

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE COMUNICAÇÕES E ARTES
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

MARIANA RAMOS CRIVELENTE

**Métodos e técnicas bibliométricas de análise de produção
científica: um estudo crítico.**

São Paulo
2019

MARIANA RAMOS CRIVELENTE

Métodos e técnicas bibliométricas de análise de produção científica: um estudo crítico.

Versão corrigida

(Versão original disponível na Biblioteca da ECA/USP)

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade de São Paulo para obtenção do título de Mestre.

Área de Concentração: Cultura e Informação. Linha de pesquisa: Organização da informação e do conhecimento.

Orientador: Prof(a). Dr(a). Nair Yumiko Kobashi

São Paulo

2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catalogação na Publicação
Serviço de Biblioteca e Documentação
Escola de Comunicações e Artes da Universidade de São Paulo
Dados inseridos pelo(a) autor(a)

Crivelente, Mariana Ramos

Métodos e técnicas bibliométricas de análise de produção científica: um estudo crítico / Mariana Ramos Crivelente ; orientadora, Nair Yumiko Kobashi. -- São Paulo, 2019. 142 p.: il. + 1 CD-ROM.

Dissertação (Mestrado) - Programa de Pós-Graduação em Ciência da Informação - Escola de Comunicações e Artes / Universidade de São Paulo.

Bibliografia

CD-ROM : Lista de autores e Lista de registros bibliográficos.

Versão corrigida

1. Bibliometria 2. Estatística 3. Produção científica 4. Indicadores bibliométricos I. Kobashi, Nair Yumiko II. Título.

CDD 21.ed. - 020

Elaborado por Alessandra Vieira Canholi Maldonado - CRB-8/6194

Dissertação de autoria de Mariana Ramos Crivelente, sob o título **Métodos e técnicas bibliométricas de análise de produção científica: um estudo crítico**, apresentada à Escola de Comunicações e Artes da Universidade de São Paulo, para obtenção do título de Mestre em Ciências pelo Programa de Pós-graduação em Ciências da Informação, na área de concentração: Cultura e Informação.

Em ____ de _____ de ____ pela comissão julgadora constituída pelos doutores:

Prof. Dr. Instituição: Nair Yumiko Kobashi (USP)

Presidente

Prof. Dr. Instituição: Maria Cláudia Cabrini Grácio (UNESP)

Prof. Dr. Instituição: Michely Jabala Mamede Vogel (UFF)

Para minha tia Rosa Maria Crivelente Horta.

Agradecimentos

À Prof(a). Nair Kobashi, minha orientadora, pela excepcional orientação e pelo incomensurável aprendizado que me proporcionou. Obrigada pelo comprometimento, paciência, respeito e por ter acreditado, apoiado e confiado em mim.

À minha família, por tudo que me proporcionaram.

Ao Caio, pelo cuidado, paciência e companheirismo.

À Laís, pelo apoio e amizade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela concessão da bolsa de mestrado para a realização desta pesquisa.

“In statistical affairs ... the first care before all else is to lose sight of the man taken in order to consider him only a fraction of the species.”
(Siméon Denis Poisson, 1835).

Resumo

CRIVELENTE, Mariana Ramos. **Métodos e técnicas bibliométricas de análise de produção científica: um estudo crítico**. 2019. 142 f. Dissertação (Mestrado em Ciências da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2019.

A dissertação “Métodos e técnicas bibliométricas de análise de produção científica: um estudo crítico” aborda os procedimentos quantitativos utilizados na avaliação da atividade científica. A pesquisa foi motivada pelo uso crescente de indicadores quantitativos bibliométricos para avaliar e classificar programas de pesquisa, instituições de ensino superior, programas de pós-graduação, a produção científica dessas instituições e os próprios pesquisadores. Procurou-se, especificamente, problematizar os estudos bibliométricos e as condições de sua aplicabilidade nesses contextos. Nessa perspectiva, foram definidas hipóteses de trabalho e sistematizado um referencial teórico para sustentar o desenvolvimento da pesquisa e a interpretação dos dados. O referencial teórico é constituído de aspectos da institucionalização da Ciência no Brasil, do ponto de vista da História da Ciência, da Quantificação da vida social, sob a perspectiva da Filosofia da Ciência e das teorias e métodos da Cientometria/Bibliometria. Na construção desse referencial foram: a) identificadas as ideias e eventos que contribuíram para a consolidação da estatística enquanto método das Ciências Sociais; b) identificados e analisados os métodos e técnicas bibliométricas; c) discutidos os aspectos problemáticos que subjazem ao uso da estatística nos estudos quantitativos da Ciência da Informação. Em seguida, foi realizado um estudo de caso de análise da produção científica de um instituto de pesquisa nacional. O estudo permitiu identificar e sistematizar os problemas e os desafios encontrados tanto na aplicação dos métodos bibliométricos quanto na interpretação dos indicadores quantitativos obtidos. Conclui-se que os estudos de avaliação da atividade científica ganham sentido se forem contextualizados e interpretados à luz da Filosofia e História da Ciência, sendo importantes, ainda, o rigor na seleção das fontes de informação, no tratamento dos dados e na composição da amostra do estudo. Espera-se que os resultados desta pesquisa contribuam para aprimorar os modelos de avaliação da atividade científica e a formulação de políticas de ciência adequadas a cada contexto.

