

**UNIVERSIDADE FEDERAL DO PARANÁ**  
**EDUARDO MICHELOTTI BETTONI**

**A CONTRIBUIÇÃO DA WEB NO USO DE INDICADORES DA PRODUÇÃO  
CIENTÍFICA NA ÁREA DE CIÊNCIAS HUMANAS E SOCIAIS APLICADAS**

**CURITIBA**  
**2011**

**EDUARDO MICHELOTTI BETTONI**

**A CONTRIBUIÇÃO DA WEB NO USO DE INDICADORES DA PRODUÇÃO  
CIENTÍFICA NA ÁREA DE CIÊNCIAS HUMANAS E SOCIAIS APLICADAS**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre, Curso de Mestrado em Ciência, Gestão e Tecnologia da Informação do Setor de Ciências Sociais Aplicadas, Departamento de Ciência e Gestão da Informação da Universidade Federal do Paraná.

**Orientadora:** Prof.<sup>a</sup> Dr.<sup>a</sup> Patricia Zeni Marchiori

**CURITIBA**

**2011**

## TERMO DE APROVAÇÃO

Eduardo Michelotti Bettoni

“A CONTRIBUIÇÃO DA WEB NO USO DE INDICADORES DA PRODUÇÃO CIENTÍFICA NA ÁREA DE CIÊNCIAS HUMANAS E SOCIAIS APLICADAS”

DISSERTAÇÃO APROVADA COMO REQUISITO PARCIAL PARA OBTENÇÃO DO GRAU DE MESTRE NO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA, GESTÃO E TECNOLOGIA DA INFORMAÇÃO DA UNIVERSIDADE FEDERAL DO PARANÁ, PELA SEGUINTE BANCA EXAMINADORA:



Prof.ª Dr.ª Patrícia Zeni Marchiori  
(Orientadora / UFPR)



Prof.ª Dr.ª Leilah Santiago Bufrem  
(Examinadora / UFPR)



Prof.ª Dr.ª Maria do Carmo D. Freitas  
(Examinadora / UFPR)

15 de abril de 2011

*À minha mãe e meus irmãos.*

## AGRADECIMENTOS

À Prof.<sup>a</sup> Patricia Zeni Marchiori, pelo tempo despendido em discussões de fundo sobre o trabalho e pela vivência em sala de aula e nos projetos.

Ao Prof. Alexandre Augusto Biz pelo constante incentivo, apoio e espaço concedido para o desenvolvimento da prática em docência e pesquisa.

À Prof.<sup>a</sup> Maria do Carmo Duarte Freitas, pela composição da banca e incentivo à vivência prática no ambiente universitário.

À Prof.<sup>a</sup> Leilah Santiago Bufrem pela composição da banca e empenho em contribuir com este trabalho.

Ao Prof. Egon Walter Wildauer e Prof.<sup>a</sup> Sonia Elisa Caregnato pela composição e contribuição na banca de qualificação.

Ao Prof. Elias Sanz Casado pelo empenho, mesmo que à distância, em contribuir com essa pesquisa.

Aos outros docentes e técnicos administrativos do DECIGI e do PPCGI pela dedicação na formação dos mestres.

Aos outros docentes e técnicos administrativos do DETUR pelas oportunidades e apoio concedidos, além de todo trabalho despendido na formação basilar.

Aos colegas do mestrado pelo apoio e auxílio durante toda esta fase.

Aos alunos dos cursos de graduação em Gestão da Informação e Turismo, por terem participado da minha formação docente.

Aos colegas do grupo o<sup>2</sup> expedição por terem me acompanhado nos momentos de descanso e reflexão.

À comunidade do *software* livre pelo trabalho voluntário na criação das ferramentas que viabilizaram tecnicamente esta dissertação, em especial: *Firefox*, *Zotero*, *LibreOffice* e *Linux Mint*.

À Capes pela concessão da bolsa que viabilizou o desenvolvimento dessa pesquisa.

## RESUMO

A destinação de recursos para a pesquisa em Ciências Humanas e Sociais Aplicadas em países em desenvolvimento é historicamente realizada por meio de critérios subjetivos e uma certa objeção aos métodos quantitativos. Isso ocorre pelo reconhecimento da existência de tendenciosidades nas fontes tradicionais de informação que proporcionam a publicação de indicadores objetivos. Mediante a evolução e popularização das tecnologias da informação e comunicação, a convergência digital nos canais de publicação científica e o surgimento de alternativas às fontes tradicionais, buscou-se verificar como emergentes ferramentas na *web* poderiam apoiar uma mudança no cenário da avaliação da pesquisa científica na área de Ciências Humanas e Sociais Aplicadas. Para tanto, foi realizado um levantamento bibliográfico documental na busca pela caracterização do comportamento na pesquisa da área, e conseqüente enumeração e descrição de indicadores de uso da produção científica mais utilizadas na ciência. Em fase posterior, foi aplicado um protocolo de descrição às ferramentas *web* de acesso gratuito a resultados provenientes de busca de informação acadêmico-científica, com vistas a verificar o seu potencial no cenário descrito. Na etapa final da pesquisa apresentou-se uma discussão com base nas fases anteriores, na qual esclareceu-se que as alternativas estudadas não estariam adequadas, no formato atual, às condições de avaliação de um comitê julgador da área. Ainda assim, servem como alicerce para a proposição de novos serviços de informação que incluam o elemento citação em suas bases, com vistas a promover uma avaliação que esteja mais focada na realidade de sua aplicação, sendo ela uma nação, disciplina, instituição ou autor. A tendência de diferentes necessidades de indicadores pelo comportamento de cada um dos extratos, sugere o uso de bases descentralizadas que permitam intercâmbio de metadados por meio da adoção de padrões, para geração de serviços customizados, sendo essa possivelmente a solução com maior viabilidade, já apresentada com grau de sucesso em outras nações.

Palavras-chave: base de dados online – ferramentas estatísticas; análise de citação; avaliação da atividade científica.

## RESUMEN

La destinación de recursos para investigación en Ciencias Humanas y Sociales Aplicadas en los países en desarrollo es históricamente basada en criterios subjetivos, con una objeción a los métodos cuantitativos. Esto ocurre debido a la existencia de parcialidad en las fuentes tradicionales de información que proporcionan la publicación de indicadores objetivos. A través del desarrollo y popularización de las tecnologías de la información y comunicación, la convergencia digital de los canales de publicación científica y, el surgimiento de alternativas a las fuentes tradicionales, se ha buscado investigar de que manera las nuevas herramientas de la web podrían apoyar un cambio en el escenario de la evaluación de la investigación científica en las Ciencias Humanas y Sociales Aplicadas. Se ha realizado una revisión de literatura para la caracterización del comportamiento en el escopo de la investigación de estas ciencias y también una descripción de los indicadores de producción científica más utilizados. A continuación, se aplicó un protocolo de descripción de las herramientas web de libre acceso a los resultados de búsqueda de información académica y científica, a fin de determinar su potencial en el escenario descrito. En la etapa final de la investigación se ha presentado un debate basado en las anteriores, en lo que se aclaró que las alternativas estudiadas no serían apropiadas, en el formato actual, a las necesidades de evaluación de un comité de destinación de recursos de estas Ciencias. Sin embargo, sirven de apoyo para la propuesta de nuevos servicios de información que incluyan la citación en sus bases de datos, objetivando una evaluación centrada más en la micro-realidad de aplicación; un país, una ciencia, una institución o un autor. La proyección de distintas necesidades de indicadores por lo comportamiento de cada una de estas partes, sugiere el uso de bases descentralizadas que permitan el intercambio de metadatos a través de la adopción de normas, para que haya la generación de servicios personalizados. El estudio concluye que posiblemente esta es la solución de más grande viabilidad, tal como se presenta, con grado de éxito en los Estados Unidos y en Europa.

Palabras-clave: base de datos en línea – herramientas estadísticas; análisis de citaciones; evaluación de la actividad científica.

## LISTA DE SIGLAS

A&HCI	- Arts and Humanities Citation Index
ABEC	- Associação Brasileira de Editores Científicos
APA	- American Psychological Association
BIREME	- Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde
C&T	- Ciência e Tecnologia
CAPES	- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CHAL	- Citation Half-Life Impact Factor
CHSA	- Ciências Humanas e Sociais Aplicadas
CNPq	- Conselho Nacional de Desenvolvimento Científico e Tecnológico
CPCI-S	- Conference Proceedings Citation Index – Science
CPCI-SSH	- Conference Proceedings Citation Index – Social Science & Humanities
CSV	- Comma-Separated Variable
CT&I	- Ciência, Tecnologia e Inovação
ECR	- Expected Citation Rate
ERIH	- European Reference Index for the Humanities
FAPESP	- Fundação de Amparo à Pesquisa do Estado de São Paulo
FAQ	- Frequently Asked Questions
FI	- Fator de Impacto
FINEP	- Financiadora de Estudos e Projetos
GS	- Google Scholar
ICI	- Índice de Citação Imediata
ISI	- Institute for Scientific Information
JCR	- Journal Citation Report
MCT	- Ministério da Ciência e Tecnologia
MEC	- Ministério da Educação
MSAcademic	- Microsoft Academic Research
MV	- Meia-Vida
OAI-PMH	- Open Archives Initiative Protocol for Metadata Harvesting
PDF	- Portable Document Format
PIB	- Produto Interno Bruto
PoP	- Harzing's Publish or Perish



PS	- Post Script Document Format
Redalyc	- Red de Revistas Científicas de América Latina y el Caribe, España e Portugal
SCI	- Science Citation Index
SciELO	- Scientific Eletronic Library Online
SSCI	- Social Science Citation Index
SSRN	- Social Science Research Network
STM	- Science, Technology and Medicine
TICs	- Tecnologias da Informação e Comunicação
TXT	- Text File Format
WIF	- Web Impact Factor
WR	- Web Ranking of World Universities

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	11
1.1 PROBLEMA E JUSTIFICATIVA.....	13
1.2 OBJETIVOS.....	17
<b>2 METODOLOGIA</b> .....	18
<b>3 CIÊNCIA MODERNA E SUAS DIVISÕES</b> .....	25
3.1 HISTÓRICO DA CIÊNCIA MODERNA.....	25
3.2 A DIVISÃO DAS CIÊNCIAS.....	26
<b>4 CANAIS DE COMUNICAÇÃO CIENTÍFICA</b> .....	31
4.1 CANAIS FORMAIS E INFORMAIS.....	32
4.2 CANAIS ELETRÔNICOS.....	35
<b>5 MEDIDAS DE INFORMAÇÃO CIENTÍFICA</b> .....	39
5.1 BIBLIOMETRIA E LEIS GERAIS DE APLICAÇÃO.....	40
5.2 CIENTOMETRIA.....	46
5.3 INFORMETRIA.....	47
5.4 WEBOMETRIA.....	47
<b>6 INDICADORES DE IMPACTO E USO NA CIÊNCIA</b> .....	53
6.1 OS ESTUDOS DE CITAÇÃO.....	54
6.2 PRINCIPAIS CRÍTICAS À ANÁLISE DE CITAÇÃO.....	59
6.2.1 A Thomson ISI, as bases de dados e a análise de citação.....	64
6.2.2 Produtos da Thomson ISI.....	67
6.3 AS MÉTRICAS APLICADAS NA AVALIAÇÃO DA CIÊNCIA.....	68
6.3.1 Contagens Básicas.....	68
6.3.2 Principais indicadores.....	70
6.3.3 Outros Indicadores.....	80
<b>7 PRODUÇÃO CIENTÍFICA EM CHSA</b> .....	84
7.1 ASPECTOS DA AVALIAÇÃO DA PRODUÇÃO CIENTÍFICA NAS CHSA.....	87
7.2 SITUAÇÃO DOS INDICADORES EM CHSA.....	91
7.3 DISCUSSÃO.....	93
<b>8 FERRAMENTAS WEB</b> .....	101
8.1 FERRAMENTAS DE BASES MULTIDISCIPLINARES.....	102
8.1.1 Ferramentas da base Google Scholar.....	102
8.1.2 Ferramentas de bases próprias.....	112
8.2 FERRAMENTAS DE BASES ORIENTADAS A DISCIPLINAS OU GRUPOS ESPECÍFICOS.....	117
8.3 DISCUSSÃO.....	121
<b>9 CONSIDERAÇÕES FINAIS</b> .....	127
<b>REFERÊNCIAS</b> .....	131
<b>APÊNDICE A – DIVISÃO DAS ÁREAS DO CONHECIMENTO PARA O CNPQ COM DETALHAMENTO EM CIÊNCIAS HUMANAS E SOCIAIS APLICADAS</b> .....	143
<b>APÊNDICE B – RELAÇÃO DE BASES DE DADOS PESQUISADAS</b> .....	144
<b>APÊNDICE C – PROTOCOLO DE DESCRIÇÃO DAS FERRAMENTAS</b> .....	145
<b>APÊNDICE D – AUTORES UTILIZADOS NA ELABORAÇÃO DO PROTOCOLO DE DESCRIÇÃO</b> .....	146



## 1 INTRODUÇÃO

A produção científica é considerada uma função primordial no ofício do pesquisador, sendo impossível dissociar essa atividade do ato de fazer pesquisa. A comunicação de resultados é tida com um reflexo do desenvolvimento das disciplinas, servindo como fonte de dados para a verificação da relevância e distribuição do investimento em pesquisa nessas áreas.

Tal postulado, ainda que válido para toda a atividade científica, apresenta variações entre as áreas da ciência, que fazem com que a maneira como a avaliação do impacto e uso das pesquisas seja pensada e aceita de maneira distinta. O comportamento dos pesquisadores, seja em relação aos canais de comunicação ou à forma e extensão dos trabalhos ou, ainda, ao modo como referenciam os estudos anteriores é distinto entre as macro áreas e até em disciplinas, sendo fatores fundamentais ao se pensar em avaliar o desempenho dos cientistas. Se por um lado as ditas “Ciências da Vida, Exatas e Engenharias” estabelecerem critérios objetivos para sua avaliação de impacto e uso da construção do saber científico, as Ciências Humanas e Sociais Aplicadas (CHSA), historicamente, se afastaram dos métodos quantitativos por visualizar uma inadequação de seu uso. A justificativa está em uma série de tendenciosidades nos serviços e métodos que provêm os indicadores de impacto e uso da produção científica, os quais sempre estiveram voltados para as outras áreas da ciência, contemplando suas especificidades.

Os primeiros sinais de mudanças desse cenário são originados da popularização do acesso à *internet* e o uso de computador no processamento de dados, a partir da década de 90, tendo como consequência alterações significativas no processo tradicional de comunicação científica, proporcionando mais facilidades de publicação e o surgimento de alternativas, assim como novos indicadores e fontes de informação para produzi-los.

Baseando-se nessa situação emergente e a possibilidade de alteração no cenário das CHSA, a presente pesquisa teve como objetivo principal verificar como ferramentas na *web* de acesso gratuito a resultados provenientes de busca de informação acadêmico-científica poderiam apoiar a utilização de indicadores de produção científica na área alvo. No desenvolvimento, foi realizado um estudo com base em levantamento bibliográfico como forma de compreender as diferenças dessas áreas em relação às outras, no que diz respeito ao processo de comunicação

científica. Ainda com esse método, buscou-se compreender e definir os indicadores quantitativos de uso e impacto da produção científica desde o início de sua aplicação, até as possibilidades de potencial nas CHSA, bem como a verificação da situação atual da área em relação ao tema.

Como resultado, verificou-se que a dificuldade central não está na aceitação de métodos quantitativos em si pelos gestores dos recursos para pesquisa em órgãos de fomento, mas sim, nas fontes de informação de onde é obtido o insumo para a publicação de indicadores. Essas, tradicionalmente, apresentam tendenciosidades que invalidam comparações igualitárias em toda a ciência, abrindo possibilidade para novos serviços de informação. Nesse sentido, procedeu-se com uma descrição e avaliação das alternativas em potencial para solucionar o problema, as chamadas ferramentas *web*.

Alinhada às descobertas anteriores, percebeu-se que o cenário do início do século XXI propiciou o surgimento de novas ferramentas analíticas e serviços de informação, no formato de fontes e bases de dados, que sugerem possibilidades para a superação das tendenciosidades e projeção do uso de indicadores quantitativos direcionados às características das CHSA e suas disciplinas.

Este trabalho foi dividido em nove capítulos. No primeiro há contextualização da pesquisa, apontamento de justificativas e definição de objetivos, norteados pelas questões de pesquisa. Pela caracterização da pesquisa, optou-se por inserir um capítulo de metodologia em seguida, com as definições dos meios para atingir o objetivo geral. Os capítulos de três a oito resultam do levantamento da literatura pertinente, conforme segue: “Ciência Moderna e suas divisões”, situando a origem do pensamento científico e formação das áreas; “Canais de Comunicação Científica”, com ponderações sobre o seu uso; “Medidas da Informação Científica” situando o histórico dessas medições; os “Indicadores de impacto e uso da produção científica”, seu uso atual e desenvolvimento e, apresentação de uma discussão na aplicação em CHSA. No capítulo nove é apresentada a descrição das “Ferramentas *web*”, precedida de uma discussão em correlação com a apresentada anteriormente. No último capítulo estão as “Considerações Finais”, bem como sugestões de trabalhos futuros.

## 1.1 PROBLEMA E JUSTIFICATIVA

Na gênese da ciência é possível identificar diferentes prioridades nas áreas de estudo, sempre vinculadas ao desenvolvimento de países ou uma região, seja para fins militares, como no período pós Segunda Guerra Mundial, como para novas descobertas na área de saúde, por exemplo, a descoberta do Raio-x. Independentemente dessas condições, a destinação de recursos sempre foi um fator decisivo para a ascensão e reconhecimento de pesquisadores, instituições e nações.

Com variação entre países, o investimento em pesquisa científica costuma ser atribuído a órgãos de fomento com a função de julgar os trabalhos mais relevantes merecedores desses incentivos, bem como verificar o andamento e os resultados. Para que isso ocorra, o principal critério de avaliação é a produção científica, ou seja, o que está sendo comunicado é o retrato dos resultados do trabalho (VAN DEN BESSELAAR; LEYDESDORFF, 2009).

No caso da realidade brasileira, a destinação de recursos para pesquisa é intermediada por fundações de amparo à pesquisa (na esfera estadual), e três órgãos principais na esfera federal: a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), a Financiadora de Estudos e Projetos (FINEP) e o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Para todos estes existem critérios para avaliação de projetos, instituições ou pessoas físicas, com vistas a definir prioridades no direcionamento de recursos.

As fundações têm como premissa fomentar o desenvolvimento científico e tecnológico através do apoio financeiro a projetos e pessoas físicas (CONPAF, 2009; FAPESP, 2009). A CAPES, por sua vez, está vinculada ao Ministério da Educação (MEC), tendo como missão avaliar a pós-graduação *strictu-sensu* no Brasil, possibilitando acesso e divulgação da produção científica, além de prover investimentos na formação de recursos humanos, com enfoque em instituições (CAPES, 2009). A FINEP atua como agência executiva vinculada ao Ministério da Ciência e Tecnologia (MCT), promovendo o desenvolvimento econômico e social do Brasil pelo estímulo à Ciência, Tecnologia e Inovação (CT&I) em empresas, universidades, institutos tecnológicos e outras instituições públicas e privadas (FINEP, 2009).

O CNPq também atua no âmbito federal como uma agência executiva do MCT, fomentando a pesquisa científica e tecnológica por meio de instrumentos como editais de projetos, pesquisa em geral e, principalmente, bolsas para pesquisadores (pessoa física). Para tal, utiliza uma divisão por áreas e subáreas da ciência como forma de avaliação. Isso ocorre devido a diferença existente entre as áreas do conhecimento, sendo que esta agência as divide de acordo com quadro apresentado no Apêndice A.

Em consulta aos critérios definidos por quatro comitês das diferentes áreas, é possível visualizar diferenças entre eles<sup>1</sup>. Nas Ciências da Vida, tendo como exemplo a Coordenação de Agropecuária e Agronegócio (COAGR), são especificados, de acordo com o tipo de bolsa, um número mínimo de artigos publicados, para os quais se aplicam critérios quantitativos, como o índice  $h^2$ . A situação é similar na Biotecnologia e Recursos Genéticos (COBRG), com referência ao mesmo índice. Nas Ciências Exatas, como a Engenharia Química, os critérios também são baseados na publicação de artigos, seguindo a classificação *Qualis/CAPES*, porém, considerando outros tipos de publicação. Já para as Ciências Humanas e Sociais Aplicadas (CHSA), nota-se a predominância de critérios subjetivos. Utiliza-se, primordialmente, o termo “produção científica”, não apresentando, porém, parâmetros quantitativos para representá-la. É possível constatar que o trato com os critérios de julgamento da produtividade científica são distintos entre as áreas.

A disciplina da ciência<sup>3</sup> que se propõe a estudar os aspectos quantitativos da produção, disseminação e uso da informação registrada, bem como oferecer padrões e modelos matemáticos para geração de indicadores de produção científica, como o utilizado por dois dos comitês supracitados, é a Bibliometria (TAGUE-SUTCLIFFE, 1992). O método com maior notoriedade dessa disciplina é o de análise de citação, popularizado na década de 1960 após o surgimento de uma reconhecida fonte de informação, o *Science Citation Index* (SCI), produto do *Institute for Scientific Information* (ISI) (SMITH, 1981). O conceito de citação é considerado uma importante forma de mensuração de impacto da informação, ou seja, um recorte dentro da aplicação da Bibliometria (GLÄNZEL; SCHOEPFLIN, 1999).

---

<sup>1</sup> Consulta ao *site* <http://www.cnpq.br>, em 15/10/2009 e novamente em 26/02/2011.

<sup>2</sup> Mais informações no capítulo 6.3.2 “Principais indicadores”.

<sup>3</sup> Adotou-se a palavra “ciência” com letra não-capitular para o conceito geral de “Atividade humana baseada em método definido, por meio do qual se produzem ou testam conhecimento” (AULETE, 2009, p. 167) e “Ciência” com a primeira letra capitular, ao tratar especificamente de alguma área, ex.: “As Ciências Exatas” (ATHAYDE, 2002).

O método de análise de citação tornou-se referência na mensuração de impacto de publicações, instituições e autores, sendo aceito pela comunidade científica internacional. Ainda assim, recebeu críticas em função de limitações que afetam e enviesam sua validade. As principais alegações se pautam na inadequação dele em certas áreas da ciência, como é o caso das Ciências Sociais Aplicadas, Artes e Humanidades. A explicação está nas tendenciosidades da fonte de informação para utilização de indicadores, os índices do ISI, que inviabilizam comparações entre diferentes áreas, considerando que os comportamentos na comunicação científica são diferentes (SMITH, 1981; MOTTA, 1983; GLÄNZEL, 1996; CRONIN, 2001; MEHO; YANG, 2007). Os reflexos desse cenário são validados na realidade nacional, conforme pôde ser comprovado na já mencionada consulta realizada aos critérios dos comitês de assessoramento do CNPq. A Associação Brasileira de Editores Científicos (ABEC) corrobora com essa reflexão, ao relatar a deficiência e falta de transparência nos critérios objetivos na destinação de recursos para as CHSA, tendo como reflexo um montante inferior de recursos revertidos para pesquisa na área (ABEC, 2009).

Aliado à popularização e ascensão dos indicadores de uso da produção científica como forma de avaliação, o surgimento e popularização das ditas “novas” Tecnologias da Informação e Comunicação (TICs), na década de 1990, mudaram as estruturas da comunicação científica. A *web*, por exemplo, possibilitou a formulação de propostas de acesso livre aos periódicos, com distintos processos de avaliação, como por exemplo, a revisão aberta (na qual os leitores podem julgar o conteúdo). As alterações também foram notadas nos canais não tidos como preferenciais, tal como livros, teses, dissertações, anais de eventos e relatórios técnicos que, paulatinamente, são convertidas em formato digital (CRONIN, 2001; MEHO; YANG, 2007).

A partir desse novo cenário, parte das alegações sobre a inadequação da análise de citação enquanto método de mensuração do impacto e relevância da pesquisa em certas áreas (como as CHSA) estão sendo eliminadas ou minimizadas. Bauer e Bakkalbasi (2005) apontam a transição de vários vetores de publicação para o suporte digital e o avanço dos mecanismos de busca e recuperação desse material, como alternativas parciais que poderiam conferir relevância aos métodos quantitativos de mensuração do impacto da produção científica nessas áreas de estudo. Contudo, deve-se superar a dificuldade de formação de um *corpus* homogêneo de produção científica, no qual haja convergência e integração, o que permitiria uma comparação mais autêntica.



As observações de Bauer e Bakkalbasi (2005) foram motivadas pelo surgimento de duas novas ferramentas de busca e recuperação de material científico-acadêmico: a *Scopus (Elsevier)*, com acesso mediante assinatura e lançada em 2004, inovou ao incluir outras formas de comunicação científica além dos periódicos, tais como apresentações em eventos científicos, *sites* e alguns livros; e o *Google Scholar (Google)*, criado em 2005, que permitiu acesso livre às buscas em diversos canais de comunicação acadêmica, ainda que omitindo os critérios para definição do escopo e caráter científico das fontes que apresenta nos resultados. Ambas, contudo, permitem a geração de indicadores de impacto da produção científica.

Além dessas, há outras iniciativas vinculadas a repositórios de acesso aberto, que são utilizadas para a geração de indicadores bibliométricos. Entre as principais estão *RePec Project* (rede com 73 países, originada em um grupo de pesquisa na Inglaterra), o *CiteReader (Social Science Research Network, Estados Unidos)*, o *CiteBase (University of Southampton, Inglaterra)* e o *CiteSeerX (Pennsylvania State University, Estados Unidos)* (CITEBASE, 2009; CITESEER X, 2009; RICHARDSON; BORCHERT, 2007).

Dentre os países latino-americanos, duas iniciativas têm destaque: o *SciELO (Scientific Electronic Library Online)*, criado para facilitar o acesso à literatura, contribuindo para superação da “ciência perdida”, entendida como um fenômeno relacionado à dificuldade de penetração das descobertas feitas nos países de terceiro mundo naqueles industrializados; e a *Redalyc (Red de Revistas Científicas de América Latina y el Caribe, España e Portugal)*, que auxilia na difusão da atividade científica editorial produzida e sobre a Ibero-América. Ambas contêm procedimentos integrados para medição de uso e impacto dos periódicos indexados (OSAREH; WILSON, 1997; REDALYC, 2009 ; SCIELO, 2009).

Excetuando a *Scopus*, os demais serviços apresentados caracterizam-se pelo acesso gratuito a resultados provenientes da busca de informação acadêmico-científica, doravante chamadas “ferramentas *web*”, enquadrando-se como fontes em potencial para geração de indicadores de produção científica. Apesar das críticas aos índices de citação do ISI na área de CHSA, tais como, a pouca diversidade de canais de comunicação, a baixa cobertura temporal da produção científica, a dispersão de periódicos e barreiras idiomáticas, as ferramentas citadas podem elucidar uma nova realidade para a avaliação da produção científica no escopo (MUELLER, 2005; BARILAN, 2008; SANTOS; KOBASHI, 2009).

Deste modo, a contribuição do presente estudo está voltada para a comunidade acadêmico-científica das CHSA, no que se refere a definição de elementos que sustentem o uso de critérios quantitativos adequados de produtividade, o que poderia servir de apoio a tomada de decisão na distribuição de recursos por órgãos de fomento.

Tendo em vista o cenário apresentado, foi possível conduzir uma discussão para verificar a validade das alternativas, por meio das seguintes questões: (a) quais são os pontos comuns entre as necessidades de diferentes disciplinas das CHSA com relação a produção científica? (b) de que forma esse comportamento influencia a aplicação de indicadores quantitativos para se medir o impacto da produção? (c) quais são as potenciais contribuições das ferramentas *web* de busca e acesso à informação acadêmico-científica como fonte desses indicadores? Partindo destes questionamentos, foi possível estabelecer objetivos e encaminhamentos para a pesquisa.

## 1.2 OBJETIVOS

Tendo as questões citadas como direcionamento, objetivou-se com o presente estudo verificar como as ferramentas *web* de acesso gratuito a resultados provenientes da busca de informação acadêmico-científica podem apoiar a utilização de indicadores de produção científica na área de CHSA. Esse objetivo geral desdobrou-se em outros específicos, a saber:

- a) identificar as características da produção científica da área de CHSA;
- b) discutir a utilização de indicadores de impacto da produção científica na área alvo;
- c) descrever as ferramentas selecionadas sob o enfoque de fonte de indicadores de impacto da produção científica e;
- d) mostrar alternativas aos atuais critérios de avaliação na destinação de recursos para pesquisa em CHSA para agências de fomento;

## 2 METODOLOGIA

Esta pesquisa configura-se de natureza qualitativa, tendo em vista características apontadas por Creswell (2007, p. 186) tais como: o uso de métodos variados e procedimentos interativos e humanísticos; a coleta de dados dos tipos bibliográfico-documental e material audiovisual; a possibilidade de novos direcionamentos de acordo com os resultados parciais e percepção do pesquisador. A orientação é fundamentalmente interpretativa, o que inclui uma descrição de cenário, pessoas ou situações para posterior análise, admitindo a influência de uma lente pessoal do autor no seu momento sociopolítico e histórico.

Com relação aos objetivos propostos, a pesquisa enquadra-se como exploratória, descritiva e, a *posteriori*, analítica. Collins e Hussey (2005) situam a exploratória como esclarecimento de ideias e aumento da familiaridade com a área de estudo, como base para uma investigação mais rigorosa em estágio posterior, constituindo uma primeira etapa da pesquisa. A descritiva, por sua vez, é utilizada para caracterizar e obter informações sobre um determinado problema ou questão. A analítica é a continuidade da anterior, na qual o pesquisador não apenas descreve como também analisa e explica o porque da ocorrência de determinados fenômenos, por meio da verificação de relações causais entre eles.

O estudo relacionou-se no plano exploratório-descritivo à identificação das características de produção científica da área de CHSA, com a discussão dos indicadores de impacto da ciência. Nesse mesmo plano esteve a descrição das ferramentas *web*. A discussão seguinte, no plano analítico, buscou a comprovação ou não do potencial dessas ferramentas como recursos para utilização dos indicadores na CHSA.

O objeto de estudo da pesquisa foi o conjunto de ferramentas *web* enquanto fonte para o uso de indicadores da produção científica na área de CHSA. Esta denominação como área “macro” da ciência é adotada em todo trabalho, alinhada à classificação proposta pelo CNPq<sup>4</sup>. Pelas possibilidades na literatura internacional, enquanto caracterização da área, lançou-se mão da produção científica de outros países.

---

<sup>4</sup> Conforme pode ser verificado no Apêndice A.

O detalhamento dos procedimentos metodológicos tem por finalidade definir a estratégia da investigação, o papel do pesquisador nesse processo (bem como sua intervenção em cada etapa), os passos na coleta e análise de dados, além de estratégias de validação, precisão de resultados e a estrutura de apresentação (CRESWELL, 2007). Optou-se pelo detalhamento do processo investigativo com base em cada um dos objetivos específicos propostos, dividindo-os entre os aspectos descritivos e analíticos.

a) Identificar as características da produção científica da área de Ciências Humanas e Sociais Aplicadas;

A pesquisa foi situada em dois tipos de locais diferentes. As bibliotecas, por meio de uma busca inicial em seus catálogos *online* e, posteriormente, no acervo daquelas pertencentes à Universidade Federal do Paraná e na Biblioteca Pública do Paraná. O segundo local foram as bases de dados nacionais e internacionais, nas quais foram recuperados recursos dos seguintes vetores: *e-books*, artigos de periódicos, anais de eventos e teses. Por meio de uma busca federada<sup>5</sup> nos sistemas enumerados no Apêndice B, foi possível acessar bases de dados, com a inserção de palavras-chave. Entendeu-se limitar àquelas que incluam resultados em português e inglês. Foram elencadas em português: comunicação científica, comunicação acadêmica, ciência perdida, ciência nos países subdesenvolvidos, Ciências Sociais Aplicadas, Ciências Humanas, Letras e Artes, Humanidades, áreas da ciência, além de diferentes combinações entre os termos. Em inglês as palavras foram: *scientific communication, little science, big science, humanities and science, arts and science, social sciences, biases in scientific communication, lost science, scientific publication, publication bias* e, tal como no primeiro caso, combinações entre os termos.

A técnica empregada nessa etapa foi a chamada bibliográfico-documental, na qual o investigador coleta documentos públicos e privados, além de literatura pertinente (CRESWELL, 2007). Essa etapa caracterizou-se por ser primordialmente descritiva. Os recursos recuperados serviram como base para redação sobre a cultura

---

<sup>5</sup> Uma busca e recuperação de informação por meio da solicitação a várias bases de dados hospedadas em diferentes sistemas online de informação, apresentando resultados consolidados (JACSÓ, 2004).

de pesquisa e divulgação de resultados na área de CHSA, estabelecendo comparações com outras. A fase analítica procedeu-se após o cumprimento do objetivo seguinte, por meio do estabelecimento de correlações.

b) Discutir a utilização de indicadores de impacto da produção científica na área alvo.

Nesta fase buscou-se a descrição dos aspectos relacionados às metrias empregadas na medição de impacto da produção científica. Assim como ocorrido na anterior, a pesquisa foi situada nas bibliotecas e sistemas de busca federada, com uso de uma sequência de palavras-chave como condução do processo de recuperação da informação pertinente, sendo dois idiomas selecionados: (português) bibliometria, informetria, infometria, cientometria, cienciometria, metrias da ciência, metrias do livro, medição na produção científica, webometria, métodos qualitativos de medição da ciência, indicadores de produção científica, indicadores de produtividade na ciência, fator de impacto, citações, análise de citação; (inglês) *bibliometrics, infometrics, informetrics, scientometrics, sciencimetrics, quantitative methods on information measures, impact factors, citation, citation analysis*.

Após a fase de coleta dos dados, enquanto técnica bibliográfico-documental, buscou-se uma categorização. Primeiramente, na definição das diferentes abordagens das metrias de informação relacionadas à medição de impacto na produção científica. Em seguida, na descrição detalhada dos diferentes indicadores bibliométricos existentes e suas principais aplicações.

O resultado da fase descritiva do primeiro objetivo específico, em conjunto com a correspondente para o segundo, embasou parte da etapa de discussão do trabalho. Portanto, foi utilizado o conhecimento adquirido com relação às peculiaridades da área de CHSA, bem como do uso, funcionalidades e aplicações dos indicadores de impacto da produção científica, como base para a sugestão de um recorte mais adequado dentre indicadores quantitativos de mensuração da ciência, como aplicação para a área.

c) Descrever as ferramentas selecionadas sob o enfoque de fonte indicadores de impacto da produção científica;

Ao contrário das fases anteriores, a técnica de coleta de dados ocorreu por meio do acesso a materiais audiovisuais que, segundo Creswell (2007), são adequados quando é necessário obter informações sobre um *software* de computador ou, no caso, um sistema de informação.

A definição do universo da pesquisa se deu por meio do estabelecimento de uma lista de critérios a serem atendidos pelas ferramentas *web*, com base na problemática do trabalho e no cumprimento dos dois primeiros objetivos, por meio da revisão de literatura, a saber:

- a) ser uma ferramenta eletrônica, isto é, disponível por meio de um terminal de computador;
- b) possibilitar a exibição de pelo menos um indicador de impacto da produção científica com base no método de análise de citação;
- c) incluir disciplinas das Ciências Humanas e/ou Sociais Aplicadas no escopo do material acadêmico-científico apresentado.

Atendendo obrigatoriamente aos três critérios citados, foram selecionadas 12 ferramentas *web* com base em pesquisas anteriores<sup>6</sup> sobre o assunto, além de uma busca na *web* com os mesmos termos em inglês apontados no objetivo anterior, acrescidos da palavra-chave *tool* (ferramenta). Apesar de compartilharem das características definidas, as ferramentas apresentaram diferenças significativas, seja em função dos objetivos de concepção, ou pelo escopo em que operam.

Como forma de facilitar a compreensão e a descrição das mesmas, utilizou-se uma adaptação da tipologia definida por Nehaus e Daniel (2008), para os quais, uma base de dados com registro de citações pode contribuir de duas formas para estudos bibliométricos:

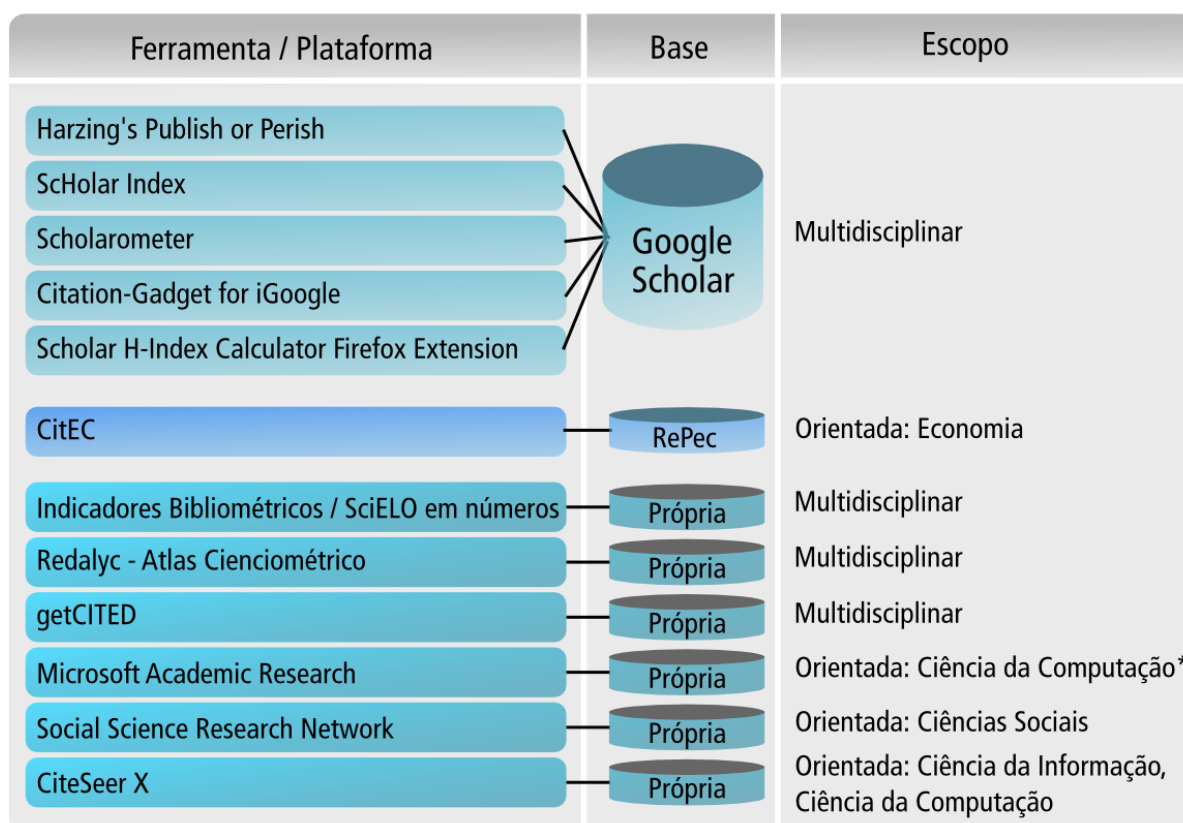
- a) como fonte de dados apenas ou;
- b) como uma plataforma que oferece ferramentas analíticas para análise bibliométrica a partir de uma fonte.

