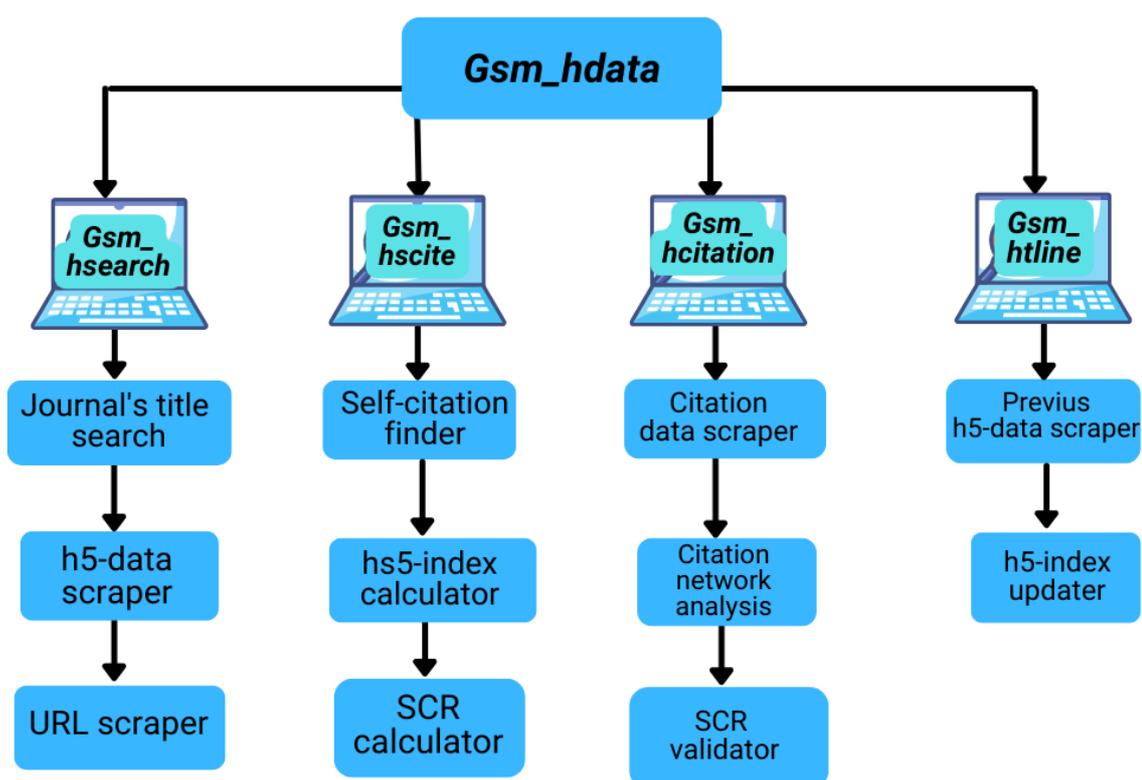
O *Gsm_hdata* é o recurso de pesquisa e de extração de dados básicos de indexação e de impacto no GSM. O *Gsm_hscite* é o recurso de identificação e de análise de autocitações. O *Gsm_hcitation* é o recurso de extração de dados de

⁹ Lattes: <http://lattes.cnpq.br/4522157219119759>. Orcid ID: <https://orcid.org/0000-0001-7027-1023>

citação dos artigos do índice h5. E o *Gsm_htline* é o recurso de extração de dados de edições anteriores do GSM.

A Figura 5 mostra a estrutura do *Gsm_hdata*, seus quatro recursos e respectivas aplicações.

Figura 5 – Estrutura da ferramenta bibliométrica *Gsm_hdata*



Fonte: Elaborado pelo autor.

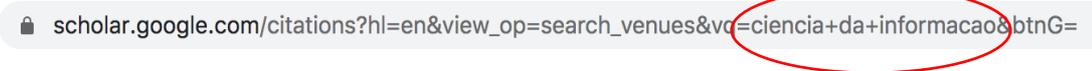
O *Gsm_hdata* desempenha funções que são básicas na maioria das bases de dados científicas, mas que não estão disponíveis no GSM. O seu funcionamento foi configurado para uso no *framework Gsm@hdata*, mas nada impede sua execução em outros contextos que empreguem dados do GSM.

4.2.1.2.1 *Gsm_hdata*

O *Gsm_hsearch* é o primeiro recurso da ferramenta *Gsm_hdata*. Executa requisições de busca e extraí dados básicos (título, índice h5, mediana h5 e URL) de periódicos indexados no GSM.

Os dados de entrada são os títulos dos periódicos, dispostos linha a linha em um documento de texto. O recurso efetua requisições contínuas de busca dos títulos no respectivo no campo do endereço URL do GSM, conforme mostra a Figura 6.

Figura 6 – Busca de títulos no endereço URL do Google Scholar Metrics

 scholar.google.com/citations?hl=en&view_op=search_venues&vc=ciencia+da+informacao&btnG=

Fonte: GSM (2022).

O mecanismo de busca do *Gsm_hsearch* é configurado com uma cadeia de caracteres a serem desconsiderados (*stopwords*), ou seja, palavras irrelevantes. Essas palavras são restritas aos idiomas inglês, português e espanhol. Correspondem aos termos mais utilizados em títulos de periódicos, tais como, “*journal*”, “*international journal*”, “*revista*”, “*revista internacional*”. Também há tratamento de pontuação e de artigos e preposições nos três idiomas.

Os dados de saída são divididos em quatro arquivos de texto com campos separados por ponto-e-vírgula. Dois arquivos apresentam os resultados encontrados e dois apresentam os resultados não encontrados. O arquivo principal apresenta a lista os termos buscados e os resultados em três padrões: “encontrado”, “contido em:”, e “não encontrado”.

O padrão “encontrado” corresponde aos casos de correspondência entre o termo buscado e o termo encontrado. O termo buscado, o título encontrado e os valores do índice h5 e da mediana h5 são apresentados respectivamente na seguinte ordem: *ciencia da informacao; Ciência da Informação; 6; 8*.

O padrão de resultado “contido em:” corresponde aos resultados que contêm o termo pesquisado inserido em uma expressão maior. Neste caso, o sistema recupera os dados dos registros nos quais o termo de busca está contido, para posterior verificação de correspondência: *ciencia da informacao; contido em; Perspectivas em Ciência da Informação; 12; 19*.

Por fim, o padrão “não encontrado” retorna resultado zero para os três campos: *ciencia da informacao; 0; 0; 0*.

Um segundo arquivo com resultados encontrados inclui os endereços URL das publicações no GSM. Mas neste caso somente com as publicações do padrão de resultado “encontrado”. Os endereços URL são relevantes no *Gsm_hdata*, pois se tornam os dados de entrada dos demais recursos. Além disso, são utilizados como elemento de identificação de periódicos em planilhas, indicados para tratamento de dados (função *procv*, por exemplo) já que geram uma identidade única.

Os dados de saída também incluem a emissão de um relatório de resultados não encontrados.

4.2.1.2.2 *Gsm_hscite*

O *Gsm_hscite* é o recurso de análise de autocitação da ferramenta *Gsm_hdata*. Executa as funções de identificar autocitações a partir de dados de citação do GSM e a de recalculá-los os indicadores de impacto. Esse recurso calcula quatro indicadores a partir das autocitações identificadas: o índice h_5 (SCHREIBER, 2009) e a mediana h_5 ; a taxa de autocitação por artigo do índice h_5 e a taxa de autocitação por periódico.

O *Gsm_hscite* pode ser executado em categorias de assuntos (*category*) ou periódico (*journal*). Os dados de entrada são os endereços URL de categorias de assunto ou de periódicos do GSM. No primeiro caso, são analisadas as 20 publicações indexadas em cada categoria de assunto requisitada. No segundo, são analisados periódicos dispostos em uma lista, cujas URL podem ter sido extraídas com o *Gsm_hsearch* ou copiadas da interface do GSM.

O *Gsm_hscite* é configurado para identificar o título principal e eventuais títulos equivalentes, considerando a não padronização de títulos no GSM. Essa configuração aumenta a sua precisão.

Títulos equivalentes podem ser indexados na lista de artigos do índice h_5 . A Figura 7 descreve o funcionamento do *Gsm_hscite* em um periódico com três variações para o título. São identificados o (1) título principal, os (2) (3) títulos equivalentes e o (4) índice e mediana h_5 .

Figura 7 – Identificação de título equivalente no Google Scholar Metrics

Revista ICONO14 Revista Científica de Comunicación y Tecnologías Emergentes			
1	4	h_5 -index:16 h_5 -median:36	
		Title / Author	Cited by Year
		Consumo mediático de adolescentes y jóvenes. Noticias, contenidos audiovisuales y medición de audiencias	67 5 2018
	2	AG Jiménez, VT Viñes, YP Ruiz ICONO 14, Revista de comunicación y tecnologías emergentes 16 (1), 22-46	
		Understanding the customer experience in the age of omni-channel shopping	66 2017
	3	A Mosquera, CO Pascual, EJ Ayensa Icono14 15 (2), 5	
		Estrategias de contenido con famosos en marcas dirigidas a público adolescente	63 2016
		AC Martínez, C del Pino Romero, VT Viñes Revista ICONO14 Revista científica de Comunicación y Tecnologías emergentes ...	

Fonte: GSM (2022).

Em um segundo momento, a lista de artigos citantes de cada artigo do h5-core é acessada (5). Os títulos (principal e equivalentes) do periódico são comparados com os títulos das fontes citantes, registrando a ocorrência de autocitação de periódico (6), conforme mostra a Figura 8.

Figura 8 – Identificação de autocitação no Google Scholar Metrics

	Quantifying the impact of the grain-for-green program on ecosystem health in the typical agro-pastoral ecotone: a case study in the Xilin Gol League, Inner Mongolia Z Wang, Q Yu, L Guo International Journal of Environmental Research and Public Health 17 (16), 5631	2020
6	Emisoras infantiles en internet. Análisis de una oferta alternativa para el niño-oyente MLB Veloso, JJP Páez ICONO 14, Revista de comunicación y tecnologías emergentes 16 (2), 100-120	2018
	The media diet, news consumption habits and disinformation of Spanish university students XS Pérez, AM López-Cepeda, J Sixto-García Revista Latina de Comunicación Social, 1056-1070	2019

Fonte: GSM (2022).

Esse processo é repetido para todas as citações recebidas pelos artigos do periódico. Por fim, são calculados o índice h₅ e mediana h₅, a taxa percentual de autocitação por artigo (*article self-cited rate* - ASR) e a taxa percentual de autocitação por periódico (*self-cited rate* - SCR) (ROUSSEAU, 1999).

4.2.1.2.3 Gsm_hcitation

O *Gsm_hcitation* é o recurso dedicado à análise de citação do h5-core. Executa a extração dos títulos das fontes e dos anos de publicação dos documentos citantes. Os dados são utilizados para analisar a distribuição de citações por ano e identificar vínculo entre os periódicos citantes e o citado.

Por meio desses dados também é possível calcular a SCR, identificando o percentual de ocorrência do título do periódico na lista de títulos das fontes citantes. O cálculo é feito em uma planilha eletrônica, individualmente para cada periódico, o que limita o seu uso em larga escala. Por outro lado, essa forma de cálculo da SCR permite testar a precisão do recurso *Gsm_hscite* na execução da mesma tarefa.

A análise de distribuição de citações por ano em conjuntos de periódicos é executada por um comando de compilação de dados (*compile_years*), que agrupa todos os anos das citações em uma coluna.

O comando de compilação possibilita a visualização dos dados em gráficos. Já as relações entre as fontes citantes e a fonte citada, bem como a ocorrência de autocitações podem ser visualizados por meio de grafos.

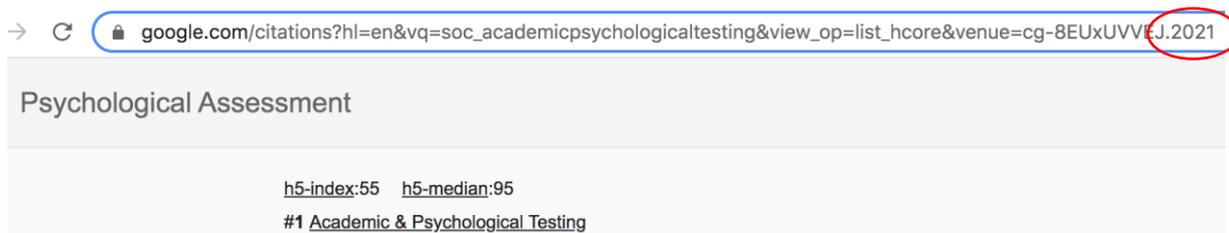
4.2.1.2.4 *Gsm_htline*

A cada atualização do GSM, que ocorre sempre no mês de julho, os dados da edição anterior são apagados do sistema. Visando a superar essa limitação, foi desenvolvido o *Gsm_htline*, quarto e último recurso da ferramenta *Gsm_hdata*, voltado para a extração de dados históricos e realização de análises longitudinais do GSM.

Os dados de entrada são os endereços URL, extraídos em listas no *Gsm_hsearch* ou copiados manualmente do GSM. O uso do recurso depende da alteração do ano identificado nos endereços URL.

A Figura 9 destaca o campo indicativo do ano da edição do GSM na URL de uma publicação indexada.

Figura 9 – Campo de ano na URL do Google Scholar Metrics



Fonte: GSM (2022).

Os endereços URL são repetidos proporcionalmente aos anos da análise pretendidos, sendo alterado o número correspondente ao ano de cada edição. O Quadro 3 mostra os formatos de entrada e de saída de dados.

Quadro 3 – Formatos de entrada e de saída do *Gsm_htline*

Dados de entrada
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=cg-8EUxUVVEJ.2021
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=cg-8EUxUVVEJ.2020
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=cg-8EUxUVVEJ.2019
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=cg-8EUxUVVEJ.2018
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=cg-8EUxUVVEJ.2017
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=EHoJPyaHbn0J.2021
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=EHoJPyaHbn0J.2020
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=EHoJPyaHbn0J.2019
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=EHoJPyaHbn0J.2018
https://scholar.google.com/citations?hl=en&vq=soc_academicpsychologicaltesting&view_op=list_hcore&venue=EHoJPyaHbn0J.2017
Dados de saída
Psychological Assessment; 55; 95
Psychological Assessment; 54; 82
Psychological Assessment; 55; 85
Psychological Assessment; 54; 79
Psychological Assessment; 48; 66
Psychological Methods; 59; 109
Psychological Methods; 47; 87
Psychological Methods; 40; 85
Psychological Methods; 38; 86
Psychological Methods; 43; 77

Fonte: Elaborado pelo autor.

Os dados de saída são o título, o índice h5 e a mediana h5 de cada edição do GSM. Caso o periódico não esteja indexado em alguma das edições, os dados de saída recebem valor zero.

4.2.1.3 Dimensão analítica do *framework*

A dimensão analítica do *framework* *Gsm@hdata* compreende a análise de dados extraídos na dimensão tecnológica a partir de conceitos incorporados pela dimensão teórica. Essa dimensão oferece subsídios para a tomada de decisão em processos de avaliação ou de classificação de periódicos, além de amparar a condução de estudos métricos envolvendo dados do GSM.

As decisões se baseiam em indicadores bibliométricos, ou já existentes, ou desenvolvidos no âmbito da dimensão teórica do *framework*. O instrumento da dimensão analítica é o protocolo de análise de citação do índice h5.

4.2.1.3.1 Protocolo de análise de citação

O protocolo de análise de citação tem por objetivo identificar discrepâncias no resultado de avaliações de impacto a partir do índice h5. Se presta a identificar o padrão médio de citação de um conjunto de periódicos e, a partir disso, identificar divergências.

O protocolo tem um nível de análise geral e um outro nível de análise específica. A análise geral é baseada em seis indicadores bibliométricos. Os três primeiros, o índice h5, o índice h_{s5} e a SCR, já são estabelecidos. Os três últimos, o VrH, o SCiR e o HgR, foram formulados especificamente no âmbito da dimensão teórica deste *framework*.

No nível de análise geral do protocolo, periódicos com características comuns (mesma área) são analisados de acordo com a distribuição das SCR. A análise de distribuição pode ser feita por meio de gráficos *box plot* ou de cálculo em planilhas eletrônicas.

A partir dessa análise são identificados os periódicos com SCR de distribuição normal e os considerados com SCR consideradas excessivas. Uma SCR excessiva pode ser caracterizada de duas formas. A primeira é por meio de um percentual fixo, por exemplo, 20%. A segunda é por meio da identificação de *outliers* em um conjunto de variáveis, visualizadas em *box plot* ou calculadas em planilhas.

Os periódicos de distribuição normal são analisados em conjunto por meio das médias dos indicadores. Os *outliers* são listados, um a um, no protocolo.

São preenchidos os índices h_5 e h_{s5} e a SCR de cada periódico *outlier*. A partir desses indicadores são calculados os três indicadores complementares:

- a) VhR – taxa percentual de variação entre os índices h_5 e h_{s5} ;
- b) $SCiR$ – índice de influência de autocitações no índice h_5 ;
- c) $HgR3$ – taxa percentual de crescimento acumulado do índice h_5 nos últimos três anos (2019-2021).

O indicador VrH mede a redução de impacto com a exclusão das autocitações. É calculado pela fórmula: $Vrh = (h_5 - h_{s5}) / h_5 * 100$, que resulta em um número percentual.

O $SCiR$ mede a influência das autocitações na variação percentual entre os índices h_5 ou h_{s5} . É calculado em razão de que, ao contrário de indicadores baseados na média de citações, o índice h é um indicador por amostragem, cujo resultado desconsidera artigos menos citados e o excesso de citações dos mais citados.

Assim, a diminuição do índice h_5 com exclusão de autocitações pode não ser proporcional ao percentual de autocitações de um periódico. O $SCiR$ é calculado pela seguinte fórmula: $SCiR = -(VrH / ScR)$.

O resultado padrão é igual a 1, ou seja, a SCR tem a mesma influência na diminuição do índice h_5 . Se o resultado for inferior a 1, as autocitações influenciam menos no impacto; se for superior a 1, as autocitações exercem mais influência que o próprio percentual de autocitações.

O HgR mede a taxa de crescimento do índice h_5 (HARZING, 2014). Foi definido em razão de que os periódicos tendem a ter crescimento anual no número de citações recebidas e, conseqüentemente, nos resultados dos indicadores. Esse crescimento segue um padrão médio de acordo com a área.

Periódicos com crescimento significativamente superior ao médio podem ser considerados decrescimento anormal. O HgR pode ser medido com diferentes janelas de tempo, entre dois e cinco anos, dependendo da disponibilidade dos dados e do tipo da análise pretendida.

Os periódicos *outliers* podem ser analisados em nível específico no protocolo. O nível específico tem a finalidade de identificar indícios de manipulação de

citações. Índícios de manipulação são identificados com base em dois critérios estabelecidos em estudos anteriores:

A Tabela 6 mostra a estrutura do protocolo de análise de citação do *framework Gsm_hdata*.

Tabela 6 – Protocolo de análise de citação do *framework Gsm@hdata*

Indicadores	h5	h_s5	SCR	VhR	SCiR	HgR
Média geral	0,0	0,0	0,0%	0,0%	0,0	0,0%
Critério definido			0,0%			
Média outliers	0,0	0,0	0,0%	0,0%	0,0	0,0%
Periódico 1	0,0	0,0	0,0%	0,0%	0,0	0,0%
Periódico2	0,0	0,0	0,0%	0,0%	0,0	0,0%
Periódico3	0,0	0,0	0,0%	0,0%	0,0	0,0%

Fonte: Elaborado pelo autor. Nota: SCR-Taxa de autocitação de periódico; VhR-variação (%) entre os índices h5 e h_s5; SCiR-índice de influência da SCR; HgR-Taxa acumulada de crescimento do índice h5.

a) Distribuição anormal de citações por ano dos documentos citantes (CHORUS; WALTMAN, 2016);

b) Vínculo institucional entre o periódico citado e um ou mais periódicos citantes (HENEBERG, 2016).

O primeiro critério se baseia na distribuição das citações recebidas pelos periódicos de acordo com o ano do documento citante. É realizada a comparação do padrão do periódico com o padrão médio da área.

Esse critério foi definido em virtude de que uma expressiva concentração de citações em um ou dois anos específicos pode ser considerado um indício de manipulação de citações.

O segundo critério visa identificar a ocorrência de excesso de citações recebidas por publicações com algum vínculo com o periódico citado, especialmente origem institucional comum, o que também pode ser considerado indício de manipulação (HENEBERG, 2016).

É necessário enfatizar que este protocolo não tem o objetivo de comprovar a ocorrência de manipulação nas publicações analisadas. É um procedimento que têm por objetivo identificar indícios, ou seja, elementos objetivos que possam embasar uma investigação mais detalhada.

4.2.2 Validação do framework *GSM@hdata*

O *framework Gsm@hdata* passou por três procedimentos preliminares de validação, sendo dois de validação parcial externa e um de validação parcial interna.

4.2.2.1 Validação parcial interna

A validação parcial interna do *framework* compreendeu três testes de precisão dos recursos da ferramenta bibliométrica *Gsm_hdata* na execução dos processos de busca, extração e análise de dados do índice h5 do GSM.

4.2.2.1.1 Teste de precisão 1

Foi realizado um primeiro teste de precisão do recurso *Gsm_hsearch* na execução de requisições de busca e identificação de periódicos indexados no GSM. Esse teste foi realizado a partir de um cenário de 100% de indexação dos títulos buscados.

Foram utilizadas as listas de periódicos das 26 subcategorias de assunto de Ciências Sociais do GSM, mostradas na Figura 10.

Figura 10 – Subcategorias de assunto de Ciências Sociais no Google Scholar

Metrics

Categories > Social Sciences > Subcategories ▾		
Subcategories	Educational Psychology & Counseling	Library & Information Science
Academic & Psychological Testing	Educational Technology	Middle Eastern & Islamic Studies
African Studies & History	Environmental & Occupational Medicine	Military Studies
Anthropology	Environmental Law & Policy	Paleontology
Archaeology	Epistemology & Scientific History	Political Science
Architecture	Ethics	Public Health
Asian Studies & History	European Law	Public Policy & Administration
Bioethics	Family Studies	Science & Engineering Education
Canadian Studies & History	Feminism & Women's Studies	Sex & Sexuality
Chinese Studies & History	Forensic Science	Social Sciences (general)
Cognitive Science	Geography & Cartography	Social Work
Criminology, Criminal Law & Policing	Health Policy & Medical Law	Sociology
Development Economics	Higher Education	Special Education
Diplomacy & International Relations	History	Sustainable Development
Early Childhood Education	Human Migration	Teaching & Teacher Education
Economic History	Human Resources & Organizations	Technology Law
Education	International Law	Urban Studies & Planning
Educational Administration	Law	

Fonte: GSM (2021).

Foram extraídos os títulos das 20 publicações de cada subcategoria, resultando em uma lista final de 1040 títulos. A lista foi executada pelo recurso *Gsm_hsearch*, conforme rotina descrita seção 2.2.2.

Foram identificados 1030 periódicos pelo padrão “encontrado”, nove pelo padrão “contido em:” e apenas um título não foi encontrado.

Do total de periódicos encontrados, apenas três foram recuperados incorretamente. Do total de periódicos listados no padrão “contido em:” cinco continham o periódico buscado e quatro não foram recuperados.

Como resultado do teste de precisão, foram corretamente localizados 1032 periódicos da lista de 1040 procurados, o que corresponde a 99,23% de precisão em um cenário de 100% de correspondência entre os títulos buscados e indexados.

4.2.2.1.2 Teste de precisão 2

Foi realizado um segundo teste de precisão, envolvendo a execução de uma rotina busca de títulos de periódicos com indexação incerta, seguida do procedimento de busca e verificação e manual previsto no framework.

Considerando a existência de procedimento de busca e verificação manual de resultados, esse teste foi realizado com auxílio de um especialista.

O *expert* convidado foi o bibliotecário Edson Mario Gavron¹⁰, especialista em recuperação da informação em bases de dados do Setor de Competência Informacional de Suporte à Pesquisa da Biblioteca Universitária da USFC.

O teste consistiu em uma simulação do procedimento realizado nesta pesquisa, a partir de uma amostra de periódicos do universo da pesquisa. Para definição do tamanho da amostra foi utilizada a técnica específica para ciências sociais proposta por Barbetta (2012), expressa pela fórmula:

$$n = \frac{N \times Z^2 \times 0,25}{(N - 1) \times e^2 + Z^2 \times 0,25}$$

Onde:

N = população

Z = constante função da probabilidade

n = tamanho da amostra

e = margem de erro

A probabilidade aceitável em pesquisas em Ciências Sociais é de 95%. A constante em função da probabilidade (Z) é 1,96 e a margem de erro de 5,0% (BARBETTA, 2012). O universo da pesquisa é composto por 10553 periódicos indexados no Latindex.

¹⁰Lattes: <http://lattes.cnpq.br/5816430697711442>. Orcid ID: <https://orcid.org/0000-0002-8761-2178>

Foi calculado em uma planilha eletrônica o tamanho da amostra n igual a 370 elementos, número que corresponde a seleção de um elemento a cada 28 do universo.

Para seleção dos elementos da amostra, todos os periódicos do universo foram ordenados alfabeticamente e receberam um número de ordem. Foram selecionados os periódicos de número de ordem múltiplo de 28 (28, 56, 84, 112, etc.) até completar o conjunto amostral de 370 periódicos.

Foi utilizado o *Gsm_hsearch* para execução de requisições de busca dos títulos dos periódicos da amostra extraídos do Latindex. Os resultados foram organizados em uma planilha eletrônica.

O procedimento de busca e verificação manual realizado pelo especialista compreendeu três processos, com as seguintes instruções:

a) Para títulos com resultado no padrão “*não encontrado*”: efetuar tentativas de busca na interface do GSM, com uso de variações, partes do título ou siglas e abreviaturas de livre escolha (REGENBRECHT; LANGLOTZ, 2015). No caso de localização na busca manual, copiar o título e os valores do índice h5 e da mediana h5. Finalizar o registro;

b) Para títulos com resultado no padrão “*contido em:*”: conferir os títulos extraídos pelo sistema. No caso de correspondência de um dos títulos (entre um 1 e 20 listados), copiar o título e os valores do índice h5 e da mediana h5. Finalizar o registro. No caso de não correspondência, realiza o procedimento “a”;

c) Para títulos com resultado no padrão “*encontrado*”: conferir os títulos buscados e encontrados com auxílio de recurso de encontrar duplicatas. No caso de correspondência, finaliza o registro. No caso de não correspondência, realiza o procedimento “a”.

A execução do recurso *Gsm_hsearch* a partir da lista de 370 títulos da amostra teve resultado a identificação de 134 títulos no padrão encontrado, 42 no padrão “contido em:” e 194 não encontrados.

Após a realização do processo de busca e verificação manual pelo especialista, foi detectado que:

a) Dos 194 títulos não encontrados, 9 foram encontrados manualmente (95,36% de precisão para títulos não indexados);

b) Dos 42 títulos com resultado no padrão “contido em:”, 21 títulos estavam corretamente contidos entre os resultados, três não estavam contidos, mas foram

encontrados manualmente e 18 estavam contidos, mas não correspondiam ao periódico procurado e não foram encontrados manualmente. Esse resultado implica em uma precisão de resposta de 92,85% para casos de discrepância entre formas de títulos buscado e indexados.

c) Dos 134 títulos não encontrados, 10 não correspondiam ao periódico correto, sendo dois localizados manualmente (92,53% de precisão para títulos indexados).

4.2.2.1.3 Teste de precisão 3

O teste 3 teve por objetivo mensurar a precisão da ferramenta *Gsm_hscite* na identificação de autocitações de periódico no índice h5. Foi realizado a partir dos 126 periódicos localizados no GSM no teste de precisão 2.

Os endereços URL dos periódicos foram inseridos no recurso *Gsm_hscite*. Foram identificadas as autocitações e, ao final, calculadas automaticamente as SCR dos periódicos por meio do recurso de compilação (*compile*), conforme descrito na seção 4.2.1.2.2.

Posteriormente, as mesmas URL foram executadas no recurso *Gsm_hcitation*, sendo extraídos os títulos das fontes citantes de cada periódico. Os dados foram analisados manualmente em planilhas eletrônicas, sendo identificadas autocitações e calculadas as SCR.

Por fim, os resultados das duas SCR foram comparados estatisticamente. A média das SCR calculadas no *Gsm_hscite* foi de 3,40%; já a média calculada manualmente a partir dos dados do *Gsm_hcitation* foi de 3,34%. Isso representa uma variação média de 0,06% entre todos os periódicos.

O coeficiente de Pearson entre o conjunto de resultados das duas variáveis foi $p=0.99$, o que indica uma correlação quase perfeita entre as variáveis.

4.2.2.2 Validação parcial externa

A validação parcial externa se deu com a publicação de resultados preliminares da pesquisa em dois artigos científicos, um no periódico *Informação & Sociedade* e outro no *Scientometrics*. Ambos são classificados no estrato A1 da área de Comunicação e Informação do sistema Qualis/CAPES.

No primeiro artigo (PINTO et al., 2020) foram analisados os periódicos científicos brasileiros indexados na edição 2020 do GSM. A lista primária de dados

de pesquisa foi extraída do diretório *UlrichsWeb*. A identificação das publicações e a extração de dados no GSM ocorreu por meio de uma primeira versão da ferramenta *Gsm_hsearch*, ainda sem essa denominação.

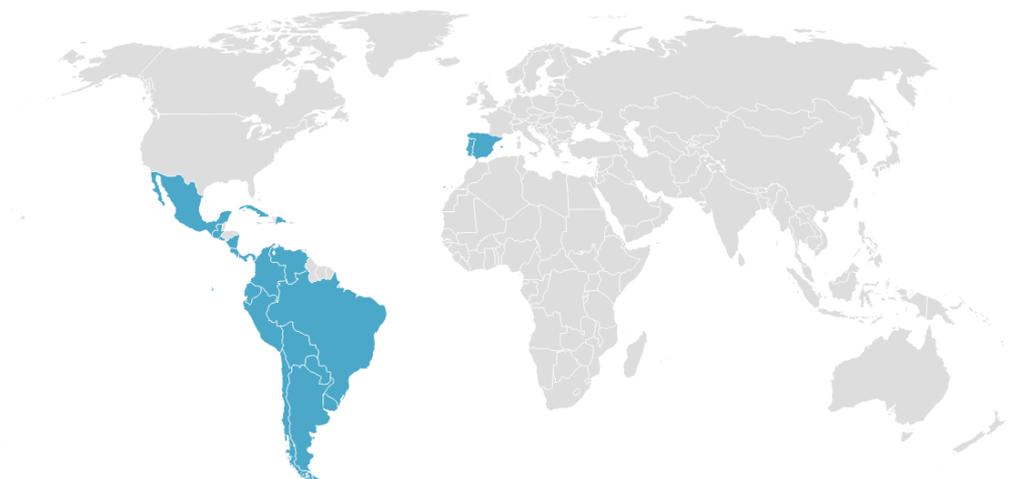
No segundo artigo (CANTO et al., 2022) foi utilizada a versão atual dessa ferramenta bibliométrica para identificação e análise de periódicos da América-Latina e do Caribe indexados na edição 2021 do GSM. Nesse caso, foram utilizados os dados de periódicos extraídos do diretório Latindex. A análise foi centrada na identificação da cobertura, das características editoriais e do impacto dos periódicos da região indexados no GSM.