Palavras-Chave: Produção científica. Indicadores bibliométricos. Bibliometria. Estatística.

Abstract

CRIVELENTE, Mariana Ramos. **Bibliometric methods and techniques of scientific production analysis: a critical study**. 2019. 142 f. Dissertation (Master's degree) – School of Communications and Arts, University of San Paulo, San Paulo, 2019.

The dissertation "Bibliometric methods and techniques of scientific production analysis: a critical study" critically addresses the quantitative procedures used in the evaluation of scientific activity. The research was motivated by the increasing use of quantitative bibliometric indicators to evaluate and classify research programs, higher education institutions, graduate programs, the scientific production of these institutions and the researchers themselves. We specifically sought to problematize bibliometric studies and the conditions for their applicability in these contexts. In this perspective, working hypotheses were defined and a theoretical framework was systematized to support the development of the research and the interpretation of the data. The theoretical framework consists of aspects of science institutionalization in Brazil, from the point of view of the History of Science, the quantification of social life, from the perspective of the Philosophy of Science and the theories and methodologies of Scientometry/Bibliometry. In building this framework, a) the ideas and events that contributed to the consolidation of statistics as a method of Social Sciences were identified; b) Bibliometric methods and techniques were identified and analyzed; c) the problematic aspects underlying the use of statistics in quantitative studies of Information Science were discussed. Then, a case study was carried out to analyze the scientific production of a national research institute. The study allowed to identify and systematize the problems and challenges encountered both in the application of bibliometric methods and in the interpretation of the quantitative indicators obtained. It is concluded that the evaluation studies of scientific activity make sense if they are contextualized and interpreted in the light of the Philosophy and History of Science, being also important the accuracy in the selection of the sources of information, in the treatment of the data and in the sample composition of the study. It is hoped that the results of this research will contribute to improving the models for evaluating scientific activity and formulating science policies appropriate to each context.

Keywords: Scientific production. Bibliometric indicators. Bibliometry. Statistics.

LISTA DE ILUSTRAÇÕES

Ilustração 1 – Quadro de Moed (2018): o modelo de avaliação desktop e abordagens alternativas	16
Ilustração 2 - Colaboração entre os autores	54
Ilustração 3 - Classificação dos autores mais produtivos	55
Ilustração 4 - Diagrama de Bradford	64
Ilustração 5 - Ordem de série versus frequência de palavras nos textos A e B	74
Ilustração 6 - Quadro de referência dos estudos de construção de indicadores bibliométricos	87
Ilustração 7 – Percurso da atividade de análise bibliométrica	89
Ilustração 8 – Pontos fortes e fracos de bases de dados selecionadas	105
Ilustração 9 - Gráfico de distribuição entre servidores públicos do Instituto Butantan, colaboradores da Fundação Butantan e bolsistas após segundo mapeamento de autores	110
Ilustração 10 - Evolução do número de artigos publicados pelo Instituto Butantan nos anos de 2007 a 2016	115
Ilustração 11 - Evolução do número de artigos publicados pelo Instituto Butantan nos anos de 2007 a 2016 por tipo de vínculo na instituição	116
Ilustração 12 - Artigos publicados nos anos de 2007 a 2016 por vínculo institucional do pesquisador no Instituto Butantan	118
Ilustrações 13 e 14 - Autoria em artigos e distribuição do corpo de servidores no Instituto Butantan	119
Ilustração 15 - Afiliação dos primeiros-autores na amostra	120
Ilustração 16 - Distribuição da primeira autoria de acordo com vínculo institucional	121

Ilustração 17 - Principais países colaboradores nos artigos publicados pelo IB	123
Ilustração 18 - Principais instituições do exterior que colaboraram nos artigos publicados pelo Instituto Butantan	124
Ilustração 19 - Vínculo do pesquisador em relação à internacionalização da pesquisa	124

LISTA DE TABELAS

Tabela 1 - Publicações recuperadas no repositório da instituição	53
Tabela 2 - Número de publicações por número de autores	58
Tabela 3 - Exemplo de aplicação da Lei de Lotka	59
Tabela 4 - Dispersão na literatura científica que trata de indexação automática	65
Tabela 5 - Zonas de Bradford na área de indexação automática	67
Tabela 6 - Ordem de série versus frequência de palavras nos textos A e B	72
Tabela 7 - Ranking de citações nos artigos publicados por Stephen P. Harter	82
Tabela 8 - Número total de artigos recuperados, por período analisado	112
Tabela 9 - Frequência relativa e frequência relativa acumulada dos artigos publicados pelo Instituto Butantan nos anos de 2007 a 2016	114
Tabela 10 - Vínculo institucional por número de artigos nos anos 2007-2016	117
Tabela 11 - Relação número de artigos por pesquisador	120