No presente estudo foi selecionada apenas a segunda opção, ou seja, as ferramentas analíticas, pelo fato de fornecerem indicadores de uso da produção científica e não apenas os dados.

---

<sup>6</sup> Osareh e Wilson (1997), Noruzi (2005), Richardsdon e Borchert (2007) e Nehaus e Daniel (2008)

Estabeleceram-se categorias para as ferramentas com base em duas variáveis: base de dados e escopo. A metade delas (6) utiliza uma fonte de dados externa, ou seja, independente do serviço da ferramenta e, dentre elas, 5 tem o *Google Scholar* como fonte em comum. O restante das ferramentas estão vinculadas à uma base própria, podendo ser multidisciplinar ou orientada à uma ou mais disciplinas das CHSA. Na Ilustração 1 são apresentadas as ferramentas de acordo com as categorias propostas.



\*Na classificação do CNPq utilizada nessa pesquisa a Ciência da Computação não está inserida nas CHSA. Ainda assim, o serviço Microsoft Academic Research inclui no grupo "Ciência da Computação" disciplinas que são da área de Ciência da Informação, pertencentes às CHSA, motivo pelo qual a ferramenta enquadrou-se no escopo da pesquisa.

ILUSTRAÇÃO 1: TIPOLOGIA DAS FERRAMENTAS ESTUDADAS

FONTE: O autor (2011)

Após a definição do escopo, foi elaborado um protocolo de descrição (Apêndice C), cuja fundamentação teórica provém dos autores listados no Apêndice D. Este foi o instrumento de coleta de dados utilizado, visando caracterizar as ferramentas *web* selecionadas, como base para posterior análise. O protocolo foi dividido em três grupos de itens a serem verificados:

- a) aspectos técnicos, suporte e interface;
- b) escopo;
- c) filtros e indicadores.

No primeiro, buscou-se situar a ferramenta enquanto um sistema de informação, estudando aspectos técnicos comuns. Neste sentido foram verificados os requisitos para funcionamento. Ainda neste grupo, buscou-se os meios de suporte: a documentação sobre o sistema, tutoriais, perguntas e respostas mais frequentes (FAQ), ajuda e suporte técnico. Coube verificar também o idioma da interface e a interoperabilidade (uso de padrões) com outros sistemas.

O grupo “b” foi criado para definição da abrangência da(s) base(s) de dados relacionadas a cada uma das ferramentas. Esse aspecto é visto como um fator crítico pois, uma das principais dificuldades com o uso das fontes de dados é a cobertura. Sendo assim, verifica-se qual é o escopo definido pelo mantenedor da ferramenta, ou ainda, algum critério que é utilizado para distinguir esse recorte da produção, do total em potencial. Ainda nesse grupo, estudou-se a periodicidade da atualização e a janela temporal de cobertura.

O grupo “c” agrega características que conferem flexibilidade e utilidade às ferramentas para visualização de indicadores de uso da produção científica. Em primeira instância, foram considerados recursos de minimização de problemas atribuídos ao método de análise de citação, como o controle de redundância, a presença de autocitação e a endogenia<sup>7</sup>. Como forma de possibilitar uma avaliação comparativa entre os resultados que as ferramentas apresentam, foram enumerados possíveis filtros: instituição, grupos de pesquisa, áreas da ciência, tipo de publicação, idioma usado na publicação e nacionalidade do autor. Sua importância está na identificação de certos comportamentos regulares e, por sua vez, podem apoiar novas análises sobre o uso da produção científica em diferentes unidades. Essas variáveis também são comuns em pesquisas que envolvem análise de citação (GLÄNZEL, 1996; VAN RAAN, 1996; CRONIN, 2001; LEYDESDORFF, 2008; LANCHO-BARRANTES; GUERRERO-BOTE; MOYA-ANEGON, 2010).

Ressalta-se pela própria característica exploratória da pesquisa, aliada ao dinamismo com que essas ferramentas surgem e são modificadas, que o estudo não pretendeu ser exaustivo na enumeração delas e sim, focado no objetivo geral

---

<sup>7</sup> Ver capítulo 6.2 “Principais críticas à análise de citação”



enquanto verificação de suas potencialidades. Complementa-se que o protocolo de descrição foi utilizado apenas em caráter de orientação do trabalho nos sistemas, não necessariamente sendo preenchido ou relatado na ordem ou na totalidade dos itens postulados, tendo em vista as variadas características das ferramentas.

A apresentação dos resultados seguiu a tipologia definida na Ilustração 1, iniciando pelas ferramentas que tem como base o *Google Scholar* e são multidisciplinares. No caso destas, optou-se por elaborar apenas uma descrição do grupo “b” - escopo, para todas, uma vez que compartilham da mesma base. Esse fator, aliado a representatividade que o *Google Scholar* tem na literatura da área, suscitaram a elaboração de um tópico específico para a base, o qual foi seguido pela descrição dos grupos “a” e “c” individualmente. Pela similaridade dos resultados foi possível elaborar um quadro consolidado ao final da descrição delas. As outras ferramentas foram apresentadas em conjunto com a descrição das respectivas bases. A discussão a respeito das ferramentas *web*, considerada a fase analítica, foi redigida ao final do capítulo, em uma correlação dos resultados obtidos nessa fase e nas anteriores.

d) Mostrar alternativas aos atuais critérios de avaliação na destinação de recursos para pesquisa em CHSA para agências de fomento.

A etapa final caracterizou-se como analítica, uma vez que não pretendeu apenas descrever uma situação, mas sim discuti-la, entendendo que o conteúdo estudado nas fases anteriores apresentou algum embasamento para a proposição de alternativas aos atuais critérios empregados na avaliação da destinação de recursos, tendo como resultado a redação dos subcapítulos.

Após a apresentação do problema, justificativa, objetivos e métodos empregados para o desenvolvimento da pesquisa, seguem os tópicos de revisão da literatura pertinente, bem como aqueles que compuseram a etapa documental da pesquisa.

### 3 CIÊNCIA MODERNA E SUAS DIVISÕES

A compreensão da história do desenvolvimento científico, bem como as posteriores divisões do saber é fundamental para entender o modo como as diferentes áreas da ciência passaram a se comunicar e compartilhar resultados. A gênese apresentada nos tópicos seguintes revela em que condições históricas houve o crescimento de certas áreas, contextualizando-as posteriormente às suas realidades de publicação científica.

#### 3.1 HISTÓRICO DA CIÊNCIA MODERNA

A Ciência Moderna, ou seja, concepção e divisão do saber que se tem no início do século XXI, iniciou-se de modo ordenado, regular e crescente após o auge da Revolução Científica no século XVII. Este movimento teve como precedentes a tradição artesanal, enquanto concepção de novos instrumentos, e a invenção da imprensa, como forma mais eficaz de distribuição da informação. Difundindo-se rapidamente na Europa, a imprensa trouxe a libertação do saber científico em relação ao esquecimento – dada a possibilidade de registro e distribuição – fundamental ao se considerar que é o reconhecimento um dos fatores motivacionais mais importantes na atividade.

A Revolução Científica caracterizou-se pela construção de artefatos tecnológicos oriundos de áreas consideradas pioneiras na ciência, como a Astronomia. Assim, o mérito da Ciência Moderna foi atribuído à tecnologia científica que, em muitos casos, desenvolvia-se antes mesmo de uma necessidade da sociedade. Os estudos do período resultaram na formação de frentes dos chamados “práticos”, devido a aplicação imediata, por isso o viés tecnológico e o trabalho intenso em áreas como a mecânica (PRICE, 1976a).

O período entre o início da revolução (século XVII) e a metade do século XX na Europa foi denominado como *Little Science*, e o período seguinte até o século XXI de *Big Science*. A divisão deu-se pois, após a Segunda Guerra Mundial, quando descobertas e melhorias se tornaram uma prioridade, primordialmente para fins militares – englobando cerca de 80% do investimento disponível na época. O montante

de recursos humanos e físicos tinha destinação definida, isto é, para áreas da Física ligadas a energia, conquista do espaço e, com menor prioridade, às Ciências Biológicas e da Vida. Na década de 1960 chegou-se a prever um colapso financeiro na chegada do próximo século, caso a taxa de crescimento da ciência se mantivesse proporcional à porcentagem de recursos empregados em relação ao Produto Interno Bruto (PIB) (WEINBERG, 1961; FURNER, 2003; ROBREDO, 2003).

Conforme ressalta Price (1976a), essa onda de crescimento tomou proporções globais por meio da troca de conhecimento e importação de tecnologias dos países não europeus, especialmente os Estados Unidos. Uma parte significativa do sucesso científico dessa nação pode ser atribuída ao investimento em ciência nesse período; enquanto importador de tecnologia e mão de obra em primeira instância e, posteriormente, desenvolvendo suas próprias academias científicas.

### 3.2 A DIVISÃO DAS CIÊNCIAS

A maneira como se desenvolveram as descobertas científicas é uma fonte parcial de compreensão da “divisão universal” da ciência. Elaborar uma classificação permeou, desde os primeiros estudos, a ideia de estabelecer relações evolutivas entre as disciplinas, bem como aspectos de proximidade e afastamento considerando características como artificialidade e naturalidade, racionalismo e subjetividade, entre outros. Meadows (1999) destaca como um dos mais expoentes pesquisadores da área, o indiano Shiyali Ranganathan, considerado o pai da Biblioteconomia. Sua pesquisa tinha como intuito oferecer uma classificação para todo conhecimento existente, trabalho que levou décadas entre observação de usuários de bibliotecas e cooperação com outros pesquisadores.

Ranganathan (1968) situou a evolução histórica de divisão do conhecimento por meio de consenso entre os filósofos. Ela se iniciou no sistema dos Védicos (antiga Índia), com as divisões: (a) Dharma: Direito, Religião, Ética e Sociologia; (b) Artha: as Aplicadas, tanto Naturais como Sociais; (c) Kama: Linguística, Artes, Literatura e as Ciências Puras; (d) Moksha: Filosofia e Experiência Mística. Em seguida, o modelo Grego, composto pela Economia Política, Direito, Artes, Matemática, Física e Teologia; serviu como base para o seguinte, chamado de Escolástico (centrado na universidade). No século XVII surge uma divisão que levaria o nome de seu autor,

Francis Bacon (Baconiana): (a) História, compreendendo História Natural, Artes e Literatura; (b) Poesia, incluindo Letras, Fábulas e Alegorias e; (c) Filosofia, Teologia Natural, Ciências Físicas, Sociologia e Economia.

No século XIX surge uma classificação como resultado das inadequações, do conceito sóciocêntrico dos Védicos, na ênfase universitária dos Escolásticos e no psicocentrismo do Baconismo – todas apontadas por cientistas naturais. Assim, Ranganathan (1968) defendeu uma classificação definitiva em três grandes áreas: As Ciências Naturais, a primeira, seria a evolução das disciplinas puras (Física, Química, Geologia e Botânica) somada a suas aplicações (Engenharias, Tecnologia, Mineração, Agricultura e Medicina). As outras duas seriam Humanidades e Ciências Sociais, relacionadas por um escala que iria do natural, em uma extremidade da primeira, ao artificial, na extremidade da segunda.

O sistema apresentado é originado do sociólogo Augusto Comte, o qual classificou as áreas em Matemática, Astronomia, Física, Química, Biologia e Física Social, com uma perceptível ênfase nas Ciências Puras. Meadows (1999) também segue essa perspectiva, chamada “em série”: Ciências Físicas, Biológicas e Sociologia, como reflexo tanto da classificação de complexidade, quanto uma sequência evolutiva. Sendo assim, supostamente a Sociologia iria se aproximar da Biologia e essa, por sua vez, das Ciências Físicas. Essa posição foi negada pela maioria dos pesquisadores das Humanidades e Sociais, haja vista que não aspiravam ser como as Ciências Físicas.

Price (1976a) aponta uma dificuldade em se considerar apenas uma história das ciências, defendendo a hipótese de uma divisão mais contemporânea após a Revolução Científica. A única disciplina que se manteve do período anterior foi a Astronomia Matemática, oriunda da convergência de descobertas gregas e babilônias, delineando-se assim, as Ciências Exatas (Astronomia, Física e Mecânica). No século XVII ascende a Química com a manipulação dos gases e, em paralelo, as Biológicas tomam forma, principalmente a partir das teorias de Darwin. O desenvolvimento de certas áreas também dependeu de incentivos de acordo com as nações, como nos Estados Unidos, em que a Astronomia (em função da posição geográfica) e as Ciências Biológicas (pela riqueza de fauna e flora do continente americano) foram mais estimuladas.

No século XIX essas áreas já haviam evoluído com notoriedade. A Física atingiu um status de racionalidade integral, enquanto a Química constituía um sistema quase perfeito para explicar as reações entre os tipos de moléculas. Já a Biologia, não tanto racional, desenvolvera um modelo sistemático e aceitável pelos estudiosos. Havia também sinergia entre as áreas, além de um sentimento coletivo de “fenômenos naturais” integralmente compreendidos, restando apenas problemas julgados sem importância. Este século também foi marcado pela estagnação, por uma suposta aproximação com o estado superior da ciência ou, fim das possibilidades de descobertas. Esse equilíbrio só foi quebrado pela descoberta do raio-x na Física, em 1895. Price (1976c) reforça que o precedente à descoberta apresentou uma situação excitante e experimental, em que a cada momento algo poderia surgir, modificando parcial ou plenamente os estudos de efeitos físicos e biológicos. Somou-se a esse movimento, um comportamento competitivo entre pesquisadores e nações em busca de méritos. A consequência foi a necessidade de melhoria da comunicação científica, com a rápida expansão do número de publicações destinadas a esse fim e tidas como um meio de comprovar o pioneirismo em dado tema.

Price (1976a) reflete que esta transição caracterizou-se pelo uso de métodos matemáticos e lógicos em todas as disciplinas, revelando-os como a mais importante forma de ordenação do mundo da observação, indissociável das raízes da ciência. Os trabalhos subsequentes obtiveram reconhecimento, desde as ditas Ciências, como é o caso da Biologia, até os estudos humanísticos, como a Economia. Com esses métodos foi possível estabelecer relações consistentes entre Física, Química e Biologia. O autor ressalta a aritmética como uma “arte divina”, viabilizando uma precisão que não pode ser alcançada em estudos de História, por exemplo.

A classificação apontada por Price (1976a) na década de 1970 considera apenas dois grupos: as “ciências” e as “não ciências”. O primeiro contempla a gênese da Ciência Moderna, as Exatas e Biológicas; o segundo refere-se a História, Filosofia e Línguas. O autor questiona a validade dos estudos do segundo grupo, baseando-se no retorno que essas pesquisas têm para sociedade. O que seria investido nos departamentos (das Universidades) de História e Línguas (como exemplo) para formação de doutores, serviria somente para formar um novo professor para alimentar a academia e assim sucessivamente. Por outro lado, nas Ciências Básicas (Física,

Química, Matemática, Astronomia e Biológicas), a sociedade absorve na indústria e outras áreas, para o emprego em Ciência e Tecnologia (C&T), cerca de 80% dos formados nesse mesmo nível.

Meadows (1999, p. 40) situa uma discussão sobre a principal divisão do saber no final do século XX: Ciências, Ciência Social e Humanidades. Ressalta-se o conceito de ciência empregado pelo autor, como um corpo coerente e sistemático de conhecimentos sobre algum tema baseado em pesquisa e estudos rigorosos, destinados a propósitos intelectuais ou pragmáticos de cunho geral sem aplicação prática imediata. Para ele, o conceito reflete uma distinção que havia no início das pesquisas científicas e foi se alterando com o passar dos séculos, sendo que novas “formas de estudar o mundo” poderiam ser aceitas, inserindo as observações e experimentos empíricos como algo aceitável, fugindo da exatidão das comprovações matemáticas. Assim, não somente a palavra ciência gera conflitos mas também o termo Humanidades. Na gênese, referia-se a estudos clássicos de Literatura e Filosofia, tendo evoluído, incluindo estudos literários, filosóficos e históricos. Para esse grupo de disciplinas também é utilizado, em alguns casos, o termo *Arts*. O autor aponta uma dificuldade de definição, principalmente em distinguir o trabalho de criação (de vários ofícios como a pintura, por exemplo) da chamada pesquisa em Humanidades.

Para Price (1976b), qualquer ciência pode ser estudada, caracterizada (e portando diferenciada), em função de sua produção documental. Dessa forma, reconhece que não é possível dissociar a comunicação científica da ciência, uma vez que são interdependentes. Meadows (1999) corrobora com essa afirmação, situando que, independentemente da época referida, a divisão da ciência sempre foi um reflexo de como os cientistas se comunicam e dividem seus espaços de pesquisa. Com isso, o autor sugere que a divisão proposta (Ciências, Ciências Sociais e Humanidades) está baseada no resultado da comunicação formal, ou seja, as áreas podem ser caracterizadas dessa forma.

É fundamental relevar que existem *caveats* nestes estudos entre a divisão das ciências e a comunicação científica, principalmente quanto à universalização dos resultados. Dentre eles está a formação dos pesquisadores, o vínculo institucional, os interesses pessoais, os contextos político e cultural, a limitação das bases de dados que cobrem os trabalhos científicos, a determinação de quais materiais são válidos,

entre outros. Apesar dessas variações, Meadows (1999) reafirma que é inegável uma verificação de regularidades na informação e comunicação, valendo-se da classificação tradicional entre as três grandes áreas.

Dessa forma é possível notar que a maneira como se desenvolveram os campos científicos no decorrer da história influenciou o grau de importância inferido a cada área, bem como a responsabilidade e forma de publicar os resultados dos trabalhos. No início da Ciência Moderna e mesmo no período pós-guerra, os investimentos e interesses se voltaram para as Ciências Puras, influenciados pelo contexto e necessidades imediatas, o que deixou de certa forma, as Ciências Humanas e Sociais em uma zona periférica da ciência. Essas relações também se refletiram na forma como cada uma passou a tornar públicos os resultados de sua construção do saber científico. Para tal compreensão, é necessária uma caracterização dos canais da comunicação científica, conforme poderá ser visto no capítulo seguinte.

#### 4 CANAIS DE COMUNICAÇÃO CIENTÍFICA

A história da comunicação científica se confunde com a da própria ciência, uma vez que o processo de construção do saber científico pressupõe a divulgação de resultados. Conforme verificado no capítulo anterior, a ciência não existe sem a comunicação, ou seja, as descobertas precisam ser compartilhadas para que tenham relevância. A matéria-prima dos pesquisadores é a soma de contribuições que outros fizeram anteriormente, retratada na presença de citações, referências e diversos tipos de notas nos trabalhos científicos. O próprio ofício do cientista tem a produção e publicação de textos como uma das principais atividades, haja vista que a motivação suprema para a profissão é o reconhecimento e o pioneirismo das descobertas, algo que só pode ser obtido por meio da comunicação (PRICE, 1976c; LEYDESDORFF, 2005).

Os primeiros registros da comunicação escrita foram os manuscritos dos gregos, seguindo uma lenta evolução até o século XV, em função da dificuldade de reprodução. Essa realidade só mudou após a invenção da imprensa na Europa, trazendo a possibilidade de cópia em larga escala e, por conseguinte, a invenção do livro. Um número maior de edições pode ser feito com menos recursos, tendo como consequência o surgimento do serviço postal, um importante marco na distribuição de impressos. Cabe ressaltar que o contexto não implicava diretamente na expansão da ciência, pois, a maioria dos livros não tinham caráter científico. Outro formato escrito surgiu no período, semelhante aos jornais dos tempos modernos. Assim como o livro, o enfoque não era científico, servindo para disseminação de notícias de interesse geral (MEADOWS, 1999).

Enquanto parte dos livros passaram a ocupar discretamente uma função na comunicação científica formal, o surgimento do canal mais reconhecido, a revista científica<sup>8</sup>, ocorreria apenas dois séculos depois da invenção do primeiro. Price (1976a, 1976b) credita o surgimento do periódico científico à expansão e sobrecarga de conteúdo proporcionada pela distribuição e popularização dos livros, causando um excesso de informação escrita, impedindo que os cientistas pudessem tomar conhecimento de tudo o que era produzido e, principalmente, julgar o que era

---

<sup>8</sup> Traduzida do conceito *Journal* do inglês, também utilizada como sinônimo de periódico científico.



relevante para si. Esse canal surge como uma solução, na metade do século XVII, por disponibilizar o conteúdo de forma sintética a partir do aprimoramento dos informes e notícias populares.

Meadows (1999) credita o aparecimento do periódico científico a movimentos ocorridos em Londres e Paris no início do século XVII. Os grupos de discussão filosófica valiam-se da comunicação científica informal (por meio oral ou intercâmbio de relatórios de andamento das pesquisas) para troca de conhecimentos, inclusive aceitando membros estrangeiros que trouxessem algo sobre progressos científicos de seus países de origem. O volume de informações chegou a um estado que somente a criação de uma publicação impressa poderia resolver o problema. Nesse contexto surgem as duas primeiras revistas científicas em 1665, na França e na Inglaterra. O conteúdo delas era mensal e avaliado por um conselho para posterior publicação, o que conferia um grau de qualidade ao canal. Desse modo, o periódico científico evoluiu para um *status* de formalidade, sendo aderido gradualmente por outras sociedades científicas.

#### 4.1 CANAIS FORMAIS E INFORMAIS

Com o crescimento dos periódicos e respectiva evolução da comunicação científica foi possível delimitar duas classes diferentes de canais em que os resultados são publicados: os formais e o informais. Para Garvey<sup>9</sup> (1979; *apud* OLIVEIRA, 1996), o primeiro grupo representa a parte pública e visível do sistema, envolvendo artigos de periódicos, livros, relatórios de encontros científicos entre outros. O segundo envolve conversas telefônicas, cartas, pré-artigos e relatórios técnicos. Meadows (1999) acrescenta ao último as conversas face-a-face, a troca de correios eletrônicos e também os encontros e congressos, com outra abordagem. Para ele, os documentos produzidos como *proceedings* são uma representação formal, mas o evento em si tem um caráter informal pelo tipo de interação (trocas e discussões) que é proporcionada.

Garvey e Griffith (1967, p. 1013) fizeram um extenso estudo sobre a diferença entre os canais formais e informais, comprovando a importância de ambos para qualquer disciplina da ciência em uma relação contrabalanceada de contribuição de

---

<sup>9</sup> GARVEY, W. D. *Communication: the essence of science*. Oxford: Pergamon Press, 1979.

cada um para preencher as lacunas do outro. Os formais são em maioria públicos, com potencial para grandes grupos, além de uma disseminação mais eficaz por mensagem (custo e distribuição). A informação nesses é armazenada e recuperada com facilidade, com potencial para geração de relatórios. Os informais são, em geral, mais restritos, apresentando dificuldade de recuperação, com informação mais dinâmica e contemporânea, em forma desestruturada e de difícil monitoramento. Apresentam ainda uma maior interação, geralmente face-a-face, algo que é considerado importante pela facilidade com que se pode julgar a relevância nestes canais.

Meadows (1999) afirma que durante o processo de comunicação científica, é provável que os pesquisadores utilizem-se dos dois tipos de canais. No início, a pesquisa é comunicada informalmente entre o círculo de pesquisadores e, a medida que avança e ganha relevância, segue para um evento científico por meio de uma apresentação. Esta etapa é muito comum nas Ciências, pela frequência dos eventos e característica de obtenção de resultados imediatos. Em geral, as Ciências Sociais e Humanidades envolvem pesquisas mais longas e, por conseguinte, empregam maior tempo antes da comunicação, ainda que parcial, dos resultados. O autor complementa uma consequência do comportamento: é mais comum que os comentários de pares ocasionem a reelaboração dos textos nestas últimas. A etapa seguinte é a publicação, geralmente em um periódico científico, ou seja, formal. O autor ainda situa que o tempo despendido para redação de um artigo de periódico é quase o dobro nas Ciências Sociais e Humanidades em relação Às Ciências.

O periódico científico estabeleceu-se como a fonte prioritária para a comunidade científica enquanto canal formal. Dentre suas vantagens está a agilidade e rapidez no processo produtivo e na distribuição, além do reconhecimento mundial enquanto formato de representação. Herschman (1970) posiciona suas três funções mais importantes: o registro público oficial, a disseminação da informação e, uma maneira de conquistar prestígio e reconhecimento. Tanto o periódico como os outros canais no processo de comunicação científica adquiriram relevância enquanto objetos de estudo da Ciência da Informação e correlatas. Snodgrass (2006) o ressalta como a forma mais objetiva de seleção em uma comunidade científica, sendo que é esta que o valida.

Ainda que os artigos de periódicos detenham esse reconhecimento, Meadows (1999) esclarece a maneira como as áreas da ciência se posicionam quanto ao seu uso enquanto canal de comunicação formal. O grupo de disciplinas agrupadas sob o rótulo *Science, Technology and Medicine* (STM), tais como Física, Robótica e Farmácia, está ligado a publicação de artigos em revistas submetidos à avaliação, haja vista que, conforme afirma Oliveira (1996), a área diminuiu significativamente a produção de livros nos séculos XIX e XX pela necessidade de agilidade nas publicações. Nas Engenharias e Tecnologia, os anais de congressos também são canais importantes. Nas Ciências Sociais e Humanidades há preferência pela publicação no formato de livro. Em alguns casos, os resultados de uma pesquisa acabam aparecendo em diferentes locais. Como exemplo, é feita a publicação por meio de teses e dissertações e, na sequência um artigo (até mais de um) como forma sintetizada de comunicar o mesmo conteúdo publicado anteriormente. Isso ocorre com maior frequência nas Ciências Sociais. Parte dos cientistas rejeitam as teses como meio formal de comunicação, ao considerar que a conseqüente redação de artigos como síntese dos resultados, equivale ao formato de publicação adequado. Considerar tanto a tese como o artigo oriundo dela como meios formais, por exemplo, seria uma aceitação de redundância.

Apesar das variações entre áreas quanto preferência, o reconhecimento enquanto canal formal é aceito basicamente para dois formatos: artigos e livros científicos. De certa forma, ambos seguem tradicionalmente um processo comum de produção. O insumo é fornecido pelos autores em potencial, os quais são submetidos à avaliação de qualidade por uma editora. Essas são classificadas em: (a) comerciais grandes, cobrindo tanto livros quanto periódicos, com enfoque no grupo STM, consideradas mais lucrativas; (b) comerciais pequenas, enfocando algum nicho, como aquelas menos valorizadas tais como como Humanidades e Ciências Sociais; (c) universitárias que, em geral, empregam mais atenção à produção de livros nas Humanidades e Ciências Sociais, devido a dificuldade de publicação nas editoras comerciais e (d) institucionais, com enfoque na pesquisa central dos institutos mantenedores. Independentemente da sua categorização, estas organizações se valem de parcerias com pesquisadores da área científica do autor (seus pares), os quais têm o papel de validar ou não a contribuição do material submetido. A última etapa é a produção do material, que pode ser impresso ou eletrônico (MEADOWS, 1999).

Desde o início da comunicação científica formal, influenciada pela invenção da imprensa e barateio no custo da distribuição, a disseminação da informação científica esteve ligada às facilidades do material impresso o que, inclusive, era a base do modelo de negócio das editoras (independentemente de área). Porém, esse mesmo motivo que impulsionou a comunicação iniciou uma etapa de estagnação e modificação, com a popularização da *web* e conseqüente ingresso de canais eletrônicos também para a ciência, sejam formais ou informais. Isso alterou – e segue alterando – o tempo para publicar, o modo de produzir, distribuir e até cobrar pela produção científica.

#### 4.2 CANAIS ELETRÔNICOS

Os primeiros computadores usados na produção científica destinaram-se ao processamento da informação bibliográfica, na década de 1960, facilitando o armazenamento e a organização de grande quantidade de informações. Eles só passaram a ser empregados em larga escala na década de 1980, haja vista que no período anterior seguia a tendência da ciência para fins militares. Dessa forma, o emprego da tecnologia da informação só chegou nas Humanidades e Sociais em período posterior. Ainda assim, o marco para a ciência seria a popularização da rede mundial de computadores, a Internet, alterando significativamente o comportamento na comunicação científica no final do século XX. O fenômeno pode ser comparado à passagem da publicação manuscrita para a impressa (MEADOWS, 1999).

O acréscimo nas possibilidades de acesso aos documentos científicos e também a outros pesquisadores, permitiu ultrapassar as barreiras geográficas e reduzir significativamente as distâncias pelo mundo virtual. O fenômeno da “rede” abre caminho para interação entre cientistas em diferentes contextos político e sociais, de nações tanto em desenvolvimento quanto desenvolvidas, e de diferentes níveis hierárquicos. Essas interações foram mais importantes em áreas com a tradição de construção coletiva do conhecimento, ou seja, aquelas em que os trabalhos com grande número de autores são mais comuns, como as Ciências Exatas e Biológicas (MEADOWS, 1999). Barreto (1998) acrescenta que na comunicação eletrônica o

tempo de resposta pode ser reduzido a praticamente zero, o formato da informação pode atender a diversas linguagens e a comunicação em rede permite um deslocamento virtual entre diferentes memórias e estoques da informação.

Ressalta-se que os canais eletrônicos em si não representam um novo meio na comunicação, eles trouxeram sim, outro formato para os mesmos canais formais e informais. Os artigos de periódicos que anteriormente eram acessíveis em papel, foram disponibilizados no formato eletrônico na rede de computadores. Meadows (1999) situa uma resistência de adesão no início, haja vista que os cientistas mais tradicionais alegavam falta de cientificidade no conteúdo *online*, descartando-o como fonte confiável. Para o autor, nesta época os periódicos eletrônicos eram criados por entusiastas (no desejo de testar a nova tecnologia), os que buscavam agilizar o processo de publicação ou, os envolvidos em disciplinas que tinham poucos recursos em papel. A transição, ainda que gradual, entre o impresso e o eletrônico, ocorreu no reconhecimento da rapidez da distribuição, incorporação de recursos multimídia, dinamismo na modificação de documentos e melhoria na gestão editorial pela facilidade de geração de relatórios.

Mueller (2006) relembra que houve um despertar em muitos pesquisadores, principalmente de países em desenvolvimento ou áreas com poucos periódicos dedicados, com relação à previsão de uma mudança mais radical no sistema de comunicação, rompendo com limitações impostas pelos centros produtores de pesquisa, conferindo maior visibilidade e penetração internacional em nações emergentes. A autora atribui a este cenário o surgimento do movimento de acesso livre à informação científica. Conforme explica Kuramoto (2006), a frente atingiu abrangência mundial como instrumento na busca pela ampliação do acesso ao conhecimento, o qual seria importante elemento na criação de fortes estruturas sociais, econômicas e técnicas em países com debilidades na circulação da informação científica. Apesar dessas vantagens, Ferreira, Marchiori e Cristofolli (2009), ao estudarem a disciplina de Comunicação e Ciência da Informação, parte da CHSA, corroboram a afirmativa de Mueller (2006) sobre um *status* ainda não reconhecido pelas publicações de acesso livre, em função de aspectos motivacionais apontados pelos pesquisadores tais como a falta de incentivo institucional, o reduzido número de periódicos científicos, a dificuldade de reconhecimento pelos órgãos de fomento e a

dificuldade de acesso a grupos de pesquisa. Todos esses se relacionam a uma percepção das condições de “recompensa”, fator primordial na opção por publicar (mesmo em canais tradicionais).

Sob esse contexto do acesso livre, Meadows (1999) explica que as editoras tiveram que tomar um novo posicionamento, principalmente aquelas voltadas a periódicos científicos. No passado elas atuavam apenas em seus países, em qualquer de área da ciência, com a distribuição do material no formato impresso. Com a popularização da Internet (facilitação da comunicação), ocorreram fusões entre editoras de diferentes nacionalidades, sendo a maioria delas, vinculadas à publicação em STM. Essas organizações adotaram novos modelos de negócio baseados na comercialização de acesso aos periódicos via assinatura, identificação de novos nichos (mais específicos) e a digitalização do material impresso que fora publicado em data anterior à publicação eletrônica. Ainda assim, é importante destacar que a inovação trouxe alguns problemas, como a frequente quebra dos direitos autorais pela facilidade de manipulação dos textos eletrônicos. Na opinião do autor, o impacto foi menor nas editoras de livros científicos, em destaque as dedicadas às publicações de Ciências Humanas e Sociais. Seja pela dificuldade de leitura de longos textos nos meios eletrônicos (os livros), ou pela própria característica de escopo e importância locais, associada as áreas supracitadas, a internacionalização desse tipo de editora foi inferior dada a solidez de seus nichos.

Bailey (2006) aponta o resultado de uma pesquisa sobre o uso dos livros eletrônicos nas áreas da Ciência. Ao contrário de Meadows (1999), ele afirma que no uso dos chamados *e-books*, as publicações mais acessadas são em Administração, Economia e Gestão (Ciências Sociais Aplicadas), seguidas de Computação e Tecnologia, Literatura e Ciências Sociais. Slater (2009) ao dar ênfase na relação de uso impresso *versus* eletrônico, cita as Ciências da Computação e Tecnologia como maior utilizadora (considerando a proporção de alunos em cada área), dos *e-books*. O autor evidencia que, além das duas citadas, as Ciências Puras e correlatas tem preferência pelos livros eletrônicos. Por outro lado, as relativas às Humanidades ainda priorizam o formato impresso. Ainda assim, concluiu que entre os resultados não há diferença significativa e as possibilidades de alteração nessas posições é algo a se considerar.

Assim como Price (1976b) aponta a produção documental como uma forma de estudar uma ciência, verificou-se a dificuldade de estabelecer um *corpus* homogêneo de produção científica de todas as áreas refletido em canais comuns, para que sejam comparadas. Investigar essas dissonâncias é fundamental para que sejam definidas políticas de destinação de recursos baseadas nos critérios de produção científica, com a devida definição da importância dentre os canais.

## 5 MEDIDAS DE INFORMAÇÃO CIENTÍFICA

A produtividade dos pesquisadores tem como principais características a quantidade de informação comunicada e a sua qualidade. No primeiro aspecto, a forma mais usual de mensurar é por meio do número de artigos publicados, ainda que seja difícil valer-se desse critério para áreas que usam outros canais. Meadows (1999) situa que há estudos direcionados para o equilíbrio entre canais, por meio de uma discussão da quantidade de artigos que equivaleria a um livro, por exemplo. Nesse sentido, também se discute se dado livro realmente retrata uma pesquisa (problema que também ocorre com alguns artigos). O autor afirma que a alta produtividade está vinculada a uma série de fatores de contexto na instituição (tanto nos docentes quanto discentes) como a riqueza, a disponibilidade de recursos de tempo, o espaço físico, e o poder de atrair pesquisadores de alta qualidade, o que reafirmaria o espaço, considerado adequado para pesquisa.

O segundo aspecto da produtividade, a qualidade, está geralmente ligado à significância que o trabalho terá para o restante da comunidade. Este critério é universalmente aceito e se expressa pelo método de análise de citação, no qual, pressupõe-se que se um trabalho for muito citado pelo restante da literatura relacionada, possui uma qualidade superior àqueles que não foram. Surge aqui uma objeção comum: o conteúdo também pode ser muito citado se apresentar resultados errôneos. Na visão de Meadows (1999), essas citações também são relevantes, ainda que apresentem problemas, justificando que isso suscita novas pesquisas e amplia a exploração do tema. Para o autor, os trabalhos não citados serão aqueles com conteúdo de menor importância.

Diodato (1994) afirma que os estudos para mensuração da comunicação científica são fundamentais para verificar as relações existentes entre estudantes, cientistas e outros profissionais, bem como o reflexo da dinâmica e evolução de uma disciplina em seu material literário. Para o autor, estão divididos em três grandes áreas; Bibliometria, Cientometria e Informetria, ainda que haja divergência e até sobreposição de conceitos entre diferentes autores dedicados ao tema. A Cientometria, supostamente a mais ampla, gera controvérsias ao pleitear interpretar que a ciência está acima apenas do que é publicado por ela. O autor sugere que há uma sobreposição de termos, defendendo a Bibliometria como um tipo de Informetria e que a Cientometria, por sua vez é um tipo de Bibliometria. A Informetria, enquanto



termo mais contemporâneo, aborda aspectos cognitivos da atividade científica e não-científica. A discussão é ampliada por Almind e Ingwersen (1997), ao sugerirem a aplicação das técnicas informétricas no então popularizado ambiente *web*, denominando esse novo subcampo de estudo como *Webometrics*. Por meio de uma ordem cronológica, os quatros conceitos são definidos a seguir.

## 5.1 BIBLIOMETRIA E LEIS GERAIS DE APLICAÇÃO

A aplicação de métodos matemáticos e estatísticos para livros e outros meios de comunicação refere-se a Bibliometria. O termo surgiu como segmento da bibliografia, na formulação de estatísticas sobre artigos científicos, com intuito de calcular a importância do que era contido nas obras. O campo evoluiu das medidas de formatos, pontos tipográficos e extensão para certos índices de comparação, como o número de vezes que um autor foi lido e então, passou a se ocupar da medida do impacto das publicações (BRAUN; GLÄNZEL; SCHUBERT, 1985; OTLET, 1986).

Broadus (1987) defende que a primeira manifestação do que viria a ser chamado de Bibliometria foi a contagem de livros e similares na biblioteca de Alexandria em 300 d.c. Hood e Wilson (2001) apontam as primeiras aplicações da Bibliometria no final do século XIX, quando Campbell (1896), valendo-se de métodos estatísticos, estudou a dispersão de temas em publicações. Shapiro (1992) relata estudos anteriores com a indexação de relatórios de casos judiciais correlacionados a outros citados. O primeiro índice de citação conhecido, em forma de livro, é datado de 1860, produzido por Henry Labatt, também com base em casos jurídicos.

O nascimento do termo Bibliometria também é tido como um tema de controvérsias. Com frequência atribui-se a origem a Pritchard (1969), como forma de substituir a Estatística Bibliográfica, termo cunhado por Hulme (1923). Por outro lado, Fonseca (1986) contesta tal paternidade, uma vez que um autor belga, Paul Otlet, teria sugerido um termo equivalente (*bibliometrie*) em 1934. Acredita-se que essa divergência tenha sido criada por uma tendência de uso prioritário da língua inglesa, sendo – não por má-fé – desconsiderado o termo cunhado em outro idioma. A Bibliometria caracteriza-se por dar luz ao processo da comunicação escrita e a natureza do desenvolvimento das disciplinas da ciência, pela contagem e análise de

suas várias facetas, com a aplicação de métodos matemáticos em livros e outros meios (NICHOLAS; RITCHEI, 1978; HOOD; WILSON, 2001). Para Diodato (1994), essa listagem deve incluir periódicos eletrônicos, correspondências de voz e vídeos.

Fonseca (1986) caracteriza a Bibliometria como produto da interdisciplinaridade no pensamento científico e humanístico, consistindo na aplicação da estatística à bibliografia. Volta-se para a contagem, cálculo e projeção que exige a participação do computador. Se por um lado os estudos bibliométricos foram inicialmente limitados à produção bibliográfica, passaram a ser também aplicados nas citações, buscando relações de impacto entre documento citante e documento citado. Araújo (2006) complementa, diferenciando a tradicional bibliografia e a Bibliometria, auferindo ao segundo termo um uso predominante de métodos quantitativos ao invés de discursivos, prerrogativa da primeira.