A validação definitiva compreende a aplicação de todos os procedimentos propostos para o *framework* *Gsm@hdata*, envolvendo as três dimensões, em um universo ampliado de pesquisa, descrito a seguir.

4.2.3 Universo da pesquisa

O universo definitivo da pesquisa contempla os periódicos editados em países Ibero-Americanos indexados no Latindex e encontrados no GSM. A Figura 11 mostra a localização geográfica dos países abrangidos.

Figura 11 – Mapa dos países Ibero-Americanos do universo da pesquisa



Fonte: Dados da pesquisa.

Foram considerados os 23 países membros da *Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura* (OEI)¹¹: Andorra, Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, Equador, El Salvador,

¹¹<https://oei.int/quienes-somos/oei>

Espanha, Guatemala, Guiné Equatorial, Honduras, México, Nicarágua, Panamá, Paraguai, Peru, Portugal, República Dominicana, Uruguai e Venezuela.

Foi identificado preliminarmente que, na maioria desses países, prevalecem as publicações acadêmicas, em espanhol e português e com aderência as áreas de humanas e sociais (CANTO et al., 2022). Em razão desse perfil da região, o GSM pode ser utilizado como fonte alternativa de avaliação de impacto, o que justifica a delimitação geográfica do universo de pesquisa.

4.2.4 Coleta de dados

Os procedimentos de coleta de dados foram realizados entre os meses de julho de 2021 e março de 2022. Foram considerados os dados mais atualizados em cada base de dados utilizada.

O Latindex foi definido como fonte primária para composição da lista geral de periódicos (*journal master list*) do universo da pesquisa. Esse diretório foi escolhido por ser o mais abrangente para publicações Ibero-Americanas e por conter todos dados necessários para caracterização do universo da pesquisa.

Foi utilizada a ferramenta *WebScraper*¹², um *plug in* para o navegador *Google Chrome*, para extração de dados de toda a coleção do diretório Latindex. Essa técnica foi utilizada em razão de que o recurso de extração próprio dessa base de dados é limitado. Foram selecionados periódicos eletrônicos ativos publicados nos países Ibero-Americanos definidos no universo da pesquisa.

Foram incluídos os três tipos de publicações periódicas contidas no Latindex: de divulgação científica e cultural, de pesquisa científica e de técnico-profissional. Foram consideradas publicações revisadas e não revisadas por pares, já que o GSM não faz essa distinção.

Foram extraídas as informações básicas dos periódicos (títulos, ISSN, editora), dados temáticos e editoriais (país, idioma, periodicidade, áreas e assuntos de pesquisa, natureza das instituições editoras, tipo de acesso, revisão por pares).

Foram realizados procedimentos de padronização dos títulos das publicações, eliminando elementos estranhos (indicadores de local, datas, abreviaturas, entre outros) presentes em alguns registros. Também foram excluídos os registros duplicados.

¹²<https://chrome.google.com/webstore/detail/web-scraper-free-web-scr/jnhgnonknehpejinehehllkplmbmhn?hl=pt-BR>

Foi utilizado o módulo *Gsm_hsearch* para pesquisa dos títulos e extração dos dados do GSM (título, índice h5, mediana h5, endereço URL). A busca foi realizada na edição 2021 do GSM, que considera artigos publicados entre 2016 e 2020 e citações registradas até julho de 2021.

Resultados com títulos semelhantes contendo o termo de busca também foram recuperados para uma análise manual de correspondência, conforme padrão de resultado “contido em:” do sistema. Para os periódicos não encontrados, também foi realizado um procedimento manual de busca, com requisições por variações do título.

Dúvidas sobre a correspondência entre o título pesquisado e o título encontrado foram sanadas por meio da análise dos dados das publicações (ISSN, editora, país). A base de dados do ISSN e o site das publicações foram consultados para eliminar dúvidas envolvendo homônimas, duplicações e variações de títulos por idioma. Em alguns casos, foram acessados os textos completos de artigos do índice h5 visando diferenciar publicações.

Os dados dos periódicos localizados manualmente no GSM foram copiados para a lista geral de pesquisa.

Para obtenção de parâmetros comparativos de visibilidade foi identificado o número de periódicos Ibero-Americanos indexados na Scopus, por meio da base de dados SJ&CR, no JCR, usando a base de dados *Clarivate InCities*, e no SciELO, utilizando os recursos da própria base.

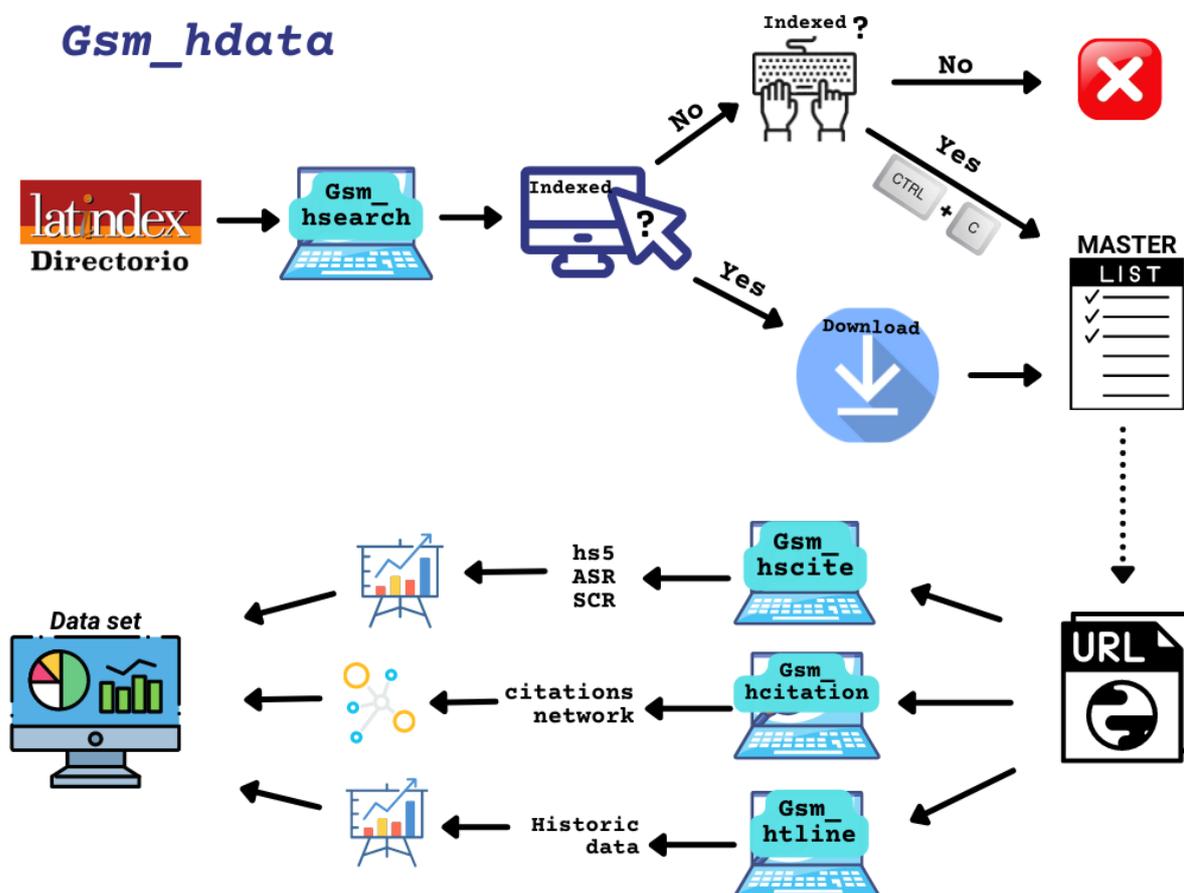
Posteriormente, foram cruzados os dados de ISSN da lista geral de periódicos indexados no GSM com os periódicos identificados nessas bases de dados, visando identificar solapamento.

Identificado o solapamento, foram extraídos os dados dos indicadores AIS, ES, nES, JCI, FI, FIws, FI5y da base de dados JCR e do CS, SNIP, SJR e índice h das bases de dados Scopus e SJ&CR.

Os dados editoriais e de impacto foram armazenados em uma planilha eletrônica.

A Figura 12 mostra o processo de coleta de dados realizado da pesquisa por meio da ferramenta bibliométrica *Gsm_hdata*.

Figura 12 – Processo de coleta de dados da pesquisa



Fonte: Elaborado pelo autor.

Foi utilizado o recurso *Gsm_hscite* para identificação e análise de autocitações nos artigos do índice h5 dos periódicos indexadas no GSM. Foram calculados o índice h_{s5} , a mediana h_{s5} , a ASR e a SCR.

Foi utilizada a ferramenta *Gsm_htline* para extração de dados de edições anteriores do GSM. O período definido foi de 2017 a 2020, completando a janela de tempo de cinco anos da edição de 2021, já que o índice h5 atual já havia sido obtido.

Os dados de todo o período foram obtidos somente de periódicos indexados com o mesmo endereço URL.

4.2.5 Tratamento e análise dos dados

Foram realizados procedimentos de verificação, limpeza e tratamento dos dados de pesquisa. Títulos homônimos ou com alto grau de semelhança foram analisados individualmente. Títulos duplicados foram eliminados.

Os dados da lista definitiva de periódicos de pesquisa foram tratados e analisados estatisticamente. A análise contemplou os aspectos de visibilidade (indexação), características editoriais (tipos de publicação, escopo e aderência temática, natureza da instituição, idiomas, periodicidade, modelo de acesso, revisão por pares) e impacto (indicadores, padrões de citação e de autocitação).

A análise de visibilidade foi centrada na comparação do número de publicações indexadas em cada base de dados selecionada na pesquisa. Os resultados foram relacionados com a aderência temática e o país de origem das publicações.

A análise de aderência temática foi baseada na classificação de sete áreas do conhecimento do Latindex: Artes & Humanidades (A&H), Ciências Agrárias (CCA), Ciências Exatas e Naturais (E&N), Ciências da Saúde (CDS), Ciências Sociais (CCS), Engenharias (ENG) e Multidisciplinar (MTD). Alguns periódicos são classificados em mais de uma área, neste caso consideramos o máximo três áreas. Periódicos classificados em mais de três áreas foram definidas somente como MTD.

A análise de escopo temático foi baseada nas categorias de assuntos do Latindex. Os descritores de assuntos foram mantidos no idioma original Espanhol, visando evitar distorções de tradução. As relações entre descritores de uma mesma publicação foram representadas por grafos e analisadas por meio da teoria de redes sociais. Os descritores foram representados por vértices e as ligações entre eles, ou seja, quando dois ou mais são atribuídos a uma mesma publicação, foram representados por arestas. Foi utilizado a ferramenta de visualização VosViewer para construção dos grafos.

Algumas classificações para dos dados do Latindex foram padronizadas.

As classificações de idiomas foram sintetizadas em três grupos: “somente inglês”, “vários idiomas” (três ou mais, incluindo o inglês) ou “idiomas regionais” (espanhol, português, catalão, euskera, entre outros).

Foram mantidas as cinco classificações de periodicidade mais comuns: anual, semestral, quadrimestral, trimestral e bimestral. As demais foram agrupadas em “outras”.

Para facilitar a visualização em análises geográficas, países com pouca representatividade estatística foram agrupados na categoria “outros países”.

Os testes de Pearson (p) e Spearman ($rho=r$) foram utilizados para identificar os níveis de correlação entre os indicadores de impacto das publicações Ibero-

Americanas indexadas em mais de uma (solapamento) das bases de dados incluídas na pesquisa. A análise foi conduzida a partir de subconjuntos de publicações que possuam cada par de indicadores analisados. Os indicadores analisados foram o índice h5 e a mediana h5 (GSM), o AIS, o ES, o nES, o FI, o FIws, o FI5y, JCI (JCR), CS (Scopus), SNIP, SJR, índice h (SJ&CR)

Esses dois testes estatísticos são frequentemente adotados em estudos métricos (CHINCHILLA-RODRÍGUEZ et al., 2015; SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015). A principal diferença entre os dois é que o coeficiente de Pearson se baseia nas mudanças dos valores das variáveis, enquanto *Spearman* analisa as posições (*ranking*) das variáveis.

Para interpretar os resultados, foi considerada uma correlação positiva perfeita para valores de ρ e r iguais a 1, em nível muito forte para valores entre 0,80 e 0,99, em nível forte entre 0,60 e 0,79, em nível moderado entre 0,40 e 0,59, fraco entre 0,10 e 0,39 e desprezível entre 0,01 e 0,09 (AKOGLU, 2018).

A técnica de regressão linear simples foi utilizada para previsão dos valores do FI e do CS a partir do índice h5 dos periódicos. Essa técnica foi definida na metodologia do Qualis para comparação de periódicos a partir de indicadores distintos em uma mesma área de avaliação (BRASIL, 2021).

Foi utilizada estatística descritiva para análise de variáveis relacionadas ao impacto (índice h5, mediana h5 e taxas de autocitação). Foram calculadas duas medidas de centralidade (média e mediana).

Para análise de distribuição foram usados gráficos em diagramas de caixa, divisão por quartis e identificação de *outliers*. Para identificação de *outliers*, usualmente é utilizado o cálculo padrão baseado no valor da média + 1,5 x IQR (intervalo interquartil). De forma diversa, nesta pesquisa foi utilizada a medida de 3 x IQR, visando a identificar somente os índices extremos (SZOMSZOR; PENDLEBURY; ADAMS, 2020).

Foi utilizado o recurso *Gsm_hscite* para identificação de autocitações a partir dos dados índice h5 de todos os periódicos do universo da pesquisa. Foi utilizado o comando de compilação de dados para calcular o índice h_s5 , a mediana h_s5 , a taxa de autocitação por artigo e a taxa de autocitação por periódico.

Para representação do fenômeno da autocitação foi escolhida a taxa de autocitação por periódico como indicador principal, considerando o seu uso recorrente em estudos relacionados.

A taxa de autocitação por periódico (*self-cited rate*) representa o percentual de autocitações em relação ao total de citações recebidas por um periódico, inclusive as autocitações (ROUSSEAU, 1999). Foi mantida a sigla original em inglês (SCR), considerando a inexistência de referência a essa métrica no idioma português.

A análise de autocitação foi dividida em duas partes, uma de análise geral e outra de análise específica. A análise geral incluiu todos os periódicos do universo da pesquisa. A análise específica foi conduzida somente com periódicos de maior impacto com alguma autocitação identificada.

A seleção dos periódicos analisados na segunda parte foi baseada em dois critérios, um de exclusão e outro de inclusão: 1) exclusão de periódicos com taxa de autocitação igual a zero; 2) inclusão de publicações com no mínimo 50 citações, aproximadamente, identificadas pela fórmula: índice h5 x mediana h5 = n , sendo $n > 49$.

Essa definição visa obter maior precisão estatística. Assegura que uma autocitação implique em percentual máximo 2% na SCR, evitando distorções em análises por valores percentuais.

Também tem o objetivo de evitar interferências ocasionadas por publicações com poucas citações recebidas, nas quais uma autocitação pode assumir percentual elevado. Por exemplo: em uma publicação com índice h5 = 2 e que tenha recebido um total de cinco citações, apenas duas autocitações representaria uma SCR de 40%.

Para facilitar a compreensão dos resultados, foram utilizadas ferramentas de visualização de dados. Os mapas foram desenhados no Datawrapper¹³, os gráficos no Flourish¹⁴, os fluxogramas no Canva¹⁵ e no Bizage Modeler¹⁶ e os grafos no VOSviewer¹⁷.

As análises estatísticas foram realizadas no módulo de análise de dados do Excel.

¹³ <https://app.datawrapper.de>

¹⁴ <https://app.flourish.studio>

¹⁵ <https://www.canva.com>

¹⁶ <https://www.bizagi.com/pt/plataforma/modeler>

¹⁷ <https://www.vosviewer.com>

O Quadro 4 apresenta uma síntese dos procedimentos metodológicos realizados.

Quadro 4 – Síntese dos procedimentos metodológicos

Objetivos	Técnica	Resultados
Comparar o nº de periódicos Ibero-Americanos indexados em bases de dados	Busca avançada (Latindex, JCR, Scopus, SJ&CR, SciELO); recurso <i>Gsm_hsearch</i> (GSM)	Definição das bases de dados de maior visibilidade da região
Identificar as características dos periódicos Ibero-Americanos indexados no GSM	Extração de dados (Latindex) e recurso <i>Gsm_hsearch</i> (GSM)	Caracterização dos periódicos da região indexados no GSM
Testar a validade do índice h5 na análise de impacto de periódicos Ibero-Americanos	Análise de correlação de Pearson e de Spearman	Confirmação da validade do índice h5
Testar a estabilidade do índice h5 na análise de impacto de periódicos Ibero-Americanos	Cálculo da taxa de crescimento do índice (recurso <i>Gsm_htline</i>)	Confirmação da estabilidade do índice h5
Identificar autocitações de periódico no índice h5	Ferramenta <i>Gsm_hscite</i>	Definição de padrões de autocitação e cálculo do índice h _s 5
Identificar padrões normais e anormais de autocitação	Protocolo de identificação de padrões de citação no índice h5 (critérios gerais)	Distinção entre periódicos normais e anormais de autocitação
Definir critérios de identificação de indícios de manipulação de citações	Protocolo de identificação de padrões de citação no índice h5 (critérios específicos)	Identificar indícios de manipulação de citação em periódicos

Fonte: Elaborado pelo autor.

5 RESULTADOS E ANÁLISE

5.1 VISIBILIDADE DOS PERIÓDICOS IBERO-AMERICANOS

Primeiramente, foram localizados os registros e extraídos os dados de 10.553 periódicos eletrônicos ativos de países Ibero-Americanos no diretório Latindex. Entre os 23 países da região, apenas Andorra e Guiné Equatorial não tiveram nenhuma publicação identificada e foram desconsiderados.

A partir da lista de títulos das publicações e utilizando a ferramenta *Gsm_hsearch*, seguido de etapas de busca e de verificação manual, foram encontradas 4.049 delas indexadas no GSM, número que corresponde a 38,4% do total do Latindex.

A Figura 13 mostra o mapa de distribuição quantitativa dos periódicos de países Ibero-Americanos indexadas no GSM.

Figura 13 – Distribuição dos periódicos em países Ibero-Americanos



Fonte: Dados de pesquisa.

Os países com maior número de publicações indexadas no GSM são Brasil, Espanha, México, Colômbia, Argentina, Chile e Cuba, representados pelos tons mais escuros no mapa.

Visando a estabelecer um parâmetro comparativo de visibilidade, a cobertura dos periódicos Ibero-Americanos também foi identificada no SciELO, no JCR e na Scopus.

A Tabela 6 mostra o quantitativo de periódicos por país em cada base de dados.

Tabela 6 – Visibilidade de periódicos Ibero-Americanos em bases de dados

País	Latindex	GSM		JCR	Scopus	SciELO
	<i>n</i>	<i>n</i>	%	<i>n</i>	<i>n</i>	<i>n</i>
Brasil	3583	1649	46%	409	391	306
Espanha	2115	945	44,7%	719	636	43
Argentina	1.061	219	20,6%	90	68	103
México	865	312	36,1%	97	112	148
Chile	633	169	26,7%	94	110	125
Colômbia	401	249	62,1%	176	114	216
Peru	275	82	29,8%	21	13	37
Equador	252	83	32,9%	17	3	22
Portugal	233	60	25,8%	50	69	59
Venezuela	192	41	21,4%	34	31	37
Cuba	178	116	65,2%	19	23	72
Costa Rica	173	53	30,6%	27	6	43
Uruguai	138	24	17,4%	15	0	23
Nicarágua	102	16	15,7%	2	0	0
Rep. Dominicana	97	5	5,2%	0	0	0
Bolívia	51	5	9,8%	2	0	34
Paraguai	46	11	23,9%	2	0	18
Panamá	37	4	10,8%	0	0	0
Guatemala	27	3	11,1%	0	0	0
El Salvador	21	3	14,3%	1	0	0
Honduras	15	0	0%	0	0	0
Total	10553	4049	38,4%	1777	1585	1286

Fonte: Dados de pesquisa. Nota: a quarta coluna corresponde ao percentual localizado no GSM em relação a quantidade encontrada no Latindex. Os números referentes ao JCR, a Scopus e ao SciELO correspondem ao total em cada fonte, independente de solapamento com o Latindex.

Os resultados indicam que o GSM é a fonte de maior visibilidade da região, com mais que o dobro do número de publicações indexadas no JCR e na Scopus e mais que o triplo em comparação com o SciELO. O JCR é a segunda fonte mais ampla com um total de 1777 publicações. O SciELO e a Scopus são as bases de dados de menor cobertura, respectivamente com 1583 e 1286 publicações Ibero-Americanas indexadas.

O GSM também possui vantagem na cobertura geográfica. Do total de 21 países da região com publicações no Latindex, 20 foram localizados no GSM. Apenas Honduras não teve nenhuma publicação encontrada, sendo desconsiderado

na pesquisa. O JCR abrange publicações de 17 países, seguido do SciELO com 15 e da Scopus com apenas 12 países da região.

Considerando o percentual de publicações em relação ao Latindex, Cuba e Colômbia são os países com maior visibilidade no GSM, respectivamente com 65,2% e 62,1% de periódicos identificados.

A cobertura quantitativa por país também varia entre as bases de dados. O Brasil possui maior cobertura no GSM e no SciELO. No JCR e na Scopus, as duas bases de dados comerciais, a Espanha possui mais publicações indexadas, seguida do Brasil e da Colômbia. Cuba, por outro lado, possui baixa visibilidade nas duas bases de dados comerciais.

5.2 CARACTERÍSTICAS DAS PUBLICAÇÕES

Os padrões de citação podem variar significativamente de acordo com as características dos documentos analisados. Aspectos como aderência e escopo temático, tipo de publicação, natureza das instituições editoras, periodicidade, idiomas podem influenciar os níveis de impacto. Em razão disso, as características dos periódicos encontradas no GSM foram identificadas a partir dos dados extraídos do Latindex.

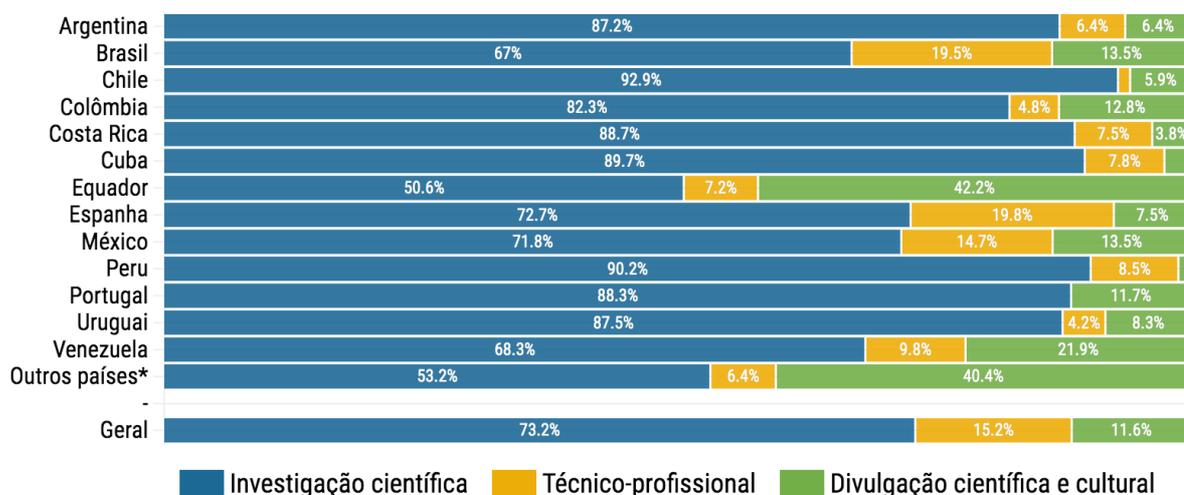
5.2.1 Tipo de conteúdo

O Gráfico 2 mostra a distribuição percentual de periódicos por país e por tipo de conteúdo da publicação, considerando as três categorias utilizadas pelo Latindex: investigação científica, divulgação científica e cultural ou técnico-profissional.

Na análise geral, os periódicos de investigação científica correspondem a quase 75% do total, seguidas dos periódicos técnico-profissionais e dos de divulgação científica e cultural.

Na distribuição por país há algumas variações. O Equador, a Venezuela e os outros países possuem maior percentual de periódicos de divulgação científica e cultural e menor de investigação científica em comparação com a média geral.

Gráfico 2 – Distribuição de periódicos por país e por tipo de conteúdo



* Bolívia, El Salvador, Guatemala, Nicarágua, Panamá, Paraguai e República Dominicana

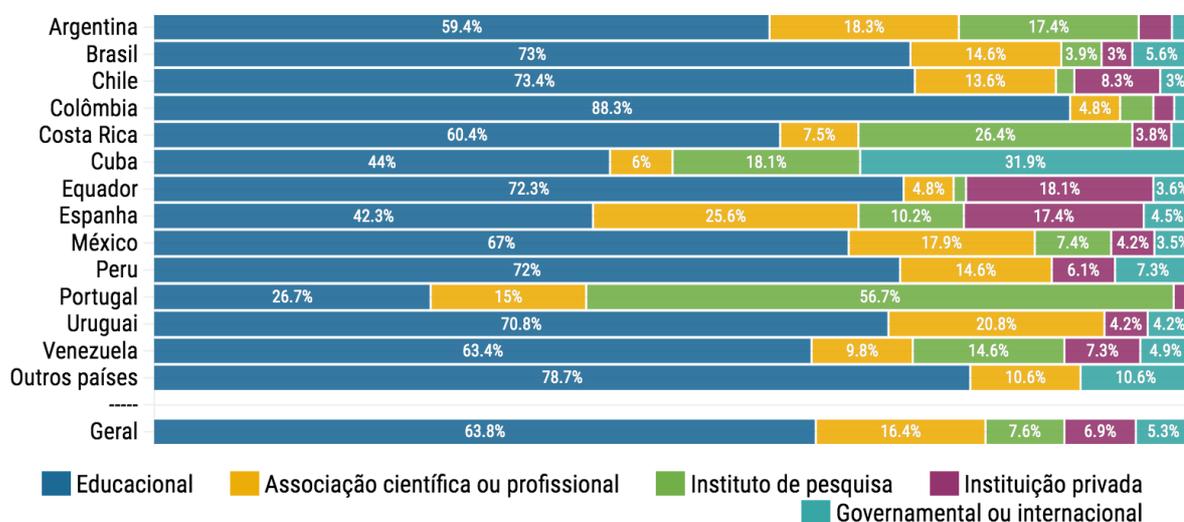
Fonte: Dados de pesquisa.

Brasil e Espanha possuem periódicos técnico-profissionais em percentual superior à média geral.

5.2.2 Natureza da instituição editora

O Gráfico 3 mostra a distribuição dos periódicos de acordo como país e a natureza da instituição editora. Mais de 63% são mantido por instituições educacionais e mais de 16% por associações científicas ou profissionais. Institutos de pesquisa, órgãos governamentais ou internacionais e instituições privadas editam um volume menor de publicações, em percentuais inferiores a 8% cada categoria.

Gráfico 3 – Distribuição de periódicos por país e por natureza da instituição editora



* Bolívia, El Salvador, Guatemala, Nicarágua, Panamá, Paraguai e República Dominicana

Fonte: Dados de pesquisa.

Algumas variações em relação ao padrão geral podem ser destacadas. Equador e Espanha possuem maior percentual de publicações mantidas por instituições privadas, incluindo editores comerciais. A Espanha também possui percentual significativo de publicações mantidas por associações científicas ou profissionais.

Esse resultado pode ser explicado pelo maior número de periódicos indexados no JCR e na Scopus. Conforme constatado em estudos anteriores, essas bases de dados indexam mais periódicos de editores comerciais (REPISO; ORDUÑA-MALEA; AGUADED, 2019).

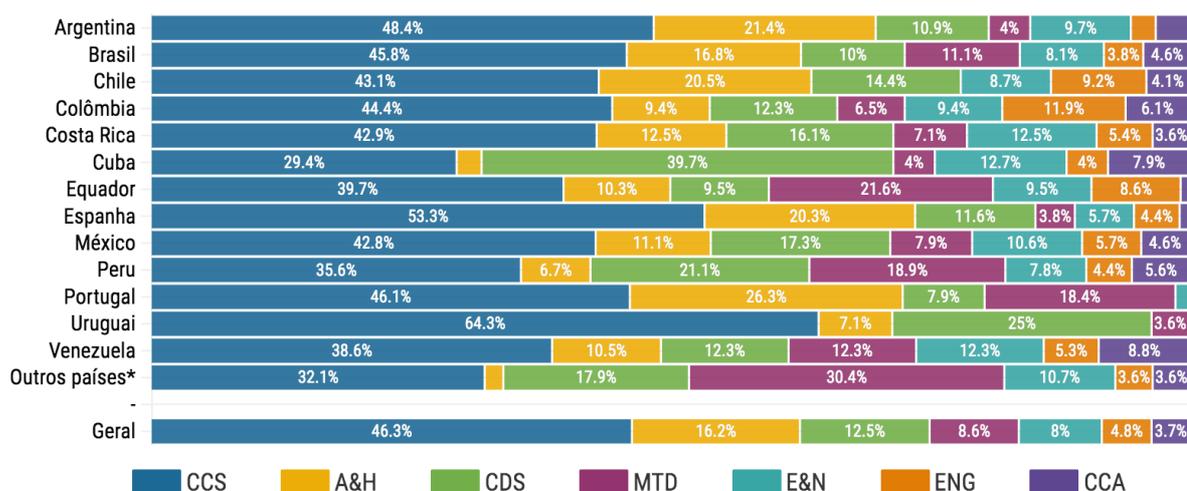
Cuba, por outro lado, possui elevado percentual de publicações mantidas por órgãos governamentais, o que pode ser explicado pelo regime político mantido nesse país.

Portugal tem a maioria das publicações editadas por institutos de pesquisa.

5.2.3 Aderência temática

O Gráfico 4 mostra a distribuição de periódicos de acordo com o país e com a área do conhecimento. Os periódicos classificados nas áreas de CCS, A&H e de CDS representam mais de 75% do total. Já os periódicos classificados nas áreas MTD, de E&N, de ENG e de CCA abrangem um percentual significativamente inferior na região.

Gráfico 4 – Distribuição de periódicos por país e por área do conhecimento



* Bolívia, El Salvador, Guatemala, Nicarágua, Panamá, Paraguai e República Dominicana

Fonte: Dados de pesquisa.

Há variações entre a média geral e os resultados de alguns países. A distribuição das publicações cubanas difere do padrão dos demais países mais produtivos. Os periódicos cubanos de CDS representam quase 40% do total. Além disso, os percentuais de periódicos cubanos de E&N e de CCA são superiores e o de CCS é inferior à média geral.

Em contraste, os periódicos de A&H e de CCS representam mais de 70% de todas as publicações da Espanha, de Portugal, da Argentina e do Uruguai.

Por fim, Bolívia, El Salvador, Equador, República Dominicana, Guatemala, Panamá e Peru têm percentuais de periódicos MTD superiores ao da média geral.

5.2.4 Escopo temático

Enquanto a análise de abrangência oferece uma visão geral acerca das áreas do conhecimento aos quais as publicações são inclinadas, a análise de escopo temático possibilita uma visão mais detalhada do enfoque científico. Os descritores de temas de pesquisa do Latindex foram utilizados para a análise de escopo temático.