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Objetivos da pesquisa	18
1.1.1 Objetivo Geral	19
1.1.2 Objetivos Específicos	19
1.2. Aspectos metodológicos da pesquisa	19
1.3 Estrutura da dissertação	21
2 REFERENCIAL TEÓRICO	23
2.1 Aspectos historiográficos da institucionalização da ciência brasileira	23
2.2 A quantificação de dados sociais	29
2.2.1 Estatística: determinismo e probabilidade	30
2.2.2 Acaso, Probabilidade e Leis estatísticas	32
2.3 Leis estatísticas	33
2.4 O conceito de normalidade	39
2.5 O contraponto	41
2.6 Lógica da inferência estatística	43
3 A BIBLIOMETRIA NA QUANTIFICAÇÃO DAS ATIVIDADES CIENTÍFICAS	46
3.1 Constituição da Bibliometria: breve histórico	47
3.2 Os métodos bibliométricos	52
3.2.1 Métodos estatísticos descritivos	52
3.2.2 Métodos bibliométricos	56
3.2.2.1 Produtividade de autores: a Lei de Lotka	56
3.2.2.2 Produtividade de periódicos: a Lei de Bradford	64
3.2.2.3 Contagem de palavras: as leis de Zipf	70
3.2.2.4 Análise de citação	75
3.2.2.4.1 Índice de Hirsch e Fator de Impacto	81
4 ESTUDO DE CASO: PRODUÇÃO CIENTÍFICA DO INSTITUTO BUTANTAN	87
4.1 Percurso metodológico do estudo de caso	87
4.2 O Instituto Butantan (IB): contextualização do objeto de pesquisa	91
4.2.1 Contexto de criação do Instituto Butantan.	93
4.3 Produção científica do Instituto Butantan: metodologia do estudo bibliométrico	

4.3.1 Moldura analítica	100
4.3.1.1 A escolha do <i>corpus</i> de análise	102
4.3.1.2 A escolha da base de dados	104
4.3.2 Recuperação dos dados	107
4.3.2.1 Mapeamento dos nomes dos servidores, colaboradores e bolsistas do Instituto e Fundação Butantan	108
4.3.2.2 Mapeamento dos artigos publicados pelos servidores, colaboradores e bolsistas do Instituto e Fundação Butantan	110
5 RESULTADOS	114
5.1 Evolução cronológica das publicações	114
5.2 Produtividade dos pesquisadores	117
5.3 Colaboração Internacional	122
6 CONSIDERAÇÕES FINAIS	127
7 REFERÊNCIAS	132
ANEXOS	142

1 INTRODUÇÃO

A Bibliometria é utilizada de forma cada vez mais intensa na avaliação das atividades científicas com base em métodos matemáticos e estatísticos. Tem especial relevância a produção de indicadores que quantificam e descrevem, de um lado, os produtos derivados das pesquisas científicas e tecnológicas e, de outro, a sua influência na produção de novas pesquisas. Nessa perspectiva, são quantificados e mensurados o crescimento da literatura científica, seus tipos e papéis, a obsolescência da informação, a pertinência e relevância dos documentos para compor acervos, o papel dos canais formais e informais na comunicação científica, os hábitos de citação e as relações intra e interdisciplinares. Considera-se, em geral, que a quantificação baseada em procedimentos bibliométricos confere um olhar “objetivo” ao desenvolvimento da ciência, concepção que se tornou rotineira na avaliação de universidades, programas de pós-graduação e revistas científicas. Porém, diversos segmentos da comunidade científica criticam as avaliações baseadas apenas em indicadores quantitativos. Tal preocupação está inscrita em diversos documentos, dentre os quais pode ser citado o Manifesto de Leiden, elaborado por um grupo renomado de pesquisadores. Foi publicado na Revista *Nature*, em 2015, a fim de expor as críticas à comunidade científica e aos gestores de políticas de ciência:

As métricas proliferaram: em geral bem intencionadas, nem sempre bem informadas e frequentemente mal aplicadas. Corremos o risco de prejudicar o sistema da ciência com as próprias ferramentas projetadas para melhorá-lo, uma vez que a avaliação é cada vez mais realizada por instituições sem o devido conhecimento das boas práticas e da interpretação adequada de indicadores (HICKS et al, 2015. p. 429, tradução nossa).¹

O Manifesto de Leiden tem tido ampla repercussão porque, além das críticas aos procedimentos utilizados em estudos bibliométricos, apresenta 10 princípios norteadores necessários ao desenvolvimentos desses estudos: 1) a avaliação quantitativa deve dar suporte à avaliação qualitativa especializada; 2) medir o desempenho de acordo com a missão da instituição, do grupo ou do pesquisador; 3)