Ainda que o precursor da análise de citação tenha sido Eugene Garfield, Fonseca (1986) relembra que no final da década de 60, sem tomar conhecimento dos estudos anteriores, o sociólogo Henri Lefebvre ressaltou a possível reconstituição das cadeias de citações, pelas quais seria possível identificar o processo que permeia uma ideia desde sua criação até sua extinção.

Geisler (2000) verifica que para a comunidade científica de C&T, a Bibliometria é considerada válida para produzir resultados das atividades de inovação. O autor define o termo como a mensuração das saídas (produtos) da publicação científica e técnica da ciência. Já Borgman e Furner (2002), entendem o tema como uma poderosa oferta de métodos e medidas – ex.: análise de citação – para estudar a estrutura e o processo da comunicação científica.

Os estudos bibliométricos podem servir como indicadores da influência de certos autores e documentos, permitindo classificar aqueles mais importantes e estabelecer relação entre eles. Algumas aplicações típicas de análises em estudos bibliométricos são as de autores mais influentes, redes de cocitação, influência de jornais, bases intelectuais, estado da arte e frequência de coautoria (RAMOS-RODRÍGUEZ; RUÍZ-NAVARRO, 2004).

As medidas derivadas da Bibliometria são tipicamente contagens de frequências de determinados eventos os quais podem ser observados e estudados para posterior formulação de probabilidades de ocorrência. Nesse campo são chamadas de distribuições ou leis bibliométricas, sendo as mais conhecidas: Lotka, Bradford e Zipf (BORGMAN; FURNER, 2002).

## a) Lei de Lotka

Na metade da década de 1920, o matemático, químico e estatístico Alfred Lotka, comprovou uma Lei básica sobre a produtividade de cientistas. Com base no periódico *Chemical Abstracts*, ele identificou que parte significativa da literatura científica era produzida por um pequeno número de autores, formulando a seguinte equação:  $y_x = 6/p^2x^a$ , em que  $y_x$  é a frequência de autores publicando um número  $x$  de trabalhos a um valor constante ( $a$ ) para cada campo da ciência (ARAÚJO, 2006, p. 13). Essa Lei também ficou conhecida como a do quadrado inverso, em que, basicamente, se tratava de uma comprovação de que apenas uma parcela dos pesquisadores são realmente produtivos (SANTOS; KOBASHI, 2009).

Price (1976a, p. 153) exemplifica que, para Lotka, se um pesquisador tiver a possibilidade de produzir um artigo em sua vida científica, a probabilidade será apenas de um quarto de que este mesmo autor produza dois artigos; um nono para produzir três; um centésimo para produzir dez (e assim sucessivamente). Afirma ainda que, por mais simples que pareça a Lei – e constatação de Lotka – é surpreendente a maneira como ela é verificada com precisão.

De uma forma geral, o número de publicações de um autor sempre foi considerada a sua métrica de produção científica. A principal contribuição de Lotka foi identificar como se dá a distribuição deste trabalho. Outros autores fizeram comprovações semelhantes em seus campos, apontando para distintos resultados do chamado expoente – o mesmo valor ( $a$ ) da fórmula anterior –. Huber (2002) buscou inserir uma terceira variável, a qual comprovou ser válida, nomeando-a “vantagem cumulativa”. Concluiu que a produtividade está diretamente ligada ao princípio de Pareto<sup>10</sup>, que os trabalhos em média apresentam uma periodicidade randômica e que a duração da carreira de um autor também segue uma distribuição contínua.

Price (1976c, p. 48) apresenta um estudo sobre requisições pela *Science Library* de Londres na década de 1960, constatando que esta distribuição de produtividade que se assemelha muito ao princípio de Pareto também é comprovada

---

<sup>10</sup> Vilfredo Pareto foi um economista e sociólogo, criador de um princípio que explicava que 20% das variáveis conhecidas seriam responsáveis por 80% dos resultados. Sua primeira comprovação mostrou que 80% da riqueza de uma nação estava concentrada em apenas 20% da população (CRAFT; LEAKE, 2002). Comparação semelhante pode ser vista no contexto da publicação científica, em que apenas uma pequena parte dos pesquisadores é responsável por um número significativo de trabalhos.

pelo uso (empréstimo) de periódicos. Em suma, 10% das edições de periódicos foram responsáveis por 80% da demanda. Cerca da metade das edições sequer foi consultada durante o ano da pesquisa.

Price (1976b) afirma que a Lei de Lotka seguiu sendo reafirmada com o passar dos anos e nem mesmo as mudanças nas características dos modelos de produção científica alteraram as constantes verificadas, principalmente no que se assemelha ao princípio de Pareto. A ampliação do número de trabalhos com mais de um autor (construção colaborativa), que altera os índices relativos ao número de autores por trabalho, não modificou o comportamento da Lei. Pelo contrário, este contexto revelou um outro tipo de movimento: aqueles pesquisadores sem destaque começaram a buscar mais a parceria com os importantes, no intento de melhorar seu trabalho tendo maior chance de sucesso, devido a chancela de um pesquisador de renome. Dessa forma, aqueles que muito produzem (e são prolíficos), estão produzindo cada vez mais (em grupo), permanecendo e reafirmando o espaço entre os mais produtivos, o que comprova as leis de distribuição.

#### b) Lei de Bradford

Outra das métricas para informação tem como objeto o periódico científico e foi formulada para identificar a extensão das ideias contidas em dado documento em relação à própria disciplina de estudo e às outras, apoiada em cálculos que atendem à proximidade e afastamento. Dessa forma, Samuel Bradford, em 1934, formulou a também conhecida Lei de dispersão (ARAÚJO, 2006).

A Lei de Bradford fala sobre uma representação espacial em duas dimensões indicando a distância entre determinados periódicos com relação à abordagem de um determinado assunto, pressupondo a formação de três zonas com diferentes concentrações. A zona nuclear contém um número  $y$  de artigos em  $n$  periódicos e é onde há uma maior concentração sobre aquele assunto. Para que se obtenha novamente o mesmo número  $y$  de artigos relacionados, buscando em outros periódicos, será necessário utilizar  $n^2$  periódicos. Da mesma forma, em uma terceira pesquisa, o usuário obterá a mesma quantidade  $y$  de artigos somente se buscar em  $n^3$  periódicos (NICOLAISEN; HJØRLAND, 2007).

Dentre suas finalidades está a administração de sistemas de informação e gestão de bibliotecas, no sentido de identificar aqueles periódicos que são mais usados em determinados temas, visando a proposição de critérios para a seleção em

uma dada coleção, de modo a equilibrar custo *versus* benefício e apoiar o estabelecimento de políticas de desenvolvimento de acervos (COUTINHO, 1988; ARAÚJO, 2006; SANTOS; KOBASHI, 2009).

Drott (1981) discorre sobre as principais falhas apontadas nos conceitos da Lei de Bradford, cuja discussão central se dá pela distinção entre a pesquisa teórica e a empírica relacionadas à Lei. Os estudiosos da primeira área entendem que a compreensão e evolução da Lei da dispersão dependem do estudo e aprimoramento de certas fórmulas matemáticas. Para os envolvidos com o empirismo, nota-se um enfoque no ponto de vista do usuário desta informação, considerando variáveis que vão além dos aspectos estatísticos, tais como as formas do estudo e as características do universo (coleção). Isso acaba, muitas vezes, invalidando o modelo teórico por considerar apenas uma “situação ideal”. Por outro lado, os estudos empíricos dificilmente passam pelas comprovações estatísticas. Esse *gap* entre o teórico e o empírico é apontado como um dos maiores problemas da aplicação dessa Lei.

### c) Lei de Zipf

A terceira Lei básica dos estudos métricos da Ciência da Informação foi elaborada por George Kingsley Zipf, no final da década de 1950. Para Araújo (2006) e Saichev, Malevergne e Sornette (2009), o tema de maior aplicação da Lei de Zipf foi em relação ao tamanho das organizações, o qual é determinado pelas vendas, receitas e número de empregados. A Lei teve origem em um *ranking* de frequência para palavras em uma determinada obra (texto longo). Hood e Wilson (2001) resumem a Lei de Zipf como um estudo da frequência ou ocorrência de palavras. Esta colocação pode ser entendida ao analisar o surgimento e motivação do pesquisador com o tema.

Segundo Wyllys (1981), Zipf estudava as mudanças fonéticas da linguagem quando passou a se interessar pela frequência da presença de palavras em dados textos. Em 1949 lançou sua obra principal, com a introdução de um conceito considerado a principal razão para o comportamento humano: a Lei do mínimo esforço. Esses autores lembram que Zipf não se sentia a vontade com as abordagens quantitativas, na matemática e na estatística, direcionando sua reflexão para considerações linguísticas, sua área de estudo.

Uma das conclusões apontadas por Wyllys (1981) é que a Lei de Zipf comprova que as pessoas tendem a utilizar com maior frequência aquelas palavras que são mais fáceis e familiares, do que aquelas tidas como não-familiares. Sendo

assim, é possível posicionar que a Lei é também conhecida como a “Lei do menor esforço”. O próprio autor questiona, na época, quais seriam aplicações possíveis para isso, além dos estudos matemáticos comprobatórios, sugerindo, talvez, suporte a um vocabulário controlado ou caracterização de diferentes áreas.

Serrano, Flammini e Menczer (2009) reconhecem o trabalho de Zipf situando a comunicação escrita como uma das mais fundamentais manifestações da linguagem humana. Com a popularização e adesão de meios eletrônicos como vetores desta linguagem, há uma facilidade em se processar e analisar muitos documentos em pouco tempo. Sendo assim, as constatações de Zipf são essenciais porque possibilitam, em diversas aplicações, compreender melhor o funcionamento da mente humana. Os autores relatam comprovações em que as proporções de ocorrência propostas por Zipf continuam seguem sendo avaliadas. Apresentam ainda a relação sobre o fenômeno estudado, proposta pelo autor: a frequência global de ocorrência de uma palavra, nas condições descritas, é inversamente proporcional ao seu *ranking* (posição na lista de frequências).

Bentley, Ormerod e Batty (2009) apontam um crescimento nas ciências estatísticas de estudos relacionados à distribuição de fenômenos sociais, abordando principalmente: *links* na internet, citações entre autores, parceiros sexuais e tamanhos de empresas (e sua extinção) em relação a muitas outras. Newman (2005) complementa com a aplicação nos seguintes temas: tamanho de terremotos, crateras da lua, número de cliques em páginas *web*, vendas de livros e algumas metrias da Ciência da Informação, como a quantidade de citações recebidas por um artigo científico.

As leis bibliométricas não representam diretamente uma medida de produtividade científica individual com aplicação e uso em comitês de avaliação. Elas se referem a constatações gerais e abrangentes sobre a história do desenvolvimento científico, auxiliando no mapeamento e verificação de certas tendências e fenômenos, que podem se comportar de maneiras distintas entre áreas da ciência ou mesmo entre nações. Essas leis antecederam medidas mais objetivas e específicas do impacto de determinados trabalhos para uma comunidade científica, que tornar-se-iam uma evidência somente após a década de 60 com o estabelecimento de índices de citações globais e estudos de aplicação. Assim, o maior reconhecimento e importância para a

avaliação da pesquisa, em Bibliometria refere-se aos chamados indicadores de impacto e uso da produção científica, apresentados no Capítulo 6 “Indicadores de Impacto e uso na Ciência”.

## 5.2 CIENTOMETRIA

O termo Cientometria foi criado no final da década de 1960, pelos russos Nalimov e Mulchenko. Eles buscaram expressar um tipo de estudo de todos os aspectos da literatura em C&T. A área começou a ganhar notoriedade com a criação do periódico *Scientometrics*, cerca de uma década depois, em que constavam todos os temas quantitativos da ciência, a comunicação e suas políticas. Uma parte significativa dos estudos na área é de difícil distinção com relação à Bibliometria, sendo que parte dos resultados desta última, são publicados na *Scientometrics*. Isso ocorre porque o produto mais tangível da ciência ainda é a publicação, objeto principal da Bibliometria. Em contraposição, existem outros produtos como as práticas de pesquisadores, as estruturas sócio organizacionais, a gestão e desenvolvimento da pesquisa, a relação com a economia, as políticas públicas, entre outras formas de relacionamento para estudo da Cientometria (HOOD; WILSON, 2001).

Para Santos e Kobashi (2009) a Cientometria preocupa-se com a dinâmica da ciência como uma atividade social, valendo-se dos seguintes objetos de estudo: produção, circulação e consumo da literatura científica. Com relação às aplicações e pesquisas na área de Cientometria, Van Raan (1997) aponta as principais categorias: o desenvolvimento de métodos e técnicas para o design, a construção e a aplicação quantitativa de indicadores em aspectos importantes da C&T, o desenvolvimento de sistemas de informação aplicados a C&T, o estudo das interações entre C&T e o estudo cognitivo dos campos da ciência e o seu processo de desenvolvimento.

### 5.3 INFORMETRIA

A primeira proposição do termo Informetria foi feita no final da década de 1980, pelo alemão Nacke, cujo objetivo era formar um conceito para cobrir a parte da Ciência da Informação responsável pela mensuração por métodos matemáticos, de forma mais ampla que a Bibliometria, considerando a informação como um fenômeno (HOOD; WILSON, 2001).

Tague-Sutcliffe (1992) define a Informetria como o estudo dos aspectos quantitativos da informação em qualquer forma, não somente em arquivos bibliográficos, tampouco em qualquer grupo social ou científico. Isso inclui a comunicação formal ou informal registrada, as necessidades de informação, além de conceitos não pensados nos primeiros trabalhos sobre Bibliometria, como, a definição das métricas de informação e a mensuração da performance na recuperação da informação. Para Ingwersen e Christensen (1997) trata-se de uma recente extensão da tradicional análise bibliométrica, envolvendo as comunidades não-acadêmicas em que a informação é produzida, comunicada e utilizada.

Hood e Wilson (2001) observam a Informetria como o termo mais genérico entre os três (Bibliometria, Cientometria e Informetria), constituindo uma macro-área. Ainda assim, seus pesquisadores costumam usar o termo que mais adequado a cada trabalho, mesmo que haja sobreposição. Nota-se uma maior utilização, sob esse ponto de vista, do termo Bibliometria, principalmente pelos pesquisadores que não são da área de Ciência da Informação.

Egghe (2005, p. 1311) afirma que o termo inclui todas as métricas da Ciência da Informação: Bibliometria (bibliografias e bibliotecas), Cientometria (políticas para ciência, análise de citação, avaliação de pesquisas) e Webometria (medidas para *web*, redes sociais da *internet* e outras como redes colaborativas).

### 5.4 WEBOMETRIA

Em meio ao eminente crescimento e popularização da Internet, Almind e Ingwersen (1997) foram pioneiros ao sugerir uma nova área dedicada a aplicação da Informetria no ambiente *web*. Usando a definição de Informetria como a pesquisa



generalizada em informação, os autores sugerem que, ao tratar da comunicação mundial em rede, um campo deveria ser criado com o nome de Webometria. Este seria complementado com outros métodos quantitativos (que não os infométricos).

Björneborn (2004, p. 12) define a Webometria segundo a Ciência da Informação, posicionando-a como estudo dos aspectos quantitativos da construção e uso de recursos informacionais, estruturas e tecnologias na *web*, sob a ótica da Bibliometria ou da Informetria. Neste sentido, engloba tanto os aspectos de elaboração, quanto de uso da *web*, por meio de quatro áreas principais: (a) análise do conteúdo de uma página; (b) análise da estrutura de ligações (*links*); (c) análise de uso (por meio de *logs*) e (d) análise da tecnologia empregada (*performance*). Para o autor, existem ainda outros termos semelhantes como *web mining*, *cyber geography*, *web intelligence* que surgiram na década de 1990, sendo que os dois que se estabeleceram foram Webometria e Cibermetria. O segundo termo seria mais amplo, selecionando não apenas páginas como também grupos de discussão, listas de *e-mails* e outros elementos da comunicação em rede. A relação entre os estudos métricos na ótica da Ciência da Informação pode ser vista na Ilustração 2.

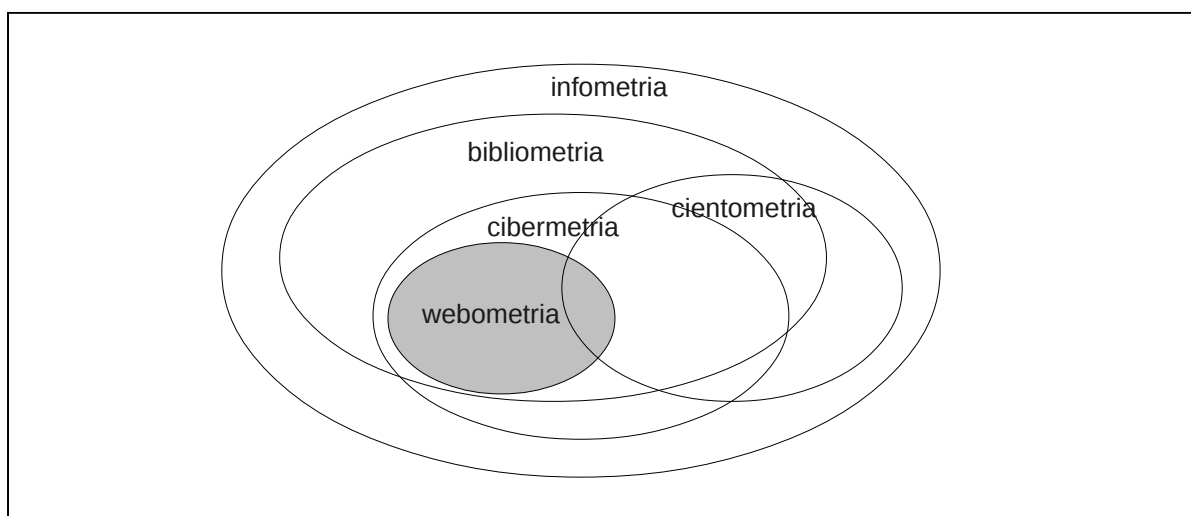


ILUSTRAÇÃO 2: RELAÇÃO ENTRE OS ESTUDOS MÉTRICOS NA CIÊNCIA DA INFORMAÇÃO  
FONTE: Björneborn e Ingwersen (2004, p. 1217).

A Ilustração 2 revela a aproximação, na teoria de Björneborn e Ingwersen (2004), do termo Webometria principalmente inserida à Bibliometria e seus métodos, ao contrário da Cibermetria que teria uma relação mais ampla, principalmente com a Informetria. Este contato entre as duas primeiras áreas é recorrente desde as

primeiras suposições feitas por Almind e Ingwersen (1997), os quais já afirmavam que o futuro da pesquisa em Bibliometria seria a informação eletrônica, pela tendência de crescimento de bases de dados e conversão de muitos formatos impressos.

A primeira pesquisa aplicada também é originada por aqueles que cunharam o termo. Almind e Ingwersen (1997) analisaram o uso da *web* na Dinamarca, comparando a seus países vizinhos e também às bases tradicionais de citação, com as seguintes variáveis: o tipo do documento *web* (origem da instituição mantenedora, presença ou não de gráficos e textos), a disciplina da ciência relacionada, o tamanho da página e sua densidade (a importância que possui em sua rede de pares). Como resultado, verificaram que não houve relação proporcional entre a produção científica nos índices de citação e o desempenho das páginas *web*, pelos seguintes motivos: os índices, ainda que possuam tendenciosidades, têm dados estruturados e confiáveis, a *web* é dinâmica e sem um padrão; a qualidade e veracidade do conteúdo *web* é duvidosa; o investimento (técnico) nas páginas de algumas instituições não é proporcional ao seu desempenho na pesquisa científica.

Os estudos subsequentes ao tema objetivaram validar os *links* como uma variável de análise válida para a informação. Björneborn (2004, p. 15) propôs uma tipologia para os *links*, pela qual seria possível aprimorar os estudos e criar indicadores com base nesses elementos. *Inlinks* seriam traduzidos como aqueles recebidos, ou seja, se um *site* A aponta para B, este último teria um *inlink* de A. Na mesma situação, se a relação for vista a partir de A, nota-se que ele possui um *outlink* para B, ou seja, uma ligação saindo para B. O termo *selflink* seria para indicar que um *site* tem uma ligação para dentro dele. *Co-link* ocorre quando dois *sites* diferentes apontam para um outro em comum. Há também uma divisão entre *links* internos (apontam para um mesmo domínio) e *links* externos (apontam para terceiros), o que não anula os tipos citados anteriormente.

Nesse sentido os *sites* de universidades se tornaram um dos principais objetos de estudo da Webometria, por meio da comparação de resultados com os índices tradicionais de uso e impacto da produção científica. Os *links* emergiram como potencial elemento para o desenho de mapas apontando o relacionamento entre diferentes pesquisadores, comunidades acadêmicas, instituições, países e regiões do planeta. Ainda, seria possível verificar a importância que cada um tem em sua rede, bem como o grau de visibilidade (BJÖRNEBORN, 2004; VANTI, 2005).

No Brasil, Vanti (2010) avaliou os *sites* das universidades da região Nordeste, como parte de uma pesquisa mais ampla a respeito de todo país. Seu trabalho deu-se por uma análise webométrica, utilizando os critérios seguintes. Primeiro, o tamanho do *site*, correspondente ao número total de páginas, calculado por um motor de busca. Em seguida, a visibilidade, que trata dos *inlinks* recebidos por um *site*, desconsiderando os *selflinks*, também medido por um motor de busca. A luminosidade, outro critério, é o indicador que representa a quantidade de *outlinks*, criado com um programa mapeador. Ainda, a densidade de rede, é o quanto um grupo de *sites* se relaciona entre si e por meio de seus *links*, calculado na divisão do número de *links* existentes entre *sites* pelo número de ligações possíveis.

Algumas relações com a Bibliometria podem ser verificadas, como o número de *links* recebidos ou enviados, análogos às citações, ou mesmo a exclusão de *selflinks*, o que se aproxima do recorte de “autocitações”, recorrente nessa área de pesquisa. Ainda assim, talvez a melhor relação expressa entre as duas metrias foi a criação do *Web Impact Factor* (WIF), também usado como critério no estudo de Vanti (2010). Quem introduziu o termo foi Ingwersen (1998), estabelecendo-o como análogo ao Fator de Impacto, tradicional e reconhecido indicador bibliométrico publicado pelo ISI para periódicos indexados em sua base de dados. O WIF é definido conforme a seguinte fórmula (THELWALL, 2000):

$$\text{WIF} = \frac{\text{Número de páginas que apontam – linkam determinado site (ou escopo)}}{\text{Nº de páginas do determinado site (ou escopo)}}$$

Vanti (2010, p. 64) considera a Webometria uma área ainda recente na Ciência da Informação, com escassa literatura teórica e prática, atribuindo a dificuldade ao contexto instável do objeto *web*. De modo geral, ela verificou que a aplicação dos critérios supracitados, principalmente o WIF, mostraram-se ainda impertinentes na medição e não condizentes com o “mundo real”. A autora conclui que os resultados são muito objetivos e, os motivos para a sua ocorrência, deveriam ser estudados com maior profundidade.

Aguillo *et al.* (2010) estudaram cinco diferentes sistemas de *ranking* entre universidades, dentre os quais um totalmente baseado em Webometria, chamado *Web Ranking of World Universities* (WR)<sup>11</sup>. Apesar de ser confundido com uma avaliação

---

<sup>11</sup> Criado pelo *Cybermetrics Lab*, alocado no *Consejo Superior de Investigaciones Científicas de España*, em Madri. <http://www.webometrics.info>

dos *websites* de universidades, esse *ranking* é formado por critérios que têm a produção científica eletrônica como principal alicerce. O valor final é obtido por meio de quatro indicadores webométricos, a saber:

- a) tamanho (número total de páginas);
- b) visibilidade (número total de outlinks);
- c) Rich Text (contagem de arquivos nos formatos: Portable Document Format-PDF, Microsoft Word-DOC, Post Script-PS e Microsoft Power Point-PPT; supostamente tipos associados a disponibilização de documentos importantes para academia);
- d) Scholar (número de documentos e citações para cada site com base no Google Scholar, visto como uma fonte de informação com amplo escopo, principalmente quanto aos repositórios institucionais e diferentes tipos de formatos de publicação, inclusive a produção científica informal) (CYBERNATICS LAB, 2010).

Por meio da pesquisa, Aquillo *et al.* (2010) encontram um resultado semelhante ao obtido por Vanti (2010). Os autores explicam que houve similaridade apenas entre os *rankings* baseados em dados bibliométricos. O índice que apresentou maior dissonância foi o WR, baseado na *web*. Atribuem o resultado aos problemas de gestão e elaboração dos portais das universidades, sendo que muitas vezes esses não correspondem, em proporção, à grandeza que aquela instituição representa. Verifica-se, nos dois casos, a dificuldade de aplicação das pesquisas pela inconsistência dos dados e o próprio reconhecimento dos objetos – *sites* das universidades – como algo importante para a própria instituição.

Na Webometria existem pesquisas voltadas não somente para universidades, mas também para repositórios institucionais (podem contemplar mais de uma instituição). Ao contrário dos critérios apresentados anteriormente, neste tipo de estudo, as metrias estão vinculadas ao registro de ações dentro dos *sites*, e conseqüente geração de relatórios. Tratam-se das estatísticas de uso na *web*. Richardon e Borchert (2007) relatam a implementação feita no *Dspace*<sup>12</sup>, na qual é possível sumarizar algumas atividades:

---

<sup>12</sup> Iniciativa de código aberto, voltada à implementação de repositórios institucionais.

- a) número de visualizações de um item;
- b) número de visualizações de uma coleção;
- c) número de visualizações por uma comunidade;
- d) sumário customizado de itens arquivados;
- e) *link* com problemas (ex.: fora do ar);
- f) acesso de usuários registrados;
- g) buscas mais populares (com maior ocorrência).

Esses critérios são mais abrangentes e ligados ao recurso tecnológico do que aqueles usados nas pesquisas anteriores. Tratam, em suma, dos registros quantitativos com pouco significado na perspectiva da avaliação de impacto, a ser apresentada no próximo capítulo. Richardson e Borchert (2007) apontam como uma fase ainda incipiente na criação e consolidação das metrias *web*, baseadas em repositórios. Os autores defendem que as dificuldades proporcionadas pelo uso do Fator de Impacto e as bases do ISI, enquanto monopólio da informação sobre uso da produção científica, ocorreram em função de um crescimento eminente do movimento de acesso aberto e criação de repositórios institucionais. Eles apontam uma necessidade de mudança do paradigma vigente na avaliação da produção, o que ocorrerá paulatinamente, seja pela necessidade de reconhecimento desses repositórios, ou pelo desenvolvimento de novos indicadores de uso da produção científica. Em hipótese, o ideal seria a busca por uma interoperabilidade de repositórios, abertos ou comerciais, pela qual se obteriam dados comuns de citação. A partir desse dado aberto, seria possível, de acordo com a necessidade de cada ator envolvido no processo, derivar uma metria adequada para cada universo de aplicação, por exemplo, o pesquisador, a universidade, o grupo de pesquisa, entre outros.

De maneira geral, apresentou-se nesse tópico uma abordagem da Webometria voltada para novas aplicações de indicadores bibliométricos e busca – convergente entre as duas – pela criação e divulgação de metrias do uso e impacto da produção acadêmico-científica em suas instâncias. Essa relação, foi estabelecida com base nas preposições de Almind e Ingwersen (1997) e Vanti (2005), para os quais, há uma relação direta entre a Bibliometria e a Webometria, ao contrário de uma terceira área, a Cibermetria, que trataria da *web* sem o enfoque da avaliação científica. Com essa abordagem, há necessidade de um detalhamento e estudo aprofundado desses meios tradicionais, ou ainda dos eminentes, de indicador o impacto na Ciência, conforme o capítulo seguinte.

## 6 INDICADORES DE IMPACTO E USO NA CIÊNCIA

O interesse em medir a ciência é originado de uma necessidade, a partir das décadas de 1950 e 1960, de avaliar as pesquisas, em função do aumento significativo da informação científica no período pós-guerra. Se até esse período o principal objeto de estudo da Bibliometria eram os textos como um todo (em livros e artigos científicos), o enfoque passa a ser nas citações (sinônimo de referências bibliográficas) (FONSECA, 1986).

A medida fundamental usada na maioria dos estudos de avaliação na ciência é a contagem de citações, ou seja, o número de vezes que um indivíduo, grupo, nação, documento ou periódico é citado e a frequência com que isso ocorre. Os índices mais elaborados da produtividade são baseados nesses elementos (BORGMAN; FURNER, 2002).

Para compreender a importância de se avaliar a ciência e seus atores, é necessário verificar o que isso representa para o sucesso de cada um deles. Pringle (2008) explica que a carreira dos cientistas é dependente da representação de sua produção. Essa, por sua vez, é comprovada por meio de métricas orientadas para o universo científico, com base no conteúdo publicado. As instituições financiadoras de pesquisa têm, em paralelo à construção do conhecimento científico, uma necessidade de definir os melhores espaços para reverter fundos, lançando mão de métodos objetivos de identificação de pesquisadores e grupos mais prolíficos. O resultado é uma reação em cadeia, na qual as instituições acadêmicas são pressionadas para mostrar resultados, o que por sua vez recai sobre os pesquisadores que precisam destacar-se com trabalhos de qualidade que melhorem a reputação das organizações às quais estão vinculados. Armbuster (2009) articula que as métricas tornaram-se populares politicamente pois apresentam simultaneamente um compromisso com competitividade, excelência e responsabilidade, o que é uma prioridade nas políticas de muitas nações.

Apesar de contraditória, a forma mais aceita de definir o impacto de determinado pesquisador, instituição, grupo de pesquisas, periódico, é por meio do método de análise de citação, pelo qual é possível elaborar desde simples medidas, como a contagem de citações, até indicadores mais complexos.

## 6.1 OS ESTUDOS DE CITAÇÃO

Os estudos de citação têm como marco inicial o trabalho de Eugene Garfield que, em 1963, introduziu o método chamado de “análise de citação”. O autor conseguiu comprovar, nos anos seguintes, que este novo objeto de estudo seria uma válida e valorizada forma de criar uma descrição histórica confiável dos campos da ciência, principalmente depois que a citação bibliográfica passou a compor a publicação científica, no início do século XX (GARFIELD; SHET; TORPIE, 1964). Para que o método tivesse aplicação, foi necessário valer-se de uma base de dados confiável, sendo utilizado o *Science Citation Index* (SCI), criado anteriormente pelo *Institute for Scientific Information* (ISI) para fins de melhoria da recuperação da informação. O SCI integra hoje um conjunto de índices de citação, de propriedade da *Thomson Reuters*<sup>13</sup>, conhecido como *Web of Science* (BAUER; BAKKALBASI, 2005; FINGERMAN, 2006; THOMSON REUTERS, 2011a)

O método popularizou-se amplamente entre os historiadores da ciência, sociólogos e especialistas no tratamento de documentos, como forma de retratar a obsolescência dos periódicos, como indicadores do fluxo da informação, na busca pesquisa a ser feita, no apontamento de tendências da ciência e em estudos da sociologia da ciência. Na década de 1980, tornou-se valorizado e relacionado ao conceito de indicador de qualidade do trabalho científico (MOTTA, 1983). Em definição semelhante, Nicolaisen (2007) afirma que o emprego da análise de citação tem três funções primordiais. Em ordem cronológica de surgimento: a busca e recuperação da informação; o desenho ou modelagem do desenvolvimento histórico da ciência e tecnologia e; a análise quantitativa e qualitativa de cientistas, publicações e instituições de pesquisa. Smith (1981) aponta a facilidade do desenvolvimento dos estudos de citação, haja vista que não seria necessária a cooperação humana (que envolve recursos e predisposição dos interessados) e, por consequência, não haveria algum tipo de contaminação das respostas.

O despontar da análise de citação ocorreu com o reconhecimento do SCI, pela criteriosidade na elaboração dos índices de citação, sendo considerado o mais importante no tema, fonte praticamente única para o fornecimento de dados para

<sup>13</sup> Durante a pesquisa foram encontradas as denominações “ISI”, “*Thomson ISI*” e “*Thomson Reuters*” na literatura pertinente. A aquisição do ISI pela *Thomson* ocorreu em 1992; dessa forma, convencionou-se utilizar “ISI” para referenciar trabalhos e discussões anteriores à 1992 e, “*Thomson ISI*” para trabalhos posteriores.

formulação de indicadores. Com este, foi possível mensurar com considerável validação dos dados, o impacto de publicações e autores, sendo aceito na definição de relevância de diversas áreas de estudo (SMITH, 1981).

Em função da importância que os índices de citação representam para a mensuração do impacto na ciência e consequentes usos derivados como a destinação de recursos para pesquisa, a análise de citação foi tema de estudos a respeito de sua validade e suas limitações.

Para Kaplan (1965), a prática de citar está relacionada à reafirmação de um conjunto de normas do comportamento científico, já que a ciência é uma instituição governada por regras internas e, dentre elas, está a necessidade de intercâmbio de informações e posterior reconhecimento, obtido por meio da citação. Garfield (1977) corrobora ao afirmar que a citação foi incorporada ao ato de fazer ciência, sendo reconhecida como essencial na validação de qualquer pesquisa.

Na busca por uma definição universal, Price (1976a) posiciona que os artigos trazem citações para justificar argumentos, referenciar trabalhos anteriores, dos quais é possível criar uma representação em rede apontando conexões entre estes pesquisadores relacionadas aos temas que são abordados. A citação seria o ato de remeter de um artigo para o outro, distinguindo-se entre citante e citado.

Nicolaisen (2007) resume o ato de citar na ciência à identificação de autores precedentes, com seus conceitos, teorias métodos e recursos, os quais inspiraram ou foram utilizados pelos pesquisadores durante a elaboração ou apresentação de suas pesquisas. Para o autor, os estudos sobre citação podem ser melhor entendidos se divididos entre os voltados para o comportamento, e aqueles direcionados para os aspectos simbólicos das citações. Aqueles sobre o comportamento ao citar podem ser entendidos por meio de duas principais teorias, a sócio construtivista e a normativa.

Na primeira, a citação possui uma natureza persuasiva, na qual os citantes utilizariam as citações apenas como forma de reafirmar seu julgamento sobre o tema, causando uma distorção e promovendo a própria representação do trabalho citado. Soma-se a essa teoria o chamado *name-dropping*, para o qual as autoridades das áreas da ciência seriam citadas apenas para que o autor citante ganhasse crédito em decorrência do reconhecimento dos citados. Ambas colocações foram posteriormente discutidas, com testes mais conclusivos para o *name-dropping*, também conhecido por ganho de credibilidade por associação. Inferiu-se que a percentagem de citação a



autoridades não sustenta a hipótese de ganho pelos citantes, dada comprovação por meio da distribuição de citações (CASE; HIGGINS, 2000; WHITE, 2004; BORNMANN; DANIEL, 2008).

Na teoria normativa a prática da citação está associada ao pagamento de um débito intelectual de um autor para outros, cujos trabalhos foram usados pelo citante. Esta referência teria tanto uma função simbólica, enquanto registro persistente de propriedade intelectual da fonte utilizada naquela alegação de conhecimento, quanto instrumental, ao informar ao leitor sobre trabalhos anteriores, que podem ser desconhecidos e possivelmente causariam interesse. A citação é vista também como uma regra do comportamento científico (KAPLAN, 1965; NICOLAISEN, 2007). Um das críticas de maior destaque a essa teoria partiu de MacRoberts e MacRoberts (1987) que inovaram ao utilizar o método de análise de conteúdo e contexto de artigos, buscando validar ou não o valor da citação enquanto representação das relações na ciência. Na sua área afim, as Ciências da Vida, invalidaram a citação enquanto representação de uso/mérito/débito intelectual ao verificarem, qualitativamente, que a maior parte das citações foi vista como desnecessária e/ou usada de maneira errada.

O trabalho pioneiro na área do comportamento, no que tange as motivações ao citar foi o de Garfield (1962, p. 85), no qual o autor elencou 15 razões, baseando-se na observação da localização no texto, na linguagem e suas variações e nas regularidades encontradas. A lista contempla aspectos tanto da abordagem normativa quanto a sócio construtivista.

- a) homenagear os pioneiros;
- b) dar crédito a trabalhos relacionados, homenageando os pares;
- c) identificar metodologias e recursos;
- d) fornecer uma leitura do pano de fundo / contexto;
- e) revisar o próprio trabalho;
- f) revisar o trabalho de outros;
- g) criticar trabalhos anteriores;
- h) sustentar declarações;
- i) fazer um alerta para trabalhos futuros;
- j) promover trabalhos pouco disseminados e/ou indexados de forma inadequada, além daqueles não citados;
- k) dar autenticidade a dados e classes de fatos (ex.: constantes da física)

- l) identificar as publicações em que ideias e fatos foram discutidos originalmente;
- m) identificar as publicações nas quais originalmente foram descritos conceitos epônimos (aqueles que levam o nome de uma personalidade) ex.: Lei de Pareto;
- n) negar ou discordar com o trabalho ou ideia de outros;
- o) discutir a primazia da declaração de outros autores.

Os estudos subsequentes ao de Garfield (1962) foram direcionados à elucidação da motivação, caracterização e uso das citações, seguindo duas tipologias: análise de contexto e conteúdo; uso de entrevistas e questionários. Com base nos estudos anteriores envolvendo estas duas tipologias, Bornmann e Daniel (2008) elaboraram um tipologia unificada com os mais importantes tipos de citações, conforme apresentado no Quadro 1.

<b>Tipo</b>	<b>Observação</b>	<b>%<sup>14</sup></b>
Afirmativo	Afirmação do trabalho citado; trabalho citado serviu de apoio para citante; trabalho citante depende de citado; trabalho citado está de acordo com ideias ou descobertas de citante.	10 a 90%
Prepositivo	Trabalho citante refere-se a conhecimento adotado em determinado contexto; trabalho citante presta agradecimento aos trabalhos pioneiros citados.	5 a 10%
Conceitual	Uso de definições, conceitos ou teorias do trabalho citado.	1 a 50%
Contrastivo	Trabalho citante contrasta seu trabalho com anteriores; trabalho citante é uma alternativa ao trabalho citado.	5 a 40%
Metodológico	Uso de materiais, equipamentos, práticas ou ferramentas do trabalho citado; Uso de métodos de análise, procedimentos e desenho do trabalho citado.	5 a 45%
Negativo	Trabalho citante disputa em alguns aspectos com o citado; trabalho citante corrige ou questiona aspectos do citado; trabalho citante avalia negativamente o citado.	1 a 15%
Superficial	Trabalho citante faz uma referência superficial ao citado; trabalho citante é citado sem comentários adicionais; trabalho citante faz uma referência redundante ao citado; trabalho citado não é aparentemente relevante para os problemas do citado;	10 a 50%
Persuasivo	O autor do trabalho citado é reconhecido como uma autoridade na área de pesquisa.	5 a 40%

QUADRO 1: TIPOLOGIA UNIFICADA DAS CITAÇÕES

FONTE: Bornmann e Daniel (2008, pp. 66-67) com base em Garzone e Mercer (2000), traduzido pelo Autor (2011).

<sup>14</sup> Bornmann e Daniel (2008, p. 66) explicam que o método para obtenção destes valores foram estudos anteriores sobre citação.