A Figura 14 mostra a rede composta por descritores de assuntos e seus relacionamentos, isto é, quando dois ou mais descritores são atribuídos a um mesmo periódico.

Foram identificados 361 descritores diferentes no total de periódicos analisados. Para composição da rede foram selecionados os 170 termos com 15 ou mais ocorrências.

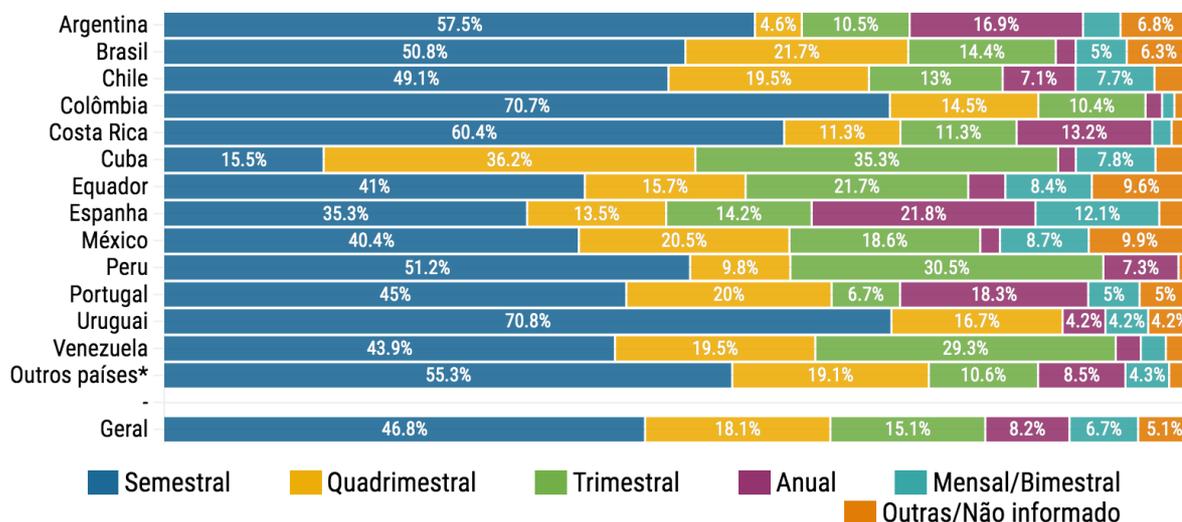
Refletindo o escopo temático predominante da região, se observa a maior ocorrência de termos referentes às áreas de CCS e de A&H. Os assuntos mais frequentes são multidisciplinaridade, educação, ciências sociais e humanidades, medicina, história, literatura, linguística, filosofia, sociologia, administração, psicologia, humanidades, saúde pública e direito.

Os descritores de maior frequência se agrupam com outros menos frequentes, estruturando oito *clusters*, ou seja, agrupamentos de acordo com a quantidade de ligações que estabelecem entre si. Cada *cluster* é representado com uma cor na rede.

O *cluster* de cor verde agrupa temas de pesquisa da área de A&H, tais como literatura, linguística, música, arte, desenho, belas artes, entre outros. O *cluster* lilás contém temas de pesquisa de ciências humanas e sociais, especialmente história,

Há também percentuais significativos de publicações quadrimestrais e trimestrais, respectivamente com 18% e 15%. As frequências mensal e bimestral, comum em periódicos mais produtivos, corresponde a menos de 7%.

Gráfico 5 – Distribuição de periódicos por país e por periodicidade



* Bolívia, El Salvador, Guatemala, Nicarágua, Panamá, Paraguai e República Dominicana

Fonte: Dados da pesquisa.

Se observam algumas variações nos padrões dos países mais produtivos da região. Cuba possui mais de 70% de publicações quadrimestrais e trimestrais e apenas 15,5% de semestrais.

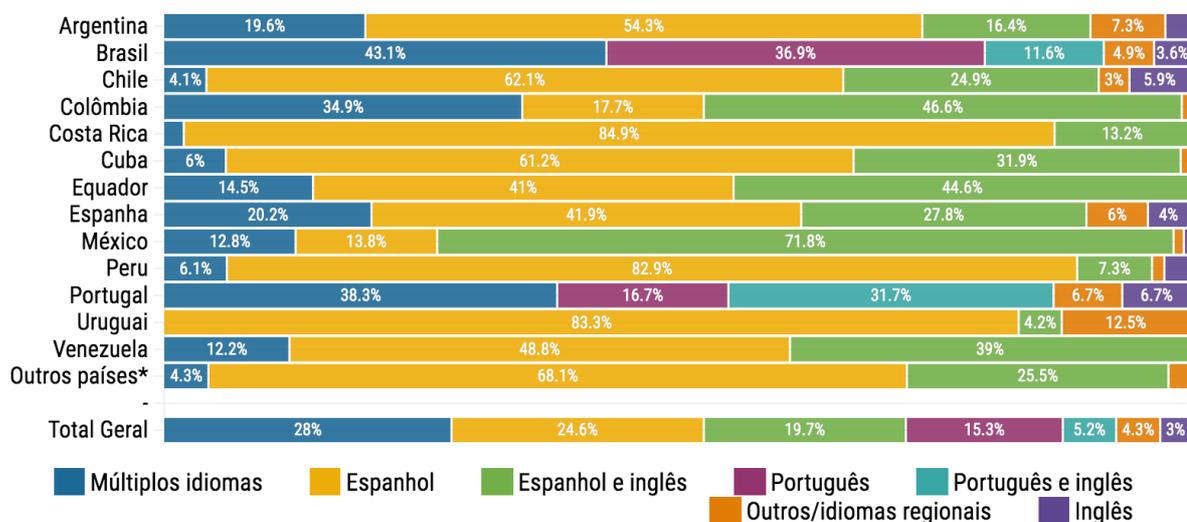
A Espanha possui menor parcela de publicações semestrais e maior de anuais e mensais/bimestrais. A Colômbia e o Uruguai, por outro lado, possuem mais de 70% de publicações semestrais.

5.2.6 Idiomas

O Gráfico 6 mostra a distribuição percentual de acordo com o país e com os idiomas das publicações. Os periódicos em múltiplos idiomas (incluindo o inglês), as somente em espanhol e as em espanhol e inglês englobam quase 75% do total. As publicações apenas em inglês representam somente 3%.

Na análise por país se destacam as publicações em espanhol, o idioma da maioria dos países da região, bem como as em espanhol e inglês. Nessa mesma lógica, as publicações em português e em português e inglês ocorrem apenas no Brasil e em Portugal.

Gráfico 6 – Distribuição de periódicos por país e por idioma



* Bolívia, El Salvador, Guatemala, Nicarágua, Panamá, Paraguai e República Dominicana

A Espanha é o país com maior percentual de publicações somente em inglês, bem como de idiomas regionais, que incluem o Catalão e o do país Basco.

5.2.7 Modelo de acesso

Quanto ao modelo de acesso, constatou que mais de 93% do total de publicações Ibero-Americanas localizadas no GSM são classificadas como de acesso aberto ou gratuito, 3,2% de acesso restrito e 1,3% de acesso por assinatura.

Esse percentual é similar na análise da maioria dos países, com poucas variações. Entre os países, a única diferença digna de nota se refere ao percentual de publicações espanholas restritas ou por assinatura, que supera os 16%.

5.3 IMPACTO DOS PERIÓDICOS

5.3.1 Validade do índice h5

Visando a testar a validade do índice h5 e da mediana h5 na mensuração do impacto dos periódicos Ibero-Americanos, foi realizada uma análise de correlação com os principais indicadores disponíveis nas bases de dados utilizadas como parâmetros comparativos.

Figura 15 – Matrix de correlação entre indicadores de periódicos Ibero-Americanos

		Spearman													
		h5	h5-m	AIS	ES	nES	FI	Flws	FI5y	JCI	CS	SJR	SNIP	h-index	m r
Pearson	h5	#	0.94	0.44	0.60	0.61	0.68	0.68	0.75	0.34	0.72	0.70	0.48	0.67	0.63
	h5-m	0.95	#	0.43	0.56	0.57	0.69	0.69	0.75	0.32	0.70	0.67	0.46	0.65	0.62
	AIS	0.49	0.48	#	0.86	0.87	0.68	0.72	0.75	0.71	0.48	0.53	0.58	0.41	0.62
	ES	0.73	0.66	0.59	#	1.00	0.62	0.64	0.69	0.63	0.63	0.66	0.54	0.70	0.68
	nES	0.73	0.66	0.59	1.00	#	0.62	0.64	0.69	0.63	0.63	0.66	0.54	0.70	0.68
	FI	0.59	0.63	0.68	0.45	0.45	#	0.98	0.93	0.78	0.90	0.77	0.62	0.68	0.75
	Flws	0.58	0.63	0.70	0.45	0.45	0.98	#	0.93	0.78	0.89	0.78	0.61	0.70	0.75
	FI5y	0.66	0.70	0.73	0.52	0.52	0.94	0.94	#	0.74	0.89	0.78	0.61	0.69	0.77
	JCI	0.27	0.27	0.61	0.30	0.30	0.71	0.72	0.68	#	0.41	0.46	0.56	0.23	0.55
	CS	0.66	0.66	0.56	0.54	0.54	0.87	0.87	0.88	0.38	#	0.85	0.64	0.78	0.71
	SJR	0.60	0.59	0.68	0.56	0.56	0.68	0.69	0.71	0.39	0.82	#	0.76	0.69	0.69
	SNIP	0.49	0.49	0.64	0.41	0.41	0.61	0.61	0.64	0.52	0.62	0.71	#	0.41	0.57
	h-index	0.65	0.61	0.45	0.75	0.76	0.54	0.56	0.61	0.12	0.70	0.65	0.35	#	0.61
	m p	0.62	0.61	0.60	0.58	0.58	0.68	0.68	0.71	0.44	0.67	0.64	0.54	0.56	0.66

Fonte: Dados de pesquisa. Índice h5 e mediana h5 extraídos do GSM (2021). AIS, ES, nES, IF, IFws, IF5y, JCI extraídos do JCR (2021). SJR, SNIP extraídos do SJR (2021) e CS e índice h5 extraídos da Scopus (2021).

A Figura 15 mostra os coeficientes de correlação entre os indicadores, com base nos testes de *Pearson* e de *Spearman*. A correlação média entre todos os indicadores é forte, com médias de $p=0.61$ e de $r=0.66$.

Os níveis de correlação são muito fortes nos dois testes entre as três configurações do FI (FI5y e FIws) e entre estes e o CS. Esses quatro indicadores também são os de maior correlação média geral, especialmente no teste de *Spearman*. Esse resultado pode ser atribuído à similaridade das fórmulas, baseadas na média de citações por artigo (SALVADOR-OLIVÁN; AGUSTÍN-LACRUZ, 2015).

Os níveis também são muito fortes entre o CS e o SJR e entre o AIS, e as duas variações do ES.

Na comparação entre os demais indicadores os níveis variam de forte a moderado. Os níveis fortes usualmente resultam da análise entre indicadores da mesma base de dados.

A comparação entre indicadores com período de análise mais longo (cinco anos) parece ocasionar coeficientes mais elevados. Isso se observa nos pareamentos entre FI5y, AIS e SJR em comparação com os resultados do pareamento destes com indicadores de período mais curto (dois anos).

O JCI, novo indicador do JCR, é o de menor nível geral de correlação. Por ser um indicador normalizado e definido com base em critérios específicos dessa base de dados, é esperado que apresente oscilações nos pareamentos com indicadores de outras fontes.

O índice h5 e a mediana h5 possuem níveis moderados de correlação geral, com médias de $p = 0,62$ e $r = 0,63$ e $p = 0,62$ e $r = 0,61$, respectivamente. Esses valores são próximos aos das médias gerais dos demais indicadores. Também são superiores ao do índice h calculado a partir dos dados de citação da Scopus, mas sem janela de tempo.

Os coeficientes mais elevados do índice h5 (nível forte) foram calculados nas comparações do teste do *Spearman* entre o FI5y, o CS e o SJR. Como os três são indicadores de período mais longo, a janela de tempo pode ter influenciado esse resultado.

No teste de *Pearson*, os coeficientes mais elevados foram obtidos na comparação entre indicadores “irmãos” (ES e nES, índice e mediana h5 e as três configurações do FI). Também são de nível forte os coeficientes entre FI e o CS e CS e SJR, assim como no teste de *Pearson*.

Níveis fracos de correlação são observados apenas com o JCI. Nos demais pareamento não são observadas discrepâncias relevantes.

O nível de correlação geral mais baixo no teste de *Pearson* pode ser reflexo dos diferentes formatos de apresentação numérica dos indicadores, que podem ampliar as variações para cima e para baixo dos valores, interferindo no coeficiente de *p*.

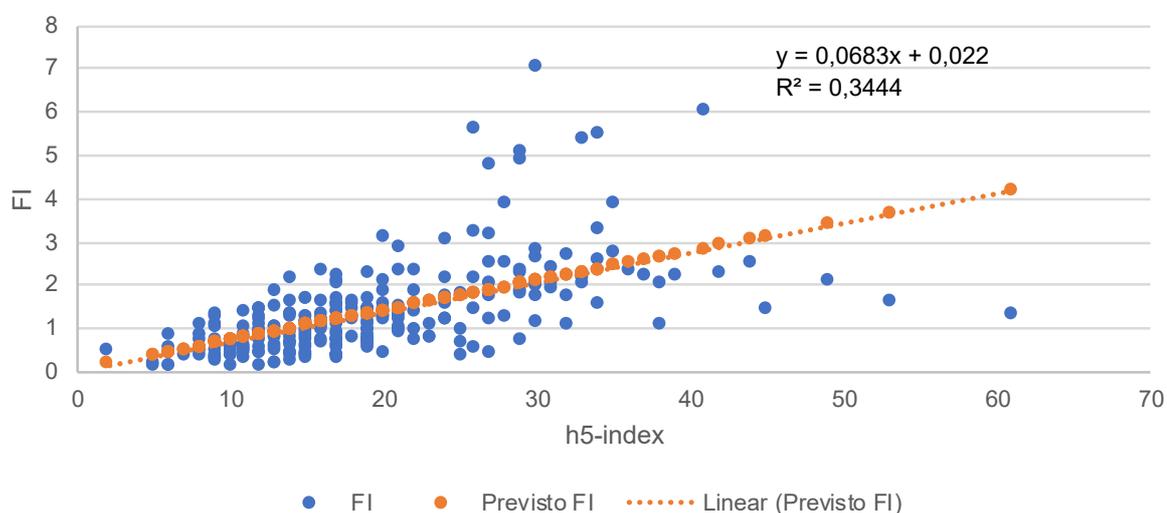
O melhor desempenho geral do teste de *Spearman* indica maior correção de posições ocupadas pelos periódicos em *rankings* a partir de diferentes indicadores.

Esses resultados sugerem que o índice h5 e a mediana h5 atendem ao requisito de validade para mensuração dos níveis de impacto, classificação ou elaboração de *rankings* de publicações Ibero-Americanas.

5.3.2 Previsão do FI a partir do índice h5

O Gráfico 7 mostra a previsão do FI com base no índice h5 em um modelo de regressão linear simples. As variáveis consideradas foram os valores dos indicadores de 263 periódicos da pesquisa que possuem ambos calculados.

Gráfico 7 - Previsão do Fator de Impacto a partir do índice h5



Fonte: Dados da pesquisa.

O modelo de regressão aponta que o FI de periódicos Ibero-Americanos que possuem índice h5 pode ser estimado por meio da fórmula: $FI = (0,0683 \times \text{“valor do índice h5”}) + 0,022$.

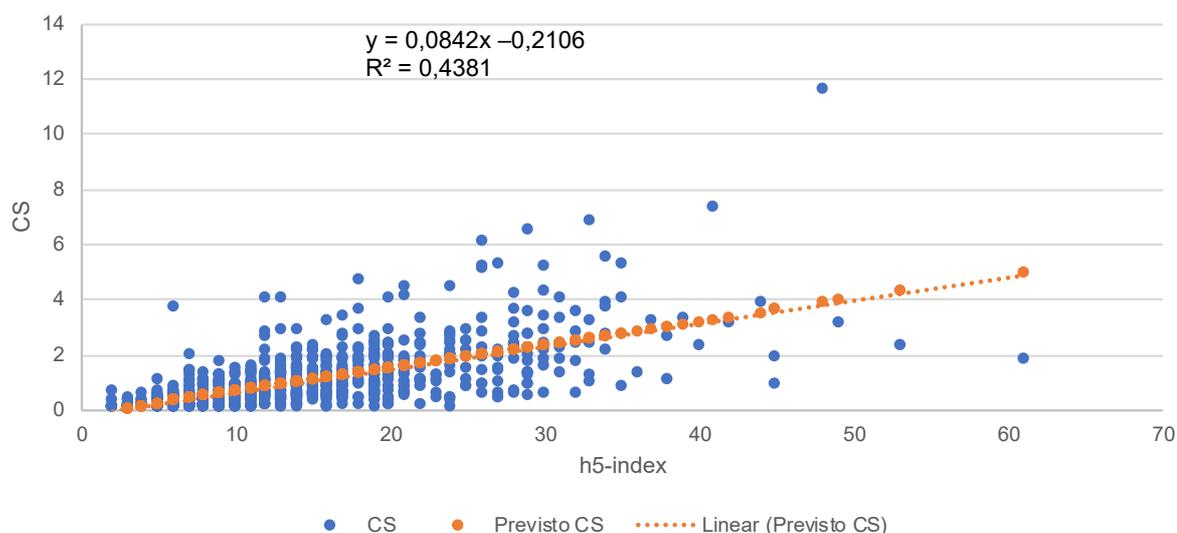
O coeficiente de determinação (R^2) calculado foi de 0,3444, o que implica em uma adequação parcial dos dados ao modelo de regressão. Deve-se considerar,

contudo, que o valor de R^2 é influenciado pelo coeficiente de correlação de *Pearson*, que conforme a Figura 15 é de nível moderado entre o FI e o índice h5 ($p = 0.59$).

5.3.3 Previsão do CS a partir do índice h5

O Gráfico 8 mostra a previsão do CS com base no índice h5 no mesmo modelo de regressão linear simples. As variáveis consideradas foram os valores dos indicadores de 978 periódicos indexados a partir das duas métricas.

Gráfico 8 - Previsão do CiteScore a partir do índice h5



Fonte: Dados da pesquisa.

O CS dos periódicos da região que possuem índice h5 pode ser estimado por meio da fórmula: $CS = (0,0842 \times \text{“valor do índice h”}) - 0,2106$.

Já o coeficiente de determinação (R^2) é igual a 0,4381. Esse valor de levemente superior que o do FI pode ser explicado pelo coeficiente de correlação mais elevado entre CS e índice h5 ($p = 0.66$).

Esses resultados sugerem que é possível estimar os valores de outros indicadores a partir do índice h5. Como esse indicador é o mais amplo para periódicos Ibero-Americanos, sua aplicabilidade vai além de sua própria medida de impacto, servindo como parâmetro para análises comparativas entre periódicos indexados em diferentes bases de dados.

O modelo de regressão linear simples utilizado na pesquisa possui precisão limitada, diretamente dependente do coeficiente de correlação de *Pearson* entre o índice h5 e o indicador que se deseja estimar. O mais adequado é utilizar essa

técnica a partir de periódicos de uma mesma área ou disciplina, visando a dirimir as diferenças de impacto existentes de acordo com o escopo temático.

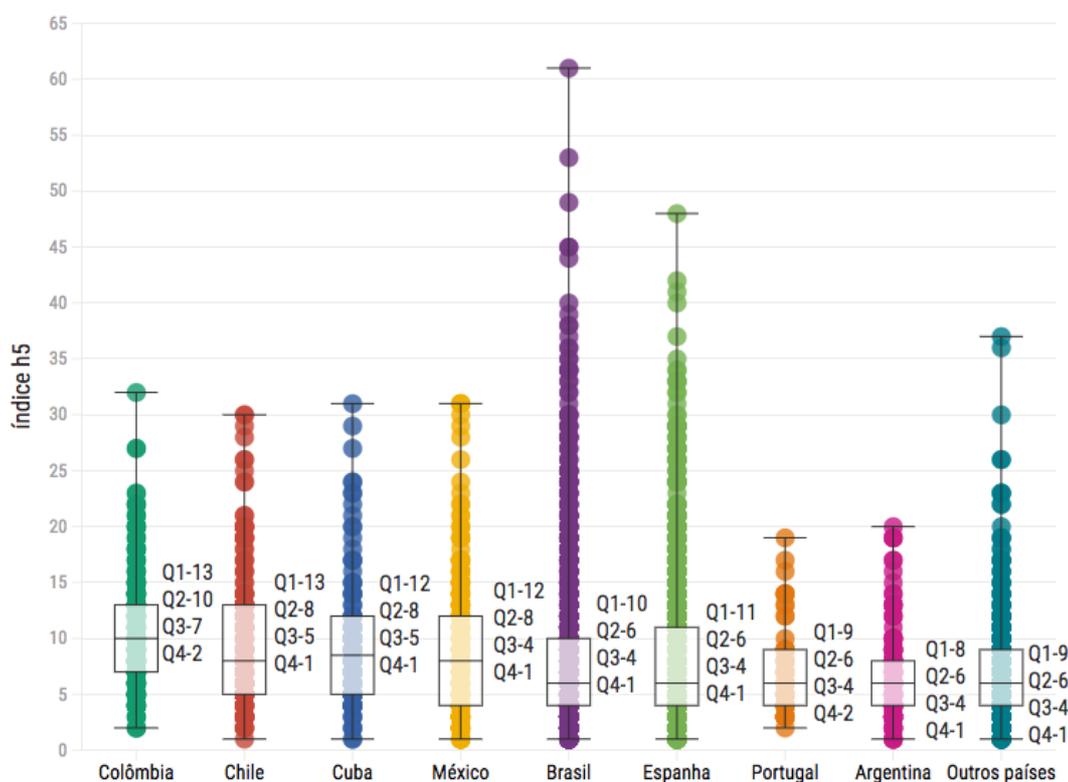
5.3.4 Impacto por país e por área do conhecimento

Constatada a validade do índice h5 para avaliação de periódicos Ibero-Americanos, as variações nos níveis de impacto foram analisadas a partir desse indicador. Os parâmetros adotados foram o país, a área do conhecimento, o idioma e a periodicidade das publicações.

O Gráfico 9 mostra a distribuição do índice h5 das publicações dos oito países mais produtivos e nos outros países em conjunto. São indicados também os limites inferiores do índice h5 em cada um dos quartis do diagrama de caixa (*box plot*).

Se observa que aproximadamente 75% de todos os periódicos possuem índice h5 com valores entre 1 e 12, considerando os limites dos quartis um e três (Q1-Q3).

Gráfico 9 – Distribuição do índice h5 dos periódicos por país



Fonte: Dados da pesquisa.

A mediana do índice h5 é mais elevada nos periódicos da Colômbia (10), Chile (8), Cuba (8) e México (8). Mas esses quatro países possuem periódicos com índice h5 máximo na faixa entre 30 e 32.

O maior volume de publicações do Brasil e da Espanha se refletem em padrões distintos de distribuição em comparação com os demais países. Embora a mediana seja inferior (6), esses dois países concentram as publicações de maior impacto, especialmente as com índice h5 acima de 35.

Os cinco periódicos brasileiros com índice h5 superiores a 40 são da área de CDS. Dos três periódicos espanhóis de maior impacto, um é de E&N e dois são de CCS.

Assim como no conjunto de outros países, Portugal e Argentina possuem publicações em níveis de impacto mais modesto. Ambos não possuem publicações com índice h5 superiores a 20. Já os outros países possuem algumas publicações em níveis mais elevados, incluindo dois periódicos, um venezuelano e outro peruano, com índice h5 superior a 35.

Tabela 7 – Média do índice h5 por país e por área

País	CDS	CCA	ENG	E&N	CCS	MTD	A&H	Geral
Colômbia	12,5	10,3	9,1	10,1	11	10,7	8,1	10,6
Chile	11	12,9	13,1	14,2	9,6	--	6,1	9,9
Cuba	12,6	8	6,7	6,8	8	10,6	6,7	9,7
México	9,3	11	9,7	8,6	9,2	7,7	4,1	8,7
Equador	6,4	7	10,2	10,2	7,7	9,1	7,5	8,3
Espanha	12	9,2	10,6	7,1	8,8	6,8	4,6	8,3
Venezuela	7,9	7,6	5,7	5,7	9,2	7,6	10	8,2
Costa Rica	8,4	13,5	9,1	6	8,4	5,5	5,4	8
Peru	10,2	9,8	8,3	5,8	7,6	5,8	6,7	7,8
Brasil	14,9	10,5	9,5	9,5	7,1	5,6	4,4	7,7
Portugal	13,3	--	14	--	7,3	7,1	5,1	7,2
Paraguai	7	--	--	--	6,3	6	10	6,9
Uruguai	11	--	--	--	5,9	2	2,5	6,8
Bolívia	8	6	6	--	7	6	--	6,6
Argentina	7,8	6,5	8,3	3,8	6,5	5	4	6,2
Rep. Dominicana	3	--	--	--	5,7	3	--	4,3
Panamá	--	--	4,5	6	3	3,7	--	4,1
Nicarágua	3	3	4,5	5	4,1	4,2	--	4,1
El Salvador	--	--	--	--	3	4	--	3,3
Guatemala	--	--	4	--	3	3	--	3,3
Geral	12,2	10,1	9,5	9	8	6,4	4,8	8,1

Fonte: Dados da pesquisa.

Complementando os dados do Gráfico 9, a Tabela 7 mostra a média do índice h5 das publicações por país e por área do conhecimento. No geral, Colômbia, Chile

e Cuba possuem as maiores médias de impacto, com valores próximos ou superiores a 10.

Na visão por área, os níveis de impacto mais elevados são o do Brasil em CDS, da Costa Rica em CCA, do Chile em ENG, de Portugal em E&N, da Colômbia em CCS e MTD, da Venezuela e do Paraguai em A&H.

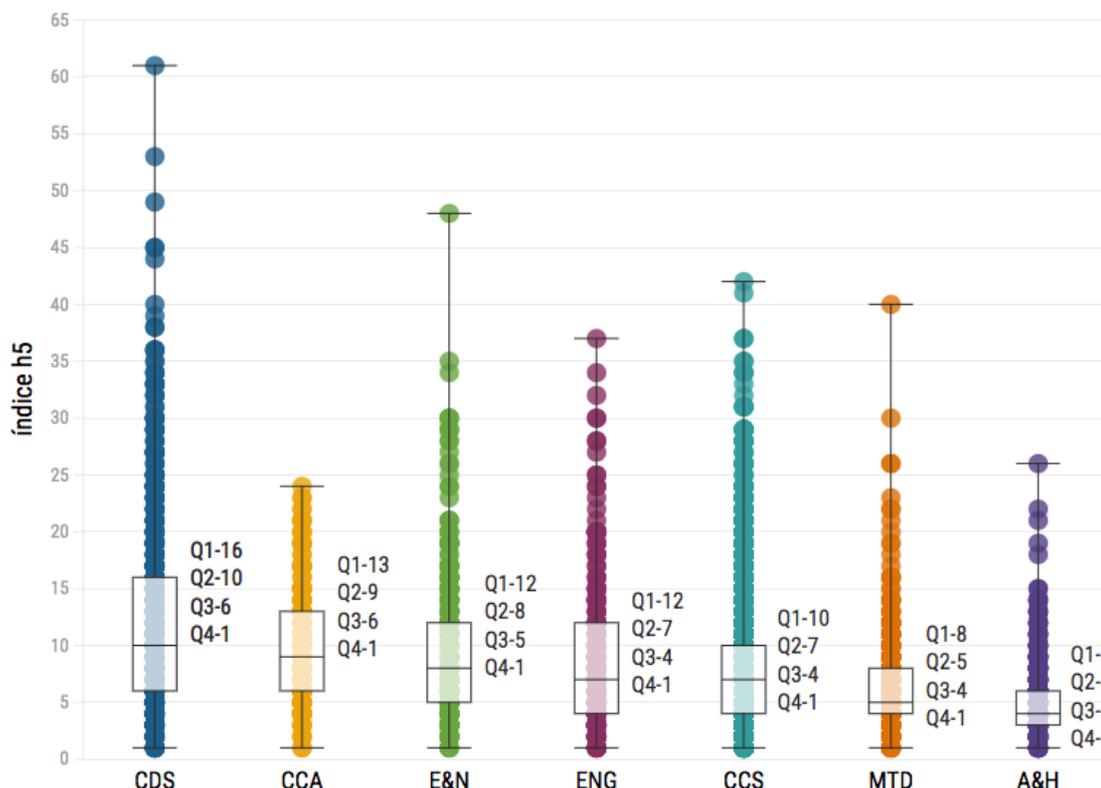
Entre os países mais produtivos, a Argentina é o que possui publicações com menor média do índice h5. Esse resultado pode ter sido em parte ocasionado pelo maior percentual de publicações de CCS e A&H desse país (Fig. 4), já que essas áreas possuem índices h5 médios inferiores que as demais.

Os periódicos com menores níveis de impacto são principalmente os dos países da América Central e do Caribe.

Os periódicos de CDS, CCA, ENG e E&N têm os maiores níveis de impacto, com índices h5 médios superiores a 9. Os periódicos de CCS, A&H e MTD, nessa ordem, possuem os menores níveis de impacto.

A variação dos níveis de impacto por área pode ser vista em detalhe no Gráfico 10, que mostra a distribuição do índice h5 em diagramas de caixa, com os respectivos limites de quartis.

Gráfico 10 – Distribuição do índice h5 dos periódicos por área



Fonte: Dados da pesquisa.

A distribuição em periódicos de CDS é mais assimétrica que as demais áreas. Além da maior mediana, essa área possui o maior número de publicações nos níveis mais elevados de impacto da região.

Além disso, a maioria dos periódicos com índice h5 em valores acima de 30 são de CDS, fator que contribui para a média mais elevada dessa área (Tabela 7).

São identificados também os limites inferiores dos quartis (Q1, Q2, Q3 e Q4) de acordo com a distribuição valores do índice h5 por área.

5.3.5 Impacto por país e natureza da instituição editora

A Tabela 8 mostra o impacto, representado pela média do índice h5, de acordo com o país e a natureza das instituições editoras. Os periódicos mantidos por associações científicas ou profissionais possuem maiores níveis de impacto, com índice h5 médio superior a 10.

Tabela 8 – Média do índice h5 por país e por natureza da instituição

País	Associação científica/profissional	Instituto de pesquisa	Governamental /internacional	Educacional	Privada
Argentina	6,9	6,9	6	5,8	8,4
Bolívia	--	--	--	6,6	--
Brasil	12,9	10,7	7,4	6,9	6,7
Chile	13,8	7,7	16	9,4	7,9
Colômbia	11,9	8,9	11,8	10,7	10,4
Costa Rica	9,5	8,8	4	7,9	5
Cuba	13,1	9,8	10,2	8,9	--
El Salvador	--	--	--	3,3	--
Equador	4,3	6	6	7,9	8,9
Espanha	9,8	8,4	7,7	8,5	6,6
Guatemala	--	--	--	3	--
México	9,8	11,2	11,7	8,4	7,1
Nicarágua	--	--	--	4,1	--
Panamá	--	--	2,5	4,5	--
Paraguai	5,2	--	4,5	8,8	--
Peru	7	--	10,7	7,7	10
Portugal	11,6	6,9	--	5,7	3
Rep. Dominicana	--	--	2	5,3	--
Uruguai	8,4	--	2	7,8	2
Venezuela	9,5	5,2	7	8	16
Geral	10,8	8,8	8,3	7,7	7,1

Fonte: Dados da pesquisa.