¹ “Metrics have proliferated: usually well intentioned, not always well informed, often ill applied. We risk damaging the system with the very tools designed to improve it, as evaluation is increasingly implemented by organizations without knowledge of, or advice on, good practice and interpretation.”

proteger a excelência da pesquisa localmente relevante; 4) manter a coleta de dados e os processos analíticos abertos, transparentes e simples; 5) permitir que os avaliados verifiquem os dados e as análises; 6) considerar as diferenças entre as áreas nas práticas de publicação e citação; 7) basear a avaliação de pesquisadores individuais no julgamento qualitativo da sua carreira; 8) evitar falsas precisões; 9) reconhecer os efeitos sistêmicos da avaliação e dos indicadores; 10) examinar e atualizar regularmente os indicadores (HICKS et al, 2015).

Diversos aspectos dos métodos bibliométricos requerem um olhar crítico, como salienta o Manifesto (HICKS et al, 2015), pois: a) os números não falam por si; b) os indicadores quantitativos devem ser interpretados contextualmente para terem sentido; c) as pesquisas devem ser avaliadas pela relevância das contribuições para a elucidação de problemas locais e regionais; d) os critérios e procedimentos de coleta e processamento dos dados de partida devem ser explícitos; e) o corpus da pesquisa, para ser validado, deve estar disponível para os avaliandos e demais interessados; f) os modelos de análise devem se adequar às especificidades de produção científica e estilos de comunicação de cada campo do saber; g) a avaliação individualizada deve levar em conta o percurso acadêmico do pesquisador e o contexto institucional no qual desenvolve as suas atividades; h) as avaliações modelam a vida dos pesquisadores, portanto, seus efeitos positivos e negativos devem ser criteriosamente pesados e i) os indicadores não são perenes, portanto, requerem atualizações constantes.

Diante das considerações acima, várias indagações podem ser enunciadas a respeito dos métodos bibliométricos: quais são os aspectos positivos e negativos da aplicação de indicadores quantitativos à avaliação? Um aspecto positivo a ser destacado é a importância e utilidade desses métodos para traçar os percursos de construção de paradigmas científicos e compará-los, no âmbito dos estudos da História da Ciência, Sociologia da Ciência e Filosofia da Ciência; esses métodos são importantes também para a autoanálise dos programas de pesquisa, na medida que podem subsidiar o planejamento de políticas científicas, gerenciar acervos de bibliotecas científicas, entre outros (GINGRAS, 2016). As críticas, por outro lado, incidem sobre os seus usos para classificar a produção científica institucionalizada, as universidades, os programas de pós-graduação, os pesquisadores individualmente e para construir políticas de ciência. O uso de tais critérios de classificação, que

acabam moldando a atividade científica de maneira nem sempre positiva, causa impactos que merecem ser cuidadosamente analisados.

Outro aspecto que merece discussão é a banalização das pesquisas baseadas em métodos bibliométricos, pesquisas conhecidas como *desktop bibliometrics* (KATZ; HICKS, 1997). Nota-se nelas a ausência: de problematização do objeto, de hipóteses de trabalho, de critérios de escolha das fontes de dados, dos princípios de composição do *corpus*, e, principalmente, não há reflexões críticas sobre os resultados obtidos. É preocupante o fato de os indicadores serem considerados retratos “verdadeiros” das atividades científicas e, embora de confiabilidade limitada, serem adotados como parâmetros de avaliação.

O termo *desktop bibliometrics*, cunhado por Katz e Hicks (1997), foi retomado por Moed (2018) no contexto do uso de indicadores bibliométricos na avaliação da atividade científica:

É claro que tanto o cálculo quanto a interpretação das métricas científicas não são feitos apenas por especialistas em bibliometria. “Desktop bibliometrics”, um termo cunhado por Katz e Hicks (1997) que se refere a uma prática de avaliação usando dados bibliométricos pressionando um botão – pode-se dizer - com base em suposições simplistas e questionáveis sobre o que os indicadores medem, está se espalhando no domínio da avaliação (MOED, 2018, p.2).²

Moed (2018) elaborou, inclusive, um quadro que compara o modelo *desktop* com as demais abordagens:

Ilustração 1 – Quadro de Moed (2018):
o modelo de avaliação *desktop* e abordagens alternativas.