Nesse quadro constam o tipo de citação, as situações em que essas se enquadram na categoria, e a variação de porcentagem encontrada em estudos sobre citação, para cada uma das categorias. Por exemplo: no tipo “Metodológicos”, correspondente ao uso de materiais, equipamentos (entre outros) de um trabalho citado, Bornmann e Daniel (2008) identificaram desde um estudo anterior em que a porcentagem de citações correspondia à 5%, variando até outro valor de 45%. Os autores esclarecem que, apesar desta variação invalidar algumas conclusões, os resultados chamam a atenção, especialmente os valores máximos atingidos pelos tipos: superficial (até 50%), persuasivo (até 40%) e negativo (15%). Nos casos, a análise de citação poderia ser considerada um método inválido para geração de indicadores de impacto, uma vez que as tipologias são vistas como desalinhadas aos pressupostos assumidos do comportamento ao citar.

Da mesma forma que os estudos apresentados visam compreender e criticar o uso da análise de citação, também apresentam dificuldades metodológicas na sua aplicação, o que cria obstáculos para a obtenção de resultados concretos sobre o tema. Os resultados apresentam uma variação para a qual é difícil chegar a conclusões de um comportamento comum. Bornmann e Daniel (2008) concluem que o processo de citação é complexo e não fornece um monitoramento ideal da performance na ciência mas, existem alegações suficientes de que as motivações não são tão diferentes ao ponto de invalidar a análise de citação enquanto método para mensuração de impacto.

Apesar das contradições e estudos nas teorias a respeito do comportamento, Nicolaisen (2007) revela que o principal objeto da Cientometria se volta para os aspectos simbólicos da citação, nos indicadores derivados desta, e a forma como eles podem refletir e representar a ciência. Nesse ponto de vista, um dos trabalhos principais da área é do próprio Garfield (1979); esclarecendo, à luz das críticas, que a citação é um indicador de utilidade. Desse modo, não pode ser interpretada ou suposta como sinal de elegância, importância relativa, qualidade, mas sim, basicamente sobre o uso. O trabalho acabou sendo aceito na elaboração das chamadas medidas de impacto e uso da ciência, haja vista que, para os interessados na avaliação, se uma determinada unidade de análise é vista como útil, pressupõe-se que ela teve um impacto no meio.

## 6.2 PRINCIPAIS CRÍTICAS À ANÁLISE DE CITAÇÃO

A abordagem de Nicolaisen (2007) sobre o comportamento ao citar e a simbologia da citação, revela apenas algumas das críticas e dúvidas com relação à sua validade como objeto da elaboração de métricas na ciência. Como resultado da primeira década de surgimento da análise de citação, Garfield (1979) elaborou um estudo analisando cada uma das principais críticas ao método, trazendo comprovações sobre a relevância ou não do que estava sendo apontado. Conforme explica Motta (1983), a importância dada a produção de medidas do desempenho científico gerou a maior parte da discussão sobre as falhas no processo, tendo o indivíduo, o pesquisador, o periódico a instituição e as nações enquanto unidade de análise. Nesse sentido, Garfield (1979) dividiu as críticas recebidas ao seu método entre (a) qualidade dos dados e, (b) as características dos dados.

### a) Críticas à qualidade dos dados

O foco, no que diz respeito a qualidade dos dados, foi a utilização somente de “primeiros-autores” nas metrias do SCI. Apresenta estudos que comprovam tanto a irrelevância, em certas áreas, como a validade em outras. Um aspecto derivado são as coautorias. Nos trabalhos com mais de um autor, seria difícil determinar a contribuição de cada um. Em outro sentido, os trabalhos com mais autores teriam mais possibilidade de serem citados se comparados ao individuais (GARFIELD, 1979).

Glänzel e Thijs (2004) estudaram o tema e comprovaram que os de múltipla autoria atingem um número maior de citações do que os individuais e, este valor é ainda maior quando a coautoria é entre pesquisadores de diferentes países. Verificaram ainda que a porcentagem de autocitação é maior nos trabalhos com mais autores e, a produtividade dos autores envolvidos nestes trabalhos não é proporcional ao número médio de citações. Para os autores, esse é um indício de que os trabalhos com múltipla autoria não são resultado de grupos de pesquisas estáveis e sim, de colaboração casual.

Outro fator atrelado à qualidade dos dados seria a precisão da contagem de citações em função de grafia igual do nome de alguns autores, contexto no qual eles seriam homógrafos. Garfield (1979) considera que dificilmente isso aconteceria em um mesmo campo da ciência, e uma alternativa ao problema seria ampliar o detalhamento sobre o autor na publicação para que não houvessem erros na criação dos índices.

## b) Críticas às características dos dados

No segundo grupo de críticas discutido por Garfield (1979) quanto a análise de citação, denominada “características dos dados”, o autor enumerou: as citações negativas, as autocitações, as citações a trabalhos de metodologia, as descobertas precoces, a obliteração, o peso individual da citação de acordo com a fonte e diferença entre as áreas, conforme explicado a seguir.

As citações negativas são aquelas em que o autor citante deseja contradizer, não apoiar ou inferir qualificação positiva. Garfield (1979, p. 362) questiona se as citações negativas são realmente sinônimos de demérito. Para ele, a crítica, assim como a comunicação, é uma das funções fundamentais do processo de construção do conhecimento científico. Para ele, trabalhos importantes foram duramente criticados antes de obter o reconhecimento. Aqueles com muitas críticas também seriam um destaque, caso contrário os cientistas não se ocupariam em citá-los. Sendo assim, de qualquer modo possui algum tipo de valor e, invalidá-las seria considerar a citação como uma medida de quantas vezes um autor esteve correto, distorcendo o real significado.

As alegações contra o uso das autocitações se pautam na possibilidade delas servirem para a manipulação de índices de citação, ainda que a prática de se autocitar seja comum. Na ocasião em que um cientista está focado no seu grupo de trabalho e detém um alto índice de autocitação, isso não significará mais do que uma especialização no assunto. Para que a autocitação realmente faça diferença, o autor deveria ter uma frequência acima do normal de publicações, o que é raro (GARFIELD, 1979).

Com relação ao possível destaque dado aos trabalhos de metodologia, Garfield (1979) explica que os pesquisadores consideram que este tipo de trabalho é menos importante do que o teórico. Para o autor, os trabalhos teóricos também são responsáveis pela abertura de novas áreas de pesquisa, comparando-os a esse possível peso maior que os metodológicos teriam. Motta (1983) revela que o assunto possui opiniões antagônicas. Se por um lado, há autores que defendem que os trabalhos sobre métodos são mais frequentemente citados, outros afirmam que isso ocorre raramente.

Outro problema é relacionado a citação de pesquisas que tratam de descobertas prematuras, também chamado Efeito Mendel. Isso ocorre quando trabalhos pioneiros não recebem citações pelo grau de inovação e contradição e

distanciamento do paradigma vigente. Apesar do estudo ser inovador, não é de domínio popular ou objeto de um grande grupo de pesquisa, fato que reduz sua possibilidade de ser citado e ter destaque dentre os pares.

Consta ainda a chamada obliteração, quando o impacto de um autor em uma área é tão importante e suas ideias tão aceitas, que já fazem parte do paradigma vigente e a citação chega a ser considerada desnecessária. Por outro lado, Garfield (1979) e Motta (1983) afirmam que isso é superado em pesquisas que comprovam, por exemplo, o destaque dado ao trabalho de Einstein (e, conseqüentemente, citações), mesmo que suas ideias já sejam um consenso na comunidade científica. Levitt e Thelwall (2008) apontam que, apesar das citações a trabalhos antigos (mesmo que nas condições apresentadas) se reduzirem com o passar dos anos, elas ainda apresentam significância nos índices.

Garfield (1979) apresenta também a alegação de que as citações não recebem uma variação ou peso, ou seja, são sempre objetivamente contadas para a geração da análise. O fato despertou interesse de pesquisadores, tendo em vista a diferença do prestígio e reconhecimento dos periódicos em que os artigos são citados. Supostamente, uma diferença de peso entre eles daria maior precisão à análise de citação. Nicolaisen (2007) corrobora ao expor a necessidade de uma qualificação da referência, ainda que o próprio Garfield (1979) tenha comprovado, na época, que a representatividade seria ínfima já que o grupo de periódicos de alto nível era pequeno para alterar os resultados. Na métrica chamada *Eigenfactor*<sup>15</sup> a citação é qualificada (ou “balanceada”) baseando-se no prestígio do periódico citante; se for um periódico muito citado, o autor do artigo citado terá maior impacto (melhores resultados) do que se fosse citado por outro periódico (PRINGLE, 2008). A solução comprova a necessidade de considerar o aspecto apesar das conclusões de Garfield (1979).

Consta também a apresentação de críticas sobre a variação de resultados entre as áreas da Ciência. Enquanto instrumento de avaliação, a análise de citação foi vista como útil já que permitia um método unificado por meio da contagem das citações. Ao comparar resultados entre as grandes áreas de estudo, notou-se a incompatibilidade da comparação, pelas características de comportamento da comunidade científica. Nesse sentido Garfield (1979) assume o problema, afirmando que não é possível usar a análise de citação para diferentes campos simultaneamente,

---

<sup>15</sup> Mais informações no capítulo 6.3.3 “Outros Indicadores”.

tendo em vista: o tamanho da área de estudo, o grau de integração e a idade da literatura, ou seja, o que é relacionado ao potencial de ser citado. A média de citações por artigo publicado, em uma determinado campo, seria o índice mais preciso que o potencial da análise de citação pode gerar.

É possível acrescentar outras temáticas a respeito da validade da análise de citação, além daquilo que foi respondido e discutido pelo criador do método. Nicolaisen (2007) afirma que a citação não pode ser vista como um elemento objetivo e sem falhas. Com relação aos erros técnicos ao citar, ele aponta uma média de ocorrência de 20% mesmo entre diferentes áreas da ciência até o final do século XX. O autor explica que, ainda, há carência de estudos práticos e empíricos na área que possam comprovar ou refutar as teorias sobre a citação.

Bornmann e Daniel (2008, p. 46-47) elaboraram uma síntese a respeito das discussões sobre as metrias em ciência, identificando uma série de fatores que alteram a possibilidade de um autor ser citado e, portanto, influenciam o uso da citação na elaboração de indicadores. Parte deles coincide com o posicionamento de Garfield (1979) Porém, os autores supracitados inovam ao criar uma tipologia para o tema, conforme pode ser visto no Quadro 2.

O uso da análise de citação pressupõe compreender e relevar todas as críticas e dificuldades apresentadas nas esferas técnica e comportamental, com influências culturais, políticas e econômicas. Ainda assim, o conceito que as vezes é tratado como “qualidade”, consolida-se, e, conforme corrigido repetidas vezes por seu precursor, limita-se a expressar o quão útil um trabalho é e, portanto, impactante em uma comunidade científica (GARFIELD, 1979).

Fatores relacionados a	Observação
Temporalidade	<ul style="list-style-type: none"> <li>• Na medida que a ciência segue crescendo exponencialmente, a velocidade do ingresso de novas obras aumenta, e as citações se concentram cada vez mais em trabalhos recentes.</li> <li>• Quanto mais citado é um trabalho, maior a probabilidade de continuar sendo citado no futuro, o que pode ser verificado por indicadores dedicados a projeções.</li> </ul>
Campo da Ciência	<ul style="list-style-type: none"> <li>• As práticas de citação variam entre as Ciências e as Ciências Sociais e também, entre os próprios subcampos de estudo.</li> <li>• Os campos menores costumam atrair menos citações do que os mais generalistas.</li> </ul>
Periódicos	<ul style="list-style-type: none"> <li>• A quantidade de vezes que um periódico é citado (e portanto medido seu impacto) depende da frequência com que é publicado.</li> <li>• A própria ordem dos artigos dentro de um periódico pode alterar a possibilidade de um trabalho ser citado, sendo que os primeiros estão mais propensos dos os últimos.</li> <li>• Acesso, a visibilidade e a internacionalização do periódico influenciarão sua possibilidade de ser citado.</li> </ul>
Artigo	<ul style="list-style-type: none"> <li>• O modo como é utilizada a citação em artigos (e suas variações), livros e outros canais, difere consideravelmente.</li> <li>• As publicações mais longas costumam ser mais citadas que as curtas.</li> </ul>
Autor-Leitor	<ul style="list-style-type: none"> <li>• O idioma e as barreiras culturais influenciam a probabilidade de citações, assim como as relações dentro de uma rede de pares e o próprio gênero do autor a ser citado.</li> <li>• As mulheres são comprovadamente menos citadas.</li> </ul>
Disponibilidade	<ul style="list-style-type: none"> <li>• O acesso físico e/ou eletrônico de uma publicação influencia a possibilidade de ser citado.</li> </ul>
Problemas Técnicos	<ul style="list-style-type: none"> <li>• A publicação incorreta de citações é um problema comum que afeta a elaboração dos índices e, por consequência, a correta contabilização.</li> <li>• Há distintas regras para a citação utilizadas pelas instituições de pesquisa e ensino.</li> <li>• Ocorrência de erros relacionados à grafia do nome dos autores e seus homônimos.</li> </ul>

QUADRO 2: FATORES QUE ALTERAM A POSSIBILIDADE DE SER CITADO

FONTE: Bornmann e Daniel (2008, p. 46-47)

Para Vanz e Caregnato (2003), a contagem de citações recebidas acaba sendo inegavelmente uma indicação de prestígio de um pesquisador em seu espaço. Nicolaisen (2007) defende que, apesar das tantas discussões sobre as teorias da citação, comprova-se, em uma grande amostra, que as críticas não se sustentam o suficiente para negar a utilidade. O autor explica que ainda há carência de estudos práticos e empíricos na área que possam comprovar ou refutar as diversas teorias sobre a citação. Outro argumento, também apresentado por outros autores, é a comparação entre a tradicional avaliação por pares e a análise de citação, que frequentemente apresenta resultados semelhantes (GARFIELD, 1979; BORNMANN; DANIEL, 2008)



Ainda que o método, de modo generalizado, possa ser considerado ideal nas derivações dos estudos de citação, é importante discutir as fontes dos dados, nas quais o método é aplicado, sendo essa uma área de estudo a parte a ser discutida no tópico seguinte.

### 6.2.1 A Thomson ISI, as bases de dados e a análise de citação

O método de análise de citação e os estudos subsequentes se popularizaram com o surgimento do ISI e consequente elaboração dos seus serviços informacionais, sendo o SCI o precursor. Com esse, foi possível mensurar o impacto de publicações e autores, sendo aceita na definição de relevância em certas áreas de estudo. Ainda assim, ressalta-se que a intenção de criar um amplo índice de citação foi a melhoria da busca e recuperação da informação científica, obtendo resultados “melhores” em função dos critérios utilizados para indexação (seleção de periódicos). No que diz respeito a avaliação da ciência, a aplicação foi posterior pela enumeração de cientistas envolvidos em uma certa área de pesquisa, ou mesmo a relação de todos os trabalhos de um pesquisador (GARFIELD, 1963; SMITH, 1981).

Se por um lado a expansão do método de análise de citação e a ascensão do SCI e posterior *Web of Science* começaram a tomar espaço no ambiente científico, trazendo benefícios e respostas para questões de valor, impacto, qualidade e significância; por outro, houve comprovação de problemas e limitações que afetam a validade dos indicadores (SMITH, 1981; MEHO; YANG, 2006; FINGERMAN, 2006).

As primeiras críticas surgiram pela existência de uma polarização no SCI, ao privilegiar as Ciências, excluindo a produção das Ciências Sociais, Artes e Humanidades. Em resposta, sucederam-se a criação de duas novas bases: *Social Sciences Citation Index* (SSCI) e *Arts and Humanities Citation Index* (A&HCI) (GARFIELD, 1974; SMITH, 1981). Ainda assim, autores como Chubin (1974<sup>16</sup>), criticaram a criação do SSCI, argumentando que, ainda que existisse uma separação por área, os índices continuavam muito quantitativos e supostamente iriam promover a prática de citar por parte dos pesquisadores, o que, segundo ele, não ocorreria (ao tomar como exemplo a pesquisa em Sociologia).

---

<sup>16</sup> Resposta publicada por Garfield (1974).

As diferenças entre as áreas do conhecimento ficaram mais acentuadas nessas bases com o passar dos anos, ampliando as dificuldades de uso do sistema de mensuração proposto com base da *Thomson ISI* em todas as ciências. É possível citar entre os principais *caveats* da metodologia e critérios em relação aos indicadores de produção científica na *Thomson ISI*: o tempo de uma publicação à outra; os canais de comunicação (esse sistema utiliza apenas os periódicos científicos e desconsidera livros, comunicação em eventos, artigos, resumos e apresentações); a questão idiomática (preferência pelo idioma inglês); e a cobertura (apenas parte dos periódicos existentes estão inclusos) (BAUER; BAKKALBASI, 2005; NEUHAUS; DANIEL, 2008).

Bollen e Van de Sompel (2008) verificaram que os resultados de fatores de impacto seriam semelhantes se aplicados em comunidades diferentes à cobertura da *Thomson ISI*, haja vista as dificuldades apontadas anteriormente quanto ao uso desses dados. Eles verificaram que o indicador pode ser aplicado a escopos menores, no sentido de refletir apenas uma realidade científica pontual, como em repositórios institucionais, por exemplo.

Nicolaisen (2007) afirma que, com o passar dos anos o SCI acabou se tornando um marco na criação de uma forma objetiva de gerar indicadores sobre a produção. A entrada paliativa desses dados tornou-os aceitos no nível da sociedade, tanto como ferramentas para elaboração de políticas e administração da ciência como na esfera interna das áreas, pela dificuldade de legitimar o controle de qualidade em certos segmentos. Nesses casos os indicadores foram acrescentados à própria revisão por pares. Os produtos decorrentes do índice, também oferecidos pela *Thomson ISI*, se tornaram referência mundial na medição de uso e impacto da ciência.

Em 2004 é criado um índice concorrente ao *Web of Science*, com proposta semelhante e impacto nas políticas científicas, chamado *Scopus*, sob propriedade da *Elsevier*. De acesso pago, também inclui índices de citação, trazendo como diferenciais a indexação de periódicos de acesso aberto além dos pagos e a inclusão de apresentações de trabalhos científicos, *sites* e alguns livros.

Mais recentemente, no final de 2008, a *Thomson ISI* adicionou dois novos índices de citação, inovando e promovendo uma revisão nas principais críticas sobre a rigidez do sistema: o *Conference Proceedings Citation Index-Science* (CPCI-S), para as ditas Ciências, e o *Conference Proceedings Citation Index – Social Science & Humanities* (CPCI-SSH), para as Ciências Sociais e Humanidades. Bar-Illan (2010) verificou o impacto do CPCI-S nas disciplinas da Ciência da Computação, uma área

em que boa parte da produção científica ocorre por meio de publicação em eventos. A autora concluiu que a inclusão do CPCI-S é relevante e altera consideravelmente os resultados da quantidade de citações recebidas na área, exemplificando que na amostra escolhida, 40% das citações eram referentes a *proceedings*.

Larsen e Ins (2010) visualizam a expansão da *Thomson ISI*, tanto no grau de ampliação da cobertura dos periódicos quanto com a criação dos índices anteriores, como uma solução parcial e insuficiente para os principais problemas do sistema. Para os autores, a produção científica segue em crescimento apresentando características novas, como uso de canais distintos aos tradicionais (ex.: anais de eventos, páginas pessoais de pesquisadores e sistema de publicação de arquivos abertos). O CPCI-S e CPCI-SSH seriam um reflexo parcial e lento da necessidade de mudança, baseada do monopólio histórico da *Thomson ISI*.

No final de 2010, o vice-presidente da *Thomson Reuters*, James Testa, anunciou um novo índice a ser comercializado em 2011, o *Book Citation Index*. Segundo ele, após o ingresso do CPCI-S e CPCI-SSH e, com base na experiência e reconhecimento obtido pelo *Web of Science*, a empresa deseja expandir para o mundo dos livros, considerados de grande importância nas Ciências Sociais e Humanas. O critério de seleção será a conformidade com padrões e conteúdo editorial, um paralelo ao que já acontece com os periódicos. Assim como o índice de *Proceedings*, ele será um produto a parte, ainda que as buscas possam ser feitas de maneira conjunta (entre os vários índices) (TESTA, 2010).

As mudanças na *Thomson ISI*, o surgimento da *Scopus*, a eminência de novas ferramentas como *Google Scholar* e o movimento do acesso aberto integram um contexto no qual há um processo de popularização e democratização das métricas, conforme já previa Pringle (2008). Nesse sentido, faz-se importante a descrição e consequente detalhamento dos indicadores de uso e impacto da ciência, bem como seus vieses e potencialidades de aplicação.

## 6.2.2 Produtos da Thomson ISI

Ao final deste tópico, são apresentados produtos da Thomson ISI, relacionados às técnicas de avaliação da performance na pesquisa científica, com base em indicadores bibliométricos. Os índices são coincidentes ou não com os apresentados na sequência do trabalho, em tópico dedicado.

Os dados utilizados para a geração de indicadores bibliométricos são provenientes dos índices de citação que integram a base *Thomson Reuters Web of Science: Science Citation Index Expanded, Social Science Citation Index, Century of Social Sciences, Arts and Humanities Citation Index, Conference Proceedings Citation Index, Index Chemicus e Current Chemical Reactions*, com mais de 10.000 periódicos nas várias disciplinas da ciência. Esclarece-se que o índice não se limita a artigos, incluindo editoriais, cartas, revisões de livros e *proceedings* de eventos (THOMSON REUTERS, 2011a). De acordo com a empresa, não existem indicadores ideais para todas situações, sendo necessário definir o intuito da avaliação para que os dados e medidas corretos sejam aplicados. Portanto, divide-os em:

- a) Produtividade:** contagem de artigos ou pesquisadores;
- b) Reconhecimento total e influência:** contagem de citações, índice-h,
- c) Reconhecimento indireto e influência indireta:** segunda geração de citações (soma das citações presentes em todos artigos que citam um dado artigo);
- d) Eficiência:** média de citações por artigo, porcentagem de artigos citados ou não citados, fator de impacto de periódicos;
- e) Impacto relativo:** porcentagem de artigos citados ou não de acordo com um universo definido (ex.: instituição), linha de base das áreas da ciência (o valor médio de citações por artigos de determinadas áreas), expectativa de citações;
- f) Especialização:** colaboração (valores de coautoria entre unidades de análise); índice de disciplinas (expressa o quão multi ou interdisciplinar são os trabalhos); frentes de pesquisa (grupo de artigos muito citados em formato de *cluster*).

Apesar das variações e objetivos ao selecionar um índice, a Thomson Reuters (2008) reforça que, de modo geral, eles são utilizados como instrumento de tomada de decisão por fundações e órgãos ligados a políticas públicas de fomento a pesquisa, como análise e representação da performance na pesquisa científica. Assim sendo e,

conforme explicado anteriormente, as iniciativas voltadas para a medição objetiva dos resultados proporcionados pela ciência, são do interesse tanto dos que dependem de recursos financeiros para o desenvolvimento, quanto os que devem destiná-los.

### 6.3 AS MÉTRICAS APLICADAS NA AVALIAÇÃO DA CIÊNCIA

As primeiras aplicações das citações enquanto indicadores são atribuídas a Eugene Garfield e Derek Solla Price durante a década de 1970, ao sugerirem que a contagem de citações poderia expressar o impacto e importância de determinado trabalho (MOTTA, 1983).

Nota-se que, de maneira geral, até o final do século XX, foi a posição dominante da *Thomson ISI*, enquanto fonte de informação sobre as citações, o fator determinante para o sucesso e o estabelecimento dos principais indicadores do uso e impacto na ciência. Para Pringle (2008), a melhoria de acesso aos dados e a convergência dos formatos de papel para o eletrônico facilitaram os estudos referentes à criação e uso de indicadores que, até o período anterior, não apresentavam complexidade, com destaque maior para o *Journal Impact Factor*, da própria *Thomson ISI*. Um exemplo é apontado por Vanz e Caregnato (2003), como a retomada de pesquisas sobre Bibliometria no Brasil no início do século XXI, haja vista a necessidade de estatísticas sobre a produção científica nacional e o eminente crescimento de uma biblioteca digital com critérios e parâmetros semelhantes aos da *Thomson ISI*. É possível comprovar a evolução dos indicadores pela coleta de materiais do período anterior à sua popularização e democratização (PRINGLE, 2008).

#### 6.3.1 Contagens Básicas

Uma das abordagens iniciais sobre indicadores tratava da análise de pertinência e uso de conteúdo científico em função do tempo. Para tanto, buscavam-se regularidades quanto ao comportamento de citação dos autores em relação a trabalhos anteriores, identificando um determinado espaço no tempo no qual um conteúdo ainda é relevante.

Meadows (1999, p. 224) argumenta que essa condição pode ser confundida com a crescente evolução do número de artigos científicos publicados, explicando o fato do uso da literatura mais recente. Trata-se, no entanto, da verificação em um determinado espaço de tempo (na Física, por exemplo, são 4,6 anos) da frequência ou aparecimento de citações de uma determinada publicação. O indicador utilizado é denominado “meia-vida”, uma analogia ao conceito da radioatividade que, aplicado à Ciência da Informação, retrata o período de tempo durante o qual metade da literatura citada no período vigente, foi publicada. O estudo também se baseia no armazenamento e processamento de referências, utilizando-as como uma forma de reforçar a aceitação delas como critério para definir se um material produzido está sendo utilizado.

Price (1976c) aponta as bibliotecas como uma fonte alternativa para os estudos de obsolescência na ciência. Para ele, o número de consultas a edições de periódicos nesses locais também pode embasar um indicador que aponte durante quanto tempo um determinado conteúdo é relevante (pelo uso) e, se apesar da unidade de análise ser maior (a edição inteira de uma revista ao invés do artigo em si), apresenta resultados semelhantes aos citados anteriormente.

Na obra de Braun, Glänzel e Schubert (1985, pp. 42-49), os autores apontam doze principais medidas da ciência para avaliar a pesquisa em nações.

- a) número de primeiros autores;
- b) número de publicações;
- c) distribuição de publicações de acordo com o subcampo científico;
- d) número de trabalhos não citados;
- e) parcela de trabalhos não citados;
- f) número de trabalhos muito citados;
- g) parcela de trabalhos muito citados;
- h) número de citações recebidas por país ou disciplina;
- i) média esperada de citações;
- j) relação entre as citações recebidas e as esperadas;
- k) média de citações no ano por país ou disciplina;
- l) relação entre o número de citações dividida pelo número de periódicos.

Para Macias-Chapula (1998), os indicadores da ciência mais conhecidos seriam o número de trabalhos, o número de citações, a coautoria, o número de patentes, o número de citações de patentes e os mapas dos campos científicos e dos

países. Meho e Rogers (2008) afirmam que os principais indicadores baseados em citação são a simples contagem e *ranking* por autor, o número de citações por documento, o fator de impacto e o índice-h.

Os dados para análise de citação não são os mesmos para a geração de todas as metrias, variando de acordo com o propósito. O ideal é a definição do objetivo da avaliação, os tipos de informação necessária e como isso será utilizado em análise futura.

### 6.3.2 Principais indicadores

Os indicadores a seguir são considerados os mais utilizados e possivelmente foram os mais discutidos desde o surgimento do primeiro deles, o fator de impacto (FI), na década de 1970. Os indicadores subsequentes apresentaram alternativas relativas ao impacto, derivando e objetivando minimizar as falhas apresentadas pelo FI. Bar-Ilan (2008) define como indicadores o índice-h, o FI e “outros indicadores”. Bollen *et al.* (2009) também consideram essas medidas e suas derivações como dominantes do campo da avaliação científica.

#### a) Fator de Impacto

O conceito de FI surgiu na década de 1950, durante a publicação dos primeiros índices de citação, o *Genetics Citation Index* e, precedido pelo SCI, já na década de 1960. O indicador de impacto de periódicos passou a ser publicado anualmente no “*Journal Citation Reports*”, no qual as publicações eram listadas de acordo com FI, devidamente classificadas por subáreas da ciência (BORGMAN; FURNER, 2002; GARFIELD, 2006).

Thomson Reuters (2008, p. 10), provedora do indicador, o define como uma contagem média de citações recebidas por um periódico. O FI é baseado no *Web of Science*, publicado no *Journal Citation Reports* (JCR), com objetivo de comparar o impacto entre periódicos. O uso do FI estaria voltado para autores, no sentido de que pudessem identificar os periódicos de maior prestígio para submeter seus trabalhos.

Para calcular o valor do FI de periódicos, Bollen *et al.* (2009) indicam a fórmula adotada pela *Thomson ISI*:

$$IF_p = \frac{\text{citações recebidas no ano } A \text{ por artigos publicados pelo periódico } p \text{ em } (A-1) \text{ e } (A-2)}{\text{artigos publicados pelo periódico } p \text{ em } (A-1) \text{ e } (A-2)}$$

O FI é baseado em dois elementos de contagem. O numerador é o número de citações recebidas no ano (o qual está sendo relatado), por artigos publicados nos dois últimos anos. O denominador é o número de artigos publicado nos mesmos últimos dois anos. Por exemplo: para calcular o FI de 2010 do periódico  $p$ , é necessário contabilizar as citações recebidas por ele em 2010, de artigos publicados em 2009 e 2008, formando o numerador. O denominador será o número de artigos publicados pelo periódico  $p$  em 2008 e 2009 (DONG; LOH; MONDRY, 2005; GARFIELD, 2006).

O FI tem uma posição dominante entre as medidas de impacto científico, tanto pelo reconhecimento e renome da fonte, os índices da *Thomson ISI*, quanto pela possibilidade de ser aplicado e interpretado além dos limites dos periódicos, seja enquanto a unidade for autores, instituições e até nações (BOLLEN *et al.*, 2009).

O uso do FI pode ser entendido de acordo com seus públicos. Para os autores, é importante saber quais periódicos têm maior prestígio e, portanto, poderão lhe proporcionar maior reconhecimento, caso tenham um artigo aceito. Nas bibliotecas, o FI é uma ferramenta de suporte à tomada de decisão na seleção das coleções mais adequadas para aquisição, ou seja, aquelas que detém o maior prestígio e, provavelmente, serão mais utilizadas. Os editores de periódicos acompanham a evolução de importância dada à certos temas com o FI, selecionando e direcionando o editorial para que seja mais atrativo. Os gestores de órgãos de fomento são atores principais no uso do FI valendo-se do mesmo para obtenção de dados objetivos que embasem o desenvolvimento de políticas públicas de direcionamento de recursos para pesquisa (GARFIELD, 2000, 2006; STREHEL, 2005).

Embora aceito em muitas dessas condições, o FI também apresenta debilidades, o que incita a derivação de estudos sobre o tema. Bordons, Fernández e Gomes (2002) apontam que os resultados do FI serão afetados inicialmente pela área da ciência, o tamanho da janela de cobertura das citações e o tipo do documento. Há disciplinas em que os autores publicam com menor periodicidade e as citações também ocorrem assim, o que, segundo os dois primeiros fatores, irá distorcer os resultados. Esses autores também estudaram a questão da diferença entre nações. Cada país tem uma orientação geral em relação à pesquisa. Por exemplo: a Inglaterra



está mais orientada para as Ciências da Vida, enquanto a Alemanha volta-se para as Ciências da Tecnologia, fatores que distorcem um resultado global comparativo. Apontam também as restrições do *Web of Science* enquanto fonte, como a incipiente cobertura nas Humanidades e, a dificuldade que periódicos de países não-centrais têm em ingressar neste índice, apontando uma produção subestimada. Inclusive, se um periódico nessa condição é aceito na base, o FI costuma ficar muito baixo pelo pouco tempo de ingresso – que deveria ser corrigido pelos analistas. Uma possível solução para o problema seria a elaboração de um índice nacional (espanhol, no caso da pesquisa) de citações e usar o FI somente neste universo, apesar das dificuldades técnicas em elaborá-lo.

Atentando a essas dificuldades, percebe-se o surgimento de outros índices com fatores de correção para aprimorar o FI e consolidar o uso das metrias na ciência. Teriam estes a ampliação ou redução da janela temporal, ou mesmo uma adaptação do índice de meia-vida aliado ao FI. Ainda cabe citar o índice-h e o índice-g, detalhados na sequência (BOLLEN *et al.*, 2009).

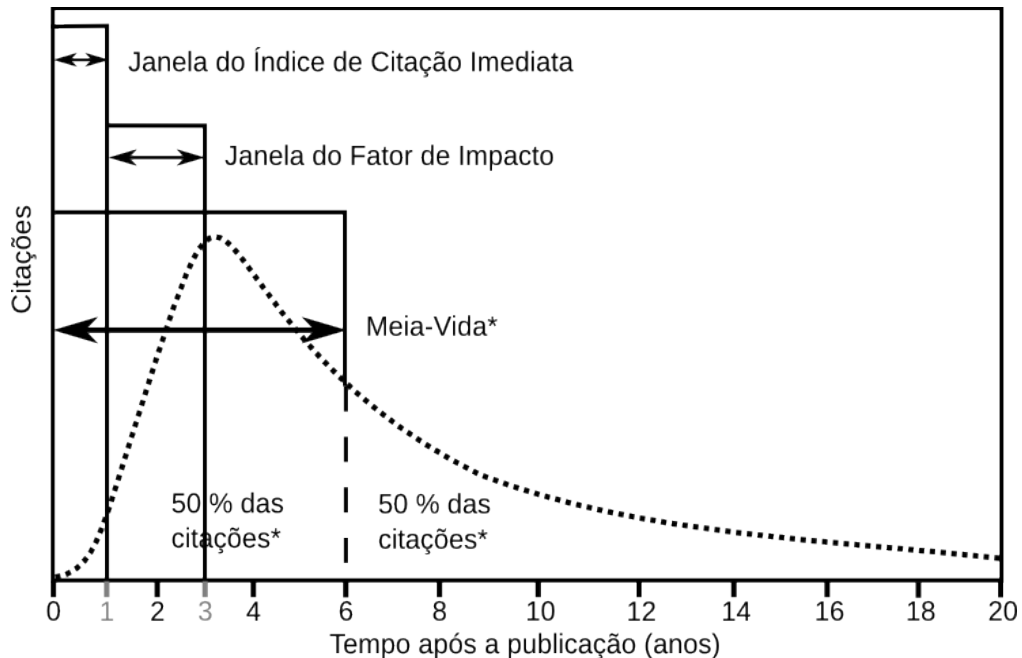
O FI pode não ser considerado ideal para medição de impacto da qualidade de artigos, porém, segundo a visão de Hoeffel (1998, p. 1225), é possivelmente o melhor em existência e, uma boa técnica de avaliação científica. Uma das comprovações é o fato de coincidir, no caso dos periódicos mais bem colocados, com aqueles em que há maior dificuldade de os autores obterem aceite.

#### b) Índice Citação Imediata e Índice de Meia-Vida

No que diz respeito ao FI, uma das alegações principais quanto a sua validade consider o efeito do ritmo de obsolescência entre as diferentes áreas da ciência. Na elaboração do indicador, Garfield (1972) estudou e definiu a janela temporal que supostamente seria a ideal pois, a maior parte das citações seriam recebidas durante os dois primeiros anos de existência do artigo. Ainda assim, utilizar uma média é inadequado, uma vez que as áreas da ciência diferem nas proporções da idade das citações, ou seja, nas taxas de obsolescência.

Como consequência desse cenário, a *Thomson ISI* passou a publicar instrumentos de medidas de obsolescência, baseados no modelo diacrônico (por meio de sucessivas observações realizadas em diversas datas) sendo eles: a velocidade de incorporação de novos conhecimentos e o ritmo de envelhecimento da literatura

(STREHEL, 2005). A relação temporal entre o FI, a meia-vida e o índice de citação imediata (referente à redução da janela temporal do FI) pode ser observada na Ilustração 3.



\*A curva de representação bem como os dados servem apenas para o entendimento do leitor sobre as relações estabelecidas, não revelando uma situação globalmente aplicável. A curva irá diferir de acordo com a disciplina da ciência estudada.

### ILUSTRAÇÃO 3: CURVA GENÉRICA DE CITAÇÕES E MEDIDAS DE IMPACTO

FONTE: Adaptado de Amim e Mabe (2003, p. 348)

No que se relaciona a velocidade de incorporação, o instrumento utilizado é o Índice de Citação Imediata, publicado pela Thomson ISI como *JCR Citation Immediacy Index*. De menor importância que o FI, é a medida do quão rápido, em média, os artigos de um periódico são citados. Para tanto, consideram-se as citações recebidas no ano que em o periódico foi publicado (BRAUN; GLÄNZEL; SCHUBERT, 1985; AMIM; MABE, 2003).

$$ICI = \frac{\text{citações recebidas pelo periódico no ano vigente}}{\text{artigos publicado no periódico no ano vigente}}$$

Este índice permite ainda identificar os periódicos que dão maior atenção às novidades da ciência. Existem algumas condições técnicas que podem distorcer e inutilizar o ICI, sendo o atraso de publicações, a frequência ou periodicidade, e a

velocidade na indexação. É por essas razões que ele tem um peso menor que o FI na avaliação do desempenho dos periódicos (BRAUN; GLÄNZEL; SCHUBERT, 1985; SANTOS, 2007).

O conceito de Meia-Vida (MV) é originado da física, baseado na ideia de meia-vida dos materiais radioativos. O ingresso na Ciência da Informação ocorreu na década de 1960, como uma forma estatística de calcular o fenômeno da obsolescência, sendo definida como: o tempo, em anos, em que metade de toda literatura ativa foi citada. O conceito de atividade está associada a ideia de ser citado. No caso do indicador da *Thomson ISI*, o conceito volta-se para a unidade de análise do periódico. Nota-se que se trata de um cálculo estimado, haja vista que, para obtenção do valor, não é necessário esperar até que a determinada publicação deixe de ser citada (EGGHE; ROUSSEAU, 2000).

Bensman (2006) explica que a publicação do índice de MV para periódicos iniciou no final da década de 1970, sob o pretexto de flexibilizar o uso do FI no relatório do JCR, tendo em vistas as diversas críticas, principalmente quanto a janela temporal e o comportamento em diferentes entre áreas. Desse modo, ficaria perceptível para o interessado no índice, que as citações podem ou não estar mais concentradas nos dois primeiros anos da publicação, pela qual o FI é calculado. O MV é passado a ser utilizado para entender e valorizar o tempo de existência de uma publicação.

De modo geral, é possível verificar que a MV é capaz de apontar o comportamento médio de citações em cada área, disciplina ou tópico, estabelecendo uma relação de comprovação ou negação do FI enquanto medida válida para ciência. Dada essa situação, Sombatsompop, Markpin e Premkamolnetr (2004) propõem o uso da MV como normalizador da janela de citação para cada área, com um indicador denominado *Cited Half-Life Impact Factor* (CHAL).

Pringle (2008) explica que em decorrência das críticas a respeito do FI e seus derivados, é proposto um índice que viria se tornar um dos principais instrumentos objetivos de avaliação, o índice-h, tendo como principal mudança a ideia de utilização de uma métrica no nível do autor e não mais do periódico.

### c) Hirsh Index (índice-h) e suas variações

O índice-h é considerado um indicador bibliométrico novo se comparado ao tempo de existência do método de análise de citação, sendo proposto em 2005 pelo físico Jorge Hirsch para quantificar o impacto individual da produção científica.