Em segundo nível de impacto seguem as publicações mantidas por institutos de pesquisa ou órgãos governamentais ou internacionais. Por fim, as publicações mantidas por instituições educacionais e as mantidas por instituições privadas são as de menores níveis de impacto na região.

Na análise por país, se destacam as médias dos periódicos chilenos, brasileiros e colombianos mantidos por associações científicas e profissionais. Periódicos mantidos por Institutos de pesquisa e instituições governamentais ou internacionais do México obtiveram níveis de impacto superiores aos das médias gerais. Os periódicos colombianos mantidos por instituições educacionais e por instituições privadas possuem desempenho médio no índice h5 superior ao geral.

5.3.6 Impacto por país e periodicidade

A Tabela 9 mostra os níveis de impacto, representado pelos valores médios da mediana h5, em suas variações por país e por periodicidade das publicações.

Tabela 9 – Média do índice h5 por país e por periodicidade

País	Mensal/ Bimestral	Trimestra I	Quadrimestral	Outras	Semestra I	Anua I
Argentina	10,6	7,5	7,5	5,5	5,9	5,9
Bolívia	--	--	--	--	6,6	
Brasil	19,2	12,4	8,1	8,3	5,5	4,4
Chile	16,5	13,4	12,5	13,2	7,5	6,4
Colômbia	18,3	13,3	13,8	13,8	9,6	7,8
Costa Rica	8,0	9,2	11,8	7,0	7,6	6,4
Cuba	16,8	10,7	8,9	5,8	7,3	6,5
El Salvador	--	--	4,0	--	3,0	--
Equador	8,6	8,6	9,5	5,5	7,4	5,3
Espanha	9,9	11,2	10,4	8,7	8,1	5,1
Guatemala	--	--	--	--	3,5	2,0
México	11,4	11,9	9,7	10,7	6,4	4,8
Nicarágua	2,0	5,5	3,5	--	4,1	--
Panamá	--	--	3,0	--	3,0	4,0
Paraguai	2,0	7,0	10,5	2,0	6,8	4,0
Peru	--	9,2	14,4	2,0	6,5	5,8
Portugal	16,7	11,5	8,0	5,0	6,6	4,4
Rep. Dominicana	--	7,5	2,7	--	--	--
Uruguai	4,0	--	6,8	6,0	8,1	4,0
Venezuela	9,0	9,6	9,5	37,0	5,3	7,0
Geral	13,5	11,5	9,3	8,6	6,7	5,2

Fonte: Dados da pesquisa.

Os dados sugerem que quanto mais frequente é o intervalo de publicação, maior é o impacto da publicação. Nesse sentido, publicações mensais ou bimestrais possuem maiores valores médios da mediana h5, seguidas das publicações trimestrais e quadrimestrais.

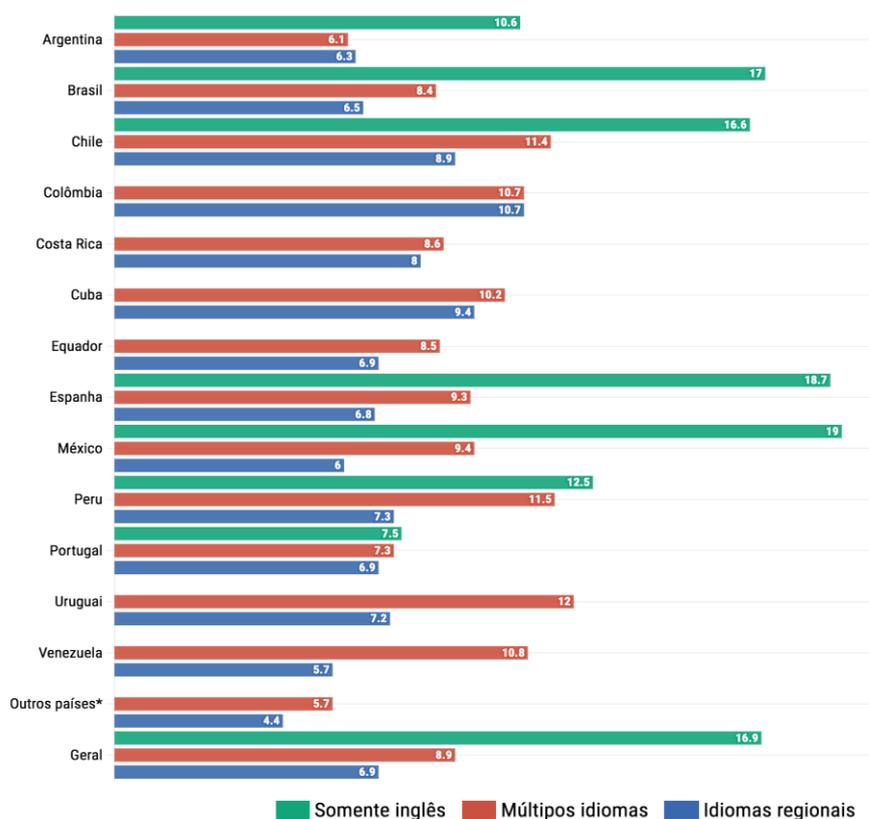
Conseqüentemente, as publicações semestrais, que são a maioria do total, e especialmente as publicações anuais possuem níveis de impacto inferiores dentre todas as periodicidades analisadas.

Esse resultado pode ser justificado pelo fato de que o índice h5 é um indicador de impacto dependente do tamanho da variável de produtividade. Logo, as publicações mais produtivas, que tendem a ter periodicidade mais frequente, possuem desempenho médio superior do que as publicações menos produtivas e de periodicidade mais longa.

5.3.7 Impacto por país e por idioma

O Gráfico 11 mostra os valores médios do índice h5 de acordo com o país e o idioma das publicações.

Gráfico 11 – Impacto por país e idioma



* Bolívia, El Salvador, Guatemala, Nicarágua, Panamá, Paraguai e República Dominicana

Fonte: Dados da pesquisa.

Os dados foram agrupados em três padrões: somente em inglês, múltiplos idiomas (incluindo o inglês) e um ou mais idiomas regionais (português, espanhol, catalão, francês, entre outros, não incluindo o inglês).

Publicações em inglês possuem maior impacto geral e em praticamente todos os países. No Brasil, na Espanha e no México o índice h5 médio das publicações é o dobro das publicações em múltiplos idiomas e, em vantagem ainda maior das publicações em idiomas regionais.

Publicações em múltiplos idiomas também possuem um nível de impacto superior ao das publicações em idiomas regionais, embora com menos diferença. Em alguns países a variação entre essas duas categorias de idiomas é inexistente (Colômbia) ou insignificante (Argentina e Cuba).

5.3.8 Crescimento anual do índice h5

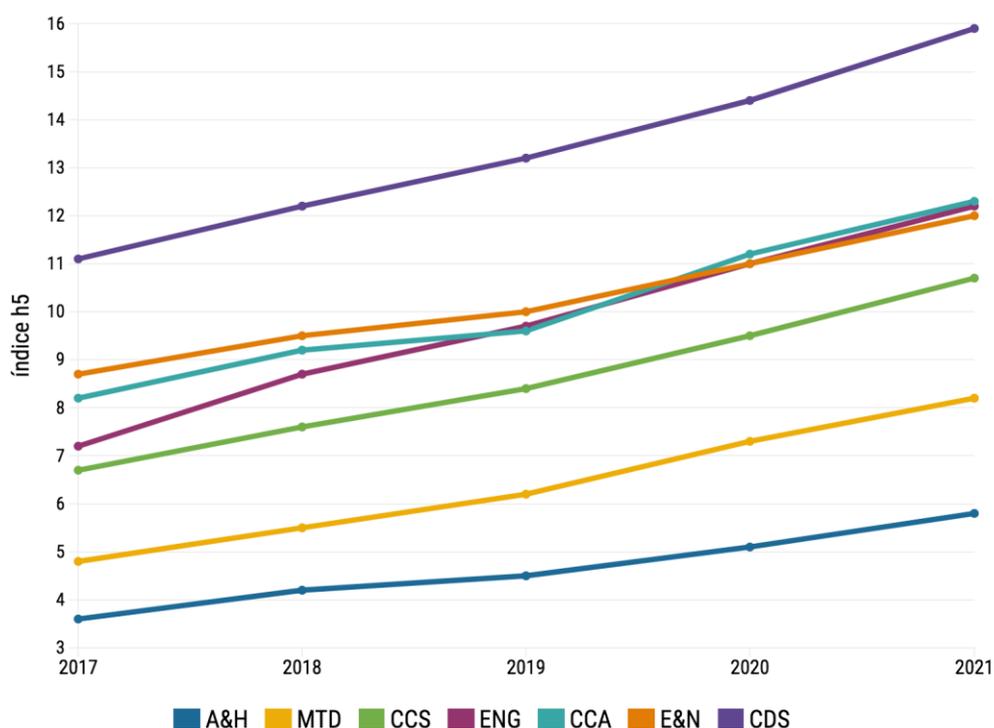
A estabilidade de um indicador está relacionada a sua suscetibilidade à variação bruscas em cada nova medição (HARZING, 2013). Uma das técnicas utilizadas para análise de estabilidade é o cálculo da taxa de crescimento anual em um conjunto de publicações.

O Gráfico 12 mostra a variação do índice h5 por área do conhecimento no período de cinco anos do índice h5. Essa análise se limitou a um conjunto de 1430 publicações que mantiveram o endereço URL em todo período, possibilitando a extração de dados por meio da ferramenta *Gsm_htline*.

Os resultados sugerem que o índice h5 tem uma tendência de crescimento constante. A partir dos valores iniciais do índice h5 em 2017, o aumento médio geral em cada ano foi de 12,7% em 2018, 8,8% em 2019, 12,2% em 2020 e 11,9% em 2021. Já o crescimento acumulado em cada área foi de 62,5% em A&H, 75,5% em MTD, 60,6% em CCS, 69,7% em ENG, 49,7% em CCA, 38,4% em E&N e 42,9% em CDS.

A taxa média de crescimento anual do índice h5 de todo o conjunto de publicações analisadas é de 11,4%. É praticamente idêntica à taxa de 11% de crescimento anual de mais de 330 periódicos brasileiros indexados no SciELO no período 2014-2018 (PACKER et al., 2019), o que indica a estabilidade do indicador em diferentes amostras e períodos.

Gráfico 12 – Crescimento do índice h5 por área do conhecimento (2017-2021)



Fonte: Dados da pesquisa.

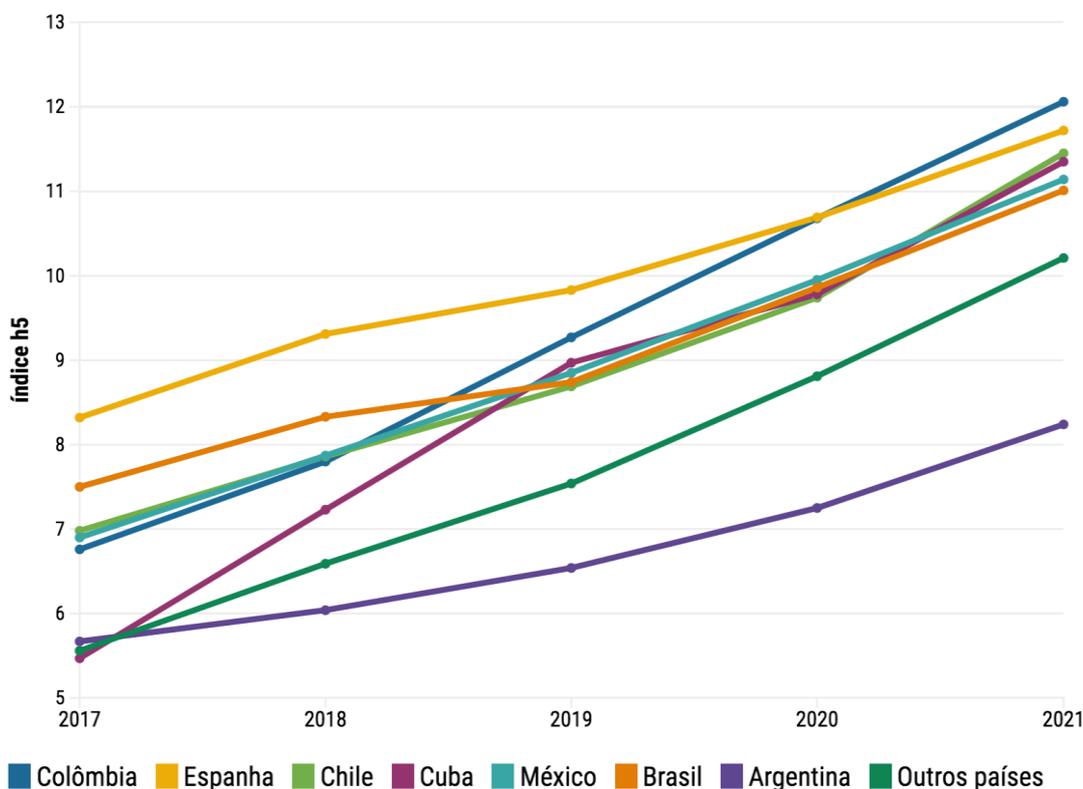
Esse ritmo crescimento manteve os níveis de impacto distintos entre a maioria das áreas ano a ano, especialmente entre CCS, MTD e A&H (menor impacto) e CDS (maior). Por outro lado, ENG, CCA e E&N, áreas que possuem níveis similares de impacto, trocaram posições, especialmente ENG devido ao crescimento acumulado superior.

O Gráfico 13 mostra o crescimento do índice h5 dos periódicos dos principais países e, de forma conjunta, dos outros países da região.

Os periódicos cubanos e colombianos registraram um aumento mais acentuado de impacto. O crescimento cubano foi de 20,3% no período, passando da última para a quarta posição no nível de impacto registrado no Gráfico 13.

De forma semelhante, os periódicos colombianos passaram da quinta para a primeira posição, graças a uma taxa média anual de crescimento de 15,6%. Os periódicos dos outros países também apresentaram uma taxa significativa de crescimento, com média anual de 16,4%.

Gráfico 13 – Crescimento do índice h5 por país (2017-2021)



Fonte: Dados de pesquisa.

Os periódicos do Chile, do México e do Brasil registraram taxas moderadas, respectivamente de 13,2%, de 12,7% e de 10,1%. Com isso, o Chile manteve a terceira posição, o México para quinta e o Brasil caiu da segunda para sexta posição.

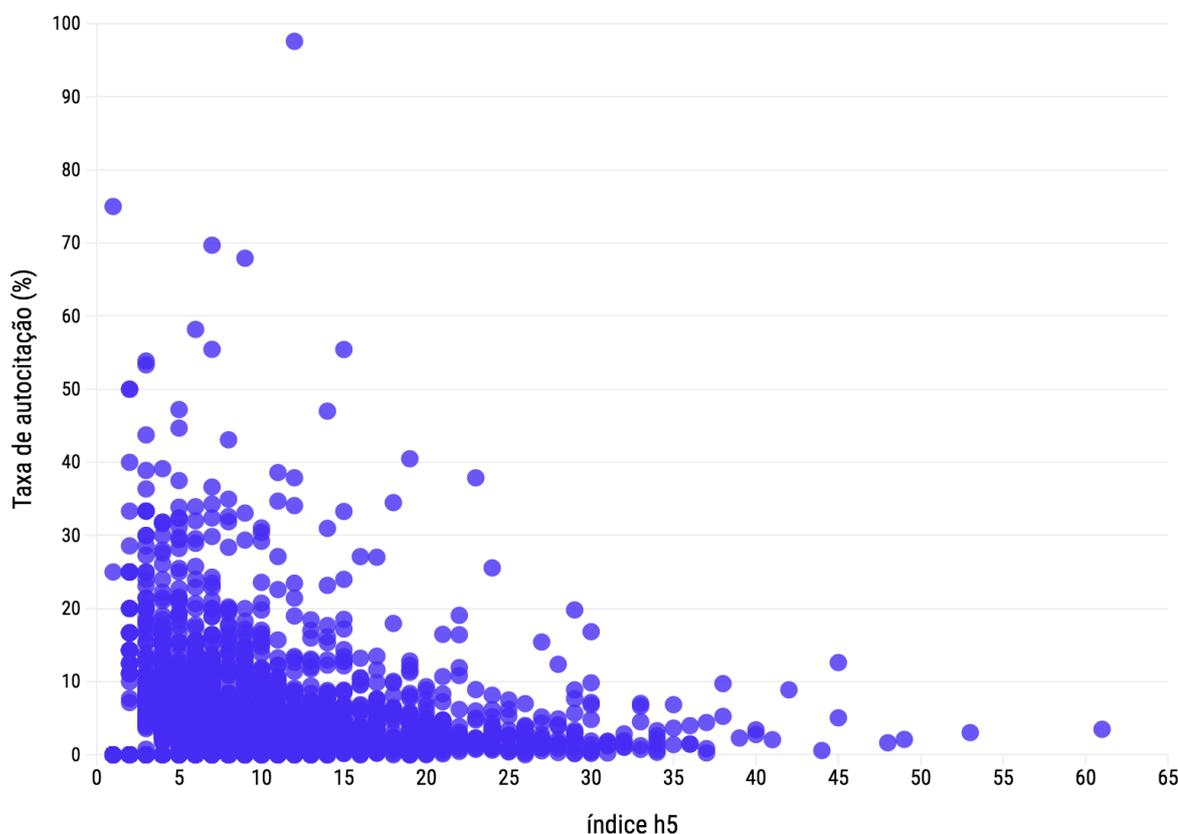
Já os periódicos da Espanha e da Argentina tiveram o menor ritmo de crescimento, com taxas de 9% e de 9,8%, respectivamente.

5.4 ANÁLISE DE AUTOCITAÇÃO NO ÍNDICE H5

O Gráfico 14 mostra a distribuição das SCR de acordo com o índice h5 de todos os periódicos do universo da pesquisa, referente a análise geral de autocitação. Não foram identificadas autocitações em 1586 periódicos, o que corresponde a 39,17% do total. Das publicações com SCR igual a zero, foi identificado que 1158 (73%) possuem índice h5 não superior a 5, ou seja, são de baixo nível de impacto.

Por outro lado, não foram detectadas SCR superiores a 20% em periódicos com índice h5 maior que 25.

Gráfico 14 – Distribuição das taxas de autocitação de acordo com o índice h5



Fonte: Dados de pesquisa.

Esses dados sugerem que SCR superiores ao limite geral de 20% fixado em estudos anteriores (TAŞKIN et al., 2021) são mais frequentes em periódicos de nível de impacto intermediário, com índice h5 entre cinco e 25.

Os valores da mediana e da média das SCR no índice h5 de todo o conjunto foram de 1,37% e 3,61%, respectivamente.

Já os valores da mediana e da média da SCR a partir dos dados do FI dos 124 periódicos que possuem esse indicador foram de 6,7% e 8,4%, ou seja, em índices equivalentes aos obtidos de outros estudos de autocitação com base no indicador do JCR (LIVAS; DELLI, 2018; SANFILIPPO et al., 2021; TAŞKIN et al., 2021; GAZNI; DIDEGAH, 2021).

A correlação de Pearson entre as SCR baseadas no índice h5 e no FI é de nível forte, com coeficiente $p = 0,69$. Esse resultado indica a validade dos dados de autocitação obtidos por meio da ferramenta Gsm_hscite, pois além de ser um valor relevante, é ainda superior ao coeficiente de $p = 0,59$ obtido entre o FI e o índice h5 (Figura 12).

Embora o tamanho dos dois conjuntos de periódicos não sejam proporcionais, esse resultado por comparação sugere que a autocitação de periódico exerce menos influência no índice h5 do que no FI.

Os resultados da análise geral com todos os periódicos possibilitaram uma visão preliminar do fenômeno da autocitação no resultado do índice h5, considerando inclusive publicações sem autocitação.

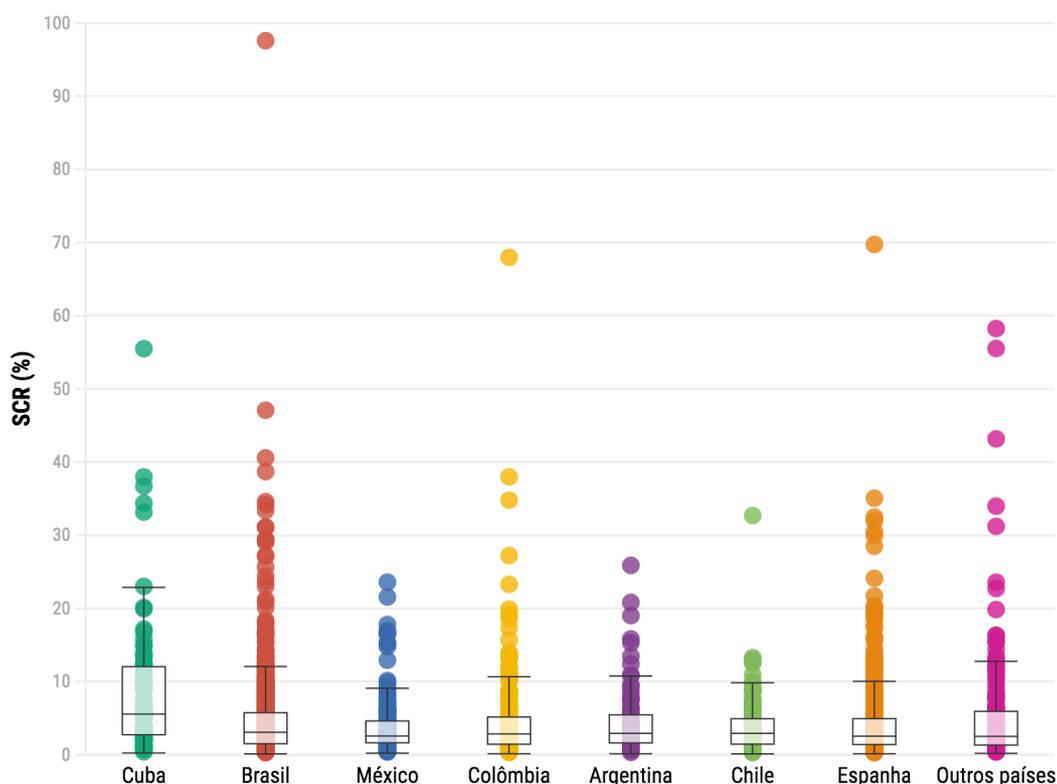
A análise específica que segue foi conduzida a partir de um subconjunto de periódicos, com no mínimo 50 citações recebidas, aproximadamente, sendo pelo menos uma autocitação ($SCR > 0$).

5.4.1 Taxa de autocitação por país

A aplicação do duplo critério resultou na identificação de um subconjunto de 1859 periódicos, que foram objeto da análise específica.

O Gráfico 15 mostra a distribuição das SCR de acordo com o país.

Gráfico 15 - Distribuição das taxas de autocitação por país



Fonte: Dados da pesquisa.

Cuba possui as SCR mais elevadas, com média de 9,8% e mediana de 5,6%. O Brasil fica em segundo, com média de 5,1% e mediana 3,9%. Ambos possuem

periódicos com SCR excessivas, especialmente as acima de 30%. O Brasil edita ainda um periódico com quase 100% de autocitações.

Colômbia e Argentina apresentam distribuição semelhante, com SCR medianas próximas a 3%. A Colômbia, contudo, possui mais *outliers*, especialmente com SCR superiores a 20%.

Chile, México e Espanha são os países com as SCR medianas mais baixas. Mas a Espanha e os outros países possuem mais *outliers*, inclusive em índices superiores a 40%.

Os dados sugerem que a autocitação ocorre com mais frequência em periódicos cubanos que nos dos demais países da região. Por outro ângulo, são observados periódicos com SCR excessivas em praticamente todos os países. Esse achado exige atenção à prática da autocitação em periódicos da região, especialmente os não indexados no JCR e na Scopus.

Periódicos indexados tendem a controlar as suas próprias SCR, já que as duas bases de dados possuem mecanismos de monitoramento de citações.

5.4.2 Taxa de autocitação por área

O Gráfico 16 mostra a distribuição de SCR de acordo com a área do conhecimento. A análise é complementada pelo cálculo das medidas de centralidade (média e mediana) por área.

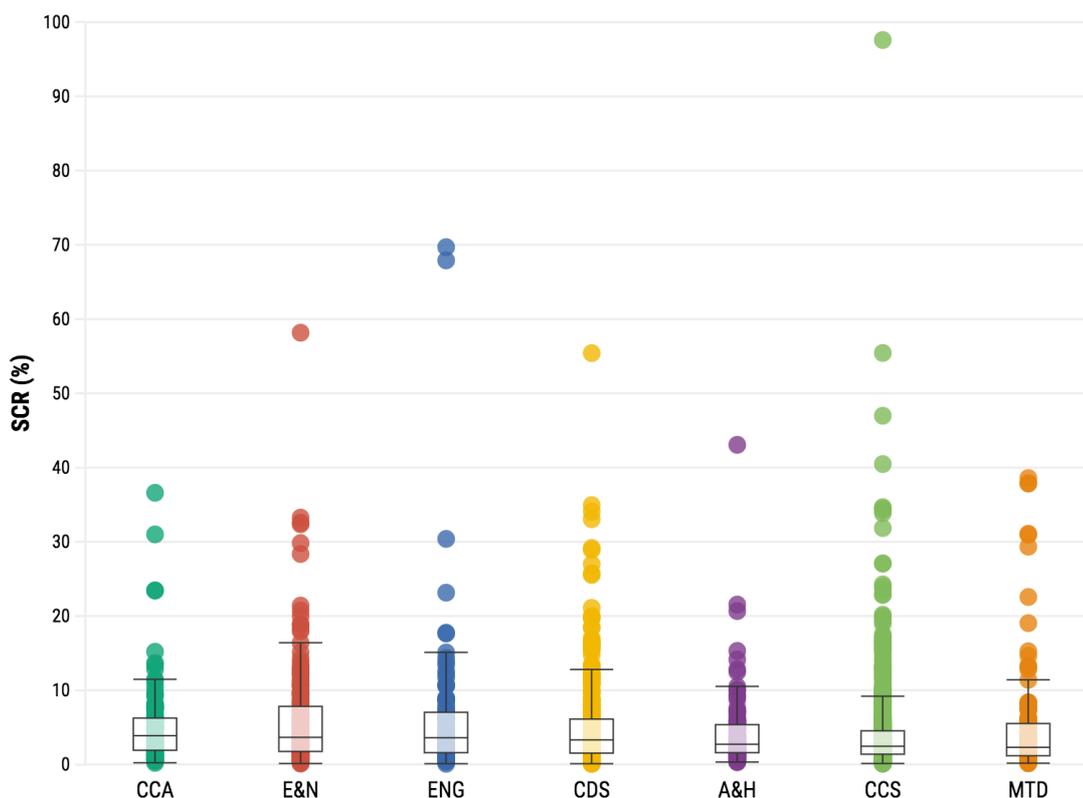
As SCR médias por área são 6,28% em ENG, 6,24% em E&N, 5,36% em CCA, 5% em MTD, 5,12% em CDS, 4,31% em A&H e 4,17% em CCS. A SCR média geral é de 4,83%.

As SCR medianas mais elevadas são as dos periódicos de CCA (3,89%), de E&N (3,68%), de ENG (3,62%) e de CDS (3,33%).

As áreas de MTD (2,33%), CCS (2,48%) A&H (2,74%) possuem as SCR medianas mais baixas.

Essa diferença entre as duas medidas de centralidade em todas as áreas pode ter sido ocasionada pela presença de *outliers*, ou seja, publicações com taxas discrepantes. SCR discrepantes, contudo, são minoria, pois 88,31% dos periódicos possuem menos de 10% de autocitações.

Gráfico 16 - Distribuição das taxas de autocitação por área



Fonte: Dados de pesquisa.

Esses resultados confirmam que as SCR calculadas com base no índice h5 tendem a ser inferiores do que as calculadas a partir do FI (TAŞKIN et al., 2021). Isso reforça a tese da menor suscetibilidade do índice há influência de autocitações (HIRSCH, 2005; WALTMAN, 2016), considerando a sua fórmula amostral, baseada em artigos mais citados.

Esse resultado pode embasar argumentos contrários à exclusão de autocitações na avaliação de impacto. Por outro lado, seria difícil defender a manutenção do desempenho de periódicos que, notadamente, tiveram aumento significativo de desempenho por força do excesso de autocitações.

A partir da premissa de que o excesso de autocitação deve ser monitorado (FRANSEN, 2007), a análise é direcionada para comparação entre periódicos com SCR normais e excessivas.

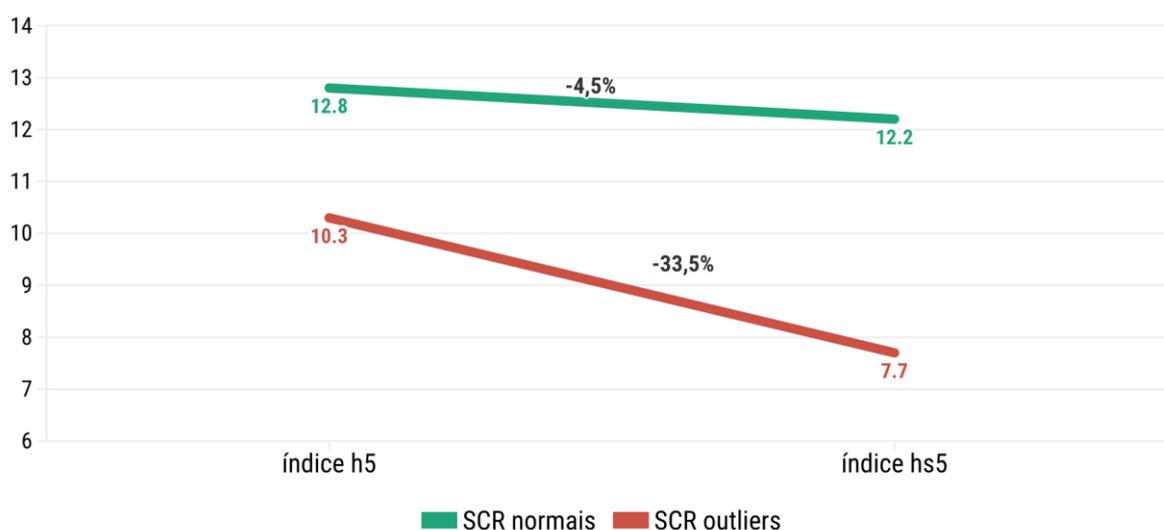
Periódicos com SCR normais são aqueles dentro de patamares médios de distribuição. Os periódicos com SCR excessivas são os estatisticamente identificados como *outliers* em conjuntos de periódicos com características em comum, como por exemplo, a mesma área do conhecimento (Gráfico 16) ou o mesmo país de origem (Gráfico 15).

A diferenciação entre SCR normais e *outliers*, portanto, decorre de análise estatística. Não é realizado no âmbito desta pesquisa juízo de valor acerca dos motivos de eventuais SCR extravagantes nos periódicos analisados.

5.4.3 Variação entre os índices h5 e h_s5

O Gráfico 17 mostra a variação média entre os índices h5 e h_s5 (VrH) em periódicos com SCR normais e *outliers* de todo o subconjunto da análise específica de autocitação.