Aspecto	Pressupostos do modelo <i>desktop</i>	Abordagens alternativas
1 Disponibilidade	Devem ser utilizados indicadores amplamente disponíveis (fator de impacto, contagem de publicações, JIFs, índice h).	Usar indicadores personalizados apropriados para o contexto de aplicação.
2 Validade	Os indicadores medem suficiente bem o que eles devem medir;	Os indicadores devem ser combinados com conhecimento especializado e medir as pré-

² “It is clear that both the calculation and the interpretation of science metrics are not merely made by bibliometric experts. “Desktop bibliometrics”, a term coined by Katz and Hicks (1997) and referring to an evaluation practice using bibliometric data– in a manner of speaking – by pushing a button, and based on simplifying, questionable assumptions on what the indicators measure, is pervading the assessment domain.”

		nenhuma validação de outras fontes é necessária.	condições de desempenho, em vez do desempenho propriamente dito.
3	Avaliação significativa	Os aspectos medidos pelos indicadores usam critérios de avaliação adequados.	É necessária uma estrutura de avaliação independente.
4	Princípio da solicitação	Quanto maior a pontuação, melhor o desempenho.	Considerar o uso de indicadores para definir padrões mínimos em vez de identificar o topo de uma classificação.
5	CrITÉrios de decisão política	A melhor performance recebe o maior suporte.	Apoiar ECSRs ³ por instituições de fomento, de acordo com o número de grupos de pesquisa emergentes.

Fonte: Moed (2018, p.11-12, tradução nossa).⁴

As abordagens alternativas parecem estar se tornando uma tendência na avaliação da pesquisa científica. A título de exemplo, a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), agência ligada ao Ministério da Educação, que avalia os cursos de mestrado e doutorado do país, anunciou que os mais de 4 mil programas em atividade deixarão de ter uma nota única e passarão a ser avaliados em cinco dimensões: Ensino e Aprendizagem, Produção de conhecimento, Inserção internacional e regional, Inovação e transferência do conhecimento e Impacto na sociedade (MARQUES, 2019).

Spinak (1998) afirmou, ainda na década de 1990, que não existe consenso sobre a medição e a avaliação da produção intelectual e acadêmica, bem como de

³ European Consortium for Sociological Research.

⁴

	Aspect	Assumptions of the desktop model	Alternative approaches
1	Availability	Widely available indicators should be used (i.e., publication counts, JIFs, h-indices)	Use tailor-made indicators appropriate for the application context
2	Validity Validade	Indicators measure what they are supposed to measure well enough; no confirmation from other sources is needed	Indicators should be combined with expert knowledge and measure pre-conditions of performance rather than performance itself
3	Evaluative significance	The aspects measured by the indicators constitute appropriate evaluation criteria	An independent evaluative framework is needed
4	Ordering principle	The higher the score, the better the performance	Consider using indicators to define minimum standards rather than identifying the top of a ranking
5	Policy decision criteria	The best performer receives the largest support	Support ECSRs by funding institutions according to the number of emerging research groups

seus impactos e influências. O autor acredita que o consenso existe dentro do sistema editorial, porém cabe às instituições de pesquisa criarem seus modelos, sem se submeter à dinâmica mercadológica das editoras.

Em pesquisas científicas, é sabido que as hipóteses que norteiam os estudos devem estar sustentadas em conhecimentos legitimados. Para orientar a presente pesquisa, foi realizada uma revisão bibliográfica preliminar que permitiu elaborar as seguintes hipóteses de trabalho:

- a abordagem bibliométrica opera com a descrição deliberadamente simplificada de situações complexas e intrincadas do mundo real. Um estudo consistente, portanto, deve contextualizar histórica, social e filosoficamente o objeto da pesquisa.
- os indicadores bibliométricos são de natureza quantitativa, elaborados com base em inferências estatísticas indutivas. Portanto, é essencial problematizar e avaliar as chamadas leis bibliométricas, que se apoiam, em larga medida, em métodos estatísticos indutivos.
- os indicadores bibliométricos são elaborados com dados obtidos em fontes bibliográficas. Como cada fonte é construída para responder a objetivos institucionais específicos, é importante que a pesquisa apresente os critérios de escolha das fontes e sua adequação aos objetivos da pesquisa, de forma clara.
- os registros das fontes bibliográficas apresentam diversos tipos de erros e inconsistências. Nessa medida, os dados de partida devem ser consolidados e padronizados antes de compor o *corpus* da pesquisa. Este último deve estar disponível para consulta e validação dos resultados.
- os resultados obtidos devem ser interpretados criticamente, de modo a contribuir para a realização de novas pesquisas e explicitados os limites de sua aplicação.

1.1 Objetivos da pesquisa

A problematização e as hipóteses de trabalho acima permitem enunciar os objetivos da pesquisa, distribuídos em objetivo geral e objetivos específicos.

1.1.1 Objetivo Geral

Analisar criticamente os métodos e técnicas bibliométricos utilizados em estudos de avaliação de produção científica, de modo a contribuir para o aprimoramento dessa vertente de pesquisa na Ciência da Informação.

1.1.2 Objetivos Específicos

- a) subsidiar a discussão dos métodos bibliométricos com a sistematização das contribuições dos estudos sociais e filosóficos da ciência;
- b) problematizar e sistematizar os métodos e técnicas bibliométricos de análise de produção científica;
- c) sistematizar os desafios enfrentados na pesquisa e indicar os procedimentos adotados para assegurar a validade dos resultados.