Objetivou também aprimorar e corrigir as deficiências associadas aos indicadores até então utilizados. Ao contrário da simples listagem de citações ordenada de modo decrescente, que pode ser facilmente influenciada por um ou mais documentos muito citados, o índice agrega tanto a quantidade quanto a qualidade do impacto das publicações e ajuda a distinguir os cientistas que publicaram o maior número de trabalhos muito citados (MEHO; ROGERS, 2008).

Em uma definição sucinta, Gálvez Toro e Amezcua (2006) sugerem que o índice- $h$  é aplicado a um pesquisador que têm  $h$  trabalhos que tenham sido citados pelos menos  $h$  vezes. Como forma de elucidar, apresenta-se a Tabela 1.

TABELA 1: EXEMPLO SIMPLIFICADO DE OBTENÇÃO DO ÍNDICE-H

Artigo	Citações recebidas	índice- $h$
1	30	1
2	25	2
<b>3</b>	<b>3</b>	<b>3</b>
4	3	4
5	1	5

FONTE: O autor (2011).

No exemplo, o pesquisador publicou 5 artigos ao total. O seu índice- $h$  será de valor 3 pois, do total de 5 trabalhos, apenas 3 obedecem à condição de ter pelo menos  $h$  citações, ou seja, 3. Hipoteticamente, se os artigos 4 ou 5 obtivessem pelo menos 4 citações, o índice- $h$  do pesquisador subiria para esse valor (4) pois, passaria a obedecer a condição.

Por uma questão de acesso e simplificação, há anos a ciência buscava um número áureo, um método quantitativo único com validade em todas as áreas, com capacidade de representação do valor de um autor e sua produção. Considerando o FI, a contagem de citações, o número de trabalho publicados, a avaliação por pares e o índice- $h$ , o último talvez seja o que mais se aproximou do ideal (GÁLVEZ TORO; AMEZCUA, 2006).

- Índice HB

Bar-Illan (2008) revela outros atributos do índice-h. É considerado dependente do tempo, por exemplo: quanto mais tempo um pesquisador está ativo, maior é a possibilidade dele obter um valor maior. Acrescenta que o índice-h não é aplicável apenas aos indivíduos. A autora cita outros estudos em que foi possível comprovar seu uso tendo periódicos como unidade de análise e, ainda, em grupo de pesquisas. Aponta ainda uma das variações do índice-h, o chamado índice-hb, voltado para avaliação de determinados tópicos na ciência.

Banks (2006), o criador do índice-hb, explica que o motivo de criar uma variação voltada para tópicos na ciência é a dificuldade que novos pesquisadores têm em se identificar com termos e temas que estão sendo muito ou pouco procurados. Com os resultados, é possível dizer se um tópico está em alta, se já esteve e continua ou mesmo se já se tornou obsoleto.

- Índice HC

No que diz respeito a questão temporal apontada, outra variação do índice-h surgiu, chamada “contemporânea” ou índice-hc. A diferença é o acréscimo de um peso relativo a idade do documento, com o objetivo de diminuir a importância daqueles trabalhos mais antigos, dado o caráter contemporâneo da ciência. Por exemplo, se o parâmetro definido fosse valor 4 para ano atual, uma publicação de 4 anos atrás seria de valor 1, e assim sucessivamente (SIDIROPOULOS; KATSAROS; MANOLOPOULOS, 2007).

- Índice HI

Outra dificuldade comumente relacionada à análise de citação está nos trabalhos com mais de um autor<sup>17</sup>. Como consequência, surge mais um indicador complementar ao índice-h para quantificar a produção científica individual. O chamado índice-hi (individual), indica o número de artigos que um pesquisador publicou sozinho em sua carreira, com pelo menos *hi* citações. Esse índice é tido como uma referência adequada para comparação entre disciplinas pois, ao normalizar o número de autores entre os trabalhos (atendendo à especificação de trabalhos individuais), supera um

---

<sup>17</sup> Mais informações em 6.1 “Os Estudos de citação”.

dos problemas de diferença entre o comportamento das áreas da ciência. O seu cálculo é dado pela divisão do índice-h original pelo número médio de autores que contribuíram para o índice (BATISTA; CAMPITELI; KINOUCI, 2006).

- Índice HI Normalizado

Ainda com relação ao índice-h individual, Harzing (2007) propõe uma extensão da solução proposta, o índice-hi-normalizado. Ao invés de dividir o índice-h total, a autora sugere primeiramente a normalização do número de citações para cada artigo, por meio da divisão delas pelo número de autores daquele artigo e, em seguida, efetua o cálculo com esses valores corrigidos. Segundo ela, esse índice atinge resultados mais precisos para os efeitos de coautoria.

- Índice HM

Uma terceira interpretação sobre índice-h no impacto da coautoria é dada por Schreiber (2008). O autor propõe o uso de um índice-hm, tendo como adaptação a contagem fracionada de artigos com base no número de autores. Com o estudo, comprovou uma diferença preocupante entre o índice-hm e o proposto por Hirsch, o que resulta em uma reflexão a respeito do quanto a coautoria distorce a medida do índice-h.

- Índice A, R e AR

Para compreender o índice-ar e seus antecedentes, é necessário resgatar o conceito de núcleo Hirsch ou, núcleo-h. Ele não é uma métrica e sim, uma definição de recorte utilizado para o cálculo de algumas métricas, tratando-se de todas as publicações que estão entre a obtenção do índice-h = 1 até o valor atingido pelo autor. Por exemplo: Se um autor tem 10 artigos publicados mas somente 6 atendem a caracterização do índice-h (os 6 já foram citados pelo menos 6 vezes), esses 6 pertencem ao núcleo-h (JIN *et al.*, 2007).

A inovação apresentada pelo índice-a é o fato de contabilizar o exato número de citações recebidas pelos artigos pertencentes ao núcleo-h, ao contrário do que ocorre no cálculo do índice-h. O índice pode ser definido como a média de citações recebidas por artigos que integram o núcleo-h. Ainda assim, verificou-se que em

algumas situações de distribuição, o índice-a não apresentava um autor com maior performance, o que motivou uma alteração na fórmula e a criação do chamado índice-r. A diferença foi o acréscimo da raiz quadrada sobre o resultado final.

Ao unir os “índice-a” e “índice-r”, com vistas a transpor o fato do índice-h não apresentar declínio, Jin *et al.* (2007) criaram o índice-ar ou, *Age-dependent indicator*. Os autores alegam que aqueles que estão sendo avaliados não podem permanecer sempre recebendo os méritos por trabalhos passados como ocorre com a utilização do índice-h. Sendo assim, a vantagem de usar o índice-ar está em reunir, ao mesmo tempo, o número atual de citações e; a idade das publicações. Ele não é um índice isolado e sim, visto como um complemento para o índice-h (dependendo, inclusive, do cálculo do antecedente para definição do núcleo-h)

- Índice AWCR, AWCRpA e AW.

Valendo-se do conceito anterior de cálculo de citações com peso para a idade da publicação referente, Harzing (2010) propõe algumas variáveis para o índice-ar. A primeira delas, o AWCR, poderia ser definido como o índice-ar com um escopo maior, ou seja, a autora sugere que o índice inclua todas as publicações, e não somente aquela integrantes do núcleo-h. Isso representaria o impacto do trabalho total do pesquisador de modo mais preciso, permitindo que artigos mais novos e pouco citados contribuíssem para o índice AWCR, e não somente aqueles contabilizando no índice-h.

Como forma de aproximar o índice proposto e o índice-h e facilitar a sua comparação, Harzing (2010) sugere o uso da raiz quadrada de AWCR, nomeando o índice de índice-aw. Tendo em vista os problemas relacionados aos trabalhos com múltiplos autores, a autora ainda propõe um terceiro índice, o AWCRpA, o qual normaliza o índice pela divisão das citações pelo número de autores na contabilização dos artigos.

Meho e Rogers (2008) apontam problemas do índice-h a serem assumidos no seu uso. Dentre eles está o fato de não considerar o total de citações que um autor tenha acumulado. Ele também não pode ser usado na comparação entre disciplinas diferentes, por apresentar variação nos resultados. Um exemplo é entre as áreas da Física, com autores que atingem o índice 50 ou maior; e da Ciência da Informação, na qual poucos autores atingiram o índice 15. Bar-Ilan (2008) generaliza a produção a

respeito do índice proposto por Hirsch, concluindo que a pesquisa nessa área ainda se encontra na “infância” e, existem perspectivas futuras para o desenvolvimento e amadurecimento do tema.

d) Egghe Index – índice-g

O índice-g foi proposto por Leo Egghe em 2006, a partir de uma modificação no índice-h. O propósito do autor foi proporcionar uma melhoria, ao atribuir ao cálculo com maior importância aos artigos muito citados (BOLLEN *et al.*, 2009).

Egghe (2006) explica que os artigos muito citados são também importantes para a determinação do índice-h, porém, quando um artigo entra no escopo dos *top* para o cálculo do índice-h, ele será utilizado apenas uma vez como determinante para o resultado. Assim sendo, as citações recebidas nos anos subsequentes não serão consideradas, mesmo que tripliquem. O autor apoia o fato do índice-h não utilizar os artigos pouco citados, mas reflete que a situação dos muitos citados deve ser considerada em termos de evolução da citação.

Costas e Bordons (2008) posicionam o índice-g como aplicado a um pesquisador que têm  $g$  trabalhos que tenham sido citados pelo menos  $g^2$  vezes. O indicador pode ser melhor compreendido por meio do exemplo na Tabela 2. No caso, o autor possuía pelo menos 4 artigos que foram citados 16 vezes ( $g^2$ ). Ele não poderia ter um índice-g superior a esse pois, não apresentava 5 trabalhos com pelo menos 25 citações cada.

TABELA 2: EXEMPLO DE IDENTIFICAÇÃO DO VALOR DO EGGHE INDEX

Artigo	Citações recebidas	Acumulado de Citações	$g^2$
1	60	60	1
2	45	105	4
3	25	130	9
<b>4</b>	<b>16</b>	<b>146</b>	<b>16</b>
5	8	154	25

FONTE: O autor (2011), com base em Egghe (2006, p. 113).

O índice-g apresenta duas importantes melhorias se comparado ao índice-h. Primeiro é o peso dado às citações recebidas que é considerado no cálculo; segundo, o índice-g de um determinado pesquisador não está limitado ou, determinado, pelo seu total de publicações. Costas e Bordons (2008) verificaram que a medida proposta por Egghe é mais adequada que o índice-h, para o tipo de pesquisador denominado como



seletivo. Suas características seriam o pequeno número de artigos publicados, com muitas citações recebidas. Para os outros tipos definidos na pesquisa: pouco produtivos, muito produtivos ou *top* pesquisadores, não houve tanta diferença ou representatividade do índice-g.

### 6.3.3 Outros Indicadores

Apesar do prestígio e uso intensivo dos indicadores apresentados até então, sendo eles os mais discutidos e estudados no método de análise de citação, existem outros identificados na literatura da área, não necessariamente derivados dos primeiros. Os elencados a seguir tratam, em resumo, de correções parciais nos problemas de citação; da adaptação do renomado algoritmo *PageRank* do Google para a avaliação na ciência e; relacionados a perspectivas futuras.

#### a) F-index

A intenção de criar o índice-f (*fractionated index*) foi a de solucionar, de maneira robusta, o problema das autocitações na contagem e elaboração de indicadores. Katsaros, Akritidis e Bozanis (2009) argumentam que a simples exclusão desse tipo de citação, conforme ocorre na maioria dos casos, não é o suficiente para definir que a medida está mais adequada, haja vista que este comportamento ao citar não é incorreto, salvo alguns casos que são usados para manipulação.

Para a definição do novo índice, foi necessária a introdução do conceito chamado *coterminal citations*, o qual expressa um padrão de citação em que um pesquisador foi coautor de vários artigos citantes de um outro artigo. A problemática é mais complexa do que o conceito de autocitação e atende pela identificação de tendências “inflacionárias” na comunicação científica que se refletem em exageros nas redes de citação. Em suma, o objetivo do índice-f foi desenvolver uma métrica capaz de não ser afetada pela existência de *coterminal citations*, ao contrário de simplesmente excluir a autocitação. No índice, é utilizado um cálculo que atribui valor e peso às citações de acordo com a sua posição na cadeia de coautorias.

Katsaros, Akritidis e Bozanis (2009) assumem que o índice-f é apenas uma pequena contribuição na discussão dos indicadores, sendo que há necessidade de evoluir nos procedimentos práticos e não somente éticos pela comunidade científica.

Um possível protocolo de caracterização de citações poderia ser adotado, por exemplo “autocitação, oposta, negativa”, o que facilitaria o desenvolvimento de novas métricas com maior precisão, desmistificando, de certo modo, a “panaceia” da Cientometria.

#### b) Y-Factor, Eigenfactor e Scimago Journal Rank

As discussões sobre o uso de algum peso nas citações são originadas da criação do FI, na década de 70. Uma das possibilidades seria a consideração da reputação do periódico de onde elas são originadas. Nesse sentido, surgem duas medidas com a intenção de usar o princípio para avaliar periódicos: o Fator Y e o Eigenfactor, ambos baseados no estudo de redes de citação (GARFIELD, 1979; NICOLAISEN, 2007; BOLLEN *et al.*, 2009).

O Fator-Y pode ser definido como um indicador híbrido entre popularidade e prestígio. Esses conceitos são frequentemente confundidos ainda que apresentem significado distinto. Utilizando a comunicação científica como analogia, exemplifica-se: um periódico *j* pode publicar muitos artigos de revisão e ser citado por estudantes de graduação e, ainda assim, ser ignorado por *experts* da área. Nesse caso, *j* obtém popularidade – é muito citado e será destaque no FI; porém, não detém prestígio, ou seja, apreciação de pares (BOLLEN; RODRIQUEZ; VAN DE SOMPEL, 2006).

Nos estudos a respeito da mensuração do prestígio, um destaque é dado àquele que também foi um marco no modelo de recuperação da informação, o algoritmo *PageRank*, aplicado a *web*. Se o conceito anterior de ranqueamento de páginas *web* baseava-se no somatório de *inlinks*<sup>18</sup>, Sergey Brin e Lawrence Page (fundadores do Google) introduziram a ideia de atribuir maior peso ao *links* originados de *sites* com maior qualidade, ou, segundo eles, aqueles com o maior número de *inlinks*. Em suma, o *ranking* estaria baseado no uso de pesos para as ligações, uma ideia análoga a atribuir peso às citações, que, de modo simplista, são entendidas como ligações (BAR-ILAN, 2007).

Bollen, Rodriguez e Van de Sompel (2006) atribuem o mérito de Brin e Page ao fato de criarem uma forma de avaliação que contabilizasse tanto a popularidade (no que diz respeito ao somatório de *links*) quanto o prestígio (uso do *status* para normalizar o *ranking*). Os autores sugerem que a situação e as dificuldades atuais

---

<sup>18</sup> Para mais informações a respeito da tipologia de *links*, ver tópico 5.4 “Webometria”.

apresentadas pelo FI deveriam ser contornadas pela adoção de um novo método de ranqueamento, o Fator-Y, baseado na integração entre o FI e um *PageRank* aplicado a uma rede de citação.

As discussões a respeito de um *PageRank* para os periódicos se aprofundaram até o surgimento de um serviço baseado na medida, o chamado Eigenfactor. Iniciado em 2006, disponibiliza essa nova métrica sob o pretexto de apresentar um *ranking* e mapeamento do conhecimento científico. Sendo de caráter gratuito, tem o Web of Science como fonte de dados, com os seguintes diferenciais: o uso do algoritmo do *PageRank*, a janela de citação de 5 anos, normalização da diferença de citação entre áreas e presença de teses e revistas populares. Os resultados são providos por meio de diagramas de redes e gráficos de evolução do valor Eigenfactor, comparado ao FI (FRANCESCHET, 2010).

Outro projeto baseado no algoritmo do *PageRank*, com vistas à avaliação da visibilidade de periódicos e nações, é o *Scimago Journal & Country Rank*. De autoria de um grupo de pesquisas espanhol, possui acesso aberto aos resultados dos *rankings*, tendo como diferencial a utilização da *Scopus*. Possibilita o recorte por áreas de pesquisa, disciplinas e países, com definição variável da janela de citação.

### c) Índice ECR e Índice C

Além dos indicadores que permitem estabelecer medidas de acordo com as citações recebidas, também existem aqueles relacionados a projeções, como é o caso do *Expected Citation Rate* (ECR). Ele indica o quanto um artigo tem possibilidade de ser citado baseado no ano de publicação, no periódico e na tipologia. Para tanto, utiliza-se um cálculo com base na comparação de outros artigos que apresentem características semelhantes e, portanto, seguirem um comportamento que possivelmente irá se repetir. Essa métrica também pode ser aplicada a grupos de pesquisa, somando o respectivo ECR de cada pesquisador (BRAUN; GLÄNZEL; SCHUBERT, 1985; THOMSON REUTERS, 2008).

Ao dividir o somatório das citações já ocorridas pelo índice-ECR é possível obter o chamado *Crown Index*, ou índice-c, também considerado um índice de desempenho, uma vez que apresenta a relação entre os resultados atuais e as perspectivas de um autor ou instituição. No caso da publicação da *Thomson ISI*, é incluída uma normalização entre áreas da ciência para que as comparações possam ser feitas entre eles.

Verifica-se que desde o início do presente século houve uma expansão significativa de tipos de indicadores, atendendo a cada momento necessidades específicas, reforçando a posição da Thomson Reuters (2008) sobre a personalização com base em necessidades especiais de aplicação. No capítulo seguinte são apresentadas as características das CHSA em sua produção científica, no sentido de permitir a construção de relações entre o enfoque dos indicadores descritos, e o comportamento nas publicações da área, conforme será discutido no final do capítulo 7 .

## 7 PRODUÇÃO CIENTÍFICA EM CHSA

No presente capítulo serão apresentadas as características da pesquisa em Ciências Humanas e Sociais Aplicadas enquanto uso de canais de publicação, métodos de avaliação, obsolescência e indicadores de produção. Ao final, propõe-se uma discussão por meio da relação entre este comportamento em CHSA e as possibilidades de indicadores apresentados no capítulo anterior.

Meadows (1999) destaca que a própria ideia de “pesquisa” nas Humanidades é contestada por parte dos cientistas de outras áreas. As alegações estão relacionadas aos critérios de cientificidade e, principalmente, afastamento do uso de métodos lógicos e tradicionais, conforme posiciona Price (1976a). Em contraposição, Meadows (1999) referencia o importante trabalho de Thomas Kuhn sobre a evolução das Ciências, como forma de legitimar e explicar o porquê da discussão de cientificidade em áreas como as Humanidades. Ele afirma que são dois os principais desafios do cientista no contexto evolutivo: diferenciar o conhecimento científico dos outros e compreender que a investigação científica está tão atrelada a aspectos sociais como qualquer outro tipo de pesquisa. Este último aspecto é essencial para compreender como os pesquisadores nas CHSA percebem a criação e divulgação do conhecimento. A diferença é que o ator (o “ser”) não é mais visto somente como uma unidade abstrata de análise, mas sim como alguém com emoções, experiência, conflitos e situado em um contexto histórico, cultural e sociopolítico.

A pesquisa nessas áreas se caracteriza por, em muitos casos, não haver uma separação clara entre o pesquisador e o objeto, como sugerem as outras ciências. Os estudos não são prioritariamente quantitativos, ou seja, não há uma busca pela determinação de leis da natureza. Desse modo, a própria ideia de evolução das Ciências proposta por Thomas Kuhn, em que ocorrem mudanças fundamentais que remetem a um direcionamento total de estudos, deve ser repensada (EUGÊNIO; FRANÇA; PEREZ, 1996). Ou seja, as CHSA apresentam novas abordagens para seus problemas sem que obrigatoriamente isso afete a permanência de estudos anteriores e os torne obsoletos. Essa substituição, alinhada ao pensamento de Thomas Kuhn, tem maior comprovação nas Ciências Exatas e da Vida.

Meadows (1999) afirma que o caráter quantitativo da pesquisa nas Ciências traz consequências tais como o maior emprego de recursos, o que justifica o fato de conseguirem um percentual maior de financiamento externo. Este, por sua vez,

baseia-se na maior relevância na publicação de resultados e, portanto, na maior propensão que mais conteúdo científico apareça em periódicos das Ciências do que nas demais áreas. A disponibilidade de recursos também causa impacto nos espaços para publicação no formato de revista. Esse tipo de canal custa dinheiro (mão de obra especializada) e, portanto, áreas com poucos recursos tem também pouco espaço para poder publicar, o que remonta à questão do uso de outros meios de publicação que não o periódico científico.

Lariviere *et al.* (2008) ressaltam que nas CHSA é mais comum que as pesquisas tenham enfoque local, com resultados que tem maior impacto e representatividade em um estado ou região. Esse fator afeta a cobertura das publicações em países com idioma principal diferente do inglês.

O conteúdo da pesquisa em CHSA é mais extenso e sua apresentação nem sempre é contemplada pelo formato de artigo de periódico. Isso tem como implicação uma velocidade de publicação mais lenta, na qual o formato livro acaba sendo mais adequado para atender a característica (MEADOWS, 1999).

O autor, quando se refere às diferentes fontes de financiamento para viabilizar o ingresso de um material científico dentro do processo de produção, priorizou aquele em formato impresso. Conseqüentemente, identifica que somente uma mudança no contexto geral poderia alterar esse quadro. No caso, seriam as facilidades associadas a expansão da rede mundial de computadores. No final de década de 1990 se percebia ainda a dificuldade por parte de alguns pesquisadores em se adaptar à nova realidade imposta pelo modelo eletrônico, em relação as competências necessárias para operar os sistemas. Outra limitação seria a de infra-estrutura, como por exemplo, recursos para compra de equipamentos, afetando mais a área de Humanidades devido à quantia empregada pelos órgãos financiadores.

O ingresso das áreas no modelo de acesso livre eletrônico também ocorreu de maneira distinta, conforme é possível notar em uma pesquisa apontada por Meadows (1999) do início da década de 1990. Cerca de dois terços dos periódicos com estas características correspondiam a Ciências Sociais e Humanidades. O dado parece incoerente, uma vez que foi verificado anteriormente que esse não é um canal prioritário para as áreas. A explicação é que o modelo trouxe algumas soluções para velhos problemas nas CHSA: alta taxa de recusa de artigos, alto custo para publicação

de impressos e pouco espaço disponível para publicação. Desse modo, o acesso livre teve grande adesão nas CHSA, ao contrário das Ciências, que não foram afetadas uma vez que seu modelo de comunicação já estava consolidado.

No detalhamento dos canais de comunicação, Larivière, Gingras e Archambault (2009) caracterizam as CHSA como aquelas que atribuem mais valor às monografias, apresentações em conferências e outros meios de publicação, tais como os livros. Acrescentam ainda que a literatura é mais fragmentada, caso em que dificilmente é formado um corpo central de revistas científicas, esclarecendo a dificuldade em se estabelecer algum tipo de análise bibliométrica de maneira satisfatória valendo-se apenas dos periódicos.

Com relação à adesão de recursos digitais, Marron (2008) afirma que, nas Humanidades os alunos utilizam mais os periódicos eletrônicos, seguidos de fóruns de discussão. O uso de *blogs* também é mais frequente para esses estudantes. No caso das Ciências Sociais, a maioria dos alunos consulta os chamados “*professional and scholarly hubs*”, ou grandes portais acadêmicos que centralizam a informação, seguido dos periódicos eletrônicos. No aspecto do recurso em si, os pre-prints<sup>19</sup> foram os mais utilizados pelos estudantes de Ciências Sociais Aplicadas. O recurso mais utilizado é o *Social Science Research Network*. Para o grupo STM, notou-se que *sites* que permitem acesso e publicação de dados de pesquisa, como o *Protein Data Bank*<sup>20</sup>, foram especialmente considerados. O autor conclui ainda que os recursos mais inovadores, tais como elementos multimídia, estão presentes nas STM, enquanto a considerada “velha tecnologia”, como as listas de discussão, são mais proeminentes no campo das Humanidades.

A própria estrutura da publicação, independente do canal, tem diferente impacto entre as áreas. Quanto mais ela for reconhecida e aceita por uma disciplina, maior a chance dos trabalhos que seguirem essa mesma linha terem maior relevância. Ao contrário das CHSA, nas Ciências estas convenções ficam mais evidentes, onde há maior diferença e menor consenso quanto aos padrões de publicação. De maneira geral, os canais da comunicação em CHSA são distintos e até certo ponto mais

---

<sup>19</sup> Um artigo científico em versão prévia, ainda não finalizado, podendo ser (ou não) publicado em um periódico com revisores. É utilizado comumente por autores da Física para facilitar a colaboração internacional nos trabalhos, podendo, mediante sugestões, elaborar uma versão final aprimorada (BROWN, 2001).

<sup>20</sup> <http://www wwpdb.org/>

complexos que nas áreas do STM, uma vez que não se tem um consenso daquelas mais comuns a serem explorados. De qualquer forma, esses fatores devem ser levados em consideração para a discussão posterior sobre indicadores.

## 7.1 ASPECTOS DA AVALIAÇÃO DA PRODUÇÃO CIENTÍFICA NAS CHSA

Meadows (1999, p. 180) relembra que o reconhecimento do trabalho é um tipo de recompensa recebida pelo pesquisador. Sendo assim, envolve o cumprimento de um requisito de qualidade. Esse controle é exercido, no caso dos periódicos, por um grupo de editores especializados na temática, selecionados pela sua representação naquela rede de pesquisa, não sendo necessariamente aqueles mais produtivos. Em épocas anteriores, o trabalho de avaliação era feito exclusivamente pelos editores. Com a ampliação das especialidades, a função de avaliar se torna um expoente, atendendo ao fato de que o primeiro não tem condições suficientes para julgar a qualidade para aquela comunidade. A expansão do papel do avaliador externo estabeleceu um certo grau de confiança atribuído aos periódicos, criando a imagem que os artigos submetidos à avaliação e publicados em periódicos eram sinônimos de pesquisa aceitável. Nas Ciências Exatas e da Vida, esse cargo costuma ser exercido pelas autoridades mais reconhecidas. Nas Ciências Sociais e Humanas, há uma certa exigência de competência para julgar de forma equilibrada e flexível os trabalhos, aceitando certos métodos empregados ainda que estejam fora dos modelos regulares. São comportamentos típicos dessas ciências, relacionados aos próprios costumes no processo de pesquisa que, em geral, possui maior variedade de referencial teórico.

Meadows (1999, p. 184) explica que as principais alegações dos avaliadores de trabalhos em Ciências Sociais são descompassos entre teoria e prática. Ao posicionar o interesse dos públicos, o autor sugere que os leitores dessa área costumam ter mais interesse prático do que os avaliadores, voltados para a teoria. Outra característica é a necessidade de julgamento por mais de um profissional, como forma de reduzir um possível viés dado por um avaliador. Sob este aspecto, aponta estudos que comprovam que não há considerável divergência entre avaliação de dois árbitros, mesmo que critérios distintos sejam utilizados.



No decorrer da avaliação, no momento em que o trabalho está sendo julgado, existem três possibilidades de retorno; a aceitação, a aceitação com restrições, portanto com correções a fazer, e a rejeição. Meadows (1999, p. 187) aponta que as Ciências Exatas detêm as menores taxas de recusa (menos de 50%), as Ciências Sociais em um número intermediário, e as Humanidades com as maiores taxas de recusa (até 90%). Para ele, isso é reflexo da concordância quanto à ausência de um referencial teórico nuclear na área. Ressalta ainda que os critérios para avaliação são menos objetivos, o que dificulta o processo assertivo entre autores e avaliadores, o que já não acontece nas Ciências, na qual são mais objetivos desde o processo de redação. O autor estende a discussão, lembrando que o processo de avaliação nas Humanidades e Sociais são mais extensos, o que, usando a notação de mesmo número de periódicos disponíveis para publicar, implicaria em uma maior recusa por parte dos editores, em função da capacidade máxima das revistas. Assim sendo, o sistema de avaliação também serve como instrumento regulador da capacidade técnica do corpo editorial *versus* submissões.

Meadows (1999, p. 188) julga que cenário sugere as razões para os pesquisadores das Humanidades considerarem os livros como um canal mais adequado para publicação. A extensão desse tipo de canal permite formular uma exposição mais completa do tema e até, supostamente, mais convincente, o que não ocorre nos periódicos pela sua limitação de tamanho.

Yankauer (1991) verificou como as diferentes disciplinas se comportam em relação à avaliação de trabalhos por pares. Concluiu que, entre 1970 e 1990, a maioria dos periódicos científicos não praticou a chamada revisão cega, sendo elas mais comum nas Ciências Sociais. A maioria dos entrevistados (acadêmicos) se mostrou a favor do uso da revisão cega por pares pela possível eliminação do viés. Ainda assim, o grupo aponta muitas falhas no processo de revisão cega por pares (como o risco de identificação), o que gera dúvida quanto a integridade do processo. Snodgrass (2006) em pesquisa mais recente notou um prevaecimento da chamada *double-blind-review*, sendo que um número cada vez maior de periódicos tem adotado esse tipo de avaliação. O autor destaca novos benefícios tais como a maior possibilidade de publicação dos autores de instituições com menor prestígio e também o maior número de autoras mulheres, diminuindo o preconceito de gênero.

#### a) Autoria e colaboração na produção científica de CHSA

A autoria e colaboração são consideradas um estudo a parte na avaliação da produção científica. Os editores consideram que aqueles que forem inseridos como autores dos trabalhos, contribuíram para o desenvolvimento do mesmo e aceitam a responsabilidade pela totalidade do conteúdo. A mudança que Meadows (1999) aponta na prática é que esses pressupostos nem sempre são confirmados. Apresenta pesquisas em que em muitos trabalhos, a presença do autor não era condizente com a sua contribuição. Isso afeta áreas em que há tradição do trabalho com múltipla autoria.

Nas Humanidades e Ciências Sociais, os trabalhos costumam ter um ou dois autores. Nas Ciências, a proporção de artigos com dois autores é a que prevalece, revelando que a necessidade de cooperação é diferente. Price (1976c) chegou a prever que, na década de 1980, se fossem seguidas as projeções de número de autores, algumas áreas não teriam mais autoria individual. Meadows (1999) relembra que em 1983, a área com menor índice (valores médios) de autores sozinhos foi a Física, com 30%.

Segundo Lariviere, Gingras e Archambault (2008), o estudo da colaboração de trabalhos pode ser explicado pela quantidade de autores nas publicações. As Ciências Sociais, ao contrário do que verificaram os autores supracitados, aproximam-se das Ciências Naturais, apresentando evolução com o aumento significativo da média de número de autores. Já no caso da Humanidades, a realidade verificada na década de 1990 se repete, pois os trabalhos possuem, na maioria, apenas um autor. Os autores concluíram que é possível estabelecer relação entre autoria e canal de comunicação, sendo que as disciplinas com maior colaboração nos trabalhos foram aquelas em que as revistas científicas são o principal meio de publicação e também, aquelas em que há predominância de pesquisa com métodos quantitativos, o que é tradição em Ciências e Engenharias.

#### b) Obsolescência, Velocidade e Intervalo entre Publicações em CHSA

Price (1976c) explica que alguns campos tendem a homenagear os seus pioneiros, com citações constantes no tempo para alguns trabalhos (como os de Einstein). Em outros com “inundação de literatura”, há tendência de citar artigos antigos com pouca frequência. Enquanto o indicador de meia-vida nas Ciências é, no geral, mais baixa (em média até no máximo 10 anos), outras áreas (e principalmente no nível das disciplinas) apresentam meias-vidas muito superiores, como é o caso da

Exegese Bíblica, que tem idade média de citações em 21,6 anos. Uma explicação é que, na primeira área os assuntos são rapidamente aceitos e/ou absorvidos e em outras, como as Humanidades, na qual se encaixa a disciplina supracitada, há maior discussão e princípios mais qualitativos para uso do conteúdo. As diferenças enquanto índice não estão somente no nível das áreas ou mesmo disciplinas, mas a própria política editorial e característica de certos periódicos pode produzir índices diferentes para mesma especificidade (MEADOWS, 1999).

Apesar dos autores de livros apresentarem uma tendência (pelo processo de concepção) a referenciar obras mais antigas e, portanto, apresentarem uma meia-vida maior, a diferença é pequena (em comparação por área) em relação aos periódicos. Nas Ciências, a meia-vida dos artigos em periódicos costuma ser baixa e, o mesmo ocorrendo para os livros. Inversamente, nas Humanidades, a meia-vida para artigos e livros são maiores comparadas às demais ciências. Nas Ciências, as informações são constantemente codificadas e absorvidas e, para se elaborar uma nova pesquisa, é fundamental saber quais são os trabalhos mais recentes sobre o assunto, caracterizando a importância dada a obras atuais. As Humanidades correspondem ao oposto, tendo a literatura antiga como matéria-prima fundamental de sua pesquisa. As Ciências Sociais encontram-se entre os dois opostos, com significativa relevância tanto dada aos materiais mais atuais quanto àqueles mais antigos (MEADOWS, 1999).

Lariviere, Gingras e Archambault (2008) estudaram a diminuição do tempo da literatura citada em função da disponibilidade *online* dos periódicos, na forma de tendência de citação dos artigos mais novos e concentração em menos periódicos. O trabalho teve a área da ciência como uma das variáveis. Foi possível concluir que as Humanidades apresentam um comportamento distinto a todas outras áreas, na comunicação, na colaboração e no uso de periódicos, caracterizando-se como uma exceção. Uma quantidade pequena de periódicos citados é apenas um reflexo da tendência, pelos autores da área, de utilizarem os livros como canal principal. Os autores revelam ainda uma mudança na concentração de citações nas áreas Biológicas e Sociais, apresentando, em 2005, o maior decréscimo até então registrado. Com base nessa condição, eles defendem que o sistema científico está se tornando mais eficiente no que tange à publicação do conhecimento, inclusive, mais democrático.

## 7.2 SITUAÇÃO DOS INDICADORES EM CHSA

O uso de indicadores de produção científica é um tema conflituoso para as CHSA desde o surgimento dos primeiros índices e métodos na década de 1960. Van Raan (2003) reconhece essa dificuldade, porém, julga que a insistente posição da comunidade científica de que as áreas são “bibliometricamente inacessíveis”, só dificulta essa aproximação na geração de critérios mais objetivos de avaliação. O tema envolve a compreensão da maneira como os trabalhos são avaliados.

Archambault *et. al.* (2006) verificaram que determinado tema nas CHSA, ao ser desenvolvido em um idioma específico, herda sua semântica, tornando-se mais interessante para o grupo de acadêmicos locais. Nesse processo, tais publicações muitas vezes não alcançam o âmbito internacional. Outros fatores culturais podem influenciar certas áreas de pesquisa, ameaçando, de certa forma, a validade dos indicadores a serem publicados, especificamente na questão idiomática, na qual há um *gap* da cobertura nacional em função da priorização das publicações em inglês nos índices tradicionais tais como o SSCI.

Glänzel e Schoepflin (1999) afirmam que, em parte dos campos do conhecimento, os indicadores bibliométricos mais típicos (baseados em artigos de periódicos) precisam ser alterados para prover dados válidos sobre a produção. Isso ocorre principalmente naqueles em que monografias, livros e relatórios são importantes canais de comunicação, transpondo a importância exclusiva do periódico, como ocorre em outros. Como forma de identificar os canais, os autores apoiaram-se em uma pesquisa no SCI e no SSCI, tendo como variáveis a porcentagem de referências a periódicos; a idade do meio referenciado, e; a porcentagem do meio referenciado. Nas CHSA, uma parte significativa dos documentos citados eram de fontes consideradas não-científicas para o SSCI: livros, teses, dissertações entre outros. Em estudos seguintes, foi possível verificar uma predisposição e apoio à aceitação dessa literatura, reconhecida a importância para CHSA e, indispensável para o uso de indicadores confiáveis para a área (LARIVIÈRE *et al.*, 2006; ALTHOUSE *et al.*, 2008; TORRES-SALINAS; MOED, 2009; LANCHO-BARRANTES; GUERRERO-BOTE; MOYA-ANEGON, 2010).

Glänzel e Schoepflin (1999) explicam que esse *gap* de cobertura é um dos maiores problemas para aplicação da bibliometria, o que não se refere a uma debilidade da técnica em si, mas da disponibilidade de bases de dados que sejam

apropriadas. Leydesdorff (2008) aponta uma inadequação do uso do FI na comparação entre duas ou mais áreas da ciência em detrimento dessas diferenças, posicionando que as verbas em pesquisa não podem ser alocadas com base nesse tipo de comparação. Aleixandre-Benavent, Valderrama-Zurián e Gonzáles-Acalide (2007) sugerem a criação de um *Factor de Impacto Español* como resultado de um consenso sobre a necessidade da análise de citação aplicada a situação nacional (Espanha).

Leydesdorff e Salah (2010) comentam a criação, em 2006, de um projeto de desenvolvimento de indicadores para as Humanidades, encomendado pelo *American Academy of Arts and Sciences*. No total, foram elencados 75 indicadores em mais de duzentas tabelas e gráficos. O modelo acabou sendo criticado por se basear nas necessidades dos acadêmicos e não nos recursos preexistentes. A proposta manteve-se afastada dos métodos tradicionais, não contemplando, por exemplo, qualquer aplicação da análise de citação.

Em 2008, o *European Science Foundation* investiu em iniciativa semelhante, o *European Reference Index for the Humanities* (ERIH). Como resultado houve uma categorização em três níveis de avaliação, julgados não necessariamente quanto a sua qualidade, mas sim quanto ao escopo e “audiência” dos periódicos. O ERIH recebeu críticas de mais de sessenta editores que julgaram esta iniciativa como algo que jamais seria considerado um *ranking*. Solicitaram, ainda, a retirada de seus periódicos do índice.

Leydesdorff e Salah (2009), no que corresponde ao uso de indicadores, reconhecem um hiato maior entre as Humanidades e as outras áreas, do que entre as Ciências Sociais e as Exatas. Elas sugerem que isso não significa que a área deve ser o elo mais frágil da cadeia de incentivos à pesquisa, tendo em vista que novas alternativas de fonte de dados surgiram para amenizar as dificuldades apresentadas pelo modelo tradicional de FI e *Thomson ISI*. Desde 2004, *Scopus* e *Google Scholar*, novas bases com canais variados, trouxeram alternativas para estudar as Artes e Humanidades com adaptações das técnicas bibliométricas. As autoras reforçam o distinto comportamento de citação dos acadêmicos dessas áreas, o que torna desvantajoso o método da análise de citação. Outro fator é o meio de divulgação da produção. Os acadêmicos da área se utilizam de diferentes vetores que, em muitos casos, não são cobertos pelas bases de dados (ex.: trabalhos de arte ou poemas). A própria importância atribuída a outros tipos de documentos (além dos periódicos) deve

ser levada em conta (ex.: livros, revisão de livros e monografias). Existem esforços internacionais no sentido de moldar as Ciências Humanas e Sociais, permitindo uma avaliação bibliométrica no futuro, por exemplo, por meio de uma lista de periódicos internacionais centrais para cada especialidade. Outra crítica é o enfoque dado ao pesquisador individual, considerado um “escopo pequeno”, e a própria definição do que realmente uma citação representa.

Com base nas assertivas verificadas neste tópico, bem como o decorrer do processo de evolução da ciência, aspectos de mensuração, indicadores e finalmente as características da área, é possível estabelecer uma discussão a respeito do tema, no intuito de apontar alternativas objetivas de avaliação às CHSA, conforme tópico seguinte.