Gráfico 17 - Variação entre os índices h5 e h_s5



Fonte: Dados da pesquisa.

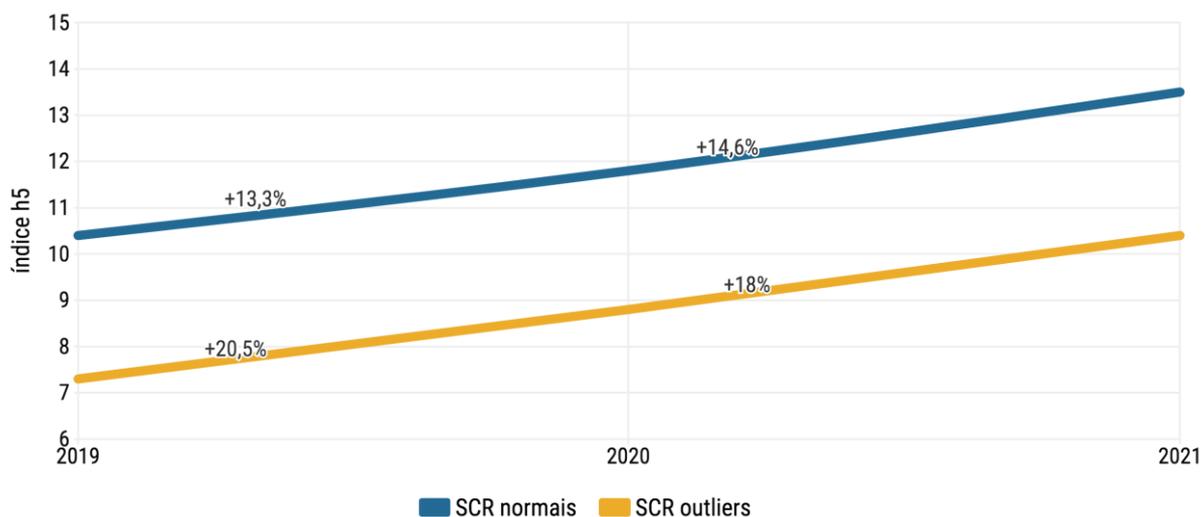
Os *outliers* possuem índice h5 médio inferior aos demais periódicos, conforme relatado em estudos a partir do FI (HENENBERG, 2016; TAŞKIN et al., 2021). A diferença se acentua na comparação pelo índice h_s5, com diminuição média de 33,5% no impacto médio dos *outliers*, ante uma queda inferior a 5% nos demais periódicos.

Esse resultado indica que o excesso de autocitações pode beneficiar periódicos e distorcer resultados de processos avaliativos a partir do índice h5 do GSM. É aconselhável monitorar autocitações e, a partir disso, fixar limites entre SCR normais e excessivas.

O Gráfico 18 mostra que os *outliers* obtiveram crescimento do índice h5 em ritmo superior aos demais nas últimas três edições do GSM (2019-2021). A vantagem acumulada foi de 10%, somadas as duas medições.

Entretanto, não é possível atribuir todo o crescimento adicional dos periódicos *outliers* somente às autocitações. Mas justificou a condução de uma análise mais profunda, especialmente por meio de comparação entre o desempenho médio de periódicos com SCR normais e *outliers*, com base nos indicadores definidos.

Gráfico 18 - Taxa de crescimento do índice h5



Fonte: Dados da pesquisa.

Foi aplicado o protocolo de análise de citação do *framework Gsm@hdata*, considerando a classificação por áreas do conhecimento. Foi observada a distribuição de SCR no *box plot* do Gráfico 16, com a diferenciação entre periódicos normais e *outliers*.

5.4.4 Análise de citação de periódicos de Ciências Agrárias

Foi constatado no Gráfico 16 que periódicos com SCR iguais ou superiores a 15,2% foram identificados como *outliers* de CCA. Do total de 108 periódicos, cinco *outliers* foram listados no protocolo de análise de citação. Os periódicos *outliers* foram identificados pela sigla da área e um número de ordem.

A Tabela 10 mostra que nos *outliers* de ordem um a quatro, o percentual de diminuição do índice h5 (VrH) foi inferior à SCR ($SCiR < 1$). Como resultado, o índice h₅ desses periódicos teve queda no valor *h* de um ou dois pontos.

O HgR3 médio dos *outliers* foi o mesmo da média da área. O periódico CCA4, contudo, teve índice de 100% nesse indicador, o que no índice h5 representou um aumento *h* de cinco para 10.

Resultado distinto foi observado em relação ao *outlier* CCA5, que teve VhR superior à SCR (SCiR = 1,2), ocasionando uma queda de três pontos no impacto isento de autocitações.

Tabela 10 – Análise de citação de periódicos de Ciências Agrárias

Área: CCA	h5	h ₅	SCR	VhR	SCiR	HgR3
Média da área	12,6	11,8	4,4%	-5,6%	1,3	29,8
Limite: estatístico	15,2%					
Média <i>outliers</i>	9,2	7,4	26%	-19,6%	0,8	29%
CCA1	10	9	15,2%	-10%	0,7	25%
CCA2	7	6	23,5%	-14,3%	0,6	-30%
CCA3	12	10	23,5%	-16,7%	0,7	50%
CCA4	10	8	31%	-20%	0,6	100%
CCA5	7	4	36,6%	-42,9%	1,2	0,0%

Fonte: Dados da pesquisa.

Esse periódico, no entanto, não teve aumento do índice h5 nos últimos três anos (HgR3), tendo as autocitações contribuído, no máximo, para a manutenção do valor *h*.

5.4.5 Análise de citação de periódicos de Engenharias

A Tabela 11 mostra a análise de citação de quatro *outliers* que tiveram SCR igual ou acima de 23,2%, percentual limite para *outliers* na amostra de 107 periódicos de ENG.

Tabela 11 - Análise de citação de periódicos de Engenharias

Área: ENG	h5	h ₅	SCR	VhR	SCiR	HgR3
Média da área	14	13,3	4,6%	-5,5%	1,2	33,4%
Limite: estatístico	17,8%					
Média <i>outliers</i>	9,5	5,8	44,6%	-39,5%	0,9	46%
ENG1	8	6	17,8%	-25,0%	1,4	100%
ENG2	14	11	23,2%	-21,4%	0,9	16,7%
ENG3	9	4	67,9%	-55,6%	0,8	80%
ENG4	7	2	69,7%	-71,4%	1,0	-12,5%

Fonte: Dados da pesquisa.

Os *outliers* tiveram diminuição média de quase 40% do índice h5 (VrH), levemente inferior à média das SCR. Essa diminuição foi mais acentuada que a percebida na análise dos periódicos de CCA.

O periódico ENG1 possui SCR no limite mínimo para *outlier*, mas registrou uma queda mais acentuada no índice h5, que foi detectada no resultado do SCiR (1,4).

Ainda foram observados resultados singulares nos indicadores dos *outliers* ENG 3 e 4, que possuem mais da metade do impacto decorrente de autocitações.

5.4.6 Análise de citação de periódicos de Ciências Exatas & Naturais

A Tabela 12 mostra os resultados da análise de citação de nove periódicos que tiveram SCR acima de 20%, limite para *outliers* na amostra de 183 periódicos de E&N.

Tabela 12 - Análise de citação de periódicos de Exatas & Naturais

Área: E&N	h5	h _s 5	SCR	VhR	SCiR	HgR3,
Média da área	12,4	11,7	5%	-5,3%	1,1	24,5%
Limite: estatístico	20%					
Média outliers	9	6,8	30,8%	-24,7%	0,8	21,7%
E&N1	8	6	20%	-25%	1,3	100%
E&N2	10	9	20,7%	-10%	0,5	n/d
E&N3	12	10	21,4%	-16,7%	0,8	50%
E&N4	8	6	28,4%	-25%	0,9	14,3%
E&N5	7	5	29,9%	-28,6%	1,0	0,0
E&N6	7	4	32,4%	-42,9%	1,3	40%
E&N7	8	5	32,6%	-37,5%	1,2	n/d
E&N8	15	13	33,3%	-13,3%	0,4	n/d
E&N9	6	3	58,2%	-50%	0,9	-40%

Fonte: Dados da pesquisa. Nota: n/d = Sem dados disponíveis.

Assim como nas áreas de CCA e ENG, os periódicos de E&N com SCR mais elevadas possuem índice h5 inferior a 10.

Os *outliers* E&N 6, 8 e 9 foram os com desempenho mais influenciado pelas autocitações, com variação de três pontos no valor *h*.

Por outro lado, não são identificadas HgR3 significativas, levando-se em conta que a média dos *outliers* é até inferior à da área.

5.4.7 Análise de citação de periódicos de Ciências da Saúde

A Tabela 13 mostra os resultados da análise de citação em periódicos de CDS. O índice estatisticamente estabelecido para identificação de *outliers* foi de

17,7%. Do total de 387 periódicos da área, 13 registraram SCR em índice igual ou superior e foram listados no protocolo de análise de citação.

A influência das autocitações não é uniforme mesmo entre os *outliers*. A maioria apresentou redução de um ou dois pontos no valor h , proporcionais à SCR média.

A média da HgR3 dos *outliers* é superior à da área, mas as individuais não são surpreendentes, exceto no periódico CDS7, com 240% de aumento de impacto acumulado.

Tabela 13 - Análise de citação de periódicos de Ciências da Saúde

Área: CDS	h5	h _s 5	SCR	VhR	SCiR	HgR3
Média da área	15,9	15,1	4,3%	-5%	1,2	24,3%
Limite: estatístico			19,7%			
Média outliers	12,3	9,5	28,8%	-22,5%	0,8	43,8%
CDS1	8	7	19,7%	-12,5%	0,6	n/d
CDS2	29	26	19,8%	-10,3%	0,5	52,6%
CDS3	9	8	20%	-11,1%	0,6	80%
CDS4	7	6	21,1%	-14,3%	0,7	16,7%
CDS5	24	16	25,6%	-33,3%	1,3	26,3%
CDS6	6	4	25,8%	-33,3%	1,3	n/d
CDS7	17	14	27%	-17,6%	0,7	240%
CDS8	6	4	28,9%	-33,3%	1,2	50%
CDS9	10	8	29,2%	-20%	0,7	25%
CDS10	9	7	33%	-22,2%	0,7	0,0
CDS11	12	9	34,1%	-25%	0,7	9,1%
CDS12	8	6	35%	-25%	0,7	-11,1%
CDS13	15	9	55,4%	-40%	0,7	-6,3%

Fonte: Dados da pesquisa. Nota: n/d = Sem dados disponíveis.

Se destacam ainda os periódicos CDS5 e 13, que tiveram maior redução de impacto com exclusão de autocitações, tanto em valores percentuais (VhR) quanto em valores h absolutos (h_{s5}).

5.4.8 Análise de citação de periódicos de Artes & Humanidades

De 121 periódicos da amostra de A&H, apenas quatro tiveram SCR iguais ou acima de 15,3%, sendo classificadas como *outliers*.

A Tabela 14 mostra que a influência das autocitações no índice h_5 dos periódicos *outliers* de A&H é inferior a observada nas demais áreas, com diminuição máxima de dois pontos nos periódicos A&H 1 e 4.

Tabela 14 - Análise de citação de periódicos de Artes & Humanidades

Área: A&H	h5	h _s 5	SCR	VhR	SCiR	HgR3
Média da área	9,1	8,8	3,6%	-4,1%	1,1	36,8%
Limite: estatístico	15,3%					
Média outliers	8,3	7	25,2%	-15,2%	0,6	25%
A&H1	14	12	15,3%	-14,3%	0,9	n/d
A&H2	6	5	20,7%	-16,7%	0,8	n/d
A&H3	5	5	21,6%	0,0%	0,0	25%
A&H4	8	6	43,1%	-25,0%	0,6	n/d

Fonte: Dados da pesquisa.

A ausência de dados impediu a análise do HgR3 de três periódicos, tendo esse indicador sido calculado apenas para um *outlier*.

A menor ocorrência de autocitações em periódicos de A&H já havia sido constatada por meio das SCR médias e medianas inferiores (seções 5.4.1 e 5.4.2).

Sendo A&H uma área ampla, que engloba disciplinas com características distintas, esse resultado corrobora a tendência de maior frequência de autocitações em áreas especializadas.

5.4.9 Análise de citação de periódicos de Ciências Sociais

Considerando a ampla amostra de 820 periódicos com média SCR de apenas 3,5%, a utilização da análise estatística para definição de *outliers* resultaria na seleção de aproximadamente 70 periódicos, muitos deles sem influência de autocitações no índice h5.

Então a identificação de *outliers* foi realizada por meio de taxa fixa de 20%, índice usual em estudos de autocitação (TAŞKIN et al., 2021).

A Tabela 15 mostra os resultados da análise de citação para os 18 considerados *outliers* a partir do limite de SCR fixado.

Se observa, primeiramente, o periódico (SSC18) com SCR mais elevada da pesquisa, de quase 100%. Por razões óbvias, é que apresenta a maior queda no índice h5 (-11 pontos).

Assim, salvo por eventual erro de indexação do GSM, o resultado indica que essa publicação não possui impacto real.

Tabela 15 - Análise de citação de periódicos de Ciências Sociais

Área: CCS	h5	h _s 5	SCR	VhR	SCiR	HgR3
Média da área	12,2	11,7	3,5%	-4,1%	1,2	31,9%
Limite: fixado	20%					
Média outliers	10,4	7,3	34,5%	-30,3%	0,9	131,3
SSC1	7	6	20%	-18,8%	0,9	16,7%
SSC2	8	6	20,2%	-37,5%	1,8	33,3%
SSC3	7	5	22,9%	-25%	1,1	133,3%
SSC4	6	4	22,9%	-14,3%	0,6	n/d
SSC5	10	8	23,6%	-25%	1,1	n/d
SSC6	15	12	24%	-20%	0,9	50%
SSC7	7	6	24,3%	-20%	0,8	75%
SSC8	16	13	27,1%	-27,3%	1,0	n/d
SSC9	11	8	27,1%	-28,6%	1,0	37,5%
SSC10	8	6	31,9%	-33,3%	1,1	14,3%
SSC11	5	4	33,9%	-28,6%	0,9	n/d
SSC12	7	5	34,3%	-25%	0,7	n/d
SSC13	18	14	34,5%	-22,2%	0,6	n/d
SSC14	11	8	34,7%	-20%	0,5	10%
SSC15	19	12	40,5%	-27,3%	0,7	533,3%
SSC16	14	9	47%	-36,8%	0,9	366,7%
SSC17	7	4	55,5%	-35,7%	0,7	n/d
SSC18	12	1	97,6%	-42,9%	0,8	n/d

Fonte: Dados da pesquisa. Nota: n/d = Sem dados disponíveis.

Se observa ainda dois periódicos com maior HgR3 dentre todos analisados, uma com mais de 300% e a outra mais de 500%. Ambas possuem SCR elevados, o que pode ser considerado um padrão discrepante, mesmo em comparação com os demais *outliers*.

5.4.10 Análise de citação de periódicos Multidisciplinares

De 138 periódicos da amostra de MTD, 13 tiveram SCR iguais ou acima de 13,5%, sendo classificados como *outliers*.

A Tabela 16 mostra que MTD é a única área com índice h5 médio dos *outliers* superior ao dos demais, o que é uma exceção do observado na literatura (HENEBERG, 2016) e nesta pesquisa.

Uma hipótese é que, sendo uma área multidisciplinar, não acompanha os padrões que diferenciam as demais áreas, haja vista que abrange periódicos de variadas disciplinas.

Tabela 16 - Padrões de citação em periódicos Multidisciplinares

Área: MTD	h5	h _s 5	SCR	VrH	SCiR	HgR3
Média da área	10,4	10	3,2%	-4%	1,1	51%
Limite: estatístico			15,2%			
Média outliers	12,6	9,2	29,2%	-25,5%	0,9	120,8%
MTD1	6	5	15,3%	-16,7%	1,1	n/d
MTD2	22	20	19,1%	-9,1%	0,5	37,5%
MTD3	11	9	22,6%	-18,2%	0,8	n/d
MTD4	9	6	29,4%	-33,3%	1,1	350%
MTD5	14	10	31%	-28,6%	0,9	133,3%
MTD6	5	4	31,1%	-20%	0,6	-16,7%
MTD7	23	13	37,9%	-43,5%	1,1	n/d
MTD8	12	8	37,9%	-33,3%	0,9	100%
MTD9	11	8	38,6%	-27,3%	0,7	n/d

Fonte: Dados da pesquisa. Nota: n/d = Sem dados disponíveis.

Já no cálculo do índice h_s5 esse padrão é ajustado, com os periódicos normais registrando impacto superior sem autocitações.

A área apresenta ainda o periódico com a segunda maior queda em valor absoluto do índice h5 após exclusão de autocitações: o MTD7, com -10 pontos.

Também se destaca o periódico MTD4, só que em virtude da HgR3 de 350%.

5.5 DISTRIBUIÇÃO DE CITAÇÕES POR ANO

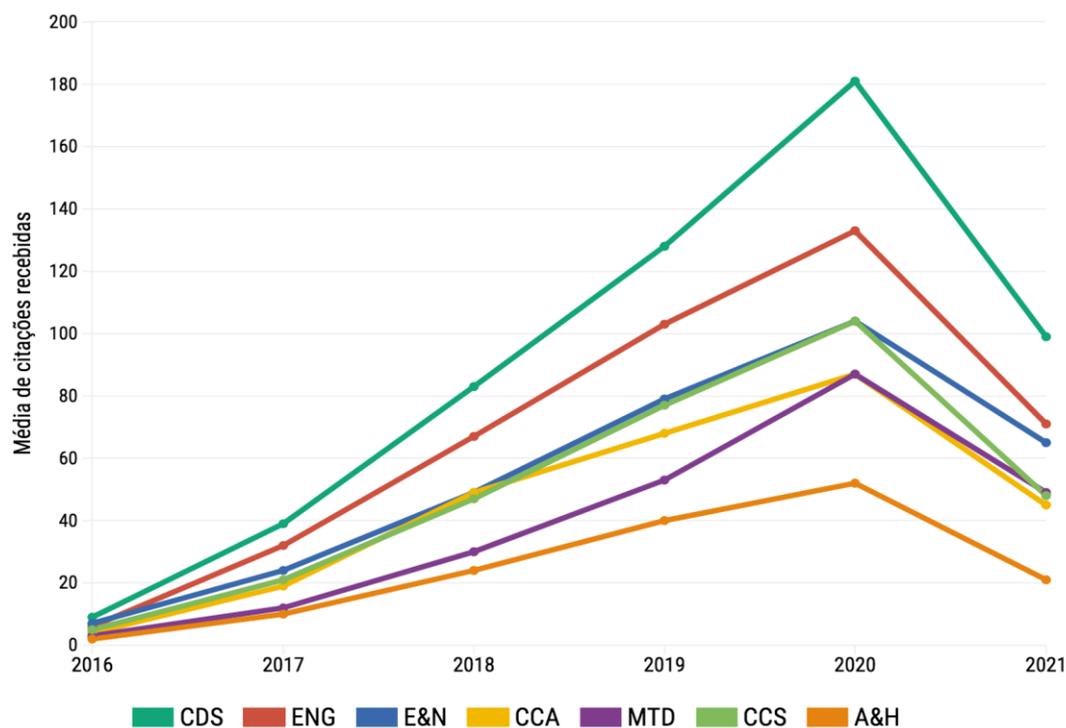
O Gráfico 19 mostra a distribuição do número médio de citações recebidas por ano no período 2016-2021, de acordo com a área.

O número médio geral de citações recebidas apresenta diferença proporcional ao impacto (índice h5 médio) de cada área, como era de se esperar.

Por outro lado, todas as áreas apresentam distribuição semelhante, com maior frequência de citações recebidas nos dois anos anteriores à edição atual do GSM.

Duas variações sutis da tendência geral podem ser observadas. A primeira é na distribuição de CCA, que parte do mesmo patamar de E&N em 2018, mas não acompanha o aumento desta nos anos seguintes. A segunda variação ocorre na curva de CCS e de MTD, que estão no mesmo nível em 2018, mas trocam de posição os anos seguintes.

Gráfico 19 - Distribuição de citações por ano e por área (2016-2021)



Fonte: Dados de pesquisa.

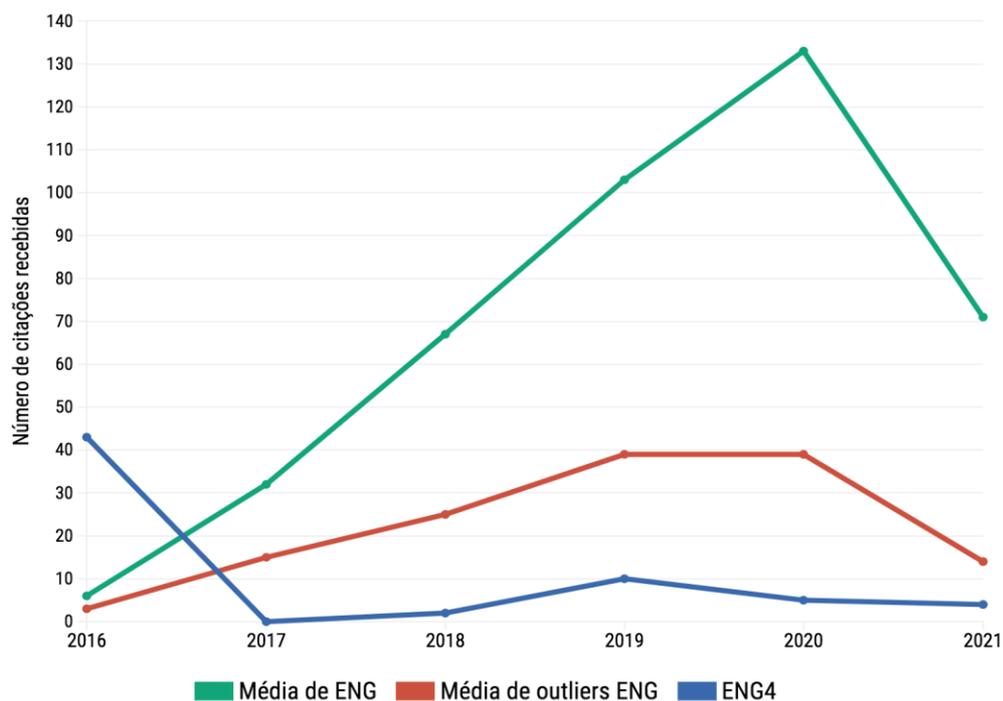
A partir da definição do padrão de distribuição de citações por ano, a identificação de periódicos com padrão discrepante da área pode ser considerado um indício de manipulação de citações, conforme consignado em estudos anteriores (CHORUS; WALTMAN, 2016).

5.5.1 Indícios de manipulação de citações

Os periódicos *outliers* listados no protocolo de análise de citação foram submetidos a análise comparativa da distribuição de citações por ano.

O Gráfico 20 mostra que a distribuição do periódico ENG4 é distinto do padrão da área e até da média dos três demais *outliers* de ENG.

Gráfico 20 - Distribuição de citações em periódicos de Engenharias (2016-2021)



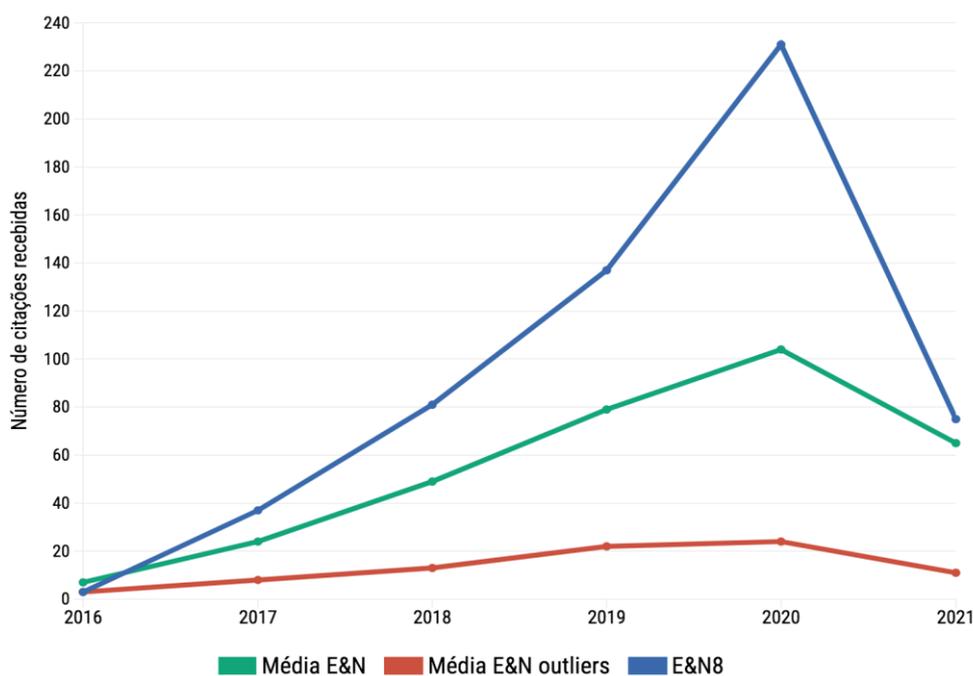
Fonte: Dados de pesquisa.

Foi identificado ainda que todas as autocitações do periódico ENG4 ocorreram em 2016, cessando totalmente nos anos seguintes.

Esse resultado pode ter sido causado pela publicação de um editorial ou de outro documento com as autocitações. Pode ainda ser oriundo um erro de indexação do GSM. Ou, indício de manipulação de citações (CHORUS; WALTMAN, 2016).

Padrão distinto, mas também digno de nota, foi detectado no padrão do periódico *outlier* E&N8, conforme mostra o Gráfico 21.

Gráfico 21 - Distribuição de citações em periódicos de Exatas & Naturais (2016-2021)



Fonte: Dados de pesquisa.

Embora a tendência de crescimento seja similar a da média da área, o *outlier* registrou mais que o dobro de aumento no número de citações em 2020.

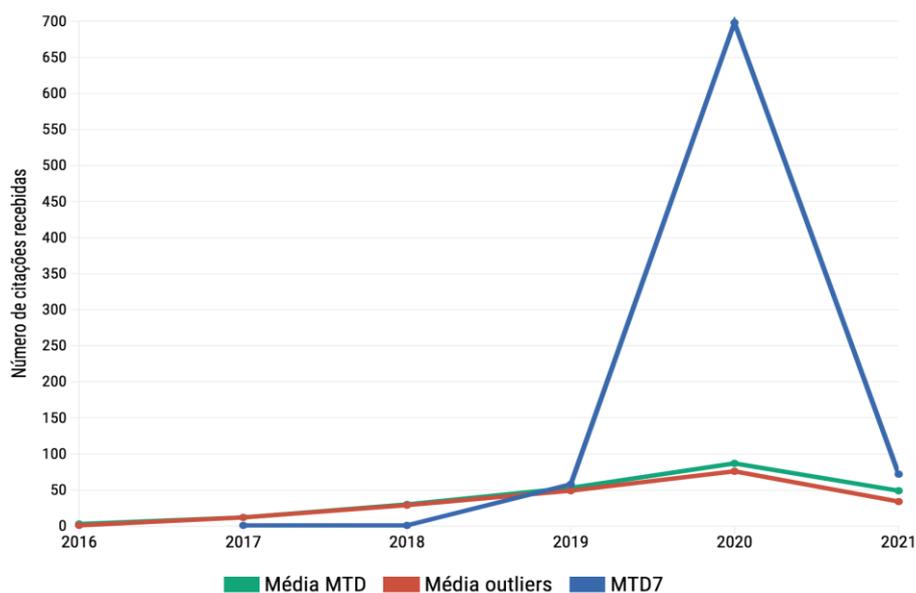
Entretanto, a VrH do E&N8 foi inferior ao seu SCR e até da média *outlier* (Tabela 11). Assim, embora a variação observada no Gráfico 21 possa ser considerada indício de manipulação de citações, o resultado de eventual prática não refletiu em aumento expressivo em indicadores de impacto, objetivo usualmente almejado em práticas dessa natureza (HENEBERG, 2016).

Essa situação tem conclusão diversa no que tange ao periódico *outlier* MTD7. O Gráfico 22 mostra uma concentração de citações recebidas no ano 2020, em intensidade sete vezes superior à das médias geral e *outliers* da área MTD.

Mas nesse caso, o periódico em destaque obteve ampla vantagem de impacto em razão do percentual expressivo de autocitações (Tabela 15), podendo ser levantada a hipótese de manipulação de citações.

O mesmo pode ser dito em relação ao periódico SSC15.

Gráfico 22 - Distribuição de citações em periódicos Multidisciplinares (2016-2021)

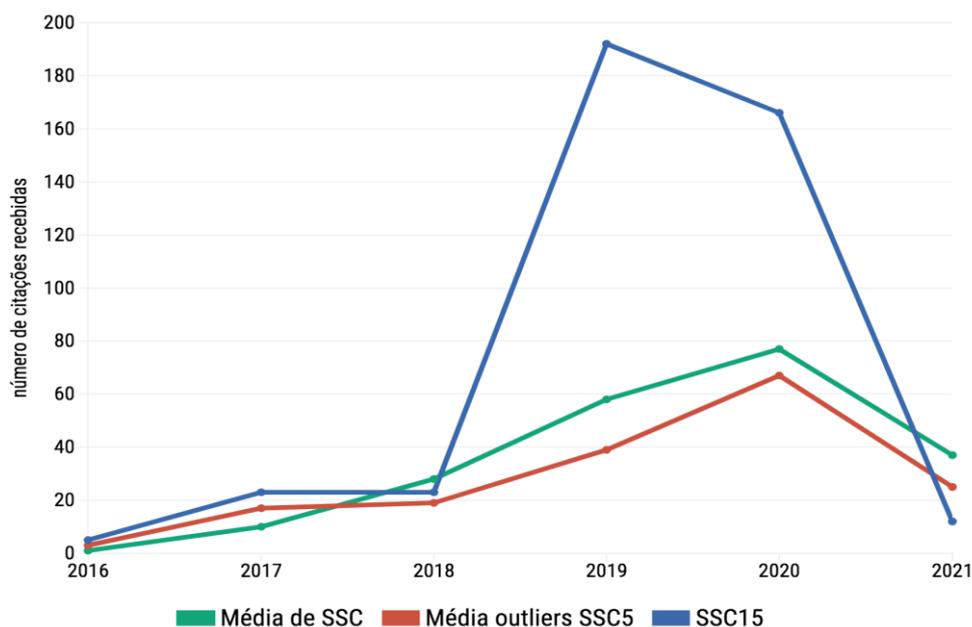


Fonte: Dados da pesquisa.

Conforme mostra o Gráfico 23, esse *outlier* não acompanha as outras médias de SSC, registrando crescimento abruoto no recebimento de citações de 2018 para 2019 e 2020.

A suspeita de manipulação de citações no SSC15 é reforçada pelo aumento substancial do índice h5 em razão das autocitações (Tabela 14).