1.2. Aspectos metodológicos da pesquisa

As obras de metodologia científica (LAKATOS, 2016; TRUJILLO FERRARI, 1974; RUIZ, 1979; BUNGE, 1980) classificam o conhecimento em 2 tipos: conhecimento intuitivo e conhecimento científico. O conhecimento intuitivo nasce da apreensão instantânea da realidade. O conhecimento científico distingue-se do conhecimento intuitivo porque recorre à razão, lançando mão de métodos, ditos científicos, para conhecer a realidade e se aproximar da verdade. Deve-se salientar que o conceito de verdade, na atividade científica, tem características próprias: a verdade é provisória por ser próprio da ciência submeter os conhecimentos a processos contínuos de validação, percurso que pode conduzir a novos resultados e a novas verdades provisórias.

Ferrater Mora afirma que existe um método quando “[...] se dispõe de, ou se segue, certo “caminho”, [...], para alcançar determinado fim, proposto de antemão.” (2004, p. 1962). O método é basicamente um conjunto de regras encadeadas racionalmente para se chegar a um fim proposto. Descartes afirma, no Discurso do Método, que as regras metódicas devem ter uma qualidade fundamental: a possibilidade de serem usadas e aplicadas por quem quer que seja; “em outros

termos, não há “métodos individuais”; os que recebem esse nome são simplesmente “costumes” ou “procedimentos” (DESCARTES apud FERRATER MORA, 2004, p.1963).

É possível falar também de métodos gerais e específicos. Os métodos gerais operam sobre os fenômenos da natureza e da sociedade em níveis de abstração elevados. Lakatos (2016) propõe quatro “métodos de abordagem”: indutivo, dedutivo, hipotético-dedutivo e dialético. Os específicos, denominados “métodos de procedimento”, são “etapas mais concretas da investigação, com finalidades mais restritas em termos de explicação geral dos fenômenos e menos abstrata” (LAKATOS, 2016, p.88). Os métodos de procedimentos das Ciências sociais, no interior da qual se situam os estudos sobre os métodos bibliométricos, são: Método histórico, Método comparativo, Método monográfico, Método estatístico, Método tipológico, Método funcionalista, Método estruturalista, Método etnográfico e Método clínico (LAKATOS, 2016, p. 93). A presente pesquisa enquadra-se no método hipotético-dedutivo, tem natureza exploratória e se apoia em métodos estatísticos.

A Bibliometria e seus métodos, objeto do presente estudo, parte de problemas construídos racionalmente e faz uso do método científico para circunscrever, delimitar, analisar o objeto de pesquisa e obter indicadores quantitativos. A produção de indicadores visa conhecer segmentos da realidade. Para isso, é necessário partir de problemas e hipóteses bem fundamentados. Dito de outro modo, o pesquisador deve definir com clareza a realidade que deseja conhecer e traçar os caminhos que permitam obter resultados válidos. Os indicadores, por sua vez, devem ser plenos de significado, adequados para descrever e avaliar a ciência produzida. Como toda pesquisa científica, a abordagem bibliométrica não é neutra. De fato, tanto a produção de ciência quanto a sua análise não são neutras por serem realizadas em contextos históricos e sociais determinados.

Esta pesquisa lança um olhar crítico sobre os métodos bibliométricos utilizados para avaliar a produção científica institucionalizada. Um estudo exploratório propõe-se a levantar problemas de pesquisa ainda pouco explorados com o objetivo de identificar novas hipóteses de pesquisa que preencham essas lacunas.

Para dar resposta adequada às indagações e hipóteses de trabalho, foi realizado um estudo de caso de aplicação de métodos bibliométricos à produção científica de um instituto de pesquisa. De acordo com Yin (2001), o estudo de caso é

uma estratégia de pesquisa que “permite uma investigação para se preservar as características holísticas e significativas dos eventos da vida real” (YIN, 2001, p.21). Sendo assim, o estudo de caso contribui para a compreensão do fenômeno social aqui proposto, preservando suas características próprias.

No caso dos indicadores científicos, Okubo (1997) chama a atenção para a necessidade de se diferenciar estudos micro e macro:

Esses indicadores científicos são igualmente adequados - desde que sejam tomadas as precauções habituais - para análises "macro" (por exemplo, uma determinada participação nacional na produção global de literatura científica durante um período especificado) e "micro" estudos (por exemplo, o papel de um determinado instituto na produção de artigos) em um campo circunscrito da ciência (OBUKO, 1997, p. 7).⁵

Em síntese, a pesquisa foi desenvolvida de acordo com os seguintes procedimentos:

- a) construção do referencial teórico, com base em estudos históricos, sociais e filosóficos da ciência;
- b) sistematização da literatura sobre a Bibliometria e seus usos na avaliação da atividade científica;
- c) realização de um estudo de caso de análise da produção científica do Instituto Butantan.