### 7.3 DISCUSSÃO

Um dos pontos de partida deste trabalho é a diferença que as áreas da Ciência apresentam, desde sua evolução, até uma convergência parcial para o início do Século XXI. Uma observação inicial sobre a história da Ciência Moderna e, um fato que não pode ser desconsiderado, é que as CHSA tardaram a ser, de fato, consideradas Ciências – ou continuam a ser desconsideradas como tal por alguns autores (MEADOWS, 1999). O historiador da Ciência Derek Solla Price defendeu, ainda na década de 1970, a forma de ordenação do mundo da observação como sempre baseado em princípios matemáticos e lógicos. Portanto havia uma divisão “Ciências” e “não Ciências”, pela qual CHSA estava no segundo grupo. Os investimentos em pesquisa também se pautaram nesse pensamento, estando a ênfase, desde a Revolução Industrial, nas Ciências Exatas, que apresentavam contribuição direta para o momento histórico. O contexto não pode ser ignorado, mesmo em uma pesquisa contemporânea. Desse modo, não se espera que na primeira década do Século XXI a situação já tenha sido equilibrada, ou que haja, de certa forma, um desmerecimento em relação a área. Trata-se de um processo evolutivo que pode ou não se equilibrar, dependendo dos interesses sociais, políticos e econômicos no desenvolvimento científico.

Aliado à esse processo, as CHSA apresentam características que lhe são peculiares, normalmente ressaltadas como a justificativa para que os critérios e entendimentos da avaliação das Ciências Exatas não pudessem ser aplicados a elas. As características identificadas, com maior influência na aplicação de indicadores são:

a) Características Gerais

- Prioritariamente qualitativa;
- Enfoque na realidade local;
- Aumento da múltipla autoria (Ciências Sociais);
- Prevalcimento de trabalhos com um autor (Humanidades);
- Citação frequente aos trabalhos mais antigos.

b) Canais de Publicação

- Pesquisas mais longas, com mais tempo na redação e menor velocidade na publicação;
- O canal preferencial de publicação é o livro, incluindo ainda anais de eventos, *eprints*, listas de discussão entre outros.
- Uso tardio de Tecnologia da Informação e Comunicação nas pesquisas; porém, na área de publicação, apresentou maior adesão que as outras áreas em detrimento da redução de custos dos sistemas e políticas de acesso aberto.

c) Sistema de Avaliação

- Julgamento por pares com exigência maior de competência de autores e árbitros;
- Interesse maior dos avaliadores pelo *corpus* teórico;
- Critérios subjetivos e alta taxa de recusa;
- Processo de avaliação mais extenso.

A partir dessa enumeração, é importante estabelecer correlações com o uso de indicadores objetivos pelas Ciências Exatas. O método básico de mensuração de impacto na Ciência, desde a década de 1960, é a análise de citação. Seu criador, Eugene Garfield, o concebeu pensando na melhoria da recuperação da informação por meio da formação das redes de citação, visualizando o potencial que o elemento citação, essencialmente entendido como um indicativo de uso, teria para a avaliação do ofício da ciência.

Apesar da aceitação e renome que o método e os consequentes serviços derivados – FI e o *Web of Science* – obtiveram, as críticas e estudos a respeito das citações sempre formaram um hiato entre o elemento e a perspectiva de objetividade desejada. Valer-se da análise de citação, pressupôs, desde a década de 1960, aceitar e assumir suas falhas. Dentre as principais está a definição de motivação, tendo em vista que, por exemplo, um trabalho poderia ser citado apenas para apoiar ideias de um autor citante, sem que correspondesse ao contexto e pensamento defendido pelo autor citado. Outra delas seria a citação persuasiva, pela busca do mérito do autor citado, ou ainda a negativa, que serviria como contraposição às ideias apresentadas. Essas falhas se remetem aos aspectos comportamentais, o que é complementado pela discussão técnica da citação. A qualidade dos dados empregados, por meio do erro do referenciamento, dificuldades com grafias de nomes semelhantes e definição do quanto uma citação representa para um trabalho de múltipla autoria, também são aspectos discutidos sobre o método.

O cenário apresentado afeta todas as ciências que de alguma forma empregam o método de análise de citação. Aquelas que optam pelo seu uso, aceitam essa situação, principalmente porque, por muitos anos, foi considerada a melhor das possibilidades para medição de resultados, além da mais confiável. Um dos pontos talvez mais divergentes de seu uso não está no método em si, mas nas fontes primárias e secundárias que alimentam os índices de citação.

O SCI (posterior *Web of Science*) é considerado o principal índice de citação e detentor de um monopólio na área. Sua concepção está diretamente atrelada ao desenvolvimento da ciência, ou seja, assim como houve uma dificuldade histórica no reconhecimento das CHSA enquanto áreas de pesquisa e, portanto, ausência de investimentos, o índice – que passou a ser publicado na década de 1960 – voltou-se para as ciências que eram vistas como prioritárias. O resultado disso foi um desenvolvimento sempre acompanhado de críticas pela pouca atenção dada às CHSA. Assim, os indicadores pressupõem o uso de uma base e, a principal e praticamente única até o início do séc. XXI sempre foi vista como tendenciosa, a área de CHSA tornou-se resistente ao uso de indicadores provenientes de métodos bibliométricos, conforme apontado por Archambault *et al.* (2006) e Van Raan (2003).



As principais críticas aos índices da *Thomson ISI* verificadas na literatura são quanto às seguintes tendenciosidades<sup>21</sup>:

- a) privilégio às Ciências Exatas e da Vida;
- b) restrição de tipos de publicação, ou seja, apenas conteúdo referente à periódicos;
- c) seleção restrita de periódicos da base;
- d) privilégio do idioma inglês;
- e) características dos indicadores produzidos.

Percebe-se uma relação dos fatores anteriores com as características das CHSA. Inicialmente, o contexto histórico como comprovação do enfoque da atividade científica em outras áreas. Em seguida, os canais de comunicação utilizados. Nas outras áreas, os documentos referentes à periódicos sempre tiveram uma representatividade maior, ao contrário das CHSA que privilegiaram os livros, monografias, teses e dissertações.

Cabe, em contraponto, situar as ações da *Thomson ISI* no intento de diminuir ou alterar esse cenário, sendo a criação do SSCI e A&HCI (década de 1970), mais recentemente o *Conference Proceedings* (2008) e, futuramente, conforme já anunciado, o *Book Citation Index* (2011), idealizados como suporte a áreas que valorizam esse outros canais de publicação.

Ainda assim, os terceiro e quarto fatores citados, sobre o processo de seleção e ingresso na base, seguem como críticos. As pesquisas em CHSA têm enfoque local e, por isso, geralmente são publicadas no idioma da região, mediante a dificuldade de replicação e até compreensão dos resultados em contextos e idiomas diferentes. Além disso, o critério de seleção do ingresso é visto como de alta exigência com baixa penetração por autores e periódicos que não estão localizados nos países com grande tradição em pesquisa, como é o caso da maioria dos em desenvolvimento. A participação de periódicos – e mais recentemente eventos científicos nos índices da *Thomson ISI* é de representatividade muito baixa nas CHSA, principalmente desses países, como é o caso do Brasil. Assim sendo, são poucos os motivos que levariam a adoção do *Web of Science* como fonte para uso de indicadores, pelos comitês de assessoramento e outros grupos de avaliação de CHSA.

---

<sup>21</sup> Conforme pode ser visto em 6.2 Principais críticas à análise de citação.

No que diz respeito a *Thomson ISI*, é cabível discutir também os principais indicadores publicados. As medidas básicas como o total de citações recebidas por um autor, artigo, periódico ou instituição, são vistas como inadequadas em uma comparação global, pois o comportamento de citações em CHSA ocorre de maneira diferente. Os trabalhos são publicados com periodicidade inferior, em função de sua extensão e tempo necessário para conclusão, portanto, a quantidade total de trabalhos é menor e, conseqüentemente, a de citações. Outro indicador importante e renomado da *Thomson ISI* é o FI. Essa talvez seja uma das medições que torna mais evidente a dificuldade de seu uso pelas CHSA. O cálculo do FI compreende uma janela temporal de apenas dois anos de citações após a publicação do trabalho, o que impossibilita seu uso em áreas em que a meia-vida pode chegar até 30 anos pois, nas CHSA, os trabalhos costumam ser citados por muito tempo após a sua publicação, seguindo uma utilização menos imediatista do que nas outras áreas.

Este cenário de inadequação histórica dos serviços da *Thomson ISI*, bem como as suas projeções, não é por si só um motivo para as CHSA descartarem os indicadores bibliométricos. Porém, Meho e Yang (2007) assim como outros autores, desconsideram o índice da *Thomson ISI* ou a busca por uma base global igualitária, pois ela nunca iria refletir o ambiente da pesquisa sociológica. Sendo assim, as soluções que apontam para o uso das bases da *Thomson ISI*, ainda que a empresa tenha proporcionado uma diversificação em seu portfólio que supostamente diminuiria as tendenciosidades, parecem ser inviáveis para a realidade de CHSA, ao menos no que diz respeito aos países em desenvolvimento. Ainda assim, isso não implica na invalidação do método da análise de citação e indicadores decorrentes, pois é possível aplicá-los em bases de dados alternativas.

Os algoritmos, explicações e comentários a respeito dos indicadores estão disponíveis para estudos e novas aplicações e não são uma propriedade da *Thomson ISI*. É o caso da adaptação do FI na criação *Scimago Journal & Country Rank*, ou da proposta de um *Factor de Impacto Español* voltado apenas para a realidade desse país.

Os indicadores também apresentam funções diferentes e sua validade pode ser relevante de acordo com a área. Considerando-se que as principais características em CHSA são a temporalidade e as questões de autoria e coautoria, o uso de contagens básicas e indicadores tradicionais, tais como total de citações recebidas, trabalhos publicados e citações por trabalhos (ainda que de caráter simples), poderiam

ser adequados em uma aplicação no nível da área ou de disciplinas. Desse modo, seria possível uma comparação semelhante, uma vez que a realidade apresentada aos envolvidos seria análoga.

O Fator de Impacto e o Índice de Citação Imediata, contudo, não seriam indicadores adequados às CHSA, mesmo que aplicados localmente ou a uma base seleta, pois a janela temporal de dois anos no caso do primeiro ou, um ano no caso do segundo, poderiam representar até um valor zero na área. Conforme identificado por Sombatsompop, Markpin e Premkamolnetr (2004), uma adaptação seria um FI que considerasse o comportamento temporal de citações de uma área, com acréscimo da variável da meia-vida, como é o caso do *Citation Half-Life Impact Factor*. Apesar da complexidade, o indicador teria aplicações na avaliação de autores, periódicos e instituições de ensino, pois seguiria o princípio do FI de identificar o quanto do potencial de citações a unidade conseguiu atingir em relação a outros elementos.

Pringle (2008) sugere que a convergência digital do formato de publicação, apresenta uma certa democratização dos indicadores de citações, os quais são usados e revistos por diversos pesquisadores, não apenas no escopo da Thomson ISI. Talvez como reflexo do cenário, surgiram, desde o início do século XXI, outros indicadores e adaptações, deslocando-se dos tradicionais FI e simples contagem, tais como o índice-h.

Para Bar-Ilan (2008), os estudos de citação subsequentes foram quase integralmente pautados neste índice e, até por isso, o índice-h foi reconhecido como um número áureo, ou uma solução que há muito tempo era esperada na medição do impacto. Ainda que receba críticas, o indicador é considerado equilibrado pois, na situação em que um autor tem um de seus trabalhos muito mais citado que outros, não há uma contaminação significativa ao aplicar tal medida de impacto, como ocorreria em outros casos (ex.: Fator de Impacto). Uma das principais críticas aos índice-h é o fato de não ser contemporâneo e não regredir, sendo que os trabalhos mais antigos tem o mesmo peso que os novos. Assim, surgiram derivações com o intuito de, por um lado, proporcionar a sua regressão, como o índice-ar (voltado para o núcleo-h) ou, valorizar a totalidade dos trabalhos, como o AWCR (que contempla todos os trabalhos já citados). Ao entender que a motivação dessas variações esteve em permitir a evolução dos trabalhos e o ressaltar o peso de inovações, não deixaria de ser adequado para a CHSA. Porém, a variável “idade de publicação” a ser utilizada no cálculo nos dois índices, seria objeto de invalidação para a área. Por mais que pareça

um valor objetivo, um trabalho que tem cinco anos em Física, tem uma representatividade quase nula para seus pares, se comprado a outro em certas disciplinas de CHSA que, com cinco anos após a publicação, ainda podem ser relevantes. Entende-se que se a busca pela regressão de importância dos trabalhos fosse considerada na criação de um indicador em CHSA – ou mesmo em outras áreas –, o ideal seria incluir uma variável que o normalizasse de acordo com a importância da obra desde a publicação. Nessa situação, a variável poderia ser a meia-vida da área ou disciplina da ciência, acrescentada ao cálculo do impacto de autores, periódicos, instituições, grupos de pesquisa entre outros.

O índice-g, proposto por Leo Egghe, teve como intuito avançar na discussão ao atribuir maior peso à quantidade de citações recebidas, o que era considerada uma falha no índice-h. Consta que sua adequação seria benéfica aos pesquisadores ditos como seletivos, ou seja, com poucas publicações, porém, com muitas citações. Ao se atentar para o primeiro aspecto, este seria um indicador ideal para as CHSA, uma vez que o total de publicações costuma ser inferior. Porém, tal cenário não implicaria, necessariamente, em um grande número de citações. Nesse caso, haveria necessidade de um estudo mais aprofundado pelo comportamento por disciplinas para definir seu uso.

Alguns outros índices mais aprimorados ainda estão em fase inicial de testes, como o índice-f (2009) e o Fator-Y (2009). Ambos tratam de um problema apontado na década de 1970, ou seja, a necessidade de uma variável peso para as citações, semelhante ao funcionamento do algoritmo do mecanismo de busca do *Google*. Se as discussões ainda são incipientes nas áreas que já utilizam tradicionalmente indicadores e têm definidos claramente os periódicos e autores nucleares, talvez as CHSA ainda precisem se posicionar nesses preceitos básicos, para que a ideia possa ser explorada no futuro. Ainda assim, percebe-se que ela será afetada positivamente com os estudos, ao lembrar que o comportamento de coautoria nos trabalhos, em parte das CHSA, tem se modificado – ampliação de trabalhos com mais de um autor – o que sugere a necessidade de definir com mais clareza o peso das contribuições.

Verificou-se que a resistência ou não uso dos indicadores bibliométricos nas CHSA não é, exclusivamente, produto da rigidez dos critérios objetivos que as outras áreas utilizam. Baseando-se naqueles apresentados nesta pesquisa, percebe-se com as devidas adaptações e ressalvas, eles poderiam ter representatividade. Porém, de fato, é a existência e coleta de fontes de informação para o uso de indicadores, a

principal dificuldade. Ou seja, a discussão do tema parece levar ao questionamento da existência, quantidade e qualidade de bases que ofereçam condições para criação de índices de citação, e não dos indicadores em si. Sob essa assertiva é que determinadas ferramentas e plataformas podem, potencialmente, oferecer elementos para a uma alternativa na área, conforme apresentado no capítulo seguinte.

## 8 FERRAMENTAS WEB

As modificações na comunicação científica e a inadequação do modelo tradicional de avaliação da ciência, aplicável às Ciências mas dificilmente aceito nas CHSA, sugerem o estudo de novas oportunidades no que compete ao processo de avaliação do impacto do ofício do pesquisador. Nesse capítulo são apresentados os resultados da coleta de dados, de carácter descritivo, feita nos sistemas previamente selecionados.

Neuhaus e Daniel (2008) são enfáticos ao relatar que o tempo em que a *Thomson ISI* era a única provedora de índices de citação já foi superado. Esse monopólio de fonte desfaz-se a medida que novas bases, repositórios, índices e ferramentas são criados. O contexto em que isso ocorre, para os autores, é a mudança de uma situação pré-publicação eletrônica, em que ter um índice de citação era algo trabalhoso e caro, para uma situação em que parte significativa da produção científica se encontra em formato digital, o qual permite uma coleta e processamento mais rápidos e baratos.

Em uma adaptação da classificação proposta pelos autores, sugere-se que essas alternativas sempre sejam entendidas como um composto de dois serviços integrados ou não: uma fonte na forma de base de dados e, uma plataforma com ferramentas analíticas que permitam análise bibliométrica a partir dessa fonte.

Atendendo aos critérios adotados neste trabalho e a classificação sugerida pelos autores supracitados, fez-se necessário, em alguns casos, distinguir o serviço da base de dados e o da ferramenta analítica, na ocasião em que são providos por sistemas ou organizações distintas. Julgou-se importante usar a divisão proposta<sup>22</sup> pelos mesmos autores, no que diz respeito ao escopo da base. Sendo assim, as ferramentas são apresentadas como pertencentes a bases de dados multidisciplinares ou, orientadas a disciplinas.

---

<sup>22</sup> Conforme apresentado previamente no Ilustração 1: Tipologia das Ferramentas Estudadas, no capítulo 2 "Metodologia".

## 8.1 FERRAMENTAS DE BASES MULTIDISCIPLINARES

Nesse tópico são apresentadas as ferramentas *web* que provêm resultados a partir de bases de dados multidisciplinares com registro de citações, sendo elas o *Publish or Perish*, *scHolar Index*, *Scholarometer*, *Citation-Gadget for iGoogle*, *H-view Visualizer* e *Scholar H-Index Calculator Firefox Extension* que utilizam a base do *Google Scholar*; e os seguintes com base própria: *Microsoft Academic Research*, *Citebase*, *getCITED*.

Esclarece-se que o *Google Scholar* em si não é uma ferramenta adequada aos critérios propostos no trabalho mas, é um componente fundamental para seis delas, servindo como um índice de citação. Desse modo, optou-se por defini-lo em separado, o que permite relatar antecipadamente o grupo “b” do Protocolo de Descrição<sup>23</sup>, relativo ao escopo.

### 8.1.1 Ferramentas da base *Google Scholar*

O *Google Scholar* (GS) foi lançado em novembro de 2004 como um serviço gratuito de varredura de documentos científicos na *web*, por meio uma diversidade de fontes. Pode ser visto ainda como uma ferramenta de busca acadêmica, criada pela maior empresa do ramo de busca na *web* no mundo, o *Google*. Entre seus resultados são apresentados os seguintes tipos de documentos acadêmicos: artigos de periódicos, livros, anais de eventos, dissertações, teses, relatórios técnicos, *preprints* e *postprints*, resumos e apresentações de slides. Estes itens, por sua vez, são provenientes de universidades, outras instituições de ensino, sociedades profissionais, grupos de pesquisas e outros repositórios (NORUZI, 2005).

Com relação às fontes e o método de varredura, Neuhaus e Daniel (2008) explicam que a extensão da cobertura do GS, principalmente do que diz respeito aos periódicos, é desconhecida, assim como a definição de quais documentos são incluídos e de que intervalo temporal. Constataram, em estudos anteriores, que o modelo de conteúdo do GS seria extraído diretamente do repositório de editores, como a *Thomson ISI* e *Scopus*. Outras coleções incluídas seriam *arXiv*, *BioMed Central*,

---

<sup>23</sup> Mais informações no capítulo 2 “Metodologia” e no Apêndice C.

*HighWire Press*, *IEEE*, *Repec*, *Springer* e *Wiley Interscience*, além de editores que firmaram parcerias (essas sim, transparentes) como a *Elsevier* e a *American Chemical Society*. Os autores ressaltam que a cobertura está além dessa lista, sendo apenas uma parcela que pode ser recuperado.

Walters (2007) discute o mesmo tema, citando que as fontes do GS podem ser divididas em três grupos. O primeiro é seu algoritmo proprietário que identifica em toda a *web* documentos que “parecem” acadêmicos no conteúdo e formato, tanto em texto completo quanto somente resumo. Um segundo grupo é o conteúdo de parceiros, coincidente com a definição anterior de Nehaus e Daniel (2008). Walters (2007) esclarece que parte desse conteúdo é tomado em acordo de redes internas e áreas restritas, não podendo ser acessado em modo completo, a não ser mediante assinatura. Cabe ao GS disponibilizar os dados identificadores e os resumos. O terceiro grupo é uma derivação; são as citações extraídas dos documentos dos dois primeiros grupos e, se caso não estejam disponíveis ou não tenham sido encontradas (nem como texto completo, nem resumo), constam apenas como “citação”.

Apesar deste cenário obscuro apresentado pelos autores, no qual dificilmente são identificados os caminhos utilizados pelo GS para coleta de dados, o *Google* apresenta uma série de recomendações para responsáveis técnicos de domínios que hospedam documentos acadêmicos. Elas poderiam se sintetizadas em:

- a) **Autores independentes:** publicar os documentos com a terminação em pdf, listados em uma página de publicação como “*publication.html*”, certificando-se que o título apareça em fonte maior que o texto na primeira página; com nome do autor logo em seguida em linha dedicada; com a presença de uma sessão “*References*” ou “*Bibliography*”.
- b) **Conteúdo:** *sites* de notícias, artigos de revistas populares, revisões de livros e editoriais não são considerados apropriados para o GS. Documentos com mais de 5MB devem ser enviados via *Google Books*, pela provável característica de extensão enquanto monografia. Os *sites* devem obrigatoriamente exibir textos completos ou resumos de modo gratuito, sem a necessidade de qualquer registro, termo de aceite ou instalação de *software*.
- c) **Estrutura para varredura:** os arquivos devem ser nos formatos HTML ou PDF; se o *site* for “pequeno”, é recomendável explicitar a parte que trata das publicações; se o *site* for “amplo” (ex.: Universidades), os *links* para *download* não devem ficar a muitos cliques da página inicial e, os menus nos formatos



*flash* e *javascript* devem ser evitados; um *site* que não apresenta erros e aceita a entrada de *robots* que varrem o conteúdo tem mais condições de ser incluído no GS.

d) **Indexação:** O GS usa um *software* para identificar os dados bibliográficos e as referências entre documentos; o emprego correto dos dados pelos provedores do conteúdo aumentam a precisão dos resultados para indexação e para o usuário final. Para os usuários de *software* de editoração / repositórios digitais, como *Eprints*, *Dspace* e *OJS*, basta uma configuração que autorize a publicação de metadados na página. São suportados ainda os formatos *Highwire Press tags*, *Eprints tags*, *BE Press tags*, *PRISM tags* e *Dublin Core tags*, tipos de denominação de dados bibliográficos. É importante também manter cada artigo em documento HTML / PDF separados, pois o sistema é incapaz de identificar mais de um por página (GOOGLE, 2011).

Noruzi (2005) explica que a utilidade do GS não está somente na recuperação da informação, e sim na sua função de índice de citação. De maneira simples, o sistema apresenta os itens e um *link* para uma sublista com todos os artigos que citaram este primeiro. A classificação de importância usa o princípio do *PageRank*, incluindo as variáveis número de citações e peso, para a qual, quando um artigo *y* recebe citação de um artigo *z* muito citado, *y* também será julgado como relevante nos resultados. Outras características do GS são a atualização do conteúdo, que é diária; a janela de cobertura inclui desde o período vigente até obras antigas, ainda que não haja transparência do período limite nesse quesito, entendendo-se que é qualquer obra que tenha sido citada por outra obra que já esteja no formato eletrônico independente de sua data de publicação ou inserção na base; e as relações entre documentos por meio de *links*, o que sobrepõe barreiras geográficas. Seus diferenciais em relação aos outros índices são: enfoque *web* pelo qual é possível captar *preprints* de *sites* de autores antes da publicação final do artigo; o custo de manutenção do sistema que é baseado no funcionamento do *Google*, ou seja, gratuito para o usuário final; a cobertura internacional; o baixo ou inexistente grau de tendenciosidades entre periódicos; a inclusão de canais de comunicação não-tradicionais; a interoperabilidade com os vários formatos de citação e, a facilidade encontrar os textos uma vez que os resultados são provenientes da *web*.

Noruzi (2005) verifica que os resultados do GS são sempre em maior número do que os encontrados no *Web of Science*. Geralmente há sobreposição (quando as referências são as mesmas) e o GS extrapola por incluir outros tipos de documento e periódicos fora do núcleo da *Thomson ISI*, além de materiais de outros idiomas que não o inglês, como francês, espanhol, português e dinamarquês.

Bakkalbasi *et al.* (2006) conduziram um estudo comparando a *Web of Science*, *Scopus* e o GS para duas disciplinas das Ciências Exatas. Os índices pagos obtiveram um melhor desempenho pela quantidade de documentos recuperados. Os autores não acreditam ser um resultado conclusivo mas ressaltam que a importância de cada índice está diretamente ligado à disciplina e ao ano de publicação estudado.

Walters (2007) fez um comparativo na área de Ciências Sociais, entre o GS e o SSCI, concluindo que o produto do Google tem pouca sofisticação para que possa ser visto como um substituto do produto da *Thomson ISI*. Ainda assim, combate algumas críticas anteriores como a suposta presença de conteúdos duplicados na base. Para ele, a ocorrência é pouca e está ligada a apenas alguns artigos principais. Os problemas associados ao GS, em síntese, são a falta de resumos (em comparação com outras bases) e dados de citação de pelo menos 30% dos artigos principais na área estudada pelo autor, o que o invalidaria enquanto índice de citação.

Meho e Yang (2007) também efetuaram uma comparação entre o *Web of Science*, *Scopus* e o GS, tendo como objeto citações a trabalhos de 25 cientistas da área de Ciência da Informação. Como resultado, relatam um aumento significativo na prolificidade dos pesquisados ao utilizar o *Scopus*. No que diz respeito ao GS, notaram importância na cobertura de anais de eventos e, principalmente, conteúdo em idioma que não o inglês. Para eles, o total de citações em cada índice não pode ser comparado mas, é fato que os resultados apresentados pelo GS são uma válida evidência do impacto internacional dos autores desta área, mais do que as bases mediante assinatura.

Mingers e Lipitakis (2010) compararam o GS ao *Web of Science* na área de Administração e Negócios na Inglaterra, concluindo que a métrica mais importante para o escopo é a contagem total de citações e que a cobertura do serviço pago ainda é pequena, quase a metade do GS. Apesar das dificuldades de garantir a qualidade de dados do GS, ou autores entenderam que o serviço é uma fonte mais adequada, na atualidade, para a mensuração com base nas citações.

Para Armbuster (2009), o GS é uma base de dados genérica da internet, com cobertura extensiva porém, com pouca satisfação dos pesquisadores por causa da falta de transparência no *corpus* e imprecisão de resultados. Ele resume que a maioria das críticas apresentadas ao GS se pautam nesses tópicos.

Marchiori (2007, p. 107-108) aponta que a qualidade do GS enquanto índice de citação está diretamente ligada ao refinamento dos seus resultados, à sua cobertura e sua permanência enquanto serviço. A autora lista problemas atribuídos ao sistema, como a ausência de um vocabulário controlado, a ausência de filtros dos resultados (por exemplo, para seleção apenas de textos completos), a falta de recursos de ordenação de resultados diferentes do *PageRank*, a impossibilidade de exportação de resultados e, por fim, a falta de transparência na divulgação de parceiros e fontes utilizadas.

Noruzi (2005) enumera as desvantagens do GS, iniciando pela cobertura talvez excessiva de conteúdo, constando: notas administrativas, materiais pessoais de estudantes, entre outros. O GS também apresenta uma certa tendenciosidade no que diz respeito aos idiomas, falhando primordialmente na indexação em Persa, Árabe, Chinês e Japonês. Há também inconsistência nas citações, como a falta de dados; a coincidência e erro na identificação de autores como iniciais homônimas e a ausência de alguns periódicos.

Se a composição da *Scopus* e da *Thomson ISI* é ideal enquanto base de dados para fins bibliométricos, seja pelas fontes de informação e impacto com base em critérios, isso pode ser considerado inexistente na apresentação do GS. Essa é a posição de Jacsó (2005), o qual julga errada a atribuição de indexador autônomo de citação dada ao sistema, uma vez que o GS seria um resultado de falta de competência e entendimento básico na formulação de um índice de citação. O autor aponta também um descompasso entre a competência que o Google tem com as buscas na *web* em relação ao GS.

Harzing e Van der Wal (2008) são defensores do uso do GS como índice de citações para geração de indicadores, atribuindo a Peter Jacsó o trabalho de maior impacto no que diz respeito às debilidades do sistema. Apesar de concordar com algumas posições do autor, verificam que parte dos testes com a base foram inconsistentes, a medida que utilizaram universos pouco aplicáveis ao contexto generalizado. Refletem ainda que o GS já passou por uma evolução e alguns dos problemas apontados inicialmente por Jacsó, em análises feitas na sequência do

lançamento do sistema, já foram parcialmente superados. Além disso, sustentam seus argumentos à favor do GS, destacando os principais problemas do *Web of Science*: a limitação geral de cobertura, a limitação de referências citadas; a contagem de citações para trabalhos fora da base, que considera os primeiros autores e a limitação para fontes em idiomas que não o inglês. Complementam, ainda, com base em uma comparação na disciplina de Administração, que o GS proporciona resultados com uma cobertura mais adequada para a área, pela variedade de fontes. Desse modo, sugerem que acadêmicos e instituições se utilizem do GS para medir impacto com uso de indicadores tradicionais.

Com um panorama a respeito desse sistema, voltado para a recuperação da informação mas também entendido e utilizado como um índice de citações, dá-se início às descrições das ferramentas *web* que dele se valem para gerar indicadores de impacto da produção científica.

a) Harzing's Publish or Perish (PoP)<sup>24</sup>

A ferramenta foi desenhada em 2006 para potencializar a apresentação profissional de pesquisadores por meio do uso de indicadores que vão além do tradicional FI, apoiando-se no GS como índice de citação. É disponibilizada gratuitamente pela *Tarma Software Research*, tendo sido concebida pela pesquisadora Anne-Wil Harzing da *University of Melbourne*, Austrália. Para o seu uso é necessário efetuar o *download* do *software* e instalá-lo localmente em computadores com *Windows*, *Linux* ou *Mac*.

Os resultados obtidos com o PoP podem ser exportados para utilização em outros programas, permitindo a criação de arquivos nos formatos BibTeX, CSV, EndNote, RIS, TXT e XLS<sup>25</sup>. A ferramenta conta também com um manual de ajuda agregado, bem como instruções básicas de uso, perguntas mais frequentes, ajuda *online* e um livro publicado.

---

<sup>24</sup> <http://www.harzing.com/pop.htm>

<sup>25</sup> O BibTeX é um tipo de arquivo utilizado para descrever, listar e processar referências em uma ferramenta homônima. O *Comma-Separated Variables* (CSV) e o XLS são tipos de arquivos de base de dados, compatíveis com *softwares* de planilha eletrônica como o *Microsoft Excel*. O EndNote é um *software* de gerenciamento pessoal de referências, com arquivo de formato próprio. O RIS é um formato de intercâmbio para programas de gerenciamento de referências. O *Text File Format* (TXT) é um formato de arquivo de textos simples (FEDER, 2011, 2011; PARSONS; OJA, 2010; REFERENCE MANAGER, 2011; THOMSON REUTERS, 2011b).

No seu uso é possível verificar indicadores tendo o autor, os periódicos, ou um agrupamento como unidade de análise, reportando os seguintes índices baseados na análise de citação: total de documentos, total de citações, total de anos em que foram publicados documentos, média de citações por ano, média de citações por documentos, média de citações por autor, média de documentos por autor, índice-g, índice h, índice-hi, índice-hi normalizado, AWCR, índice-aw, índice AWCRpA, índice-e e índice hm.

Como variação dos resultados é possível efetuar filtragem por autoria, coautoria, um ou mais periódicos, uma ou mais áreas da ciência, janela temporal, somente documentos com citações, presença ou ausência de palavras-chave. A ferramenta permite também o tratamento de dados por meio da seleção e expurgo de obras pertinentes para o cálculo.

#### b) *scHolar Index*<sup>26</sup>

Criado em 2007 pelo pesquisador Nicolas Roussel do *Institut National de Recherche en Informatique et en Automatique*, na França, é um *site* que faz interface ao GS, efetuando pesquisas na base e sintetizando os dados de acordo com o interesse de exibição de indicadores. A interface é apenas em inglês, apresentando como suporte ao usuário, uma pequena explicação sobre a finalidade. Os resultados não podem ser tratados nem exportados para outro formato que não o apresentado no *site*. Possibilita a filtragem por autoria, área da ciência, idioma do documento citado, janela temporal e mínimo de citações. Apresenta os indicadores: total de citações, média de citações por trabalho, índice-h e índice-g.

#### c) *Scholarometer*<sup>27</sup>

A ferramenta foi concebida em 2009 pela *School of Informatics and Computing* da *Indiana University Bloomington*, como uma extensão<sup>28</sup> a ser instalada nos navegadores *web*, *Mozilla Firefox* e *Google Chrome*. Nesse sentido, é dependente da compatibilidade com um dos dois *softwares*. Além da ferramenta instalada, apresenta recursos de exibição de redes de autoria e disciplinas no próprio *site*. Outro recurso

---

<sup>26</sup> <http://interaction.lille.inria.fr/~roussel/projects/scholarindex/index.cgi>

<sup>27</sup> <http://scholarometer.indiana.edu/>

<sup>28</sup> As extensões são complementos que permitem ao usuário adicionar novas funções e mudar o modo como o navegador *web* é usado (MOZILLA, 2011).

adicional é a disponibilização de complementos para serem inseridos no *site* do pesquisador, para divulgar seus resultados. O idioma da interface é apenas em inglês, apresentando possibilidade de exportação de resultados. Apresenta documentação sobre o sistema nos formatos de “perguntas mais frequentes” e “Ajuda”

Oferece um controle de tratamento do nome de autores em duas etapas. Primeiro durante a solicitação, ou seja, é obrigatória a inserção de um marcador referente à área de conhecimento junto do nome do autor. Na própria base do *Scholarometer* há um registro com todas as solicitações, como uma espécie de histórico para que aqueles que venham a pesquisar o mesmo autor, tenham menor dificuldade em “limpar” os dados. Ao digitar um marcador, o sistema já exibe todos os que já foram utilizados. O segundo controle é feito após a exibição dos resultados, quando os trabalhos aparecem listados e é possível fazer um refino, retirando itens redundantes ou pertencentes a outros autores.

O sistema apresenta os seguintes filtros: janela temporal, mínimo e máximo de citações, trabalhos em coautoria, presença ou ausência de determinada palavra-chave, constando em um periódico, hospedado em e, publicado em idioma específico (ou mais de um). Os indicadores apresentados são: número de artigos, número de citações, índice-h, índice-g, índice-hm.

#### d) *Citation Gagdet for iGoogle*<sup>29</sup>

Foi criado em maio de 2007 por Jan Feyreisl, com apoio do *Google Project Hosting*. Trata-se de um complemento que pode ser adicionado à página principal de usuários do serviço de contas do *Google (iGoogle)*, operando em qualquer computador que tenha acesso à internet e um navegador *web*. Apesar da interface do *iGoogle* variar de acordo com a preferência do usuário, os dados da ferramenta são em inglês. De caráter simplificado, não apresenta possibilidade de intercâmbio de dados ou mesmo qualquer tipo de tratamento dos resultados. A ferramenta também não apresenta instruções e possibilidade de filtros, a não ser o caso de trabalho com até dois autores, quanto é possível calcular indicadores com base nessa coautoria. Os indicadores são índice-h, total de citação e total de trabalhos.

---

<sup>29</sup> <http://code.google.com/p/citations-gadget/>

e) *Scholar H-Index Calculator Firefox Extension*<sup>30</sup>

O Scholar H-Index Calculator Firefox Extension foi criado em 2009, pelo pesquisador Agelin Bee, com base no sistema de extensões do navegador *Mozilla Firefox*. A ferramenta funciona diretamente no *site* do GS, incluindo na página normal dos resultados, uma área com indicadores, além de ferramentas para tratamento manual de resultados. Por estar vinculada a essa página, tem um limite de documentos que contabiliza (cem), o que a torna frágil no caso da análise de pesquisadores prolíficos. Disponibiliza uma instrução para os usuários entenderem o funcionamento. No que diz respeito aos filtros, como se baseia na página de resultados do GS, são os campos avançados da busca, ou seja: ocorrência ou ausência de palavras, autoria, periódico e janela temporal, além da possibilidade de normalização de trabalhos de múltipla autoria. Os indicadores disponibilizados são o total de citações, o índice-h, o índice-g e o índice-e.

Devido à similaridade dos resultados encontrados nas ferramentas *web* que utilizam o GS como índice de citação, foi possível criar um quadro (Quadro 3, p. 111) consolidado com filtros e indicadores, apresentando na sequência.

---

<sup>30</sup> <https://addons.mozilla.org/en-US/firefox/addon/45283/>

Ferramenta	Publish or Perish	scHolar Index	Scholar-ometer	Citation Gadget iGoogle	Scholar H-Index Calculator
<b>FILTROS</b>					
Área da ciência	•	•	•		•
Autocitação					•
Autoria	•	•	•	•	•
Coautoria	•		•	•	•
Hospedado em repositório			•		
Idioma do documento		•	•		
Janela temporal	•	•	•		•
Máximo de citações			•		
Mínimo de citações	•	•	•		
Palavras presentes/ausentes	•		•		•
Título de periódico	•		•		
<b>INDICADORES</b>					
AWCR	•				
AWCRpA	•				
Índice-aw	•				
Índice-e	•				•
Índice-g	•	•	•		•
Índice-h	•	•	•	•	•
Índice-hc	•				
Índice-hi	•				
Índice-hi-normalizado	•				
Índice-hm	•		•		
Média de autores/documento	•				
Média de citações/ano	•				
Média de citações/autor	•				
Média de citações/documento	•	•			
Média de documentos/autor	•				
N.º anos publicados	•				
N.º citações	•	•	•	•	•
N.º documentos	•	•	•	•	•

QUADRO 3: RESULTADO CONSOLIDADO DAS FERRAMENTAS BASEADAS NO GOOGLE SCHOLAR

FONTE: O autor (2011).



No Quadro 3 é possível visualizar inicialmente uma diferença significativa entre os resultados, com destaque para duas ferramentas – Publish or Perish e Scholarometer – que apresentaram maior conformidade com as variáveis avaliadas no protocolo de descrição. A análise dos resultados de todas as ferramentas, inclusive as subsequentes, é apresentada no subcapítulo 8.3 “Discussão”.

O GS é apontado como uma das principais alternativas, no que diz respeito aos índices de citações. É por esse motivo que foram encontradas mais de uma opção de ferramentas para tratamento de seus dados. Ainda assim, existem outras alternativas que se enquadraram nos critérios preestabelecidos, conforme apresentado no tópico seguinte.

#### 8.1.2 Ferramentas de bases próprias

Ainda que não com a mesma dimensão ou proposta de funcionamento que o GS, existem outras iniciativas com escopo multidisciplinar que constituem índices de citações e ferramentas derivadas para aplicação do método de análise de citação. Ao contrário do GS, elas não se caracterizam como um *Autonomous Citing Index*<sup>31</sup>. São modelos de publicação eletrônica que, dentre suas políticas, contém o aporte de indicadores, no caso do SciELO e da Redalyc. Uma terceira opção apresentada é uma iniciativa que se baseia na contribuição dos autores, o *getCITED*.