Gráfico 23 - Distribuição de citações em periódicos de Ciências Sociais (2016-2021)

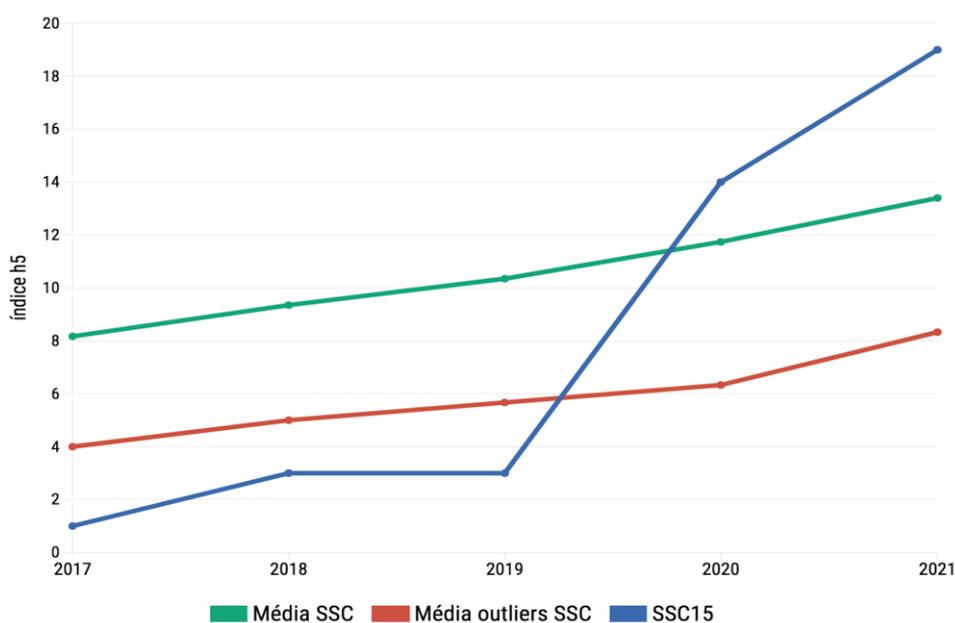


Fonte: Dados da pesquisa.

Além disso, a taxa de crescimento acumulada de três anos desse periódico não tem comparação na pesquisa, motivando, inclusive a análise em período estendido de cinco anos.

O Gráfico 24 mostra o índice h5 do periódico CCS15 no período de cinco anos, em comparação com os valores médios da área e dos demais *outliers* de CCS.

Gráfico 24 – Índice h5 de periódicos de Ciências Sociais (2017-2021)



Fonte: Dados da pesquisa.

O volume de citações (e especialmente autocitações) recebidas pelo periódico em destaque a partir de 2019 resultou em um aumento no índice h5 superior ao da área.

O conjunto de padrões anômalos desse periódico justificou a sua seleção para análise do último parâmetro utilizado para identificação de manipulação de citações: a existência de vínculo dentre fontes citantes e citada.

Foi utilizada a técnica de análise de redes sociais para construção de uma das citações recebidas pelo periódico CCS15, conforme mostra a Figura 16.

Além de um nodo de alta densidade representando as autocitações, outros dois nodos em destaque representam publicações vinculadas a mesma instituição do periódico analisado.

O cálculo dos dados da rede aponta que o “periódico vinculado a” efetuou 30 citações (10% do total) e o “periódico vinculado b” 45 citações (24%). Somadas,

Caso a justificativa seja totalmente aceita, a publicação seguiria o procedimento regular de classificação. Caso seja parcialmente aceita, a publicação pode ter sua classificação ajustada, visando corrigir o excesso de impacto ocasionado pelo padrão anormal.

Por fim, no caso de rejeição da justificativa, a publicação seria excluída do processo ou classificada como não científica, dependendo dos critérios do sistema de avaliação.

6 DISCUSSÃO

6.1 LIMITAÇÕES DO GOOGLE SCHOLAR METRICS

A realização desse estudo possibilitou uma análise detalhada do funcionamento do GSM, tendo sido constatado que a maior parte das limitações do sistema mencionadas em estudos anteriores ainda persistem (JACSÓ, 2012; LÓPEZ-CÓZAR; CABEZAS-CLAVIJO, 2012, 2013; LÓPEZ-CÓZAR; ORDUÑA-MALEA; MARTIN-MARTIN, 2019).

A não divulgação de uma lista geral de periódicos e a inexistência de filtros por ISSN, país ou região dificultam procedimentos de busca e de análise de cobertura geográfica do GSM (GU; BLACKMORE, 2017). A falta de padronização de entradas por título pode exigir múltiplas tentativas de busca de periódicos, com uso de variações do título, sigas e abreviaturas (REGENBRECHT; LANGLOTZ, 2015).

A falta de um recurso de extração dos dados também continua sendo um obstáculo para o uso do GSM, especialmente em estudos métricos em larga escala. A supressão de dados de edições anteriores, que não ocorre em outras bases de dados, torna complexa a realização de estudos longitudinais com base no índice h5.

Erros de indexação e de duplicação de entradas ainda são detectados. Não foi aferir, contudo, se são em menor proporção que o relatado em estudos anteriores do GSM. O não tratamento de autocitações e a possibilidade de manipulação de citações permanece sendo duas questões preponderantes para

A ferramenta bibliométrica *Gsm_hsearch* contornou parte considerável dessas limitações, com aceitável nível de precisão de resultados em seus quatro módulos. O *Gsm_hsearch* viabilizou a realização do estudo a partir do amplo universo de pesquisa. Foi automatizada a maior parte do procedimento de coleta de dados com baixo custo computacional.

A resposta do módulo *Gsm_hsearch* obteve alto nível de precisão, principalmente nos casos de igualdade entre a grafia do título pesquisado e o exibido no GSM. Em casos de semelhança entre os títulos, a resposta do sistema em padrão “contido em” permitiu a identificação a partir dos títulos incluídos na lista de resultados, sem necessidade de consulta ao GSM.

Em alguns casos, contudo, foi inevitável efetuar a busca manual na própria interface do GSM. Em outros, foi necessário também acessar o texto completo de artigos para identificação de dados adicionais da publicação (ISSN, editor, país).

Esses procedimentos foram realizados também para distinguir publicações homônimas ou com alto grau de semelhança de títulos (REGENBRECHT; LANGLOTZ, 2015).

Essas limitações no mecanismo de busca do GSM poderiam ser minimizadas com a inserção de um campo de ISSN. No entanto, passados 10 anos do seu lançamento, a Google parece querer manter o GSM um sistema minimalista (MARTÍN-MARTÍN et al., 2014), sem incorporar os recursos básicos presentes em praticamente todas as bases de dados de periódicos.

É mais provável ser o caso de uma estratégia de *design* dos produtos da família Google Scholar do que uma limitação técnica.

Durante os procedimentos de busca e de verificação manual foram encontrados inúmeros títulos com entrada duplicada, conforme reportado em estudos anteriores (LÓPEZ-CÓZAR; MARTÍN-MARTÍN, 2019; COSTA; CANTO; PINTO, 2020). Em alguns casos também foi constatado que um ou mais artigos listados no *h5-core* não pertenciam a publicação exibida pelo sistema, erro de atribuição que pode distorcer o resultado do índice h5.

Foi em razão desses erros que o *framework Gsm_hdata* inclui o procedimento de verificação manual dos dados. Somente a melhoria do sistema de controle bibliográfico do GSM, além da inclusão de um campo de ISSN, reduziria a possibilidade de erros e aumentaria do módulo *Gsm_hsearch*.

6.2 COBERTURA DO GOOGLE SCHOLAR METRICS

A escolha do diretório Latindex como fonte primária dos dados de pesquisa se mostrou acertada. É a fonte mais abrangente dentre as utilizadas nesse estudo, o que é justificável por se tratar um amplo diretório, com requisitos mínimos de inclusão de publicações.

Os níveis de padronização e de atualização dos dados dos periódicos Ibero-Americanos também são maiores em comparação com os dados de periódicos brasileiros extraídos no diretório Ulrich`sWeb, conforme divulgado nos resultados preliminares de pesquisa (PINTO et al., 2020; CANTO et al., 2022).

Foi constatado que o GSM é a fonte bibliométrica mais abrangente para analisar os periódicos da região a partir de um indicador de impacto reconhecido (índice h). O número de periódicos identificados no GSM é mais que o dobro do número do JCR e da Scopus e mais que o triplo do SciELO (Tabela 6).

O GSM permite a avaliação de periódicos de áreas de menor visibilidade, especialmente de ciências sociais e de humanidades, mas que são preponderantes no contexto Ibero-Americano.

A comparação entre o JCR, a Scopus e o SciELO evidenciou a ampliação de cobertura regional promovida pela Clarivate, especialmente pela criação do ESCI, se refletiu na atual cobertura do JCR. A Scopus, que sempre foi considerada uma fonte mais ampla para publicações regionais e de idiomas não inglês perdeu essa vantagem, já que na última atualização a Clarivate passou a incluir as coleções de periódicos emergentes e de humanidades no JCR.

O SciELO, por outro lado, não tem acompanhado o ritmo de crescimento da visibilidade dos periódicos Ibero-Americanos, mantendo a sua coleção com praticamente o mesmo tamanho nos últimos anos (Gráfico 1). O caminho de periódicos da região, que usualmente passava pelo SciELO para chegar na WoS, atualmente pode encontrar um caminho direto por meio do ESCI (GONZALEZ-PARDO; REPISO; ARROYAVE-CABRERA, 2020).

Não seria difícil acreditar na existência de uma cláusula contratual de “não expansão” na criação do SciELO Citation Index. Mas essa questão é alheia aos objetivos dessa pesquisa, podendo ser explorada em estudos futuros.

Mas mesmo com a ampliação de visibilidade internacional patrocinada pela Clarivate, o GSM ainda é a única fonte de avaliação de impacto para periódicos de alguns países, pois República Dominicana, Panamá, Honduras e Guatemala não possuem periódicos indexados nas outras bases de dados. Em outros países com poucos periódicos indexados, o GSM tem cobertura mais ampla, tais como o Peru, o Equador e Cuba (Tabela 6).

Mas o GSM promove uma ampliação de cobertura mesmo em países de maior visibilidade internacional, tais como Brasil e a Espanha. Esses dois países são os com maior número de periódicos identificados, abrangendo mais de 65% do da região indexado do GSM.

O Brasil possui um volume superior de periódicos que a Espanha, tanto no Latindex quanto os localizados no GSM nesse estudo. No caso do Brasil, esse resultado amplia o achado preliminar, que identificou 1070 publicações a partir de uma lista mais restrita de títulos extraída do Ulrichsweb (PINTO et al., 2020).

No caso da Espanha, por outro lado, o número encontrado nesse estudo foi inferior ao último relatório de periódicos espanhóis indexados no GSM, que listou

1448 periódicos na edição de 2019 do GSM (LÓPEZ-CÓZAR; MARTÍN-MARTÍN, 2019). O país ibérico, por outro lado, possui maior visibilidade no JCR e na Scopus, resultado que pode ser em parte explicado por algumas distinções entre os periódicos espanhóis e os demais da região.

O Brasil possui 73% dos periódicos mantidos por instituições de ensino e apenas 5% são de instituições privadas, categoria que inclui os editores comerciais. Na Espanha são menos de 45% de periódicos acadêmicos, quase a mesma proporção de periódicos mantidos por associações científicas ou profissionais e as instituições privadas (Gráfico 3).

A origem institucional das publicações pode ser uma das razões da maior indexação da Espanha na Scopus e no JCR, já que periódicos mantidos por editores comerciais são maioria nessas bases de dados (REPISO; ORDUÑA-MALEA; AGUADED, 2019). O impacto médio mais elevado dos periódicos espanhóis também poderia explicar a vantagem na indexação (Tabela 7).

É provável que outros fatores não abordados nessa pesquisa também podem ter influência nas diferenças dos resultados entre Brasil e Espanha.

A localização da Espanha, mais próxima de grupos editoriais, laboratórios, universidades e centros de pesquisa europeus a melhor estrutura acadêmica e científica são alguns fatores que podem explicar a sua vantagem de visibilidade. O incentivo à internacionalização de periódicos parece ter se iniciado antes na Espanha que no Brasil, refletindo nos números atuais (NAVAS-FERNÁNDEZ; ABADAL; RODRIGUES, 2018).

Somados aos periódicos do Brasil e da Espanha, os do México, da Colômbia, da Argentina, do Chile e de Cuba englobam mais de 95% do universo da pesquisa. Isso justificou uma análise mais detalhada desses países individualmente, e dos demais países de forma conjunta.

6.3 CARACTERÍSTICAS DOS PERIÓDICOS IBERO-AMERICANOS

Os periódicos analisados compõem um conjunto suficiente para representar o cenário editorial Ibero-Americano, ainda que seja limitado aos periódicos mais produtivos. É provável que a cobertura de mais de 4 mil periódicos da região pelo GSM não seja mais ampla devido critério de produtividade mínima de 100 artigos em cinco anos exigido para indexação, conforme já pontuado por López-Cózar e Martín-Martín, 2019 e Costa, Canto e Pinto (2020).

A indexação de periódicos técnico-profissionais e de divulgação científica e cultural pode ser considerada mais uma vantagem do GSM em comparação com as bases de dados comerciais. Essa política favorece principalmente os países de menor produção e estrutura de editoração de periódicos, que contam com menos publicações de investigação científica (Gráfico 2).

Por outro lado, a indexação conjunta de periódicos de tipos conteúdos distintos requer a utilização de outra fonte para diferenciação, considerando que o GSM não faz distinção entre publicações científicas e não científicas, revisadas por pares e não revisadas.

A baixa presença de editores comerciais é uma característica da editoração científica da região. A relação entre esse fator e baixa indexação em bases de dados já foi relatada em estudos anteriores (REPISO; ORDUÑA-MALEA; AGUADED, 2019), assim como a preponderância do acesso aberto entre as publicações da região (MINNITI; SANTORO; BELLI, 2018).

Periódicos de ciências sociais e de artes e humanidades foram a maioria dos analisados, seguidos de uma parcela relevante de periódicos de ciências da saúde (Gráfico 10). Tanto os resultados da análise de escopo, quanto de aderência temática por meio de análise de descritores de assuntos (Figura 14) confirmaram essa característica da região, ao encontro do que já foi relatado em estudos pretéritos (ALPERIN; ROZEMBLUM, 2017), especialmente na América Latina.

Como consequência, um menor percentual de periódicos de ciências exatas, naturais e engenharia foram identificados, o que remete para as deficiências da ciência básica e tecnológica da região. As variações desse padrão em alguns países sugerem políticas diferenciadas, a exemplo de Cuba, que dá maior ênfase à edição de periódicos das ciências exatas e naturais (Gráfico 4).

Os países de menor produtividade, por outro lado, tendem a editar periódicos multidisciplinares, o que pode indicar a manutenção de publicações de escopo mais amplo visando a divulgação de resultados de pesquisas de diversas áreas.

A maioria dos periódicos analisados são de periodicidade menos frequente, considerando que quase 70% edita no máximo três fascículos por ano (semestrais e quadrimestrais). O percentual significativo de periódicos não indexados na JCR e na Scopus explica esse resultado, pois essas fontes indexam mais periódicos mensais e bimestrais (REPISO; ORDUÑA-MALEA; AGUADED, 2019). Já fontes com menos critérios de indexação possuem maior proporção periódicos de periodicidade

semestrais (GONZALEZ-PARDO; REPISO; ARROYAVE-CABRERA, 2020).

Embora os dados de periodicidade não reflitam diretamente o número de artigos por número, uma frequência mínima semestral parece ser suficiente para atender aos critérios de indexação GSM de 100 artigos em um período de cinco anos. Isso implica uma produtividade anual de 20 artigos (10 por fascículo), parâmetro que pode ser almejado por periódicos que buscam a indexação no GSM.

Os idiomas espanhol e português ainda são a maioria dos periódicos da região (Gráfico 6). O inglês, contudo, é utilizado em quase 60% do total de periódicos, o que sugere um estágio intermediário de internacionalização idiomática da região (CHINCHILLA-RODRÍGUEZ et al., 2015).

6.4 VALIDADE E ESTABILIDADE DO ÍNDICE H5

Os resultados da matriz de correlação entre os indicadores (Figura 15) confirmam a validade do índice h5 para avaliação de impacto dos periódicos Ibero-Americanos. Os coeficientes obtidos pelo índice h5 foram similares aos registrados pelo índice h calculado pela Scopus.

O maior nível de correlação do índice h5 foi obtido no teste de Spearman na comparação com o FI5y ($\rho = 0,75$). É provável que o período de cinco anos de ambos tenha contribuído para esse resultado. Esse coeficiente foi o mesmo registrado entre FI5y e o AIS, dois indicadores do JCR. Foi ainda superior ao registrado entre FI5y e outros indicadores tradicionais (SJR, SNIP), calculados a partir de dados da Scopus.

Embora o índice h5 utilize dados de citação distintos do JCR e da Scopus, os resultados produzidos não apresentam discrepâncias que apontem para a sua invalidade. Pelo contrário, reforçam a sua validade, especialmente a sua adequação para classificação ou elaboração de rankings de periódicos, considerando o melhor desempenho no teste de Spearman.

A análise longitudinal conduzida a partir dos dados extraídos pela ferramenta *Gsm_htline* advoga pela estabilidade do índice h5 nas atualizações anuais do GSM (Gráficos 12 e 13). A taxa de crescimento médio de 11%, inclusive foi a mesma obtida em outra análise com uma amostra de mais de 300 de periódicos brasileiros indexadas no SciELO (PACKER et al., 2019).

Essa igualdade entre os resultados sugere que o índice h5 é um indicador estável em análises em larga escala, ou seja, que não apresenta variações bruscas

em cada medição anual. Também indica a validade da ferramenta *Gsm_htline* para extração e análise de dados de edições anteriores do GSM.

Destaca-se que esse uso do recurso possibilitou a análise histórica do índice h5 mais ampla da literatura consultada. Foram extraídos dados completos dos últimos cinco anos de 1430 periódicos, mais de um terço do universo da pesquisa.

6.5 ANÁLISE DE IMPACTO COM BASE NO ÍNDICE H5

A análise de impacto relevou a influência das mencionadas características dos periódicos na frequência de citações recebidas e, conseqüentemente, no desempenho do índice h5.

Foi confirmada a relação entre o idioma inglês e um nível de impacto mais elevado. Os periódicos somente em inglês têm índice h5 médio significativamente superior ao dos de múltiplos idiomas. Também há variação de impacto entre ambos e os somente em idiomas regionais (Gráfico11).

Foram identificadas diferenças de impacto conforme a aderência temática. Os periódicos de ciências da saúde, de ciências agrárias e de engenharias apresentaram índice h5 médio superior que os de artes, humanidades e ciências sociais os menores.

Essa diferença de impacto entre áreas é notória na literatura (HENEBERG, 2016; WALTMAN, 2016; LÓPEZ-CÓZAR et al., 2019;). Mas se esses resultados não surpreendem, apresentam aspecto inovador em razão do uso do índice h5 como parâmetro de impacto e em larga escala, já que a análise de impacto levou em conta todos os periódicos indexados no GSM.

Usualmente, estudos de impacto da região são baseados somente em periódicos indexados na WoS/JCR e Scopus, o que reduz a representatividade e mantém o viés dessas fontes (COLLAZO-REYES, 2014; VÉLEZ-CUARTAS; LUCIO-ARIAS; LEYDESDORFF, 2016; IBÁÑEZ-MARTI, 2018).

O fato de o índice h ser baseado apenas em artigos mais citados, portanto, não restringe a sua aplicação em nível micro (autor, periódico). Pode ser utilizado também em níveis *meso* e *macro* (país, região), tal qual foi utilizado nesta pesquisa.

Os resultados apontam que os critérios de avaliação com base no índice h5 devem considerar os diferentes níveis de impacto de cada área do conhecimento, da forma como é feito sistema Qualis no Brasil (Gráfico 10).

Os periódicos da Colômbia, do México e de Cuba têm o índice h5 médio mais alto da região. Isso pode ser atribuído a uma distribuição mais uniforme do impacto e, no caso dos dois primeiros países, ao maior percentual de periódicos que incluem o inglês como idioma, mesmo que de forma conjunta com outros idiomas (múltiplos).

Por outro ângulo, Brasil e Espanha possuem os periódicos de índice h5 mais elevado, principalmente os superiores a 30, que são raros na região. É provável que o maior número de periódicos desses dois países pode ter favorecido a menor média, considerando o volume de publicações de baixo impacto ($h5 < 5$), especialmente no Brasil.

Entre os países mais produtivos, a Argentina tem o menor desempenho geral no índice h5. As razões podem ser a forte aderência temática às ciências sociais e às humanidades e a baixa proporção periódicos em inglês.

6.6 INFLUÊNCIA DE AUTOCITAÇÕES NO ÍNDICE H5

A identificação e a análise de autocitações no âmbito do *framework Gsm_hdata* implicou na resolução de questões pertinentes, uma de natureza técnica e outra de natureza teórica.

A questão de natureza técnica solucionada pelo *framework* diz respeito não tratamento de autocitações pelo GSM (MARTIN-MARTIN et al. 2014). O *Gsm_hscite*, neste contexto, representa um recurso bibliométrico inovador, pois possibilita a identificação de autocitação no índice h5 com considerável precisão. Também calcula quatro métricas de autocitação, um feito sem equivalentes, nem mesmo no PoP, a ferramenta de análise mais utilizada em estudos baseados em dados do GS (HARZING; VAN DE WAL, 2008; HARZING, 2013; JACSÓ, 2009; AMARA, LANDRY, 2012; MINGERS; MEYER, 2017; MARTÍN-MARTÍN et al., 2018).

A questão de natureza teórica diz respeito ao fato de que a maioria das pesquisas sobre autocitações foi conduzida a partir de dados do FI (FRANDSEN, 2007; ROUSSEAU, 2009; CHORUS; WALTMAN, 2016; SANFILIPPO et al., 2021; TAŞKIN et al., 2021) ou teve como objeto a autocitação de autor (WALTMAN, 2016), tornando a autocitação de periódico no índice h um profundo *gap* de pesquisa.

Foi constatado que o índice h5 sofre menor influência de autocitações de periódico em comparação com o FI. A configuração do índice h resulta na desconsideração de parte das citações recebidas pelo periódico, incluindo autocitações, conforme foi alegado por Hirsch (2005). Assim, os resultados da média

e da mediana das SCR são inferiores aos de estudos baseados em autocitações no FI (LIVAS; DELLI, 2018; SANFILIPPO et al., 2021; TAŞKIN et al., 2021; GAZNI; DIDEGAH, 2021).

O percentual médio de autocitações de todos os periódicos analisados ficou abaixo de 4%, sendo que quase 90% possui menos de 10% de autocitações. O resultado do teste de Pearson entre os índices h5 e h_s5 obteve coeficiente de 0,97, ou seja, quase perfeito. Aponta para uma influência mínima das autocitações em uma medida geral de impacto com base no índice h5.

Algumas constatações corroboraram achados de estudos anteriores. As áreas com SCR mais altas são as com menos periódicos na pesquisa, ou seja, as áreas mais especializadas (FRANSEN, 2007; ROUSSEAU, 2009; LIVAS; DELLI, 2018; SANFILIPPO et al., 2021). Essa tendência pode ter origem no fato de que áreas mais especializadas formam redes citação de menor densidade, o que aumenta a probabilidade de identidade entre fontes citante e citada (autocitação).

Os periódicos com SCR mais elevadas são de nível de impacto médio mais baixo (HENENBERG, 2016; TAŞKIN et al., 2021). Isso sugere que as autocitações podem mesmo ser uma estratégia de autopromoção, adotadas por publicações mais recentes e ainda sem reconhecimento, conforme levantado por González-Sala, Osca-Lluch e Haba-Osca (2019).

Uma hipótese levantada é que após um período inicial com mais autocitações, o alcance de um nível intermediário de impacto pode aumentar a proporção de citações externas e, por consequência, diminuir a SCR.

De forma geral, os dados sustentam a desnecessidade de prévia exclusão de autocitação de periódico em análises de impacto a partir do índice h5 do GSM, como defende parte da literatura (WALTMAN, 2016).

Sob outra perspectiva, foi demonstrado que a não exclusão antecipada de autocitações não pode levar à desconsideração total desse parâmetro (FRANSEN, 2007). É necessário monitorar SCR, visando a evitar distorções, pois em todas as áreas foram identificados periódicos com percentuais estatisticamente discrepantes.

Foi constatado que a resistência do índice h5 às autocitações diminui com taxas discrepantes. Em nível de *outlier*, as autocitações podem interferir decisivamente nos níveis de impacto, conferindo vantagem na comparação com periódicos com SCR moderadas.

Isso justifica o monitoramento de autocitações no índice h5 e, se necessário, ajustes nos resultados em determinados casos.

6.7 MANIPULAÇÃO DE CITAÇÕES

A identificação de SCR excessivas também foi um ponto de partida para análise de outros padrões de citação a partir de dados do índice h5. Os parâmetros adicionais, definidos como critérios específicos do protocolo, foram amparados na distribuição de citações por ano e na identificação da existência de vínculo entre os periódicos citantes e o periódico citado (HENENBERG, 2016).

Foram identificados periódicos com padrões anormais na distribuição de citações recebidas, concentrando volume desproporcional de citações em um ano específico. A identificação de padrões distintos de distribuição de citações por ano, aliada a distorções em outros indicadores do protocolo de análise de citações foram consideradas indícios de manipulação de citações (HENENBERG, 2016; CHORUS; WALTMAN, 2016).

Em alguns casos, foram identificados periódicos sem real impacto, tendo em vista que a ampla maioria (Tabela 15) ou quase a totalidade das citações recebidas eram autocitações.

Um dos periódicos analisados (SSC15) apresentou todos os elementos de configuração de manipulação de citações abrangidos pelo *framework*. Além de excesso de autocitações, de crescimento desproporcional do índice h5, e de distribuição anormal de citações (Tabela 15), demonstrou ter alta densidade de citações oriundas de dois periódicos mantidos pela mesma instituição (Figura 16).

Seria forçoso classificar essa publicação como de caráter acadêmico ou científico, pois a lista de referências não é um elemento acessório de um trabalho científico. É um critério relevante de cientificidade, que fornece o elemento principal de análise de impacto: a citação.

7 SÍNTESE DO FRAMEWORK GSM@HDATA

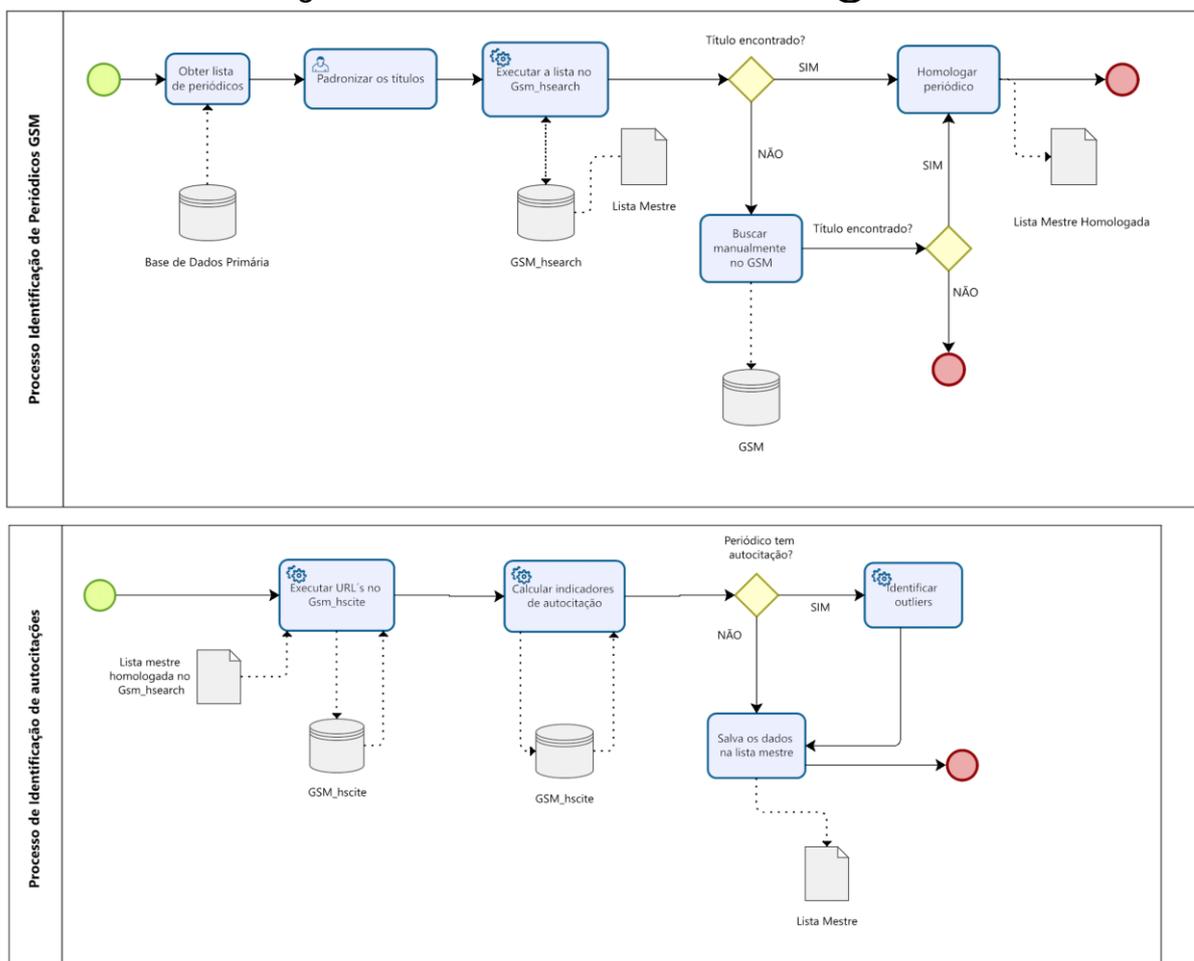
A finalização de todos os procedimentos metodológicos descritos a partir do conjunto de periódicos Ibero-Americanos representou a validação definitiva do *framework Gsm@hdata*.