No Capítulo 4.1, estão expostos os procedimentos metodológicos específicos do estudo de caso, cujo objeto empírico é a produção científica do Instituto Butantan.

1.3 Estrutura da dissertação

A dissertação é constituída de seis capítulos, que passamos a descrever de forma resumida.

Na Introdução, estão expostas as principais questões que motivaram o estudo, o objeto da pesquisa, os aspectos epistemológicos da abordagem, as hipóteses de trabalho, os objetivos, a metodologia geral e a estrutura da dissertação.

⁵ "These scientific indicators are equally suitable – providing the customary precautions are taken – for “macro” analysis (e.g. a given country’s share in global output of scientific literature over a specified period) and “micro” studies (e.g. a given institute’s role in producing papers in a highly circumscribed field of science)."

No Capítulo 2 é apresentado o referencial teórico da pesquisa: aspectos da institucionalização da Ciência no Brasil, sob o olhar da História da Ciência e os usos da estatística nas pesquisas das áreas de Ciências sociais e humanas, à luz da Filosofia da Ciência.

No Capítulo 3 são apresentados e problematizados os métodos bibliométricos, tanto quanto seus limites para avaliar a atividade científica, com foco na Estatística descritiva e nos Métodos bibliométricos.

O Capítulo 4 é dedicado ao Estudo de caso de análise da produção científica de um instituto de pesquisa (Instituto Butantan) por meio de abordagem bibliométrica. Este estudo permitiu desvendar muitos dos desafios teóricos e técnicos enfrentados na produção de indicadores bibliométricos.

No Capítulo 5 estão sistematizados e discutidos os resultados do estudo de caso.

Finalmente, no Capítulo 6, são apresentadas as conclusões da dissertação e a indicação de novos temas de estudo.

2 REFERENCIAL TEÓRICO

O quadro teórico da presente pesquisa é constituído de 2 itens: aspectos historiográficos da institucionalização da Ciência no Brasil e a quantificação da vida social, na perspectiva da Filosofia da Ciência.

2.1 Aspectos historiográficos da institucionalização da ciência brasileira

O objetivo deste item é revisar algumas abordagens sobre a institucionalização da ciência no Brasil. Daremos destaque à implantação das políticas de saúde pública no Brasil, tendo em vista o interesse especial deste tema para a análise bibliométrica da produção científica do Instituto Butantan, apresentado no Capítulo 4.

A História da Ciência tinha, inicialmente, como objeto privilegiado, as ciências naturais ou ciências duras. Para os estudiosos da ciência interessava analisar os produtos dessas atividades científicas. Todavia, na década de 1970, essa disciplina sofreu uma profunda renovação. Thomas Kuhn, físico e filósofo estadunidense, publicou, em 1962, a obra *Estrutura das Revoluções Científicas*, na qual questionou a ideia de que o conhecimento científico se impõe pelo seu valor de “verdade”. De acordo com o autor, os próprios cientistas eram os criadores da noção de ciência como verdade. O trabalho de Kuhn acabou por relativizar toda a concepção de ciência vigente até a década de 1960. Segundo Kuhn (1997), o que caracteriza a comunidade científica é, na verdade, o trabalho com os paradigmas científicos, os quais contêm princípios teóricos, e também os princípios de ação, estabelecidos socialmente.

Após o trabalho de Thomas Kuhn, a ciência e as práticas científicas passaram a receber novos olhares. Dominique Pestre (1996) sistematizou esse período de transição em dois momentos. Em um primeiro, questionou a ideia de ciência e em um segundo momento, a partir desse questionamento, identificou a emergência de novos objetos científicos. Segundo Pestre (1996) a renovação da História das Ciências teve início em meados dos anos 1970, quando o sociólogo britânico David Bloor propôs quatro categorias para abordar a História das Ciências: causalidade, simetria, imparcialidade e flexibilidade. Causalidade seria a ideia de que fatores sociais interferem nas mudanças na ciência. Simetria estaria relacionada às abordagens que

evitam disparidades dos dados e anacronismos. A imparcialidade sugere que os estudiosos da ciência sejam imparciais no tratamento dos fatos. A flexibilidade diz respeito à exigência de abordagem crítica. Apenas simetria e imparcialidade permanecem aceitas até hoje (BLOOR apud PESTRE, 1996).

Por volta da metade dos anos 1980, há um novo período de renovação das concepções de ciência. Ela passa a ser vista como uma construção para entender os fenômenos e não mais como a descoberta ou a obtenção de uma “verdade” revelada. É nesse contexto que Bruno Latour (1996) propõe a valorização do papel do cientista, em particular dos que se dedicam às ciências sociais e humanas.

O objetivo de Latour é apresentar um projeto de História Social das Ciências, que rompe a divisão entre ciência e política. Ciência, para Latour, é um conjunto de atividades colocadas para o cientista, dentre as quais estão: a coleta de dados, o diálogo com os pares para convencê-los, a construção de alianças políticas e a persuasão do público mais amplo.