##### a) Indicadores Bibliométricos SciELO e SciELO em números<sup>32</sup>

Dentre as ferramentas inseridas em modelos eletrônicos de publicação científica, destaca-se o SciELO, criado para facilitar o acesso à literatura, contribuindo para superação da “ciência perdida” - como é chamado um fenômeno relacionado à dificuldade de penetração das descobertas feitas nos países de terceiro mundo naqueles industrializados. O projeto é originado da cooperação entre a FAPESP e o Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (BIREME),

---

<sup>31</sup> Um sistema que capaz de criar um índice de citação por meio da varredura de literatura em formato eletrônico (LAWRENCE; LEE GILES; BOLLACKER, 1999).

<sup>32</sup> <http://www.scielo.org/>

instituições nacionais e internacionais da comunicação científica e editores, tendo o marco inicial de operação o ano de 1998. Este modelo apoia-se em três componentes, sendo:

- a) o desenvolvimento de alianças entre os atores da comunicação científica;
- b) a aplicação da metodologia na operação de coleções nacionais, como já ocorre na Argentina, Brasil, Chile, Colômbia, Cuba, Espanha, México, Portugal e Venezuela;
- c) a metodologia.

Por meio do item “c”, o SciELO possibilita a edição e publicação de periódicos com organização e recuperação dos dados e produção de indicadores estatísticos de uso e impacto da literatura. Sua base, de onde é possível obter esses indicadores, define uma série de critérios, cujo objetivo é contribuir para formação e desenvolvimento de uma coleção-núcleo de periódicos com um padrão de qualidade semelhante ao de revistas científicas internacionais de destaque.

O escopo do SciELO pode ser resumido como uma coleção de revistas científicas que publicam com periodicidade artigos resultantes de pesquisa científica original, com um processo de revisão e aprovação obtido com intermédio dos pares. Há definição de critérios específicos para admissão e também permanência, sendo que para o segundo grupo, tratam-se de indicadores divididos entre uso (números de acesso) e impacto (análise de citação). O SciELO é ainda um provedor de dados do modelo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), o qual permite a interoperabilidade com outros sistemas que compartilham desse protocolo.

Em números<sup>33</sup>, a base do SciELO integra cerca de 750 periódicos, o que atinge a faixa de 300.000 artigos e um total de 6 milhões de citações. Contempla todas as grandes áreas da ciência e está presente nos seguintes países: Argentina, Brasil, Chile, Colômbia, Cuba, Espanha, México, Portugal, Venezuela; além de perspectiva de implantação em outros seis países.

É nesse contexto que se inserem duas áreas do sistema, “Indicadores Bibliométricos da Rede SciELO” e “SciELO em números”, ambas relacionadas ao uso de indicadores. Atendendo a um dos componentes do modelo de publicação, as áreas de indicadores apoiam-se nos dados da base de indicadores de SciELO, uma vez que

---

<sup>33</sup> Consulta efetuada ao “SciELO em números” em 2/3/2011, <http://www.scielo.org>

os periódicos integrantes procedem a marcação de artigos citados como requisito de ingresso e permanência no sistema. As interfaces de acesso a todo sistema estão disponíveis em português, espanhol ou inglês.

A área de Indicadores Bibliométricos é onde são disponibilizados os dados no formato de planilhas eletrônicas compatíveis com o programa *Microsoft Excel*. Não há documentação, ajuda, tutorial ou suporte a respeito do tema. Os dados são referentes ao período entre o ano 2000 e 2008.

As planilhas eletrônicas estão divididas entre três coleções: Indicadores de Publicação, Indicadores de Coleção e Indicadores de Citação (Quadro 4).

Tipo de Indicador	Conteúdo
Produção	<b>Total da rede SciELO por:</b> Ano, periódico, assunto e país de afiliação do autor. <b>País de afiliação do autor por:</b> Ano, periódicos, país de publicação da revista, assunto. <b>Número de coautores por:</b> periódico, assunto.
Coleção	<b>Periódico por:</b> ano, assunto, indicadores gerais.
Citação	<b>Ano de citação por:</b> Idade do documento, tipo de documento. <b>Periódico citante por:</b> Idade do documento citado, tipo de documento citado, <b>Assunto do periódico citante por:</b> idade do documento citado, tipo de documento citado, periódico citado. <b>País de afiliação do autor citante por:</b> idade do documento citado, tipo do documento citado, periódico citado.

QUADRO 4: INDICADORES BIBLIOMÉTRICOS REDE SCIELO

FONTE: SciELO (2010)

O “SciELO em números” foi inaugurado como parte de uma reestilização do portal eletrônico, no ano de 2007. Ao contrário do conteúdo citado anteriormente, está destinado à exibição dos indicadores e não somente a disponibilização dos dados. A funcionalidade está disponível no navegador *web* mas os resultados não podem ser exportados para outros sistemas. Há uma instrução de uso na área de “Novidades” do *site* novo, mas não um conteúdo relacionado durante as operações com a ferramenta. Ao contrário dos Indicadores Bibliométricos SciELO, essa nova ferramenta tem cobertura de publicações em todo o período de existência da rede, ou seja, de 1997 até 2011 (com atualização semanal).

O SciELO em números está dividido entre três categorias de indicadores. O “Uso do *Site*” permite verificar número de acesso às revistas, aos artigos por idioma, aos fascículos, títulos mais visitados e artigos mais visitados do mês - visibilidade. A categoria “Citações” pode ser filtrada por país, janela temporal e periódico, possibilitando acesso aos indicadores: fascículos, artigos, citações concedidas,

citações recebidas, média de artigos por fascículo, FI com janela temporal de dois ou três anos e vida média. A categoria “Coautoria” permite verificar essa ocorrência ao selecionar nações, janela temporal e periódico.

#### b) Redalyc

A *Red de Revistas Científicas de América Latina y el Caribe, España e Portugal* (Redalyc)<sup>34</sup> foi aberta em 2002 com o intuito de manter um local de exposição de periódicos *online*, de livre acesso, contribuindo com a difusão da atividade científica produzida e a respeito da Ibero-América, haja vista a escassa visibilidade que a região tem no meio científico. Inicialmente voltada para Ciências Humanas e Sociais, em 2006 abriu espaço para ingresso das outras áreas, integrando uma base multidisciplinar. Os periódicos que compõem o escopo da Redalyc seguem a metodologia *Latindex*<sup>35</sup> para consolidar o acervo com publicações que cumpram requisitos e critérios editoriais, tendo como premissas a revisão por pares e apresentação de resultados com originalidade. No total são cerca de 730 periódicos, com 170 mil artigos de texto completo. A Redalyc também tem a função de provedor de dados por meio do protocolo OAI-PMH, pelo qual é possível efetuar intercâmbio dos dados com sistemas compatíveis.

No que diz respeito às funções e facilidades da Redalyc, o sistema provê acesso a uma gama de indicadores bibliométricos, com intuito de melhorar a comunicação entre pares e permitir a avaliação dinâmica do conteúdo da base, principalmente voltado para editores científicos que podem utilizar para fins de tomada de decisão.

A Redalyc apresenta dois serviços voltados para publicações de estatísticas: os indicadores de uso, baseados em métodos webométricos por *logs* no sistema, contabilizando cliques, acessos e *downloads*; e o *Atlas Cienciométrico*, que é um somatório dessas funções com a aplicação do método de análise de citação, em um painel consolidado. Os indicadores e respectivos filtros são apresentados de acordo com as categorias e descrições:

---

<sup>34</sup> <http://atlasrscf.redalyc.uaemex.mx/acibr/>

<sup>35</sup> Um sistema regional de informação *online* para revistas científicas da América Latina, Caribe e Espanha, no qual a Redalyc está inserida e herda características como objetivo e enfoque.

- a) **Resumo:** *downloads* mensais por ano, total de publicações e total de autores; totais por instituição de ensino.
- b) **Produção:** total de publicações, por área, tipo e janela temporal.
- c) **Colaboração:** total de autores, média de autores por artigo, médio de autores por artigos e por ano, e total de trabalhos em colaboração
- d) **Uso:** *downloads* mensais.
- e) **Citações:** total de citações, citações à publicações periódicas ou não, citações a publicações nacionais ou não, citações por ano, citações por país, citações por área, citações por instituição, média de citações por artigos, autocitações.
- f) **Revistas:** colaboração (artigos em colaboração, média de autores por trabalho por ano, distribuição autoral e total de autores), trabalhos (média de artigos por número, trabalhos por área, total por tipologia, e total de trabalhos), uso (distribuição entre países, média de *downloads* mensais).
- g) **Redes:** representação das redes de cooperação entre universidades, áreas da ciência e periódicos.

A ferramenta caracteriza-se como um painel de resultados, voltada àqueles diretamente envolvidos com a comunicação científica, o que se comprova pela necessidade cadastro com vinculação institucional e ausência de documentação ou maior informações sobre a totalidade desses indicadores, possivelmente por se voltar a um público-alvo que detém conhecimento sobre o assunto.

b) *getCITED*<sup>36</sup>

Foi criada em 2000 com o intuito de facilitar a disseminação e a discussão na pesquisa acadêmica por meio de uma base de dados e um fórum de discussões. O diferencial está na ação dos autores como contribuidores, ou seja, todo o conteúdo é inserido por eles. Este preenchimento permite a inserção dos dados dos documentos científicos e também de suas citações, o que forma um índice de citação autônomo, pelo qual é possível gerar alguns indicadores.

O *getCITED* tem caráter multidisciplinar, apresenta-se como um *site* na *web*, somente no idioma inglês. A base é composta pelo conteúdo que qualquer autor registrado insere, o que dificulta a definição de um critério ou escopo que a

---

<sup>36</sup> <http://www.getcited.org/>

caracterize. De maneira estimada<sup>37</sup>, são 40 mil citações, com 392 mil identidades (pesquisadores registrados), 13 mil instituições e 3,3 milhões de documentos acadêmico-científicos (artigos de periódicos, livros, capítulos de livros, anais de eventos, relatórios entre outros).

Os itens que se encontram na base podem ser editados por qualquer usuário registrado, ou seja, a ideia é que os dados sejam revistos e avaliados pelos próprios pares (e também alterados por eles). Com a ausência de critérios para inserção, não é possível definir com precisão a janela temporal de cobertura. O sistema possui ajuda, fórum e instruções para auxiliar o usuário. Os indicadores gerados são por autor, instituição ou por trabalho, sendo o total de citações recebidas, visualizações e citações para outros documentos, além de outros baseados no princípio da rede como: total de comentários, total de citações nos comentários, e qualidade com base no *ranking* de notas atribuídas pela comunidade. Além desses indicadores, há uma área de relatórios genéricos, na qual são divulgados índices de “10” ou “100” autores, instituições ou documentos mais citados.

## 8.2 FERRAMENTAS DE BASES ORIENTADAS A DISCIPLINAS OU GRUPOS ESPECÍFICOS

Além das ferramentas pautadas em bases multidisciplinares, existem aquelas que são orientadas para determinadas disciplinas ou grupo delas, com finalidade semelhante de auxiliar na recuperação da informação e publicação de estatísticas e indicadores a respeito da rede de publicações. Conforme divisão do tópico anterior, nesse caso também existem aquelas “próprias”, ou seja, que tem uma base para aplicação (MSAcademic, SSRN, CiteSeer X) e as que utilizam uma base externa (CitEc).

---

<sup>37</sup> Consulta efetuada em 2/3/2011.

a) *Microsoft Academic Research (MSAcademic)*<sup>38</sup>

A ferramenta foi criada em 2009 pela *Microsoft Research Asia*, como uma busca acadêmica, direcionada para disciplinas da Ciência da Computação<sup>39</sup>, com cerca de 8 milhões de documentos, 5 milhões de autores e atualização semanal<sup>40</sup>. O serviço se propõe a prover um modo inovador de explorar artigos, anais de eventos, periódicos e autores, por meio da formação de uma rede de pesquisadores, editores e outros usuários. A interface do sistema é no idioma inglês, em um *site*. Apresenta documentação sobre o funcionamento e a construção do sistema, bem como tutoriais para auxiliar o usuário na navegação.

Assim como ocorre nas ferramentas baseadas no GS, os resultados são gerados com a inserção de palavra-chave no buscador, podendo ser (pela busca avançada), direcionadas por autor, eventos científicos, periódicos, ano de publicação ou busca conjugada. O sistema sugere que os usuários contribuam com a melhora da base ao inserir e modificar informações a respeito dos dados apresentados. Existem ainda recursos para visualização de redes de colaboração e citação, bem como navegação por documentos citados a partir dos resultados da busca.

Os filtros disponíveis são: autoria, janela temporal, evento científico, periódico, instituição ou subdisciplina. Cabe ressaltar que o sistema tem controle próprio de redundância para os filtros, ou seja, o usuário precisa selecionar a respectivo objeto da unidade de análise, por exemplo, o nome do autor, em uma lista fornecida pelo *MSAcademic* antes ou depois da busca inicial. Os indicadores apresentados pela ferramenta são: o total de publicações, o total de citações recebidas, o índice-g, o índice-h, o total de coautores, a janela temporal de publicação e a representação da rede de coautoria.

b) *Social Science Research Network (SSRN)*<sup>41</sup>

A SSRN é uma iniciativa *online* que congrega um conjunto de redes especializadas na disseminação da pesquisa em Ciências Sociais. O conteúdo é obtido por meio de um processo de submissão direta ou via as redes menores, ou

---

<sup>38</sup> <http://academic.research.microsoft.com/>

<sup>39</sup> Por uma falta de uniformidade na divisão das subáreas da ciência entre país e instituições, nessa ferramenta são consideradas algumas disciplinas relativas à Ciência da Informação que, por sua vez, inclui-se nas CHSA no caso da realidade brasileira.

<sup>40</sup> Consulta efetuada em 2/3/2011.

<sup>41</sup> <http://www.ssrn.com/>

ainda na formação de parcerias com editores e instituições. A biblioteca *online* do SSRN é dividida em textos completos e resumos, o que soma cerca de 600 mil trabalhos no total, apresentando a interface no idioma inglês<sup>42</sup>.

As estatísticas e indicadores fornecidos pelo SSRN utilizam um mecanismo externo chamado *CiteReader*, o qual identifica, quantifica e analisa citações por meio de um processo de varredura em documentos no formato PDF. Para tal, há o estabelecimento de normas editoriais viabilizando o processo. A exibição dos resultados é feita por meio de uma classificação, no qual são elencados autores, documentos ou instituições, com possibilidade de filtro por disciplinas das Ciências Sociais, conforme segue:

- **Janela temporal de 12 últimos meses:** Total de *downloads*, total de publicações e total de *downloads* por artigo
- **Janela temporal contabilizando todo tempo de publicação:** Total de *downloads*, total de citações, total de documentos, total de *downloads* por documento, total de citações por documento e *Eigenfactor*.

As mesmas estatísticas que estão disponíveis na área da classificação, ficam disponíveis para o usuário como resultado de qualquer busca que tenha sido efetuada na base por meio de palavras-chave ou autoria.

#### c) *CiteSeer X*<sup>43</sup>

A ferramenta é uma biblioteca digital voltada para a literatura científica nas áreas de Ciências da Computação e Ciências da Informação, além de um mecanismo de busca e recuperação da informação. Foi criada em 1997, sob a denominação *CiteSeer*, no *NEC Research Institute*, pelos pesquisadores Steve Lawrence, Lee Giles e Kurt Bollacker. Em 2003 o projeto foi transferido para a *Pennsylvania State University's College of Information Sciences and Technology*.

O sistema é denominado como *Autonomous Citation Indexing*, ou seja, uma base que cria índice de citação por meio de uma varredura em diversas fontes, tal como ocorre com o GS. O sistema disponibiliza seu código, provendo algoritmos, metadados, técnicas e *softwares* que podem embasar o desenvolvimento de soluções análogas em outras bibliotecas digitais. Os dados disponíveis no sistema podem ser recuperados e acessados pelo padrão OAI-PMH. Ainda assim, o *CiteSeer X* não

<sup>42</sup> Consulta efetuada em 2/3/2011.

<sup>43</sup> <http://citeseer.ist.psu.edu/>



define a maneira como o algoritmo faz a varredura e tampouco o que é científico ou não, baseando-se na presença de elementos em documentos *web*. A interface é em inglês, estando disponível por meio de um *site*. Há instruções de uso sobre o sistema mas não sobre os indicadores e resultados.

Aos resultados provenientes de buscas na base são acrescentadas algumas metrias da informação, comportando-se de modo diferente se a busca é por autor ou por documento. No primeiro caso são exibidos o número de citações, o número de trabalhos e o índice-h. No segundo, apenas o total de trabalhos e o de citações. É possível filtrar os resultados pela autoria, periódico, palavras-chave, janela temporal e número mínimo de citações.

#### d) *CitEc*<sup>44</sup>

A ferramenta *CitEc*, de autoria do pesquisador José Manuel Barrueco da *Universidad de Valencia*, foi criada em 2004 como implementação do sistema *CiteSeer* de indexação automática de citações, na biblioteca digital do *Research Papers in Economics Project* (Repec). Esta base é formada pelo trabalho de contribuidores em 73 países, na busca por uma composição descentralizada de documentos científicos da pesquisa na área de Economia. A inserção de materiais é feita por meio das parcerias com instituições de pesquisa e editoras científicas, por exemplo: *Elsevier*, *Wiley Blackwell*, *Springer* e *American Economic Association*. Nesse sentido, a base conta com cerca de 1 milhão de documentos<sup>45</sup>, entre artigos, *softwares*, livros e capítulos, somente de fontes registradas.

O *CitEc* apoia-se na extração de dados feita em cada arquivo contido na base *Repec*, como forma de criar um índice que aponte dados estatísticos de citação. Os dados resultantes da varredura não são disponibilizados diretamente para os usuários, apenas os índices já consolidados. A janela temporal dos resultados é entre 2004 e 2008, com documentos do tipo *preprint* ou artigos de periódicos registrados. O sistema apresenta-se no idioma inglês, com visualização dos resultados no navegador *web*, contando com uma pequena descrição das funcionalidades. O único indicador apresentado é o FI, sendo possível filtrar por país de origem do periódico e ano base do cálculo.

---

<sup>44</sup> <http://citec.repec.org/>

<sup>45</sup> Consulta efetuada em 2/3/2011.

Tendo conhecimento dos pontos fortes e fracos das alternativas, mediante sua possível aplicação na área de CHSA e considerando os aspectos levantados em capítulos anteriores no que diz respeito as peculiaridades e necessidades da área, foi possível conduzir uma discussão sobre o tema, conforme apresentado no tópico seguinte.

### 8.3 DISCUSSÃO

A história de evolução e desenvolvimento da pesquisa científica moldaram a maneira como cada área do conhecimento representa a sua contribuição, demonstrada por meio da enumeração das características da produção na academia, distintas e portanto, influentes nos resultados de comparações globais. Conforme discussão apresentada no capítulo 7.3, os indicadores de impacto e uso da produção científica devem ser pensados e utilizados de maneira distinta entre as áreas. Nesse sentido, foram apresentadas as tendências que a fonte dos indicadores – os índices de citação tradicionais – detém e os invalidaram até então no caso das Ciências Humanas e Sociais Aplicadas. A discussão incluiu ainda a viabilidade do uso de parte deles, como o índice-h e Fator de Impacto, desde que os fatores de temporalidade e densidade das citações nas CHSA fossem considerados. Neste capítulo foram apresentadas as alternativas *web* para uma possível melhoria da situação do uso de indicadores quantitativos e, conforme segue, elaborada uma discussão afinal a respeito do tema.

Com base nas ferramentas *web* estudadas, percebeu-se que algumas delas podem abrir caminho para uma solução na área de CHSA. Considerando o princípio de um sistema de avaliação em potencial como o somatório de uma base adequada com uma ferramenta estatística que permita tratar os dados, sugerido por Nehaus e Daniel (2008), sugere-se um cenário dividido, genericamente, em duas categorias:

- a) bases de dados com índices de citação gerados por meio de varredura, com um ou mais algoritmos dedicados à verificação de citações;
- b) bases de dados com índices de citação provenientes de marcação de citações, uma a uma, trabalho a trabalho, ou por meio do intercâmbio de metadados com uso de um protocolo padronizado.

No primeiro grupo o destaque e alvo principal de pesquisas na área é o *Google Scholar*, que apoia-se no potencial técnico do motor de busca *Google*, e não em um estudo aprofundado sobre a recuperação da informação científica. Ainda assim, é capaz de transpor as principais tendenciosidades no uso das bases tradicionais; reduzindo ou eliminando as barreiras idiomáticas, igualando a exibição dos resultados e importância na cobertura de periódicos e incluindo outros canais de comunicação científica como livros, monografias, anais de eventos, *preprints* entre outros.

A “impureza” dos dados encontrados no GS é o principal problema identificado ao ser constantemente comparado com o *Web of Science*. O produto da *Thomson ISI* possui uma política e construção de um serviço por trás de sua base, diferente de um algoritmo que faz uma varredura por aproximação. Esclarece-se, com base em Harzing e Van der Wal (2008), que a parcela de documentos “impuros” ou desconsiderados já é bem inferior ao início do funcionamento do GS (em 2004), haja vista que parte dos registros da base já é originário de sistemas de publicação eletrônica, repositórios institucionais, ou mesmo editoras parceiras, por meio da coleta de metadados. Um argumento ainda seria que essa origem também não traduz confiabilidade, considerando que parte dos sistemas não têm uma política editorial adequada ou qualidade dos dados/metadados empregados.

Considerar ou não o GS como uma fonte de informação para indicadores de uso e impacto é sinônimo de investimento contínuo no tratamento dos dados recuperados, com enfoque na necessidade de avaliação de cada interessado. A colocação está alinhada ao que Armbuster (2009) defende a respeito de uma realidade com muitos dados e indicadores possíveis, na qual é necessário definir para qual função serão empregados, trabalhando-os de maneira personalizada.

A lacuna entre o todo – a base inteira –, e as necessidades específicas dos interessados em avaliar o impacto da produção científica em CHSA pode ser preenchida parcialmente pelas ferramentas *web* apoiadas no GS. O *scholar Index*, *Citation-Gadget for iGoogle* e *Scholar H-Index Calculator*, oferecem interfaces para que o usuário final (que efetua buscas no GS) possa acessar rapidamente poucos indicadores, pois há poucas funções ou filtros existentes se comparadas ao *Scholarometer* e *Publish or Perish*.

Estes dois últimos contam com extensa documentação, apoio técnico sobre os indicadores, possibilidade de intercâmbio para outras ferramentas estatísticas e, principalmente, facilidades com o tratamento dos dados e indicadores, conforme apresentado anteriormente no Quadro 3 (p. 111). São ferramentas que, ao serem exploradas, podem extrapolar os limites das dificuldades elencadas para o uso de indicadores pelas CHSA como, por exemplo, a superação da barreira idiomática com filtro por nacionalidade, a superação da cobertura de documentos pela característica da base, a seleção da unidade de análise pela autoria, instituição, ou área da ciência.

Porém, cabe um questionamento após a apresentação dessas ferramentas; o quanto viável é utilizá-las na avaliação do uso da produção científica em CHSA, no contexto dos órgãos de fomento? Conforme relatado, com uso de uma série de filtros e tratamentos, seria possível discuti-las na perspectiva de um aprimoramento / complementação nos métodos de avaliação. Contudo, há outra questão subsequente: Quais seriam os recursos necessários para que este tratamento fosse feito e, a relação investimento-resultados que justifique o emprego de fundos para essa finalidade intermediária?

Percebeu-se que estas questões também permeiam os estudos internacionais sobre o uso do GS, sendo que as pesquisas até o momento se direcionam para o mapeamento da atividade científica, ou, para outros estudos cientométricos, como forma de compreender a evolução de disciplinas e suas relações e não, necessariamente, como uma ferramenta empregada na avaliação por órgãos de fomento. A explicação estaria na necessidade de constante refinamento dos dados e, a própria permanência do serviço no mercado que, em sendo gratuito, não apresenta garantias a longo prazo de sua estabilidade, conforme apontado por Marchiori (2007).

No que diz respeito às outras ferramentas *web* de bases de dados que são preenchidas com a técnica de varredura automática, não foram localizados trabalhos comparativos na literatura de referência. Ainda assim, é possível tecer considerações a respeito de seu potencial.

O MSAcademic está focado na formação de redes de pesquisadores para uma área da ciência específica, as Ciências da Computação (ainda que contemple disciplinas de Ciência da Informação). O seu diferencial em relação ao GS é a atuação conjunta entre editores científicos e autores; os primeiros por meio de parcerias transparentes com a base de dados, no fornecimento de metadados das publicações, e os outros como reguladores da qualidade dos dados apresentados, por meio da

interação e edição. Ainda assim, não apresenta aplicação direta nas CHSA, por ter um escopo que permeia mais as Ciências Exatas do que esta, além de uma tendenciosidade com relação ao idioma inglês. A ferramenta serve como uma possível sugestão de modelo, na qual há interação dos vários atores do processo da produção científica.

Assim como o MSAcademic, o *CiteSeer X* se volta para disciplinas que permeiam duas áreas, as CHSA e as Exatas, o que a invalida como solução direta para a questão da base. Ainda assim, também tem atributos que sugerem adaptações e potencialidades na concepção de um serviço de informação análogo. O *CiteSeer X* é, além de um índice de citação, o provedor de um algoritmo de código aberto voltado à varredura automática de bases de dados independentes, permitindo o reconhecimento de referências bibliográficas. Isso significa que, mediante o estabelecimento de uma padronização de formatos no depósito de documentos em uma base, é possível usar esse algoritmo para criar automaticamente um índice de citação e, portanto, publicar indicadores de impacto. Esse é o caso da ferramenta CitEc, que usa o código do *CiteSeer X* para publicar seus próprios indicadores a partir da base Repec. A facilidade de tratar os dados em uma base controlada (ao contrário do GS), é proporcional à diminuição de custos na formulação de índices de citação, considerando possíveis políticas implantadas e o uso de um algoritmo gratuito de varredura. A solução pode atender, inclusive, bases já existentes. É o caso do SSRN, que implementou um algoritmo chamado *CiteReader* (fornecido por uma empresa, de código fechado) para criar um índice de citação em sua base já existente.

A Redalyc também apresenta um modelo de solução alinhado à varredura automática por meio de um algoritmo, mediante o estabelecimento de critérios obrigatórios a serem atendidos pelos editores de periódicos, como o uso de algum padrão de referências, por exemplo, o *American Psychological Association* (APA), a ser definido na política editorial do periódico submetido à base. Nota-se que, nesse caso, trata-se de estabelecer e verificar padrões, com subsequente redução de custo da elaboração do índice de citação. Partindo do emprego dessa técnica pela Redalyc, é criada a ferramenta *Atlas Cienciométrico*; um serviço de gestão editorial, voltado para o uso em instituições, ou mesmo, gestores de órgãos que regulam a pesquisa nos países em que estão inseridos.

Cabe uma ressalva a respeito desse modelo. Há duas iniciativas semelhantes à esses padrões sugeridos pela Redalyc: *Counting Online Usage of Networked Electronic Resources* (COUNTER), focado na criação de códigos de prática para periódicos e bases de dados online em países europeus, objetivando o aumento da interoperabilidade entre as diferentes bases, e o *Metrics from Scholarly Usage of Resources* (MESUR), iniciativa norte-americana que sugere um modelo semântico do processo de comunicação científica que, associado a serviços de informação, é uma ferramenta em potencial no trato de dados bibliográficos, de citação e uso da informação proveniente de diferentes fontes (COUNTER, 2010; MESUR, 2010).

A solução da Redalyc mostrou-se como um exemplo alternativo pois, iniciou-se com o intuito de facilitar o acesso à publicações nas CHSA em países de idiomas latino-americanos, projetando-se posteriormente para outras áreas. Sua proposta já enfrenta duas tendenciosidades principais do *Web of Science*: o idioma e a área da ciência. A ferramenta em si – *Atlas Cienciométrico* – tem funcionalidades de filtros e indicadores similares ao *Publish or Perish* e *Scholarometer*, com a diferença que o conteúdo da base passa por um criterioso processo de aceitação. Uma dificuldade verificada é o ainda presente enfoque apenas na literatura de periódicos.

Até então foram apresentadas as possíveis soluções que criam índices mediante varredura, cabendo definir aquelas em que o índice de citação é construído por meio da marcação individual de documentos citados. Nesse grupo situa-se, inicialmente, o *getCITED*. Ao contrário de todas as outras ferramentas analisadas, está baseado no preenchimento de uma base de citação por meio da colaboração entre autores. Sendo assim, tem como princípio a desvinculação com o mercado editorial, que, segundo justificativa de seus criadores, teria o monopólio das redes científicas que deveria estar, supostamente, nas mãos dos autores. Apesar do diferencial, a proposta sofre com problemas tais como a validade dos dados ali inseridos, principalmente por não passarem por nenhum processo de revisão ou critério. Ainda nesse sentido, é difícil definir os limites do serviço, ou seja, o escopo. Dentre as ferramentas analisadas, é a disponibilizada há mais tempo (desde 2000), porém, verificou-se durante a pesquisa que as atualizações podem estar suspensas desde o ano de 2006.

A base de dados SciELO também utiliza a técnica de construção do índice de citação por meio da marcação individual de referências, sendo esta a única característica que compartilha com o *getCITED*. A proposta da primeira é semelhante

a da Redalyc, porém, iniciando com enfoque à produção científica dos países em desenvolvimento com direcionamento para as Ciências da Saúde. Os benefícios do SciELO também são semelhantes a Redalyc: valorização do enfoque local, criteriosidade e qualidade dos dados, melhor condição de ingresso pelos periódicos e reconhecimento nacional.

O SciELO apresenta duas ferramentas que possibilitam a geração de indicadores. Os “indicadores bibliométricos”, que são apenas os dados para serem processados, apresentando cobertura temporal restrita (2000 a 2008); e o “SciELO em números”, uma sessão inaugurada com o portal novo (2007). Este último contém dados semanalmente atualizados e já apresenta indicadores consolidados. Os dados do SciELO já são utilizados em sistemas da avaliação e acompanhamento da produção científica no Brasil, como o Currículo Lattes. Nesse entendimento, ele apresenta maior propensão à se enquadrar enquanto proposta. A base pode ser vista como uma iniciativa semelhante ao *Web of Science*, dada a sua reconhecida criteriosidade na seleção de periódicos e no escopo de países de língua não inglesa, principalmente aqueles em desenvolvimento. Os pontos desfavoráveis de seu uso são: a restrição do canal de comunicação, o periódico e; a dificuldade de ingresso e alinhamento das áreas de CHSA, devido às características herdadas da Ciências Exatas que embasaram sua criação. Um reflexo disso é a presença do indicador FI, com respectivas janelas de cobertura de dois ou três anos.

De modo geral, as ferramentas *web* analisadas apresentam finalidades, enfoque, e técnicas para tratamentos e operação dos dados de maneira distinta o que dificulta, inclusive, a criação de uma classificação comum. Não há uma solução única que satisfaça as necessidades apresentadas na revisão de literatura neste trabalho. Porém, diversas características e potencialidades identificadas em parte delas podem sugerir a projeção de novos serviços que contemplem os objetivos elencados. Nesse sentido, são apresentadas considerações finais no capítulo seguinte.

## 9 CONSIDERAÇÕES FINAIS

A presente pesquisa teve como objetivo geral uma abordagem qualitativa voltada para verificação do uso de ferramentas *web* de acesso gratuito a resultados provenientes da busca de informação acadêmico-científica, como alternativa a utilização de indicadores de uso da produção científica nas Ciências Humanas e Sociais Aplicadas.

O estudo partiu do questionamento de possíveis regularidades existentes no comportamento dos pesquisadores de disciplinas que compõem as CHSA no que diz respeito à publicação de resultados da atividade científica. Para obter tal resposta, inicialmente foi elaborado um panorama histórico sobre a área, esclarecendo as diferenças evolutivas entre ela e as outras disciplinas da ciência. Na sequência, houve a caracterização da produção científica, no que diz respeito aos canais de comunicação, o método de avaliação e aspectos gerais. Nesse sentido, entende-se que o primeiro objetivo foi atingindo, haja vista que foram verificadas diferenças significativas, como a importância dada aos livros enquanto canal, e o comportamento ao citar e publicar dos pesquisadores em CHSA.

Um segundo questionamento tratou da influência desses resultados anteriores na aplicação de indicadores quantitativos para medição do impacto da produção científica na área alvo. Houve a necessidade inicial de uma revisão de literatura cobrindo desde as primeiras medidas científicas, até a ascensão da análise de citação como principal método empregado na avaliação do ofício da ciência. Percebeu-se que durante o período anterior a popularização das TICs (até o final da década de 1990), os indicadores de uso da produção científica permaneceram estagnados enquanto área de estudo e evolução, tendo grande aceitação nas Ciências Exatas e da Vida, por meio de um reconhecido índice de citação, o SCI.

No período posterior, a facilitação do processamento e obtenção de uma quantidade representativa de dados por meio do computador e *internet*, possibilitaram a proposição de novas metrias, tais como o índice-h, índice-g e *Eigenvector*. A evolução delas foi acompanhada por novas bases de dados com índices de citação, como a *Scopus* e o *Google Scholar*, em resposta às críticas sofridas pelo sistema *Thomson ISI* quando às tendências apresentadas.



Sob esse cenário, percebeu-se que a principal dificuldade do uso de métodos qualitativos nas CHSA não está atrelado à definição de um indicador ideal e sim, de um índice de citação proveniente de uma base de dados mais adequada e menos tendenciosa que a da *Thomson ISI*. Excetuando-se essa conclusão, ainda que em um nível menor de importância, percebeu-se que as medidas como Fator de Impacto e índice-h, podem sim ser aplicadas às CHSA, desde que certos componentes centrais do comportamento da área sejam considerados, como a temporalidade e frequência de citações. A solução viria, supostamente, da inclusão da variável obsolescência por área da ciência, na medição de impacto de autores e periódicos. Ainda assim, entende-se que o objetivo específico relacionado foi atendido parcialmente, dada a complexidade da construção dos indicadores e a própria situação, ainda emergente, das CHSA com o seu uso.

Como terceiro e último questionamento, correspondentes aos terceiro e quarto objetivos específicos, buscou-se identificar o potencial das ferramentas *web* de busca e acesso à informação científica como fonte para os indicadores, o qual foi julgado como atendido por meio das discussões que se seguem. Verificou-se que nenhuma das ferramentas estudadas seria uma alternativa, exatamente como se apresentam, aos atuais critérios de avaliação propostos por agências de fomento em CHSA. O propósito do trabalho também não foi o de criar um novo modelo, ou de sugerir que um existente seja usado e sim, de mostrar que desde o início do século XXI, diversas discussões a respeito de novas bases de dados, do emprego intensivo de tecnologia, e da customização de serviços de informação, sugerem que, mesmo áreas que não tem tradição do uso de indicadores, possam a vir adotá-los.

Ao considerar as características e funções das ferramentas, recorreu-se novamente à *Armbuster (2009)* quando posiciona que a dificuldade não está mais na tecnologia nem no método empregado e sim, na construção de serviços de informação; no caso, bases de dados com valor para a academia e, conseqüentemente, para aqueles que a fomentam. Isso indica a necessidade de personalização, ou seja, a busca por indicadores por área da ciência ou, com maior efetividade, por disciplinas, com bases de citações comuns ou não. Essas soluções, ao incluir o contexto das CHSA, se enquadram em uma dessas três possibilidades:

- a) base controlada, com inserção de dados de citação pela marcação individual, alinhada à algum padrão de intercâmbio de metadados;
- b) base controlada, com varredura de dados de citação por algoritmos, alinhada à algum padrão de intercâmbio de metadados;
- c) uso de toda a *web*, como uma grande base, com varredura por algoritmos e posterior tratamento intensivo dos dados.

Os sistemas de armazenamento digital, os quais podem ser usados como bases controladas, estão em plena expansão e, têm ganhado aceitação enquanto ferramenta para o estudo das relações na comunidade científica segundo a visão de Bollen e Van de Sompel (2008). Os autores julgam que tais sistemas são uma alternativa promissora como fonte de informação para indicadores de impacto. Outra vantagem seria a elaboração de estudos com usuários da própria comunidade como, por exemplo, a coleta de *logs* de acesso a um tipo de informação, selecionando-se um grupo, ou uma instituição. Isso refletiria o impacto percebido para aquela unidade de análise. A vantagem da utilização dessas bases, e não de toda a *web* como no caso do GS, é que as chamadas medidas de visibilidade, dentro dos estudos da Webometria, poderiam ser trabalhadas paralelamente àquelas enfocadas no método de análise de citação – uso e impacto. Nesse trabalho foram constatados dois exemplos dessa união, o *Atlas Cienciométrico* da Redalyc e o SciELO em números o que, de certa forma comprova a importância desse desenvolvimento.

Em uma análise prévia das três possibilidades, fundamentando-se na discussão apresentada, o uso de bases controladas com criação de índices de citação por meio da varredura de um algoritmo, operando em padrões de publicação preestabelecidos, como é o caso dos exemplos internacionais COUNTER e MESUR, parece a menos custosa e mais efetiva resposta ao problema desta pesquisa. Desse modo, mesmo que as bases de dados fossem descentralizadas, como é caso de revistas eletrônicas e repositórios de teses e dissertações sob a ingerência de instituições de ensino e pesquisa, o uso de protocolos comuns possibilitaria a elaboração de estudos cientométricos, bibliométricos e webométricos de acordo com a necessidade de determinada agência de fomento ou interessado na avaliação.

Em consequência da elaboração deste estudo, foi possível elencar possibilidades de trabalhos futuros, não contemplados pelo escopo definido ou pelas dificuldades apresentadas durante o tempo disponível para a pesquisa, servindo como sugestão para outros pesquisadores como continuidade.

- a) Estudo quali-quantitativo sobre o comportamento e uso da citação em CHSA, tendo como base uma tipologia de citações, como a apresentada no Quadro 1;
- b) Estudo comparativo de indicadores obtidos por meio das bases *Google Scholar*, *Scopus* e *Web of Science* tendo disciplinas ou instituições brasileiras como unidade de análise, valendo-se das ferramentas *Publish or Perish* e *I* ou *Scholarometer*;
- c) Estudo comparativo de resultados obtidos por diferentes ferramentas *web*, na recuperação e tratamento de dados oriundos da base *Google Scholar* para uma área piloto em CHSA;
- d) Diagnóstico da situação atual do uso de protocolos intercâmbio de metadados em periódicos e bancos de teses e dissertações em universidades brasileiras, com enfoque nas disciplinas de CHSA;
- e) Estudo comparativo de resultados obtidos a respeito da produção científica no formato livro, de autores brasileiros em determinada disciplina de CHSA, entre as bases *Google Books* (sistema de busca de livros do *Google* com metadados e recuperação de citações por meio do *GS*) e *Book Citation Index* da *Thomson ISI* (lançamento em 2011);
- f) Estudos voltados para análise de tópicos emergentes com base na ocorrência de palavras em obras, com uso de ferramentas emergentes, como o *Google Books Ngram Viewer*, um serviço piloto do *Google* que apresenta esses resultados a partir de sua própria base de dados com enfoque em livros.