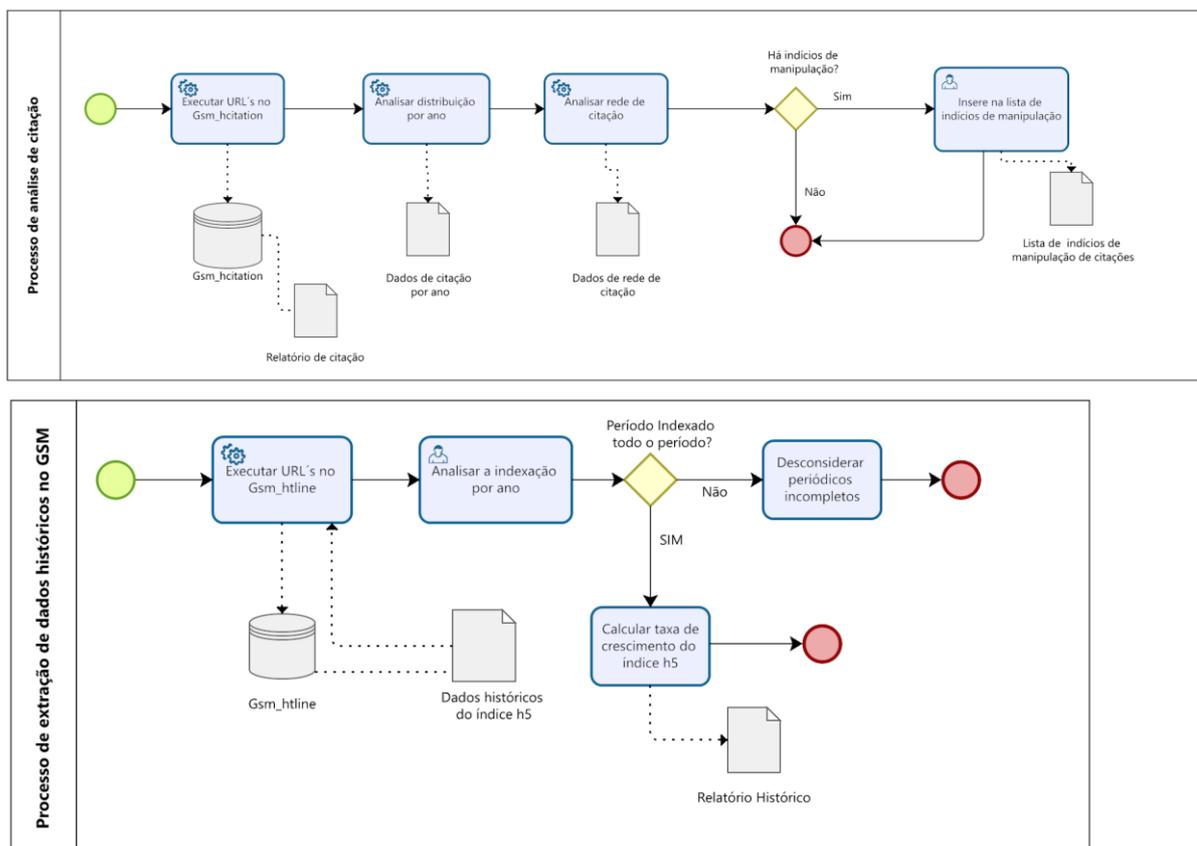
Das três dimensões do *framework*, a tecnológica foi evidenciada pelo funcionamento e pela precisão da ferramenta bibliométrica *Gsm_hdata*.

Os processos foram realizados essencialmente por meio dos quatro recursos, o *Gsm_hsearch*, o *Gsm_hscite*, o *Gsm_hcitation* e o *Gsm_hline*.

Os quatro processos foram modelados visando a apresentação de uma síntese do *framework Gsm@hdata*, conforme mostra a Figura 17.

Figura 17 – Síntese do *framework Gsm@hdata*





Fonte: Elaborado pelo autor.

8 CONSIDERAÇÕES FINAIS

A realização deste estudo possibilitou uma análise detalhada do funcionamento do Google Scholar Metrics, de suas limitações, mas, principalmente, de seu potencial de utilização.

O *framework* *Gsm_hdata* solucionou parte relevante dos problemas de funcionamento dessa fonte alternativa de avaliação.

As rotinas computacionais realizadas pela ferramenta bibliométrica *Gsm_hdata* possibilitaram a realização de um estudo bibliométrico e centimétrico de larga escala e longitudinal, algo raro em estudos envolvendo o Google Scholar Metrics.

A definição do índice h5 como parâmetro em sistemas nacionais de avaliação é um sinal de que o Google Scholar Metrics tem aumento a sua relevância no cenário da avaliação científica Ibero-Americano.

No Brasil, as vantagens desse sistema parecem ter superado as limitações. O resultado da avaliação do Qualis Periódicos prevista para ser divulgado o final de 2022 deve evidenciar o principal mérito do Google Scholar Metrics: a sua ampla cobertura.

A possibilidade de avaliação de impacto de centenas ou até de milhares de periódicos que não possuem outros indicadores tem um apelo especial na região, que há muito clama por outras formas de medir a sua ciência.

O Qualis deve pôr o índice h5 na posição de maior destaque desde a sua criação em 2012. Em caso de êxito na sua utilização no complexo sistema brasileiro baseado em critérios distintos por áreas, é esperado que isso influencie outros países.

Mesmo no México e na Colômbia, que também elegeram o índice h5 como espécie de “plano b” para avaliar periódicos não indexados no JCR e na Scopus, a escolha do Publish or Perish como fonte de cálculo pode ser revista.

Mas o potencial do índice h5 vai além dos países com maior produção de periódicos. Outros países poderiam considerar o uso do Google Scholar Metrics, acrescentando a análise de impacto à de visibilidade, que usualmente é baseada no Latindex e no SciELO.

Nesse sentido, embora o *framework* *Gsm@hsearch* possa ser utilizado a partir de qualquer lista de periódicos, o Latindex se mostrou a fonte primária de dados adequada para a região. Em razão de ser um diretório construído por meio de

uma cooperação regional, indexa parcela relevante das publicações acadêmicas e científicas Ibero-Americanas, incluído dados editoriais com controle bibliográfico.

Essas características preenchem duas deficiências do Google Scholar Metrics: a ausência de uma lista dos periódicos indexados e de seus dados editoriais, mesmos os mais elementares como o ISSN.

A ponte entre o Latindex e o Google Scholar Metrics foi edificada por meio do *Gsm_hsearch*, o primeiro recurso da ferramenta *Gsm_hdata*. Quase infalível em casos de identidade entre os títulos buscado e indexado, a sua configuração também conseguiu contornar variações de formas de entrada por título com significativa precisão.

O funcionamento do recurso *Gsm_hdata* representou o alcance de um importante objetivo inicial. Permitiu divulgação de dois estudos preliminares baseados em parâmetros de visibilidade e de impacto.

Mas até esse ponto, o protótipo de *framework* não enfrentava a problemática das autocitações. Pois se em indicadores de outras bases de dados essa questão está controlada, no índice h5 implicava numa incógnita, que foi em grande medida esclarecida neste estudo.

A análise de autocitações por meio do recurso *Gsm_hscite* é um dos resultados mais relevantes da pesquisa. Permitiu um achado teórico no campo dos estudos métricos, ao identificar os limites da influência das autocitações de periódico no índice h, inclusive comparativamente com o Fator de Impacto.

O *Gsm_hscite* representou também um avanço tecnológico, ao efetuar o cálculo de novas métricas de autocitação no índice h5, nos moldes do que pode ser feito em avançadas bases de dados, mas que não pode ser feito no Publish or Perish, *software* que é referência em análise de dados do Google Scholar.

Esse recurso atende inclusive ao anseio da coordenação da área 31 – Comunicação e Informação da CAPES, que em relatório da avaliação de meio termo em 2019 expressou a necessidade do desenvolvimento de uma ferramenta com as funcionalidades do *Gsm_hscite*.

As novas métricas de autocitação, primordialmente o índice h_s5 e a taxa de autocitação por periódico, podem ser utilizadas em sistemas de avaliação baseados no índice h5.

Os dois novos indicadores oferecem evidências objetivas para aplicação de correções no impacto de periódicos que apresentarem distorções pelo excesso de autocitações.

As correções podem envolver o rebaixamento de posição em rankings, estratos ou quartis, o ajuste do valor do índice h5, ou ainda a classificação de periódicos como não científicos, ante a preponderância de autocitações em relação às citações externas.

Os avanços na seara bibliométrica não se restringiram à análise de autocitações. A perspectiva de um estudo longitudinal baseado no índice h5, possível por meio do recurso *Gsm_hpline* também representou os avanços da pesquisa.

Pois passados dez anos do lançamento do Google Scholar Metrics, muitas questões relacionadas ao seu tamanho e ao seu crescimento estavam sem respostas, ou com respostas baseadas em estimativas.

Os resultados produzidos por meio do *framework Gsm@hdata* apresentaram medidas reais de tamanho e de ritmo de crescimento da coleção de periódicos Ibero-Americanos no *Google Scholar Metrics*.

Entretanto, devido aos objetivos e à metodologia empregada, não foram investigadas questões relacionadas aos erros do sistema.

É uma lacuna da pesquisa que é deixada como sugestão de estudos futuros. Tanto o percentual de duplicações de títulos, quanto a influência dos variados tipos de erros do GSM no cálculo do índice h5 são perguntas ainda sem respostas.

O recurso *Gsm_hcitation* pode auxiliar investigações dessa natureza, já que extraí dados para posterior tratamento e análise.

Os resultados alcançados sugerem, entretanto, que os erros não afetam a confiabilidade do índice h5, já que a sua validade e estabilidade foi atestada, pelo menos no contexto dos periódicos Ibero-Americanos.

Uma hipótese a ser testada em estudos futuros seria a de que, dado o padrão de cálculo amostral, o índice h5 não absorveria eventuais erros em contagens de citações, tal qual ocorrem com as autocitações.

Outro caminho que poderia ser trilhado é o da aplicação do *Gsm@hdata* em outras regiões de menor visibilidade científica. O sentido mais natural de continuidade desta pesquisa ou de sugestão de estudos futuros é o da aplicação do

framework em periódicos do continente africano, que já possuem uma infraestrutura de indexação no *African Journals Online*¹⁸.

A utilização futura do *framework* não precisa ser completa. A dimensão tecnológica, por exemplo, pode ser utilizada para atualização do *Journal Scholar Metrics* com baixo custo computacional. Embora estagnada nos dados de 2015, essa relevante contribuição do *EC3 Research Group*¹⁹ da Universidade de Granada ainda é a fonte secundária de dados mais abrangente do *Google Scholar Metrics*.

Poderia ainda ser construído um recurso similar ao *Journal Scholar Metrics*, mas com escopo geográfico Ibero-Americano. Seria um meio de hospedar devidamente os dados desta pesquisa e de facilitar a atualização anual, que inclusive acaba de ocorrer no *Google Scholar Metrics* (28 de junho de 2022).

Uma extensão idealizada para ferramenta *Gsm_hdata*, mas que ficou de fora desta tese foi o *Gsm_hdisplay*. Seria uma espécie de *plug in* para ser exibido na página de periódicos de acesso aberto hospedados no *Open Journal Systems*, mostrando os índices h5 e h_s5, as medianas h5 e h_s5 e a taxa de autocitação, com botão de seleção por ano de análise.

Estudos futuros também têm espaço no campo do desenvolvimento de *software*. A construção de uma interface para a ferramenta bibliométrica *Gsm_hdata*, com possibilidade de execução em outros sistemas operacionais poderia ser considerada.

Por fim, a aplicação lógica do *framework* seria em sistemas de classificação de periódicos de países de menor visibilidade, tanto os que não avaliam o parâmetro impacto, quanto o que já o fazem, mas não com a mesma profundidade, ignorando autocitações, por exemplo.

¹⁸ <https://www.ajol.info/index.php/ajol>

¹⁹ <https://ec3-research.com/>

9 CONCLUSÃO

A avaliação de impacto de periódicos Ibero-Americanos com base no índice h5 pode ser realizada em larga escala e com baixo custo computacional se utilizado o *Gsm@hdata*. O framework também garante mais confiabilidade e menor suscetibilidade dos resultados às distorções eventualmente ocasionadas por excesso de autocitações e por práticas de manipulação de citações.

Mas o *framework* conduz uma avaliação baseada em apenas um, dentre diversos parâmetros de representação de impacto científico de periódicos disponíveis na atualidade.

Nenhum parâmetro, individualmente, consegue representar com fidelidade todas as dimensões de qualidade e de impacto da ciência. Somente um processo combinado de avaliação, iniciado com a revisão por pares, complementado pelo uso de múltiplos indicadores que atentem para as diferenças entre disciplinas e áreas do conhecimento, pode refletir o amplo espectro do impacto científico de periódicos.

REFERÊNCIAS

- AGUADO-LÓPEZ, E. et al. Ibero-America in mainstream science (Thomson Reuters/Scopus): A fragmented region. **Interciencia**, v. 39, n. 8, p. 570–579, 2014. Disponível em: <https://www.redalyc.org/articulo.oa?id=33931820006>. Acesso em: 10 jun. 2022.
- ALPERIN, J. P.; ROZEMBLUM, C. La reinterpretación de visibilidad y calidad en las nuevas políticas de evaluación de revistas científicas. **Revista Interamericana de Bibliotecología**, v. 40, n. 3, p. 231–241, 1 set. 2017. Disponível em: <https://revistas.udea.edu.co/index.php/RIB/article/view/327794>. Acesso em: 10 jun. 2022.
- AMARA, N.; LANDRY, R. Counting citations in the field of business and management: Why use Google Scholar rather than the Web of Science. **Scientometrics**, v. 93, n. 3, p. 553–581, 2012. Disponível em: <http://link.springer.com/10.1007/s11192-012-0729-2>. Acesso em: 10 set. 2020.
- AYLLÓN, J. M.; MARTÍN-MARTÍN, A.; ORDUÑA-MALEA, E. Índice H de las Revistas Científicas Españolas según Google Scholar Metrics (2011-2015). **Economía y Empresa**. n. October, p. 1–39, 2016. Disponível em: <https://plataformarevistascomunicacion.org/wp-content/uploads/2016/11/%C3%8Dndice-H-de-las-Revistas-Cient%C3%ADficas-Espa%C3%B1olas-seg%C3%BAAn-Google-Scholar-Metrics.pdf>. Acesso em: 10 set. 2020.
- BARBETTA, P. A. **Estatística aplicada às ciências sociais**. 8. ed. rev. Florianópolis: Ed. da UFSC, 2012. 315 p.
- BARBOSA, G. R. **Avaliação da pesquisa em humanidades**. Relatório Técnico: CAPES, 1/2020. Disponível em: <https://www.gov.br/capes/pt-br/centrais-de-conteudo/23072020-dav-aph-pdf>. Acesso em: 10 maio 2022.
- BARNES, C. The h -index Debate: An Introduction for Librarians. **The Journal of Academic Librarianship**, v. 43, n. 6, p. 487–494, 1 nov. 2017. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S0099133316301732>. Acesso em: 10 set. 2020.
- BARTNECK, C.; KOKKELMANS, S. Detecting h-index manipulation through self-citation analysis. **Scientometrics**, v. 87, n. 1, p. 85–98, 11 abr. 2011. Disponível em: <http://link.springer.com/10.1007/s11192-010-0306-5>. Acesso em: 10 set. 2020.
- BERGSTROM, C. T.; WEST, J. D.; WISEMAN, M. A. The Eigenfactor™ metrics. **Journal of Neuroscience**, v. 28, n. 45, p. 11433–11434, 2008. Disponível em: <https://europepmc.org/article/pmc/6671297> . Acesso em: 10 set 2020.
- BICAS, H. E. A. Ineditismo, originalidade, importância, publicidade, interesse e impacto de artigos científicos. **Arquivos Brasileiros de Oftalmologia**, v. 71, n. 4, p. 473–474, ago. 2008. Disponível em:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-27492008000400001&lng=pt&nrm=iso&tlng=pt. Acesso em: 10 jun. 2022.

BORNMANN, L. *et al.* A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. **Journal of Informetrics**, v. 5, n. 3, p. 346–359, jul. 2011. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1751157711000071>. Acesso em: 10 set. 2020.

BORNMANN, L. *et al.* The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute. **Journal of the Association for Information Science and Technology**, v. 67, n. 11, p. 2778–2789, 1 nov. 2016. Disponível em: <http://doi.wiley.com/10.1002/asi.23627>. Acesso em: 10 set. 2020.

BORNMANN, L. *et al.* Diversity, value and limitations of the journal impact factor and alternative metrics. **Rheumatology International**, v. 32, n. 7, p. 1861–1867, 23 jul. 2012. Disponível em: <https://link.springer.com/article/10.1007/s00296-011-2276-1>. Acesso em: 10 set. 2020.

BORNMANN, L. h-Index research in scientometrics: A summary. **Journal of Informetrics**, v. 8, n. 3, p. 749–750, jul. 2014. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1751157714000649>. Acesso em: 10 set. 2020.

BORNMANN, L.; DANIEL, H.-D. The state of h index research. Is the h index the ideal way to measure research performance? **EMBO Reports**, v. 10, n. 1, p. 2–6, 2009. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1751157711000071>. Acesso em: 10 set. 2020.

BRAUN, T.; GLÄNZEL, W.; SCHUBERT, A. A Hirsch-type index for journals. **Scientometrics**, v. 69, n. 1, p. 169–173, 20 out. 2006. Disponível em: <http://link.springer.com/10.1007/s11192-006-0147-4>. Acesso em: 10 set. 2020.

BRASIL. **Portaria n.145**, de 10 de setembro de 2021. Consolida as disposições sobre o Qualis Periódicos, seus objetivos e finalidade. Disponível em: <https://in.gov.br/en/web/dou/-/portaria-n-145-de-10-de-setembro-de-2021-344468240>. Acesso em: 24 set. 2021.

CASTELLÓ-COGOLLOS, L.; SIXTO-COSTOYA, A.; LUCAS-DOMÍNGUEZ, R.; AGULLÓ-CALATAYUD, V. *et al.* Bibliometrics and indicators of scientific activity (XI): Other useful resources in the evaluation: Google scholar, microsoft academic, Ifindr, dimensions and Lens.org. **Acta Pediatrica Espanola**, 76, n. 9-10, p. 123-130, 2018. Disponível em: <https://digital.csic.es/handle/10261/249457> Acesso em: 10 jun. 2022.

CAMACHO, O. *et al.* An approach to measure the impact of scientific regional journals: The latindex case. **RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao**, v. 2019, n. E20, p. 13–23, 2019. Disponível em: https://www.researchgate.net/publication/334668365_An_Approach_to_measure_the

[Impact of Scientific Regional Journals The Latindex case](#). Acesso em: 10 jun. 2022.

CAMPANARIO, J. M. Journals that rise from the fourth quartile to the first quartile in six years or less: Mechanisms of change and the role of journal self-citations. **Publications**, v. 6, n. 4, 2018. Disponível em: <http://www.mdpi.com/2304-6775/6/4/47>. Acesso em: 10 set. 2020.

CANTO, F. L. et al. Latin American and Caribbean journals indexed in Google Scholar Metrics. **Scientometrics**, v. 127, n. 2, p. 763–783, 20 fev. 2022. Disponível em: <https://link.springer.com/article/10.1007/s11192-021-04237-x>. Acesso: 10 set. 2020.

CATALINI, C.; LACETERA, N.; OETTL, A. The incidence and role of negative citations in science. **Proceedings of the National Academy of Sciences**, v. 112, n. 45, p. 13823–13826, 10 nov. 2015. Disponível em: <https://pnas.org/doi/full/10.1073/pnas.1502280112>. Acesso em: 10 jun. 2022.

CHEIN, F. **Introdução aos modelos de regressão linear**: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas. ENAP: Brasília, 2019. Disponível em: https://repositorio.enap.gov.br/bitstream/1/4788/1/Livro_Regress%C3%A3o%20Linear.pdf. Acesso em: 10 jun. 2022.

CHINCHILLA-RODRÍGUEZ, Z. et al. Benchmarking scientific performance by decomposing leadership of Cuban and Latin American institutions in Public Health. **Scientometrics**, v. 106, n. 3, p. 1239–1264, 12 mar. 2016. Disponível em: <http://link.springer.com/10.1007/s11192-015-1831-z>. Acesso em: 10 jun. 2022.

CHORUS, C.; WALTMAN, L. A large-scale analysis of impact factor biased journal self-citations. **PLoS ONE**, v. 11, n. 8, p. 1–11, 2016. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0161021>. Acesso em: 10 jun. 2022.

CIENTÍ, D. D. et al. Revista española de Documentación Científica, Vol 36, No 1 (2013). v. 36, n. 1, p. 1–12, 2020. Disponível em: <https://redc.revistas.csic.es/index.php/redc/issue/view/81>. Acesso em: 10 junho 2021.

CLARIVATE ANALYTICS. **Web of Science**. Disponível em: <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>. Acesso em: 10 maio 2022.

COLLAZO-REYES, F. et al. Publication and citation patterns of Latin American & Caribbean journals in the SCI and SSCI from 1995 to 2004. **Scientometrics**, v. 75, n. 1, p. 145–161, 18 abr. 2008. Disponível em: <https://link.springer.com/article/10.1007/s11192-007-1841-6>. Acesso em: 10 maio 2022.

COLLAZO-REYES, F. Growth of the number of indexed journals of Latin America

and the Caribbean: the effect on the impact of each country. **Scientometrics**, v. 98, n. 1, p. 197–209, 15 jan. 2014. Disponível em: <https://link.springer.com/article/10.1007/s11192-013-1036-2>. Acesso em: 10 maio 2022.

COLLAZO-REYES, F.; LUNA-MORALES, M. E.; LUNA-MORALES, E. Change in the publishing regime in Latin America: from a local to universal journal, *Archivos de investigación Médica*. **Scientometrics**, v. 110, n. 2, p. 695–709, 28 fev. 2017. Disponível em: <https://link.springer.com/article/10.1007/s11192-016-2207-8>. Acesso em: 10 maio 2022.

COLÔMBIA. Publindex. Sistema de Indexación y Homologación de Revista Especializadas de CTI. Disponível em: <https://minciencias.gov.co/content/publindex>. Acesso em 10 maio. 2021.

COSTA, H.; CANTO, F. L. DO; PINTO, A. L. Google Scholar Metrics e a proposta do novo Qualis: impacto dos periódicos brasileiros de Ciência da Informação. **Informação & Sociedade: Estudos**, v. 30, n. 1, 31 set. 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/50676>. Acesso em: 10 maio 2022.

COTHRAN, T. Google Scholar acceptance and use among graduate students: A quantitative study. **Library and Information Science Research**, v. 33, n. 4, p. 293–301, 2011. Disponível em: <http://dx.doi.org/10.1016/j.lisr.2011.02.001>. Acesso em: 28 set. 2020.

CRESWELL, John W. **Projeto de pesquisa**: métodos qualitativo, quantitativo e misto. 2. ed. Porto Alegre: Artmed, 2007. 248 p.

CRESPO-GASCÓN, S.; TORTOSA, F. S.; GUERRERO-CASADO, J. Producción de revistas científicas en América Latina y El Caribe en Scopus, Journal Citation Reports y Latindex en el área de los recursos naturales: su relación con variables económicas, ambientales y de inversión en investigación. **Revista española de Documentación Científica**, v. 42, n. 1, p. 224, 5 mar. 2019. Disponível em: <https://redc.revistas.csic.es/index.php/redc/article/view/1030>. Acesso em: 28 set. 2020.

CRONIN, B. **The citation process**: the role and significance of citation in scientific communication. London: Taylor Graham, 1984.

CUETO, M. A história das ciências e o Qualis Periódicos. **Hist. cienc. saude-Manguinhos**, Rio de Janeiro, v. 26, n. 4, p. 1083-1084, dez. 2019. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-59702019000401083&lng=pt&nrm=isso. Acesso em: 28 jan. 2020.

CRONIN, B. **The citation process**: the role and significance of citation in scientific communication. London: Taylor Graham, 1984.

DELGADO, E.; REPISO, R. The impact of scientific journals of communication: Comparing google scholar metrics, web of science and scopus. **Comunicar**, 21, n.

41, p. 45-52, 2013. Disponível em: <https://www.revistacomunicar.com/indice-en/articulo.php?numero=41-2013-04>. Acesso em: 10 jun. 2022.

DE FILIPPO, D.; ALEIXANDRE-BENAVENT, R.; SANZ-CASADO, E. Toward a classification of Spanish scholarly journals in social sciences and humanities considering their impact and visibility. **Scientometrics**, v. 125, n. 2, p. 1709–1732, 17 nov. 2020. Disponível em: <https://link.springer.com/article/10.1007/s11192-020-03665-5>. Acesso em: 28 jan. 2020.

DE PABLOS COELLO, J. M.; MARTÍN, C. M.; LÓPEZ, M. T. Google, the change of paradigm of the scientific metrics. **Ilu**, 18, p. 225-235, 2013. Disponível em: <https://go.gale.com/ps/i.do?id=GALE%7CA497671475&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=11370734&p=IFME&sw=w&userGroupName=ufsc.br>. Acesso em: 10 jun. 2022.

DE WINTER, J. C. F.; ZADPOOR, A. A.; DODOU, D. The expansion of Google Scholar versus Web of Science: A longitudinal study. **Scientometrics**, v. 98, n. 2, p. 1547–1565, 2014. Disponível em: <https://link.springer.com/article/10.1007/s11192-013-1089-2>. Acesso em: 28 jan. 2020.

DING, J.; LIU, C.; KANDONGA, G. A. Exploring the limitations of the h-index and h-type indexes in measuring the research performance of authors. **Scientometrics**, v. 122, n. 3, p. 1303–1322, 25 set. 2020. Disponível em: <http://link.springer.com/10.1007/s11192-020-03364-1>. Acesso em: 10 set. 2020.

ELKINS, M. R. et al. Correlation between the Journal Impact Factor and three other journal citation indices. **Scientometrics**, v. 85, n. 1, p. 81–93, 2010. Disponível em: <https://link.springer.com/article/10.1007/s11192-010-0262-0>. Acesso em: 10 set. 2020.

ELSEVIER. **Content Coverage Guide**. p. 1–24, 2010. Disponível em: https://www.elsevier.com/_data/assets/pdf_file/0017/114533/Scopus_GlobalResearch_Factsheet2019_FINAL_WEB.pdf. Acesso em: 10 set. 2020.

FALAGAS, M. E. et al. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. **The FASEB Journal**, v. 22, n. 2, p. 338–342, 20 fev. 2008. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1096/fj.07-9492LSF>. Acesso em: 28 jan. 2020.

FERRER-SAPENA, A. et al. The Impact Factor as a measuring tool of the prestige of the journals in research assessment in mathematics. **Research Evaluation**, v. 25, n. 3, p. 306–314, 2016. Disponível em: <https://ideas.repec.org/a/oup/rseval/v25y2016i3p306-314.html>. Acesso em: 28 jan. 2020.

FLADUNG, Rainer. **Scientific communication: economic analysis of the electronic journal market**. Stuttgart: Ibidem-Verlag, 2007.

FLATT, J.; BLASIMME, A.; VAYENA, E. Improving the Measurement of Scientific

Success by Reporting a Self-Citation Index. **Publications**, v. 5, n. 3, p. 20, 1 ago. 2017. Disponível em: <https://www.mdpi.com/2304-6775/5/3/20>. Acesso em: 20 jan. 2022.

FONG, E. A.; WILHITE, A. W. Authorship and citation manipulation in academic research. **PLOS ONE**, v. 12, n. 12, p. e0187394, 6 dez. 2017. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0187394>. Acesso em: 20 jan. 2022.

FRANSEN, T. F. Journal self-citations-Analysing the JIF mechanism. **Journal of Informetrics**, v. 1, n. 1, p. 47–58, 2007. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1751157706000071>. Acesso em: 20 jan. 2022.

GARFIELD, Eugene. 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, nov. 1972. Disponível em: <http://science.sciencemag.org/content/178/4060/471>. Acesso em: 20 jan. 2022.

GARFIELD, E. The **Mechanization of Indexing**. Sep. 29, 1952. Disponível: <http://www.garfield.library.upenn.edu/papers/mechanizationofindexing09291952.html>. Acesso em: 20 jun. 2022.

GAZNI, A.; DIDEGAH, F. Journal self-citation trends in 1975–2017 and the effect on journal impact and article citations. **Learned Publishing**, v. 34, n. 2, p. 233–240, 2021. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1348>. Acesso em: 03 jun. 2022.

GUÉDON, J.-C. Acesso Aberto e divisão entre ciência predominante e ciência periférica. In: FERREIRA, S. M.; TARGINO, M. G. (Orgs.). **Acessibilidade e visibilidade de revistas científicas eletrônicas**. São Paulo: Editora São Paulo, 2010.

GIMÉNEZ-TOLEDO, E. La evaluación de las Humanidades y de las Ciencias Sociales en revisión. Disponível em: **Revista española de Documentación Científica**, v. 41, n. 3, p. 208, 2018. Disponível em: <https://redc.revistas.csic.es/index.php/redc/article/view/1012>. Acesso em: 10 maio 2022.

GONZALEZ-PARDO, R.; REPISO, R.; ARROYAVE-CABRERA, J. Revistas iberoamericanas de comunicación a través de las bases de datos Latindex, Dialnet, DOAJ, Scopus, AHCI, SSCI, REDIB, MIAR, ESCI y Google Scholar Metrics. **Revista Española de Documentación Científica**, v. 43, n. 4, p. 276, 2020. Disponível em: <https://redc.revistas.csic.es/index.php/redc/article/view/1302>. Acesso em: 10 maio 2022.

GONZÁLEZ-PEREIRA, B.; GUERRERO-BOTE, V. P.; MOYA-ANEGÓN, F. A new approach to the metric of journals' scientific prestige: The SJR indicator. **Journal of Informetrics**, v. 4, n. 3, p. 379–391, jul. 2010. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1751157710000246>. Acesso em: 28 jan.

2020.

GONZÁLEZ-SALA, F.; OSCA-LLUCH, J.; HABA-OSCA, J. Are journal and author self-citations a visibility strategy? **Scientometrics**, v. 119, n. 3, p. 1345–1364, 2019. Disponível em: https://ideas.repec.org/a/spr/scient/v119y2019i3d10.1007_s11192-019-03101-3.html. Acesso em: 10 maio 2022.

GREGORIO CHAVIANO, O. Evaluación y clasificación de revistas científicas: reflexiones en torno a retos y perspectivas para Latinoamérica, **Revista Lasallista de Investigación**, 2018. Disponível em: http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S1794-44492018000100166&lng=es&nrm=iso. Acesso em: 10 maio 2022.

GU, X.; BLACKMORE, K. Characterisation of academic journals in the digital age. **Scientometrics**, v. 110, n. 3, p. 1333–1350, 3 set. 2017. Disponível em: <http://link.springer.com/10.1007/s11192-016-2219-4>. Acesso em: 10 set. 2020.

GUSENBAUER, M. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. **Scientometrics**, v. 118, n. 1, p. 177–214, 10 jan. 2019. Disponível em: <http://link.springer.com/10.1007/s11192-018-2958-5>. Acesso em: 10 set. 2020.

HARTLEY, J. On the need to distinguish between author and journal self-citations. **Scientometrics**, v. 81, n. 3, p. 787–788, 2009. Acesso em: 10 maio 2022.

HARZING, A. W. A preliminary test of Google Scholar as a source for citation data: A longitudinal study of Nobel prize winners. **Scientometrics**, v. 94, n. 3, p. 1057–1075, 2013. Disponível em: https://ideas.repec.org/a/spr/scient/v94y2013i3d10.1007_s11192-012-0777-7.html. Acesso em: 10 maio 2022.

HARZING, A.-W. A longitudinal study of Google Scholar coverage between 2012 and 2013. **Scientometrics**, v. 98, n. 1, p. 565–575, 5 jan. 2014. Disponível em: <http://link.springer.com/10.1007/s11192-013-0975-y>. Acesso em: 10 set. 2020.

HARZING, A.; VAN DER WAL, R. Google Scholar as a new source for citation analysis. **Ethics in Science and Environmental Politics**, v. 8, n. 1, p. 61–73, 3 jun. 2008. Disponível em: <http://www.int-res.com/abstracts/esep/v8/n1/p61-73/>. Acesso em: 10 set. 2020.

HENEBERG, P. From excessive journal self-cites to citation stacking: Analysis of journal self-citation kinetics in search for journals, which boost their scientometric indicators. **PLoS ONE**, v. 11, n. 4, p. 1–20, 2016. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0153730>. Acesso em: 10 maio 2022.

HERNANDEZ SAMPIERI, Roberto; FERNÁNDEZ COLLADO, Carlos; BAPTISTA LUCIO, Pilar. **Metodologia de pesquisa**. 5. ed. Porto Alegre: Penso, 2013.

HIRSCH, J. E. An index to quantify an individual's scientific research output. **Proceedings of the National Academy of Sciences**, v. 102, n. 46, p. 16569–

16572, 15 nov. 2005. Disponível em: www.pnas.org/cgi/doi/10.1073/pnas.0507655102. Acesso em: 10 set. 2020.

HOU, J. Exploration into the evolution and historical roots of citation analysis by referenced publication year spectroscopy. *Scientometrics*, Dordrecht, v. 110, n. 3, p. 1437-1452, 2017. Disponível em: <https://link.springer.com/article/10.1007/s11192-016-2206-9>. Acesso em: 02 dez. 2017.

HUANG, M.; CATHY LIN, W. The influence of journal self-citations on journal impact factor and immediacy index. **Online Information Review**, v. 36, n. 5, p. 639–654, 21 set. 2012. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/14684521211275957/full/html>. Acesso em: 10 set. 2020.

IBAÑEZ-MARTI, J. J. La Ciencia en Latinoamérica: Tendencias y patrones. **Revista de la Facultad de Ciencias**, v. 7, n. 1, p. 23–39, 1 jan. 2018. Disponível em: <https://revistas.unal.edu.co/index.php/rfc/article/view/69409>. Acesso em: 10 maio 2022.

IOANNIDIS, J. A generalized view of self-citation: Direct, co-author, collaborative, and coercive induced self-citation. **Journal of Psychosomatic Research**, v. 78, n. 1, p. 7-11, 2015. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0022399914003882>. Acesso em: 02 dez. 2017.

JACSÓ, P. Calculating the h-index and other bibliometric and scientometric indicators from Google Scholar with the Publish or Perish software. **Online Information Review**, v. 33, n. 6, p. 1189–1200, 27 nov. 2009. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/14684520911011070/full/html>. Acesso em: 10 set. 2020.

JACSÓ, P. Google Scholar duped and deduped – the aura of “robometrics”. **Online Information Review**, v. 35, n. 1, p. 154–160, 22 fev. 2011. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/14684521211241503/full/html>. Acesso em: 28 jan. 2020.

JACSÓ, P. Google Scholar Metrics for Publications. **Online Information Review**, v. 36, n. 4, p. 604–619, 3 ago. 2012a. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/14684521211254121/full/html>. Acesso em: 10 set. 2020.

JACSÓ, P. Metadata mega mess in Google Scholar. **Online Information Review**, v. 34, n. 1, p. 175–191, 23 fev. 2010. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/14684521011024191/full/html>. Acesso em: 10 set. 2020.

JACSÓ, P. Using Google Scholar for journal impact factors and the h-index in nationwide publishing assessments in academia – siren songs and air-raid sirens. **Online Information Review**, v. 36, n. 3, p. 462–478, 15 jun. 2012b. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/14684521211241503/full/html>.

Acesso em: 10 set. 2020.

JAMES, C. *et al.* CiteScore metrics: Creating journal metrics from the Scopus citation index. **Learned Publishing**, v. 32, n. 4, p. 367–374, 19 out. 2019. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/leap.1246>. Acesso em: 23 dez. 2020.

JUKEMURA, José; DINIZ, Márcio Augusto. Avaliação do Qualis Periódicos: balanço do *upgrade* do Qualis na medicina III. **Revista do Colégio Brasileiro de Cirurgiões**, v. 42, vsupl. 1, p. 54-56, 2015. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-69912015000800054&lng=en&nrm=iso. Acesso em: 23 dez. 2020.

LESHEM, S.; TRAFFORD, V. Overlooking the conceptual framework. **Innovations in Education and Teaching International**, v. 44, n. 1, p. 93–105, fev. 2007. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/14703290601081407>. Acesso em: 10 maio 2022.

LIU, X. Z.; FANG, H. A comparison among citation-based journal indicators and their relative changes with time. **Journal of Informetrics**, v. 14, n. 1, p. 101007, 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1751157719302901>. Acesso em: 10 maio 2022.

LIVAS, C.; DELLI, K. Journal Self-Citation Rates and Impact Factors in Dentistry, Oral Surgery, and Medicine: A 3-year Bibliometric Analysis. **Journal of Evidence-Based Dental Practice**, v. 18, n. 4, p. 269–274, 2018. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30514441/>. Acesso em: 10 maio 2022.

LOAN, F. A.; NASREEN, N.; BASHIR, B. Do authors play fair or manipulate Google Scholar h-index? **Library hi tech**, v. ahead-of-p, n. ahead-of-print, 2021. Disponível em: <https://en.x-mol.com/paper/article/1437536354485202944>. Acesso em: 10 maio 2022.

KAPLAN, N. The norms of citation behaviour: prolegomena to the footnote. **American Documentation**, v. 16, n. 3, p. 179-84, 1965. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/asi.5090160305/abstract>. Acesso em: 02 dez. 2017.

KIM, J. A.; HUH, S.; CHU, M. S. Correlation among the citation indices of Korean scientific journals listed in international databases. **Science Editing**, v. 1, n. 1, p. 27–36, 2014. Disponível em: <https://www.semanticscholar.org/paper/Correlation-among-the-citation-indices-of-Korean-in-Kim-Huh/b1077c9c77009acdcf50c67b004459f32b1eb8a6>. Acesso em: 10 maio 2022.

KOUSHA, K.; THELWALL, M. Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. **Journal of the American Society for Information Science and Technology**, 58(7), 1055–1065. Disponível em: <http://dx.doi.org/10.1002/asi.20584>. Acesso em: 10 maio 2022.

KOUSHA, K.; THELWALL, M. Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines.

Scientometrics, v. 74, n. 2, p. 273–294, 2008. Disponível em:

<http://link.springer.com/10.1007/s11192-008-0217-x>. Acesso em: 23 dez. 2020.

KRAMPL, A. Journal Citation Reports. **Journal of the Medical Library Association**, v. 107, n. 2, p. 280-283, Apr. 2019. Disponível em:

<http://jmla.pitt.edu/ojs/jmla/article/view/646>. Acesso em: 23 jan. 2020.

LATINDEX. **Latindex**: Sistema Regional de Información en línea para Revistas Científicas de América Latina, el Caribe, España y Portugal. Disponível em:

<https://www.latindex.org/latindex/>. Acesso em: 10 maio 2022.

LARIVIÈRE, V.; HAUSTEIN, S.; MONGEO, P. The oligopoly of academic publishers in the digital era. **PLOS One**, v. 10, n.6, 2015. Disponível em:

<http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0127502&representation=PDF>. Acesso em: 20 jan. 2020.

LARIVIÈRE, V.; SUGIMOTO, C. R. **The Journal Impact Factor**: a brief history, critique, and discussion of adverse effects. arXiv, Springer Handbook of Science and Technology Indicators, 2018. Disponível em:

<https://arxiv.org/ftp/arxiv/papers/1801/1801.08992.pdf>. Acesso em: 20 jan. 2020.

LEYDESDORFF, L.; BORNMANN, L. The operationalization of “Fields” as WoS Subject Categories (WCs) in evaluative bibliometrics: the cases of “Library and Information Science” and “Science & Technology Studies”. **Journal of the**

Association for Information Science and Technology, v. 67, n. 3, p. 707-714, 2016. <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.23408>. Acesso em: 10 set. 2020

LÓPEZ-CÓZAR, E.; CABALLERO, R. R. El impacto de las revistas de comunicación: comparando Google Scholar Metrics, Web of Science y Scopus. **Comunicar**, p. 45–52, 2013. Disponível em: <http://rabida.uhu.es/dspace/handle/10272/6971>. Acesso em: 28 jan. 2020.

LÓPEZ-CÓZAR, E.; CABEZAS-CLAVIJO, Á. Google Scholar Metrics: an unreliable tool for assessing scientific journals. **El Profesional de la Información**, v. 21, n. 4, p. 419–427, 1 jul. 2012. Disponível em:

<https://recyt.fecyt.es/index.php/EPI/article/view/epi.2012.jul.15>. Acesso em: 10 set. 2020.

LÓPEZ-CÓZAR, E.; CABEZAS-CLAVIJO, Á. Ranking journals: could Google Scholar Metrics be an alternative to Journal Citation Reports and Scimago Journal Rank?

Learned Publishing, v. 26, n. 2, p. 101–113, 1 abr. 2013. Disponível em:

<http://doi.wiley.com/10.1087/20130206>. Acesso em: 10 set. 2020.

LÓPEZ-CÓZAR, E.; MARTÍN-MARTÍN, A. Índice H de las revistas científicas españolas según Google Scholar Metrics (2013-2017), 2018. Disponível em:

<https://digibug.ugr.es/handle/10481/54306>. Acesso em: 10 jun. 2022.

LÓPEZ-CÓZAR, E.; MARTÍN-MARTÍN, A. Índice H de las revistas científicas españolas según Google Scholar Metrics (2014-2018). n. October, 2019. Disponível em: <https://www.revistacomunicar.com/pdf/documentos/2019-INDICE-H-2014-2018.pdf>. Acesso em: 10 jun. 2022.

LÓPEZ-CÓZAR, E.; ORDUÑA-MALEA, E.; MARTÍN-MARTÍN, A. Google Scholar as a Data Source for Research Assessment. **Springer Handbooks**. , 2019. p. 95–127. Disponível em: http://link.springer.com/10.1007/978-3-030-02511-3_4. Acesso em: 10 set. 2020.

LÓPEZ-CÓZAR, E.; ROBINSON-GARCÍA, N.; TORRES-SALINAS, D. The Google scholar experiment: How to index false papers and manipulate bibliometric indicators. **Journal of the Association for Information Science and Technology**, v. 65, n. 3, p. 446–454, mar. 2014. Disponível em: <https://ideas.repec.org/a/bla/jinfst/v65y2014i3p446-454.html>. Acesso em: 10 set. 2020.

MACHIN-MASTROMATTEO, J. D. Latin American triple-A journals 2: The Venezuelan case. **Information Development**, v. 35, n. 5, p. 839–845, 16 nov. 2019. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0266666919877962>. Acesso em: 10 maio 2022.

MACROBERTS, M. H., MACROBERTS, B. R. Problems of Citation Analysis: A Critical Review. **Journal of the American Society for Information Science**, v. 40, n. 5, p. 342-349, 1989. Disponível em: <https://ideas.repec.org/a/bla/jamest/v40y1989i5p342-349.html>. Acesso em: 10 maio 2022.

MARTÍN-MARTÍN, A. *et al.* Can we use Google Scholar to identify highly-cited documents? **Journal of Informetrics**, v. 11, n. 1, p. 152–163, 1 fev. 2017. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S175115771630298X>. Acesso em: 10 set. 2020.

MARTÍN-MARTÍN, A. *et al.* Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. **Journal of Informetrics**, v. 12, n. 4, p. 1160–1177, 1 nov. 2018. Disponível em: <https://linkinghub.elsevier.com/retrieve/pii/S1751157718303249>. Acesso em: 10 set. 2020.

MARTÍN-MARTÍN, A.; AYLLÓN, J.; ORDUÑA-MALEA, E. Google Scholar Metrics 2014: a low cost bibliometric tool. **arXiv.org**, 2014. Disponível em: <https://arxiv.org/abs/1407.2827>. Acesso em: 10 maio 2022.

MARTÍN-MARTÍN, A.; ORDUNA-MALEA, E.; LÓPEZ-CÓZAR, E. Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a MTdy comparison. **Scientometrics**, v. 116, n. 3, p. 2175–2188, 26 set. 2018. Disponível em: <http://link.springer.com/10.1007/s11192-018-2820-9>. Acesso em: 10 set. 2020.

MARTÍN-MARTÍN, A.; ORDUNA-MALEA, E.; LÓPEZ-CÓZAR, E. Journal Scholar Metrics: building an Arts, Humanities, and Social Sciences journal ranking with

Google Scholar data. **22nd International Conference on Science, Technology & Innovation Indicators (STI)**, n. September, 2017. Disponível em: https://www.researchgate.net/publication/263856149_Google_Scholar_Metrics_2014_a_low_cost_bibliometric_tool. Acesso em: 20 jun. 2021.

MÉXICO. CRMCYT. **Sistema de Clasificación de Revistas Mexicanas de Ciencia y Tecnología**. Disponível em: <https://conacyt.mx/servicios-en-linea/crmcyt/>. Acesso em: 10 set. 2020.

MICHEL, M. H. **Metodologia e pesquisa científica em CCS**: um guia prático para acompanhamento da disciplina e elaboração de trabalhos monográficos. 2. ed. São Paulo: Atlas, 2009.

MINASNY, B.; HARTEMINK, A. E.; MCBRATNEY, A.; JANG, H. J. Citations and the h index of soil researchers and journals in the Web of Science, Scopus, and Google Scholar. **Peerj**, 1, p. 16, 2013. Disponível em: <https://peerj.com/articles/183/>. Acesso em: 10 jun. 2022.

MINGERS, J.; MEYER, M. Normalizing Google Scholar data for use in research evaluation. **Scientometrics**, v. 112, n. 2, p. 1111–1121, 22 ago. 2017. Disponível em: <http://link.springer.com/10.1007/s11192-017-2415-x>. Acesso em: 10 set. 2020.

MINNITI, S.; SANTORO, V.; BELLI, S. Mapping the development of Open Access in Latin America and Caribbean countries. An analysis of Web of Science Core Collection and SciELO Citation Index (2005–2017). **Scientometrics**, v. 117, n. 3, p. 1905–1930, 2018. Disponível em: https://ideas.repec.org/a/spr/scient/v117y2018i3d10.1007_s11192-018-2950-0.html. Acesso em: 10 maio 2022.

MOED, H. F. Measuring contextual citation impact of scientific journals. **Journal of Informetrics**, v. 4, n. 3, p. 265–277, jul. 2010. Disponível em: https://econpapers.repec.org/article/eeeinfome/v_3a4_3ay_3a2010_3ai_3a3_3ap_3a265-277.htm. Acesso em: 10 maio 2021.

MONTANARI, F.; PACKER, A. L. Critérios de Seleção de Periódicos para Indexação e Publicação nas Coleções da Rede SciELO. IN: PACKER, A.L., et al., orgs. **SciELO – 15 Anos de Acesso Aberto**: um estudo analítico sobre Acesso Aberto e comunicação científica. Paris: UNESCO, 2014, 188 p. Disponível em: <http://dx.doi.org/10.7476/9789237012376>. Acesso em: 10 maio 2022.

MORENO-GIL, S. et al. The dissemination of tourism scientific research in Latin American journals. A bibliometric study. **Anatolia**, v. 00, n. 00, p. 1–16, 2020. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/13032917.2020.1795892>. Acesso em: 10 maio 2021.

NAVAS-FERNÁNDEZ, M.; ABADAL, E.; RODRIGUES, R. S. Internationality of Spanish scholarly journals indexed in Web of Science and Scopus. **Revista Española de Documentación Científica**, n. 3, p. 209, 2018. Disponível em:

<https://redc.revistas.csic.es/index.php/redc/article/download/1014/1608?inline=1>.

Acesso em: 10 jul. 2021.

NEDERHOF, A. Bibliometric monitoring of research performance in the Social Sciences and the Humanities: a review. **Scientometrics**, v. 66, n. 1, p. 81-100, 2006. Disponível em: <https://link.springer.com/article/10.1007/s11192-006-0007-2>. Acesso em: 02 dez. 2017.

OKAGBUE, H. I. et al. Curve Estimation Models for Estimation and Prediction of Impact Factor and CiteScore Using the Journal Percentiles: A Case Study of Telecommunication Journals. **International Journal of Online and Biomedical Engineering (iJOE)**, v. 15, n. 14, p. 31, 26 out. 2019. Disponível em: <https://online-journals.org/index.php/i-joe/article/view/11373>. Acesso em: 10 maio 2021.

OKAGBUE, H. I.; TEIXEIRA DA SILVA, J. A. Correlation between the CiteScore and Journal Impact Factor of top-ranked library and information science journals. **Scientometrics**, n. 0123456789, p. 3–7, 19 abr. 2020. Disponível em: <https://doi.org/10.1007/s11192-020-03457-x>. Acesso em: 10 set. 2020.

ORDUÑA-MALEA, E. *et al.* **About the size of Google Scholar**: playing the numbers. 2014. Disponível em: <http://arxiv.org/abs/1407.6239>. Acesso em: 10 set. 2020.

ORDUNA-MALEA, E.; AYTAC, S.; TRAN, C. Y. Universities through the eyes of bibliographic databases: a retroactive growth comparison of Google Scholar, Scopus and Web of Science. **Scientometrics**, v. 121, n. 1, p. 433–450, 24 out. 2019. Disponível em: <http://link.springer.com/10.1007/s11192-019-03208-7>. Acesso em: 10 set. 2020.

ORDUÑA-MALEA, E.; LÓPEZ-CÓZAR, E. Google Scholar Metrics evolution: an analysis according to languages. **Scientometrics**, v. 98, n. 3, p. 2353–2367, 23 set. 2014. Disponível em: <http://link.springer.com/10.1007/s11192-013-1164-8>. Acesso em: 10 set. 2020.

ORTEGA, J. L. How is an academic social site populated? A demographic study of Google Scholar Citations population. **Scientometrics**, v. 104, n. 1, p. 1–18, 9 jul. 2015. Disponível em: <http://link.springer.com/10.1007/s11192-015-1593-7>. Acesso em: 10 set. 2020.

PACKER, A.L.; MENEGUINI, R. O SciELO aos 15 anos: raison d'être, avanços e desafios para o futuro. IN: PACKER, A.L., et al., orgs. **SciELO – 15 Anos de Acesso Aberto**: um estudo analítico sobre Acesso Aberto e comunicação científica. Paris: UNESCO, 2014, 188 p. Disponível em: <http://dx.doi.org/10.7476/9789237012376>. Acesso em: 10 maio 2022.

PACKER, A. L.; COP, N.; SANTOS, S. A Rede SciELO em Perspectiva. IN: PACKER, A.L., et al., orgs. **SciELO – 15 Anos de Acesso Aberto**: um estudo analítico sobre Acesso Aberto e comunicação científica. Paris: UNESCO, 2014, 188 p. Disponível em: <http://dx.doi.org/10.7476/9789237012376>. Acesso em: 10 maio 2022.

PACKER, A. et al. A coleção SciELO Brasil aos 20 anos. **Informe SciELO Brasil**, set. 2019. Disponível em: <https://www.scielo20.org/redescielo/wp-content/uploads/sites/2/2018/09/Informe-SciELO-Brasil-atualizada-1.pdf>. Acesso em: 20 maio 2022.

PALACIOS, M. The uncertainty of colombian scientific journals with Publindex. **Colombia médica**, v. 48, n. 4, p. 152–154, 1 out. 2017. Disponível em: <https://colombiamedica.univalle.edu.co/index.php/comedica/article/view/3728>. Acesso em: 10 maio 2022.

PAVAS, A.; ARZOLA DE LA PEÑA, N. A Review of the Citation Indicators of the Ingeniería e Investigación Journal. **Ingeniería e Investigación**, v. 40, n. 1, p. 3–7, 5 ago. 2020. Disponível em: http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0120-56092020000100003. Acesso em: 10 maio 2022.

PITOL, S. P.; DE GROOTE, S. L. Google Scholar versions: Do more versions of an article mean greater impact? **Library Hi Tech**, v. 32, n. 4, p. 594–611, 2014. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/LHT-05-2014-0039/full/html>. Acesso em: 10 maio 2022.

PRICE, D. S. **O desenvolvimento da ciência**: análise histórica, filosófica, sociológica e econômica. Rio de Janeiro: Livros Técnicos e Científicos, 1976.

REGENBRECHT, H.; LANGLOTZ, T. **Initiative for an H-index based rating of conferences and journals in HCI and related fields**. 81-82. 2015 Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85051557213&doi=10.1145%2f2808047.2808050&partnerID=40&md5=5b53cdf7dd0679b9c3ad28d19d7ec759>. Acesso em: 10 maio 2022.

REPISO, R.; ORDUÑA-MALEA, E.; AGUADED, I. Revistas científicas editadas por universidades en Web of Science: características y contribución a la marca universidad. **El Profesional de la Información**, v. 28, n. 4, p. 1–10, 26 jun. 2019. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2019.jul.05>. Acesso em: 10 maio 2022.

RODRÍGUEZ MORALES, A. J.; OCHOA OROZCO, S. A.; TRISTÁN, P. M. Impact of Colombian health journals: Comparing Publindex versus Google Scholar Metrics, SciELO and SCOPUS. **Revista Cubana de Información en Ciencias de la Salud**, 25, n. 1, p. 24-35, 2014. Disponível em: <https://pesquisa.bvsalud.org/portal/resource/pt/lil-705671>. Acesso em: 10 jun. 2022.

RODRIGUES, R. S.; ABADAL, E. Scientific journals in Brazil and Spain: Alternative publishing models. **Journal of the Association for Information Science and Technology**, v. 65, n. 10, p. 2145–2151, out. 2014. Disponível em: <http://diposit.ub.edu/dspace/bitstream/2445/58637/1/634857.pdf>. Acesso em: 10 maio 2022.

ROUSSEAU, R. Temporal differences in self-citation rates of scientific journals. **Scientometrics**, v. 44, n. 3, p. 521–531, 1999. Disponível em: <https://akjournals.com/view/journals/11192/44/3/article-p521.xml>. Acesso em: 10 maio 2022.

SALVADOR-OLIVÁN, J. A.; AGUSTÍN-LACRUZ, C. Correlación entre indicadores bibliométricos en revistas de Web of Science y Scopus. **Revista General de Información y Documentación**, v. 25, n. 2, p. 341-359–359, 17 dez. 2015. Disponível em: <https://revistas.ucm.es/index.php/RGID/article/view/51241>. Acesso em: 10 maio 2022.

SANFILIPPO, F. et al. Self-citation policies and journal self-citation rate among Critical Care Medicine journals. **Journal of Intensive Care**, v. 9, n. 1, p. 1–5, 2021. Disponível em: <https://jintensivecare.biomedcentral.com/articles/10.1186/s40560-021-00530-2>. Acesso em: 10 maio 2022.

SCHREIBER, M. The influence of self-citation corrections and the fractionalised counting of multi-authored manuscripts on the Hirsch index. **Annalen der Physik**, v. 18, n. 9, p. NA-NA, 17 ago. 2009. Disponível em: <http://doi.wiley.com/10.1002/andp.200910360>. Acesso em: 10 set. 2020.

SPIROSKI, M. Current biomedical scientific impact (2013) of institutions, academic journals and researchers in the Republic of Macedonia. **Prilozi (Makedonska akademija na naukite i umetnostite. Oddelenie za medicinski nauki)**, 35, n. 3, p. 67-80, 2014. Disponível em: <https://sciendo.com/article/10.1515/prilozi-2015-0010>. Acesso em: 10 jun. 2022.

STOCKHAMMER, E.; DAMMERER, Q.; KAPUR, S. The Research Excellence Framework 2014, journal ratings and the marginalisation of heterodox economics. **Cambridge Journal of Economics**, v. 45, n. 2, p. 243–269, 6 mar. 2021. Disponível em: http://www.postkeynesian.net/downloads/working-papers/PKWP1715_fCgwlkp.pdf. Acesso em: 10 maio 2022.

SZOMSZOR, M.; PENDLEBURY, D. A.; ADAMS, J. How much is too much? The difference between research influence and self-citation excess. **Scientometrics**, v. 123, n. 2, p. 1119–1147, 10 maio 2020. Disponível em: https://ideas.repec.org/a/spr/scient/v123y2020i2d10.1007_s11192-020-03417-5.html. Acesso em: 10 jun. 2022.

TAŞKIN, Z. et al. Self-Citation Patterns of Journals Indexed in the Journal Citation Reports. **Journal of Informetrics**, v. 15, n. 4, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1751157721000924>. Acesso em: 10 maio 2022.

TEIXEIRA DA SILVA, J. A. The Google Scholar h-index: useful but burdensome metric. **Scientometrics**, v. 117, n. 1, p. 631–635, 20 out. 2018. Disponível em: <http://link.springer.com/10.1007/s11192-018-2859-7>. Acesso em: 10 set. 2020.

TEIXEIRA DA SILVA, J. A.; DOBRÁNSZKI, J. Multiple versions of the h-index: cautionary use for formal academic purposes. **Scientometrics**, v. 115, n. 2, p. 1107–

1113, 20 maio 2018. Disponível em: <http://link.springer.com/10.1007/s11192-018-2680-3>. Acesso em: 10 set. 2020.

THELWALL, M.; WILSON, P. Regression for citation data: An evaluation of different methods. **Journal of Informetrics**, v. 8, n. 4, p. 963–971, out. 2014. Disponível em: <https://arxiv.org/abs/1510.08877>. Acesso em: 10 maio 2022.

TORRES-SALINAS, D.; RUIZ-PÉREZ, R.; LÓPEZ-CÓZAR, E. *Google Scholar* como herramienta para la evaluación científica. **El Profesional de la Información**, v. 18, n. 5, p. 501–510, 1 set. 2009. Disponível em: <https://recyt.fecyt.es/index.php/EPI/article/view/epi.2009.sep.03>. Acesso em: 10 set. 2020.

VAGARINHO, J. P. Como identificar a originalidade num artigo científico ou numa tese de doutoramento? How to identify originality in a scientific manuscript or in a doctoral thesis? **Educar em Revista**, v. 35, n. 73, p. 181–207, 2019. Disponível em: <https://www.scielo.br/j/er/a/WdQsHy6Q6cfs65Lcgg36MP/?lang=pt&format=pdf>. Acesso em: 10 maio 2022.

VANCLAY, J. K. Impact factor: Outdated artefact or stepping-stone to journal certification? **Scientometrics**, v. 92, n. 2, p. 211–238, 2012. Disponível em: <https://link.springer.com/article/10.1007/s11192-011-0561-0>. Acesso em: 20 jan. 2020.

VAN NOORDEN, R. Brazilian citation scheme ousted: Thomson Reuters suspends journals from its rankings for 'citation stacking'. **Nature**, v. 500, n. 7464, p. 510–1, 2013. Disponível em: <https://www.nature.com/news/brazilian-citation-scheme-ousted-1.13604>. Acesso em 10 set. 2020.

VASEN, F.; VILCHIS, I. L. Sistemas nacionales de clasificación de revistas científicas en América Latina: tendencias recientes e implicaciones para la evaluación académica en ciencias sociales. **Revista Mexicana de Ciencias Políticas y Sociales**, v. 62, n. 231, p. 199–228, set. 2017. Disponível em: http://www.scielo.org.mx/scielo.php?pid=S0185-19182017000300199&script=sci_abstract. Acesso em: 10 maio 2022.

VÉLEZ-CUARTAS, G.; LUCIO-ARIAS, D.; LEYDESDORFF, L. Regional and global science: Publications from Latin America and the Caribbean in the SciELO Citation Index and the Web of Science. **El Profesional de la Información**, v. 25, n. 1, p. 35, 20 jan. 2016. Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2016.en.e.05>. Acesso em: 10 maio 2022.

VERA-BACETA, M.-A.; THELWALL, M.; KOUSHA, K. Web of Science and Scopus language coverage. **Scientometrics**, v. 121, n. 3, p. 1803–1813, 12 dez. 2019. Disponível em: <http://link.springer.com/10.1007/s11192-019-03264-z>. Acesso em: 10 set. 2020.

VÍIU, G.-A. A theoretical evaluation of Hirsch-type bibliometric indicators confronted with extreme self-citation. **Journal of Informetrics**, v. 10, n. 2, p. 552–566, 1 maio 2016. Disponível em:

<https://linkinghub.elsevier.com/retrieve/pii/S1751157716300542>. Acesso em: 10 set. 2020.

WALTERS, W. H. Information Sources and Indicators for the Assessment of Journal Reputation and Impact. **Reference Librarian**, 57, n. 1, p. 13-22, 2016. Article.

Disponível em:

<https://www.tandfonline.com/doi/abs/10.1080/02763877.2015.1088426?journalCode=wref20>. Acesso em: 10 jun. 2022.

WALTERS, W. H. Citation-Based Journal Rankings: Key Questions, Metrics, and Data Sources. **IEEE Access**, v. 5, n. Section V, p. 22036–22053, 2017. Acesso em: 10 maio 2022.

WALTMAN, L. A review of the literature on citation impact indicators. **Journal of Informetrics**, v. 10, n. 2, p. 365–391, maio 2016. Disponível em:

<http://dx.doi.org/10.1016/j.joi.2016.02.007>. Acesso em: 10 maio 2022.

WALTMAN, L. et al. Some modifications to the SNIP journal impact indicator.

Journal of Informetrics, v. 7, n. 2, p. 272–285, abr. 2013. Disponível em:

<https://ideas.repec.org/a/eee/infome/v7y2013i2p272-285.html>. Acesso em:

WALTMAN, L.; TRAAG, V. A. Use of the journal impact factor for assessing individual articles need not be wrong. arXiv, preprint **ArXiv**:1703.02334, 2017.

Disponível em: <https://arxiv.org/ftp/arxiv/papers/1703/1703.02334.pdf>. Acesso em: 20 jan. 2020. Acesso em: 10 maio 2022.

WALTMAN, L.; VAN ECK, N. J. The relation between Eigenfactor, audience factor, ho and influence weight. **Journal of the American Society for Information Science and Technology**, v. 61, n. 7, p. 1476–1486, 15 abr. 2010. Disponível em:

<http://doi.wiley.com/10.1002/asi.21354>. Acesso em: 10 maio 2022.

XU, L.; DING, K.; LIN, Y. Do negative citations reduce the impact of cited papers?

Scientometrics, v. 127, n. 2, p. 1161–1186, fev. 2022. Disponível em:

<https://dl.acm.org/doi/abs/10.1007/s11192-021-04214-4>. Acesso em: 10 maio 2022.

YANG; WANG. Visualizing information science: Author direct citation analysis in China and around the world. **Journal of Informetrics**, v. 9, n. 1, p. 208-225, 2015.

Disponível em:

<http://www.sciencedirect.com.ez46.periodicos.capes.gov.br/science/article/pii/S1751157715000024?via%3Dihub>. Acesso em: 10 set. 2020.

YU, T.; YU, G.; WANG, M. Y. Classification method for detecting coercive self-citation in journals.

Journal of Informetrics, v. 8, n. 1, p. 123–135, 2014. Disponível em:

<https://ideas.repec.org/a/eee/infome/v8y2014i1p123-135.html>. Acesso em: 10 maio 2022.

ZARIFMAHMOUDI, L.; JAMALI, J.; SADEGHI, R. Google Scholar journal metrics:

Comparison with impact factor and SCImago journal rank indicator for nuclear medicine journals. **Iranian Journal of Nuclear Medicine**, 23, n. 1, p. 8-14, 2015.

Disponível em:

https://irjnm.tums.ac.ir/pdf_10759_f5cbbd6ee667518f0b37e8a753c6a7dd.html.

Acesso em: 10 jun. 2022.

ZHANG, Y.; LUN, H. Is Google Scholar useful for the evaluation of non-English scientific journals? The case of Chinese journals. **Learned Publishing**, v. 32, n. 2, p. 154–162, 23 abr. 2019. Disponível em:

<https://onlinelibrary.wiley.com/doi/abs/10.1002/leap.1208>. Acesso em: 10 set. 2020.

ZHOU, Y. Self-citation and citation of top journal publishers and their interpretation in the journal-discipline context. **Scientometrics**, v. 126, n. 7, p. 6013–6040, 2021.

Disponível em: <https://dl.acm.org/doi/abs/10.1007/s11192-021-03995-y>. Acesso em: 10 maio 2022.