## REFERÊNCIAS

- ABEC. Fórum de Ciências Humanas. In: ENCONTRO NACIONAL DE EDITORES CIENTÍFICOS, 2009, Águas de Lindóia. **Anais...** Águas de Lindóia: 2009. Disponível em: <<http://www.abecbrasil.org.br/includes/palestraslindoia/sex/tarde/Humanas2.pdf>>. Acesso em: 14/11/2010.
- AGUILLO, I. F.; BAR-ILAN, J.; LEVENE, M.; ORTEGA, J. L. Comparing university rankings. **Scientometrics**, v. 85, n. 1, p. 243-256, 2010.
- ALEIXANDRE-BENAVENT, R.; VALDERRAMA-ZURIÁN, J. C.; GONZÁLES-ALCAIDE, G. El factor de impacto de las revistas científicas: limitaciones e indicadores alternativos. **El Profesional de la Información**, v. 16, n. 1, p. 4-11, 2007.
- ALMIND, T. C.; INGWERSEN, P. Informetric analyses on the world wide web: methodological approaches to webometrics. **Journal of Documentation**, v. 53, p. 404--426, 1997.
- ALTHOUSE, B. M.; WEST, J. D.; BERGSTROM, T. C.; BERGSTROM, C. T. Differences in Impact Factor Across Fields and Over Time. **Departmental Working Papers**. 2008. Disponível em: <<http://www.escholarship.org/uc/item/76h442pg>>. Acesso em: 25/10/2010.
- AMIM, M.; MABE, M. A. Impact factors: use and abuse. **Medicina**, v. 63, n. 4, 2003. Disponível em: <[http://www.scielo.org.ar/scielo.php?script=sci\\_arttext&pid=S0025-76802003000400011&lng=es&nrm=iso&tlng=en#](http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S0025-76802003000400011&lng=es&nrm=iso&tlng=en#)>. Acesso em: 23/2/2011.
- ARAÚJO, C. Bibliometria: evolução histórica e questões atuais. **Em Questão**, v. 12, n. 1, 2006. Disponível em: <<http://www.seer.ufrgs.br/index.php/EmQuestao/article/viewArticle/16>>. Acesso em: 21/2/2010.
- ARCHAMBAULT, É.; GAGNÉ, É. V. The use of bibliometrics in the Social Sciences and Humanities. **Science Metrix**. 2004. Disponível em: <[http://www.science-metrix.com/pdf/Science-Metrix\\_Use\\_Bibliometrics\\_SSH.pdf](http://www.science-metrix.com/pdf/Science-Metrix_Use_Bibliometrics_SSH.pdf)>. Acesso em: 14/8/2010.
- ARCHAMBAULT, É.; VIGNOLA-GAGNÉ, É.; CÔTÉ, G.; LARIVIÈRE, V.; GINGRASB, Y. Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. **Scientometrics**, v. 68, n. 3, p. 329-342, 2006.
- ARMBRUSTER, C. Whose Metrics? On Building Citation, Usage and Access Metrics as Information Service for Scholars. **SSRN eLibrary**. 2009. Disponível em: <[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1464706](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1464706)>. Acesso em: 2/12/2009.
- ATHAYDE, P. **Manual para redação acadêmica**. Belo Horizonte: Keimelion, 2002.
- AULETE, C. **Minidicionário contemporâneo da língua portuguesa**. 2 ed. Rio de Janeiro: Lexicon, 2009.

BAILEY, T. P. Electronic Book Usage at a Master's Level I University: A Longitudinal Study. **The Journal of Academic Librarianship**, v. 32, n. 1, p. 52-59, 2006.

BAKKALBASI, N.; BAUER, K.; GLOVER, J.; WANG, L. Three options for citation tracking: Google Scholar, Scopus and Web of Science. **Biomedical Digital Libraries**, v. 3, n. 1, p. 7, 2006.

BANKS, M. G. An extension of the Hirsch index: Indexing scientific topics and compounds. **Scientometrics**, v. 69, n. 1, p. 161-168, 2006.

BAR-ILAN, J. Manipulating search engine algorithms: the case of Google. **Journal of Information, Communication and Ethics in Society**, v. 5, n. 2/3, p. 155-166, 2007.

BAR-ILAN, J. Informetrics at the beginning of the 21st century: a review. **Journal of Informetrics**, v. 2, n. 1, p. 1-52, 2008.

BAR-ILAN, J. Web of Science with the Conference Proceedings Citation Indexes: the case of computer science. **Scientometrics**, v. 83, n. 3, p. 809-824, 2010.

BARRETO, A. Mudança estrutural no fluxo do conhecimento: a comunicação eletrônica. **Ciência da Informação**, v. 27, n. 2, p. 122-127, 1998.

BATISTA, P. D.; CAMPITELI, M. G.; KINOUCI, O. Is it possible to compare researchers with different scientific interests? **Scientometrics**, v. 68, n. 1, p. 179-189, 2006.

BAUER, K.; BAKKALBASI, N. An Examination of Citation Counts in a New Scholarly Communication Environment. **D-Lib Magazine**, v. 11, n. 09, 9.2005. Disponível em: <<http://www.dlib.org/dlib/september05/bauer/09bauer.html>>. Acesso em: 7/10/2009.

BENSMAN, S. Garfield and the Impact Factor. In: **Annual review of information science and technology**. Information Today, Inc., 2006. Disponível em: <[http://books.google.com.br/books?id=8\\_HOSmPie7wC&printsec=frontcover#v=onepage&q&f=false](http://books.google.com.br/books?id=8_HOSmPie7wC&printsec=frontcover#v=onepage&q&f=false)>. Acesso em: 11/8/2010.

BENTLEY, R. A.; ORMEROD, P.; BATTY, M. An evolutionary model of long tailed distributions in the social sciences. **arXiv.org**. 2009. Disponível em: <<http://arxiv.org/abs/0903.2533>>. Acesso em: 7/6/2010.

BJÖRNEBORN, L. **Small-world link structures across an academic web space**. 399 f. Dissertation (PhD in Department of Information Studies) – Royal School of Library and Information Science, Copenhagen, 2004. Disponível em: <<http://vip.db.dk/lb/phd/phd-thesis.pdf>>. Acesso em: 21/1/2011.

BJÖRNEBORN, L.; INGWERSEN, P. Toward a basic framework for webometrics. **Journal of the American Society for Information Science and Technology**, v. 55, n. 14, p. 1216-1227, 2004.

BOLLEN, J.; RODRIQUEZ, M. A.; VAN DE SOMPEL, H. Journal status. **Scientometrics**, v. 69, n. 3, p. 669-687, 2006.

- BOLLEN, J.; VAN DE SOMPEL, H. Usage impact factor: The effects of sample characteristics on usage-based impact metrics. **Journal of the American Society for Information Science and Technology**, v. 59, n. 1, p. 136-149, 2008.
- BOLLEN, J.; VAN DE SOMPEL, H.; HAGBERG, A.; CHUTE, R. A Principal Component Analysis of 39 Scientific Impact Measures. **PLoS ONE**, v. 4, n. 6, p. e6022, 2009.
- BORDONS, M.; FERNÁNDEZ, M.; GÓMEZ, I. Advantages and limitations in the use of impact factor measures for the assessment of research performance. **Scientometrics**, v. 53, n. 2, p. 195-206, 2002.
- BORGMAN, C. L.; FURNER, J. Scholarly communication and bibliometrics. **Annual Review of Information Science and Technology**, v. 36, n. 1, p. 2-72, 2002.
- BORNMANN, L.; DANIEL, H. What do citation counts measure? A review of studies on citing behavior. **Journal of Documentation**, v. 64, n. 1, p. 45-80, 2008.
- BRAUN, T.; GLÄNZEL, W.; SCHUBERT, A. **Scientometric indicators: a 32 country comparative evaluation of publishing performance and citation impact**. Philadelphia: World Scientific Publishing, 1985.
- BROADUS, R. N. Early approaches to bibliometrics. **Journal of the American Society for Information Science**, v. 38, n. 2, p. 127-129, 1987.
- BROWN, C. The E-volution of preprints in the scholarly communication of physicists and astronomers. **Journal of the American Society for Information Science and Technology**, v. 52, n. 3, p. 187-200, 2001.
- CAMPBELL, F. **The Theory of the National and International Bibliography: with Special Reference to the Introduction of System in the Record of Modern Literature**. London: Library Bureau, 1896.
- CAPES. **Mestrados e Doutorados reconhecidos** Disponível em: <<http://www.capes.gov.br>>. Acesso em: 28/9/2009.
- CASE, D. O.; HIGGINS, G. M. How can we investigate citation behavior? A study of reasons for citing literature in communication. **Journal of the American Society for Information Science**, v. 51, n. 7, p. 635-45, 2000.
- CITEBASE. **Citebase** Disponível em: <<http://www.citebase.org>>. Acesso em: 12/2/2009.
- CITSEER X. **About** Disponível em: <<http://citseerx.ist.psu.edu/>>. Acesso em: 12/2/2009.
- COLLINS, J.; HUSSEY, R. **Pesquisa em administração**. 2 ed. Porto Alegre: Bookman, 2005.
- CONPAF. **CONPAF** Disponível em: <<http://www.conpaf.org.br>>. Acesso em: 11/4/2009.

COSTAS, R.; BORDONS, M. Is g-index better than h-index? An exploratory study at the individual level. **Scientometrics**, v. 77, n. 2, p. 267-288, 2008.

COUNTER. **Online Usage of Electronic Resources**. Disponível em: <<http://www.projectcounter.org/>>. Acesso em: 11/1/2011.

COUTINHO, E. As armadilhas da lei de Bradford. **Rev. de Biblioteconomia de Brasília**, v. 16, n. 2, 1988. Disponível em: <<http://164.41.105.3/index.php/RBB/article/view/461>>. Acesso em: 22/2/2010.

CRAFT, R. C.; LEAKE, C. The Pareto principle in organizational decision making. **Management Decision**, v. 40, n. 8, p. 729 - 733, 2002.

CRESWELL, J. **Projeto de Pesquisa: métodos qualitativo, quantitativo e misto**. 2 ed. Porto Alegre: Artmed, 2007.

CRONIN, B. Bibliometrics and beyond: some thoughts on web-based citation analysis. **Journal of Information Science**, v. 27, n. 1, p. 1-7, 2001.

CYBERNATICS LAB. **Ranking web of world universities** Disponível em: <<http://www.webometrics.info>>. Acesso em: 24/1/2011.

DELL'ORSO, F. Bibliography Management Software: Analysis and Comparison of some packages. 1. may. 2008. Disponível em: <<http://www.burioni.it/forum/dellorso/bms-dasp1/text/index.html>>. Acesso em: 18/11/2010.

DIODATO, V. P. **Dictionary of bibliometrics**. Routledge, 1994.

DONG, P.; LOH, M.; MONDRY, A. The "impact factor" revisited. **Biomedical Digital Libraries**, v. 2, n. 1, p. 7, 2005.

DROTT, M. Bradford's Law: Theory, Empiricism and the Gaps Between. **Library Trends**, v. 30, n. 1, p. 41-52, 1981.

EGGHE, L. Expansion of the field of informetrics: Origins and consequences. **Information Processing & Management**, v. 41, n. 6, p. 1311-1316, 2005.

EGGHE, L. Theory and practise of the g-index. **Scientometrics**, v. 69, n. 1, p. 131-152, 2006.

EGGHE, L.; ROUSSEAU, R. Aging, obsolescence, impact, growth, and utilization: definitions and relations. **Journal of the American Society for Information Science**, v. 51, n. 11, p. 1004-1017, 2000.

EUGÊNIO, M.; FRANÇA, R.; PEREZ, R. C. Ciência da Informação sob a ótica de Thomas Kuhn: elementos de reflexão. **Perspectivas em Ciência da Informação**, v. 1, n. 1, p. 27-39, 1996.

FAPESP. **Bolsas**. Disponível em: <<http://www.fapesp.br>>. Acesso em: 11/4/2009.

FEDER, A. **BibTeX**. Disponível em: <<http://www.bibtex.org/>>. Acesso em: 16/3/2011.

FERREIRA, S. M. S. P.; MARCHIORI, P.; CRISTOFOLLI, F. Percepção e motivação para publicar em revistas tradicionais e de acesso aberto: um estudo nas ciências da comunicação. **Comunicação & Sociedade**, v. 31, n. 52, p. 79-125, 2009.

FINEP. **Formas de atuação**. Disponível em: <<http://www.finep.gov.br>>. Acesso em: 11/4/2009.

FINGERMAN, S. Web of science and Scopus: Current Features and Capabilities. **Science and Technology Librarianship**, p. 1, 2006.

FONSECA, E. N. D. **Bibliometria**. Editora Cultrix, 1986.

FRANCESCHET, M. The difference between popularity and prestige in the sciences and in the social sciences: A bibliometric analysis. **Journal of Informetrics**, v. 4, n. 1, p. 55-63, 2010.

FURNER, J. Little Book, Big Book. **Journal of Librarianship and Information Science**, v. 35, n. 2, p. 115 -125, 2003.

GÁLVEZ TORO, A.; AMEZCUA, M. El factor h de Hirsch: una actualización sobre los métodos de evaluación de los autores y sus aportaciones en publicaciones científicas. **Index de Enfermería**, v. 15, n. 55, 2006. Disponível em: <[http://scielo.isciii.es/scielo.php?script=sci\\_arttext&pid=S1132-12962006000300009&lang=pt](http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1132-12962006000300009&lang=pt)>. Acesso em: 16/2/2011.

GARFIELD, E. Can citation indexing be automated? **Essays of an Information Scientist**, v. 1, p. 84-90, 1962.

GARFIELD, E. Science Citation Index. In: FIRST SYMPOSIUM ON MACHINE METHODS IN SCIENTIFIC DOCUMENTATION, 1963. **Anais...** 1963. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.141.878>>. Acesso em: 15/11/2010.

GARFIELD, E. Citation Analysis as a Tool in Journal Evaluation. **Science**, v. 178, n. 4060, p. 471 -479, 1972.

GARFIELD, E. The social science citation index. **The American Sociologist**, v. 3, n. 9, p. 164-165, 1974.

GARFIELD, E. To cite or not to cite: a not of annoyance. **Essays of an Information Scientist**, v. 3, n. 35, p. 215, 1977.

GARFIELD, E. Is citation analysis a legitimate evaluation tool? **Scientometrics**, v. 1, n. 4, p. 359-375, 1979.

GARFIELD, E. Use of Journal Citation Reports and Journal Performance Indicators in measuring short and long term journal impact. **Croatian Medical Journal**, v. 41, n. 4, p. 368-374, 2000.

GARFIELD, E. The History and Meaning of the Journal Impact Factor. **JAMA: The Journal of the American Medical Association**, v. 295, n. 1, p. 90 -93, 2006.



GARFIELD, E.; SHET, I.; TORPIE, R. **The use of citation data in writing the history of science**. Philadelphia: The Institute for Scientific Information, 1964.

GARVEY, W. D.; GRIFFITH, B. C. Scientific Communication as a Social System: The exchange of information on research evolves predictably and can be experimentally modified. **Science**, v. 157, n. 3792, p. 1011-1016, 1967.

GARZONE, M.; MERCER, R. Towards an Automated Citation Classifier. In: AI '00, 2000, London. **Anais...** London: Springer, 2000. Disponível em: <<http://portal.acm.org/citation.cfm?id=726271>>. Acesso em: 13/2/2011.

GEISLER, E. **The metrics of science and technology**. Greenwood Publishing Group, 2000.

GLÄNZEL, W. A bibliometric approach to social sciences. National research performances in 6 selected social science areas, 1990–1992. **Scientometrics**, v. 35, n. 3, p. 291-307, 1996.

GLÄNZEL, W.; SCHOEPFLIN, U. A bibliometric study of reference literature in the sciences and social sciences. **Information Processing & Management**, v. 35, n. 1, p. 31-44, 1999.

GLÄNZEL, W.; THIJS, B. Does co-authorship inflate the share of self-citations? **Scientometrics**, v. 61, n. 3, p. 395-404, 2004.

GOOGLE. **About Google Scholar**. Disponível em: <<http://scholar.google.com/intl/en/scholar/about.html>>. Acesso em: 3/3/2011.

HARZING, A. **Reflexions on the h-index**. 2007. Disponível em: <<http://perso.telecom-paristech.fr/~maitre/LTCI/bibliometrie/harzing1.pdf>>. Acesso em: 25/2/2011.

HARZING, A. **Publish or Perish User's Manual**. Disponível em: <<http://www.harzing.com/pophelp/metrics.htm>>. Acesso em: 25/2/2011.

HARZING, A.; VAN DER WAL, R. Google Scholar as a new source for citation analysis. **Ethics in Science and Environmental Politics**, v. 8, p. 61-73, 2008.

HERSCHMAN, A. The Primary Journal: Past, Present, and Future. **Journal of Chemical Documentation**, v. 10, n. 1, p. 37-42, 1970.

HOEFFEL, C. Journal impact factors. **Allergy**, v. 53, n. 12, p. 1225-1225, 1998.

HOOD, W.; WILSON, C. The Literature of Bibliometrics, Scientometrics, and Informetrics. **Scientometrics**, v. 52, n. 2, p. 291-314, 2001.

HUBER, J. C. A new model that generates Lotka's law. **Journal of the American Society for Information Science and Technology**, v. 53, n. 3, p. 209-219, 2002.

HULME, E. **Statistical bibliography in relation to the growth of modern civilization**: two lectures delivered in the University of Cambridge. London: Butler and Tanner, 1923.

INGWERSEN, P. The calculation of web impact factors. **Journal of Documentation**, v. 54, n. 2, p. 236-243, 1998.

INGWERSEN, P.; CHRISTENSEN, F. H. Data set isolation for bibliometric online analyses of research publications: Fundamental methodological issues. **Journal of the American Society for Information Science**, v. 48, n. 3, p. 205-217, 1997.

JACSÓ, P. Thoughts about federated searching. **Information Today**, v. 21, n. 9, p. 17-30, 2004.

JACSÓ, P. As we may search: comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. **Current Science**, v. 89, n. 9, p. 1537-1547, 2005.

JIN, B.; LIANG, L.; ROUSSEAU, R.; EGGHE, L. The R- and AR-indices: Complementing the h-index. **Chinese Science Bulletin**, v. 52, n. 6, p. 855-863, 2007.

KAPLAN, N. The norms of citation behavior: prolegomena to the footnote. **American Documentation**, v. 16, n. 3, p. 179-184, 1965.

KATSAROS, D.; AKRITIDIS, L.; BOZANIS, P. The f-index: Quantifying the impact of coterminal citations on scientists ranking. **Journal of the American Society for Information Science and Technology**, v. 60, n. 5, p. 1051-1056, 2009.

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

LANCHO-BARRANTES, B. S.; GUERRERO-BOTE, V. P.; MOYA-ANEGON, F. What lies behind the averages and significance of citation indicators in different disciplines? **Journal of Information Science**, v. 36, n. 3, p. 371-382, 2010.

LARIVIÈRE, V.; ARCHAMBAULT, É.; GINGRAS, Y.; VIGNOLA-GAGNÉ, É. The place of serials in referencing practices: comparing natural sciences and engineering with social sciences and humanities: Research Articles. **Journal of the American Society for Information Science and Technology**, v. 57, p. 997-1004, 2006.

LARIVIÈRE, V.; GINGRAS, Y.; ARCHAMBAULT, É. The decline in the concentration of citations, 1900-2007. **Journal of the American Society for Information Science and Technology**, v. 60, n. 4, p. 858-862, 2009.

LARSEN, P. O.; INS, M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. **Scientometrics**, v. 84, n. 3, p. 575-603, 2010.

LAWRENCE, S.; LEE GILES, C.; BOLLACKER, K. Digital libraries and autonomous citation indexing. **Computer**, v. 32, n. 6, p. 67-71, 1999.

LEVITT, J. M.; THELWALL, M. The most highly cited Library and Information Science articles: Interdisciplinarity, first authors and citation patterns. **Scientometrics**, v. 78, n. 1, p. 45-67, 2008.

LEYDESDORFF, L. The Evaluation of Research and the Evolution of Science Indicators. **Current Science**, p. 1510--1517, 2005.

LEYDESDORFF, L. Caveats for the use of citation indicators in research and journal evaluations. **Journal of the American Society for Information Science and Technology**, v. 59, n. 2, p. 278-287, 2008.

LEYDESDORFF, L.; SALAH, A. A. Maps on the basis of the Arts & Humanities Citation Index: the journals Leonardo and Art Journal versus "Digital Humanities" as a topic. **Journal of the American Society for Information Science and Technology**, p. n/a-n/a, 2010.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, v. 27, n. 2, 1998. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-19651998000200005](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19651998000200005)>. Acesso em: 5/10/2009.

MACROBERTS, M.; MACROBERTS, B. Another test of the normative theory of citing. **Journal of the American Society for Information Science**, v. 38, n. 4, p. 305-306, 1987.

MARCHIORI, P. Z. In Google we trust? In: GIANNASI-KAIMEN, M. J.; CARELLI, A. E. (EDS.) **Recursos informacionais para compartilhamento da informação: redesenhando acesso, disponibilidade e uso**. Rio de Janeiro: E-papers, 2007.

MARRON, N. Current Models of Digital Scholarly Communication. 11.2008. **Association of Research Libraries**. Disponível em: <<http://www.arl.org/bm~doc/current-models-report.pdf>>. Acesso em: 15/3/2010.

MEADOWS, A. **A comunicação científica**. Brasília: Brique de Lemos/livros, 1999.

MEHO, L. I.; ROGERS, Y. Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. **Journal of the American Society for Information Science and Technology**, v. 59, n. 11, p. 1711-1726, 2008.

MEHO, L. I.; YANG, K. Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. **Journal of the American Society for Information Science and Technology**, v. 58, n. 13, p. 2105-2125, 2007.

MESUR. **Metrics from scholarly usage of resources** Disponível em: <<http://www.mesur.org>>. Acesso em: 11/1/2011.

MINGERS, J.; LIPITAKIS, E. A. E. C. G. Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. **Scientometrics**, v. 85, n. 2, p. 613-625, 2010.

MOTTA, D. F. Validade da análise de citação como indicador de qualidade da produção científica: uma revisão. **Ciência da Informação**, v. 12, n. 1, p. 53-59, 1983.

MOZILLA. **About Mozilla add-ons: complementos para o Firefox**. Disponível em: <<https://addons.mozilla.org/pt-BR/firefox/about>>. Acesso em: 16/3/2011.

- MUELLER, S. A publicação da ciência: áreas científicas e seus canais preferenciais. **Datagramazero**, v. 6, n. 1, 2005.
- MUELLER, S. P. M. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, p. 27-38, 2006.
- NEUHAUS, C.; DANIEL, H. Data sources for performing citation analysis: an overview. **Journal of Documentation**, v. 64, n. 2, p. 193-210, 2008.
- NEWMAN, M. E. J. Power laws, Pareto distributions and Zipf's law. **Contemporary Physics**, v. 46, n. 1, p. 323-351, 2005.
- NICHOLAS, D.; RITCHIE, M. **Literature and bibliometrics**. C. Bingley, 1978.
- NICOLAISEN, J. Citation analysis. **Annual Review of Information Science and Technology**, v. 41, n. 1, p. 609-641, 2007.
- NICOLAISEN, J.; HJØRLAND, B. Practical potentials of Bradford's law: a critical examination of the received view. **Journal of Documentation**, v. 63, n. 3, p. 359-377, 2007.
- NORUZI, A. Google Scholar : the new generation of citation indexes. **Libri**, v. 55, n. 1, p. 170-180, 2005.
- OLIVEIRA, M. Canais formais de comunicação do conhecimento atropológico produzido no Brasil. **Ciência da Informação**, v. 25, n. 3, 1996. Disponível em: <<http://www.brapci.ufpr.br/download.php?dd0=8768>>. Acesso em: 3/8/2010.
- OSAREH, F.; WILSON, C. Third World Countries (TWC) research publications by disciplines: a country-by-country citation analysis. **Scientometrics**, v. 39, n. 3, p. 253-266, 1997.
- OTLET, P. O livro e a medida bibliometria. In: FONSECA, E. N. **Bibliometria: teoria e prática**. São Paulo: Cultrix, 1986.
- PARSONS, J. J.; OJA, D. **New perspectives on computer concepts 2011**. Cengage Learning, 2010.
- PRICE, D. D. S. **A ciência desde a Babilônia**. 3 ed. São Paulo: Universidade do Estado de São Paulo, 1976a.
- PRICE, D. D. S. **O desenvolvimento da ciência**. Rio de Janeiro: Livros Técnicos e Científicos, 1976b.
- PRICE, D. D. S. A general theory of bibliometric and other cumulative advantage processes. **Journal of American Society for Information Science**, v. 5, p. 292--306, 1976c.
- PRINGLE, J. Trends in the use of ISI citation databases for evaluation. **Learned Publishing**, v. 21, n. 2, p. 85-91, 2008.
- PRITCHARD, A. **Statistical bibliography: an interim bibliography**. London: North-Western Polytechnic School of Librarianship, 1969.

RAMOS-RODRÍGUEZ, A.; RUÍZ-NAVARRO, J. Changes in the intellectual structure of strategic management research: a bibliometric study of the Strategic Management Journal, 1980–2000. **Strategic Management Journal**, v. 25, n. 10, p. 981-1004, 2004.

RANGANATHAN, S. Choice of scheme for classification. **Library**, v. 5, n. March, p. 1-69, 1968.

REDALYC. **Red de Revistas Científicas de América Latina y el Caribe, España y Portugal**. Disponível em: <<http://redalyc.uaemex.mx>>. Acesso em: 3/9/2010.

REFERENCE MANAGER. **RIS Format Specifications**. Disponível em: <[http://www.refman.com/support/risformat\\_intro.asp](http://www.refman.com/support/risformat_intro.asp)>. Acesso em: 16/3/2011.

RICHARDSON, J.; BORCHERT, M. Usage metrics for open access repositories. In: AUSWEB, 2007, Australia. **Anais...** Australia: 2007. Disponível em: <<http://ausweb.scu.edu.au/aw07/papers/refereed/richardson/paper.html>>. Acesso em: 19/11/2009.

ROBREDO, J. **Da Ciência da Informação revisitada aos sistemas humanos de informação**. Brasília: Thesaurus Editora, 2003.

SAICHEV, A.; MALEVERGNE, Y.; SORNETTE, D. **Theory of Zipf's Law and Beyond**. 1º ed. Springer, 2009.

SANTOS, N. Índices bibliométricos: exemplos de sua aplicação em pediatria. **Acta Pediatria Portuguesa**, v. 1, n. 38, p. 39-43, 2007.

SANTOS, R. N.; KOBASHI, N. Bibliometria, Cientometria, Infometria: conceitos e aplicações. **Pesquisa Brasileira em Ciência da Informação**, v. 2, n. 1, p. 155-172, 2009.

SCHREIBER, M. To share the fame in a fair way, hm modifies h for multi-authored manuscripts. **New Journal of Physics**, v. 10, n. 4, p. 040201, 2008.

SCIELO. **Scientific Electronic Library Online** Disponível em: <<http://www.scielo.br>>. Acesso em: 3/9/2010.

SERRANO, M. A.; FLAMMINI, A.; MENCZER, F. Beyond Zipf's law: Modeling the structure of human language. **arXiv.org**. 2009. Disponível em: <<http://arxiv.org/abs/0902.0606>>. Acesso em: 22/2/2010.

SHAPIRO, F. R. Origins of bibliometrics, citation indexing, and citation analysis: The neglected legal literature. **Journal of the American Society for Information Science**, v. 43, n. 5, p. 337-339, 1992.

SHAPLAND, M. **Evaluation of Reference Management Software on NT** (comparing Papyrus with ProCite, Reference Manager, Endnote, Citation, GetARef, Biblioscape, Library Master, Bibliographica, Scribe, Refs). Disponível em: <<http://eis.bris.ac.uk/~ccmjs/rmeval99.htm>>. Acesso em: 18/11/2010.

SIDIROPOULOS, A.; KATSAROS, D.; MANOLOPOULOS, Y. Generalized Hirsch h-index for disclosing latent facts in citation networks. **Scientometrics**, v. 72, n. 2, p. 253-280, 2007.

SLATER, R. E-books or print books, "big deals" or local selections--What gets more use? **Library Collections, Acquisitions, and Technical Services**, v. 33, n. 1, p. 31-41, 2009.

SMITH, L. C. **Citation Analysis**. Disponível em: <<http://hdl.handle.net/2142/7190>>. Acesso em: 9/3/2010.

SNODGRASS, R. Single- versus double-blind reviewing: an analysis of the literature. **SIGMOD Rec.**, v. 35, n. 3, p. 8-21, 2006.

SOMBATSOMPOP, N.; MARKPIN, T.; PREMKAMOLNETR, N. A modified method for calculating the Impact Factors of journals in ISI Journal Citation Reports: Polymer Science Category in 1997-2001. **Scientometrics**, v. 60, n. 2, p. 217-235, 2004.

STREHEL, L. O fator de impacto do ISI e a avaliação da produção científica: aspectos conceituais e metodológicos. **Ciência da Informação**, v. 34, n. 1, 2005. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/viewArticle/599/538>>. Acesso em: 21/1/2011.

TAGUE-SUTCLIFFE, J. An Introduction to Informetrics. **Information Processing and Management**, v. 28, n. 1, p. 1-3, 1992.

TESTA, J. **Thomson Reuters Announces Book Citation Index Scheduled for Release in 2011**. 25. oct. 2010. Disponível em: <<http://newsbreaks.infotoday.com/NewsBreaks/Thomson-Reuters-Announces-Book-Citation-IndexScheduled-for-Release-in-Q--70969.asp>>. Acesso em: 27/2/2011.

THELWALL, M. Web impact factors and search engine coverage. **Journal of Documentation**, v. 56, n. 2, p. 185-189, 2000.

THOMSON REUTERS. **Using bibliometrics: a guide to evaluating research performance with citation data**. Disponível em: <<http://science.thomsonreuters.com/news/2008-07/8465001/>>. Acesso em: 10/2/2011.

THOMSON REUTERS. **Web of Science**. Disponível em: <[http://wokinfo.com/products\\_tools/multidisciplinary/webofscience/](http://wokinfo.com/products_tools/multidisciplinary/webofscience/)>. Acesso em: 26/2/2011.

THOMSON REUTERS. **EndNote**. Disponível em: <<http://endnote.com/>>. Acesso em: 16/3/2011.

TORRES-SALINAS, D.; MOED, H. F. Library Catalog Analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics. **Journal of Informetrics**, v. 3, n. 1, p. 9-26, 2009.

VAN DEN BESSELAAR, P.; LEYDESDORFF, L. Past performance, peer review and project selection: a case study in the social and behavioral sciences. **Research Evaluation**, v. 18, n. 4, p. 273-288, 2009.

VAN RAAN, A. Scientometrics: state of the art. **Scientometrics**, v. 38, n. 1, p. 205-218, 1997.

VAN RAAN, A. F. J. Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. **Scientometrics**, v. 36, n. 3, p. 397-420, 1996.

VAN RAAN, A. The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. **Technikfolgenabschätzung**, v. 12, n. 1, p. 20-29, 2003.

VANTI, N. Os links e os estudos webométricos. **Ciência da Informação**, v. 34, n. 1, 2005. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0100-1962005000100009&lang=pt](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-1962005000100009&lang=pt)>. Acesso em: 24/1/2011.

VANTI, N. Mapeamento das instituições federais de ensino superior da região nordeste na web. **Informação & Informação**, v. 15, n. 1, 2010. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/4704>>. Acesso em: 24/1/2011.

VANZ, S. A.; CAREGNATO, S. E. Estudos de citação: uma ferramenta para entender a comunicação científica. **Em Questão**, v. 9, n. 2, p. 295-307, 2003.

WALTERS, W. H. Google Scholar coverage of a multidisciplinary field. **Information Processing & Management**, v. 43, n. 4, p. 1121-1132, 2007.

WEINBERG, A. M. Impact of Large-Scale Science on the United States: Big science is here to stay, but we have yet to make the hard financial and educational choices it imposes. **Science**, v. 134, n. 3473, p. 161-164, 1961.

WHITE, H. D. Reward, persuasion, and the Sokal Hoax: A study in citation identities. **Scientometrics**, v. 60, n. 1, p. 93-120, 2004.

WYLLYS, R. Empirical and Theoretical Bases of Zipf's Law. **Library Trends**, v. 30, n. 1, p. 53-64, 1981.

YANKAUER, A. How blind is blind review? **American Journal of Public Health**, v. 81, n. 7, p. 843-845, 1991.

**APÊNDICE A – DIVISÃO DAS ÁREAS DO CONHECIMENTO PARA O CNPQ COM  
DETALHAMENTO EM CIÊNCIAS HUMANAS E SOCIAIS APLICADAS**

<b>Áreas do Conhecimento</b>	<b>Coordenações de Programas de Pesquisa</b>
<b>Ciências da Vida</b>	Agropecuária e Agronegócio – COAGR Biotecnologia e Recursos Genéticos – COBRG Gestão de Ecossistemas – COGEC Oceanográficas e Impactos Ambientais – COIAM Biociências – COBIO Saúde – COSAU
<b>Ciências Exatas e da Terra e Engenharias</b>	Ciências Exatas – COCEX Ciências Químicas e Geociências – COCQG Pesquisa, Desenvolvimento e Aplicações – COADP Engenharia – COENG Capacitação Tecnológica e Competitividade – COCTC Energia – COENE
<b>Ciências Humanas e Sociais Aplicadas</b>	<b>Ciências Humanas e Sociais - COCHS</b> Comitês de Assessoramento: AC - Artes, Ciência da Informação e Comunicação DC - Divulgação Científica FI - Filosofia HI - História LL - Letras e Linguística PS - Psicologia e Serviço Social  <b>Ciências Sociais Aplicadas e Educação – COSAE</b> Comitês de Assessoramento: AE - Administração e Economia CS - Antropologia, Arqueologia, Ciência Política, Direito, Relações Internacionais e Sociologia ED - Educação SA - Arquitetura, Demografia, Geografia, Turismo e Planejamento Urbano e Regional

FONTE: Adaptado de CNPq (2011).



## APÊNDICE B – RELAÇÃO DE BASES DE DADOS PESQUISADAS

Nome da base	Localização
ArXiv.org	<a href="http://arxiv.org">http://arxiv.org</a>
Brapci	<a href="http://www.brapci.ufpr.br">http://www.brapci.ufpr.br</a>
CiteSeer X	<a href="http://citeseerx.ist.psu.edu/">http://citeseerx.ist.psu.edu/</a>
E-LIS	<a href="http://eprints.rclis.org">http://eprints.rclis.org</a>
EBSCOhost	<a href="http://search.ebscohost.com/">http://search.ebscohost.com/</a>
Google Scholar	<a href="http://scholar.google.com/">http://scholar.google.com/</a>
Periódicos Capes	<a href="http://www.periodicos.capes.gov.br">http://www.periodicos.capes.gov.br</a>
Redalyc	<a href="http://redalyc.uaemex.mx">http://redalyc.uaemex.mx</a>
Scielo.br	<a href="http://www.scielo.org/">http://www.scielo.org/</a>
Sciencedirect	<a href="http://www.sciencedirect.com/">http://www.sciencedirect.com/</a>
Scientific Commons	<a href="http://www.scientificcommons.org/">http://www.scientificcommons.org/</a>
Scirus	<a href="http://www.scirus.com">http://www.scirus.com</a>
Scopus	<a href="http://www.scopus.com/search/form.url">http://www.scopus.com/search/form.url</a>
Social Science Research Network	<a href="http://www.ssrn.com/">http://www.ssrn.com/</a>
Web of Knowledge	<a href="http://www.isiknowledge.com/">http://www.isiknowledge.com/</a>
Willey Interscience	<a href="http://www3.interscience.wiley.com/">http://www3.interscience.wiley.com/</a>

FONTE: O autor (2010).

## APÊNDICE C – PROTOCOLO DE DESCRIÇÃO DAS FERRAMENTAS

Nome da Ferramenta:

URL:

<b>A) Aspectos Técnicos, Suporte e Interface</b>
<ul style="list-style-type: none"> <li>• Idiomas da interface;</li> <li>• Sistema operacional e requisitos para o funcionamento;</li> <li>• Interoperabilidade (forma de exibição dos resultados, obediência a algum padrão de metadados, existência de funcionalidade para intercâmbio de dados com outros sistemas).</li> <li>• Documentação sobre o sistema;</li> <li>• Tutoriais, ajuda e suporte técnico (fóruns, FAQs, entre outros).</li> </ul>
<b>B) Escopo</b>
<ul style="list-style-type: none"> <li>• Escopo (conteúdo acessado pela ferramenta; critério utilizado para o recorte em relação a toda a produção);</li> <li>• Periodicidade de atualização;</li> <li>• Janela temporal de cobertura.</li> </ul>
<b>C) Filtros e Indicadores</b>
<ul style="list-style-type: none"> <li>• Tratamento do nome dos autores (agrupamento ou controle de redundância)</li> <li>• Filtro para autocitação ou endogenia;</li> <li>• Filtro por instituições, grupos de pesquisa ou organizações;</li> <li>• Filtro por áreas da ciência;</li> <li>• Filtro por tipo de publicação;</li> <li>• Filtro por idioma usado na publicação;</li> <li>• Filtro por nacionalidade do autor;</li> <li>• Indicadores disponíveis.</li> </ul>

FONTE: O autor (2010), adaptado conforme descrito em “2 Metodologia” (p.18).

## APÊNDICE D – AUTORES UTILIZADOS NA ELABORAÇÃO DO PROTOCOLO DE DESCRIÇÃO

A) Aspectos Técnicos, Suporte e Interface	Autores Consultados
<ul style="list-style-type: none"> <li>• Idiomas da interface;</li> <li>• Sistema operacional e requisitos para o funcionamento;</li> <li>• Interoperabilidade (forma de exibição dos resultados, obediência a algum padrão de metadados, existência de funcionalidade para intercâmbio de dados com outros sistemas).</li> <li>• Documentação sobre o sistema;</li> <li>• Tutoriais, ajuda e suporte técnico (fóruns, FAQs, entre outros).</li> </ul>	(SHAPLAND, 1999; DELL'ORSO, 2008)
B) Escopo	
<ul style="list-style-type: none"> <li>• Escopo (conteúdo acessado pela ferramenta; critério utilizado para o recorte em relação a toda a produção);</li> <li>• Periodicidade de atualização;</li> <li>• Janela temporal de cobertura.</li> </ul>	(VAN RAAN, 1996; THELWALL, 2000; CRONIN, 2001; BORGMAN; FURNER, 2002; BAKKALBASI et al., 2006; ARCHAMBAULT et al., 2006; MEHO; YANG, 2007; HARZING; VAN DER WAL, 2008; ARMBRUSTER, 2009)
C) Filtros e Indicadores	
<ul style="list-style-type: none"> <li>• Tratamento do nome dos autores (agrupamento ou controle de redundância)</li> <li>• Filtro para autocitação ou endogenia;</li> <li>• Filtro por instituições, grupo</li> <li>• Filtro por tipo de publicação;</li> <li>• Filtro por idioma usado na publicação;</li> <li>• Filtro por nacionalidade do autor;</li> <li>• Indicadores disponíveis.</li> </ul>	(VAN RAAN, 1996; GLÄNZEL, 1996; ALMIND; INGWERSEN, 1997; ARCHAMBAULT; GAGNÉ, 2004; MEHO; YANG, 2007; LEYDESDORFF, 2008; BARILAN, 2008)

FONTE: O autor (2010), adaptado conforme descrito em “2 Metodologia” (p.18).