A História das Ciências, assim redefinida, considera a relação ciência-sociedade como objetos de investigação, sendo necessário realizar operações de tradução da ciência para a política e vice-versa (LATOURE, 1996, p. 135). A operação de tradução consiste em harmonizar dois interesses: o político e o científico. Desnudar esses dois eventos sob diferentes prismas é a tarefa da História Social das Ciências.

O desenvolvimento da ciência moderna é tradicionalmente atribuído aos países da Europa. Dois autores se destacam na defesa dessa hipótese. Needham (1954) se debruçou sobre a emergência da ciência moderna dentro dos limites territoriais da Europa Ocidental. O segundo nome importante desta vertente é Bassalla (1967), que se dedicou ao estudo das modalidades de difusão dessa ciência para o resto do mundo.

Na visão de Bassalla (1967), o colonizador era o protagonista central da difusão da ciência para as colônias, sendo estas meras receptoras do conhecimento produzido na Europa Ocidental. O modelo de difusão da ciência por ele proposto é constituído de três fases: na primeira fase, a colônia (“não-científica”), seria apenas a fornecedora de dados para a ciência europeia. Se destacam, nesse período, os viajantes e naturalistas que participam das grandes expedições de exploração científica e de missões religiosas. Na segunda fase, chamada de ciência colonial, os cientistas autóctones, embora dependentes da cultura científica europeia, de suas

instituições, instrumentos e publicações, expandem os estudos do centro produtor para a periferia colonizada. Na terceira fase, completa-se o processo de transplante, tendo início as lutas das colônias para consolidar a tradição cultural e científica autóctone. Tal transição tem como força motriz o nacionalismo, tanto político quanto cultural. O sucesso desse esforço, porém, deveria aliar a resistência às crenças filosóficas e religiosas locais com a concomitante valorização do papel do cientista.

A visão eurocêntrica foi questionada por Kapil Raj (2007), com base na crítica dos valores morais e políticos da ciência moderna, caracterizados como alienantes e desumanizadoras. Desse modo, combate a ideia de que a ciência moderna é um fenômeno inerente à Europa Ocidental, tanto quanto a concepção sobre sua difusão, que teria beneficiado os povos (não-europeus) pelo contato com os europeus e o capitalismo.

Raj (2007) reexaminou a natureza do fazer científico no espaço globalizado da pré-modernidade. Dá destaque ao papel dos contatos interculturais na circulação dos conhecimentos especializados que constituíram a ciência entre a Europa Ocidental e o Sul da Ásia, nos séculos XVII e XIX. O autor enfatizou, de modo especial, a construção do conhecimento científico nas próprias zonas de contato, por serem locais legítimos de produção de conhecimentos. Ao evidenciar que a produção do conhecimento da Europa Ocidental estava intimamente ligada às redes de comércio transatlântico, Raj (2007) ressignifica o processo de construção da ciência europeia. Contribui também para mostrar que a ciência não se resume a construir um sistema de proposições formais das descobertas, afirmando o caráter instrumental, prático e político da atividade científica.

Kapil Raj (2007) demonstra que as academias e universidades da Europa estavam empenhadas em produzir conhecimentos por meio do comércio transoceânico. As companhias de comércio eram dependentes de expertise científica para realizar as expedições ultramarinas. As corporações de comércio eram, de fato, as principais empregadoras de cientistas e homens de saber, cuja missão era investigar o potencial econômico dos produtos encontrados nessas regiões.

A forte associação entre produção do conhecimento e comércio oceânico em larga escala instiga a reflexão sobre a ocorrência de relações dialéticas entre centro e periferia. Nesta concepção, a construção do conhecimento é uma via de mão dupla. Desenvolve-se justamente nas zonas de contatos interculturais, onde os atores,

colonizador e colonizado, contribuem de diferentes formas para moldar o conhecimento científico. Decorre desse raciocínio a ideia de que a certificação, a legitimação e a reapropriação do conhecimento devem ser historicizadas e ressignificadas de modo a combater ideias pré-concebidas sobre a ciência moderna.

Portugal, por exemplo, introduziu reformas profundas para superar sua defasagem científico-cultural frente às potências europeias - Inglaterra e França- e manter o império ultramarino e suas possessões, sobretudo o Brasil (VARELA, 2001). As Reformas Pombalinas marcaram o início da racionalização das práticas administrativas, em moldes europeus, que foi estendido para as Américas sob a influência das metrópoles coloniais (KURY, 2004).

[...] o Iluminismo europeu não foi apenas um movimento no campo das ideias e da filosofia, mas um conjunto de transformações na esfera das sociabilidades e da circulação de textos impressos, bem como uma reunião de práticas administrativas, executadas, em geral, pelo Estado e visando racionalizar o funcionamento da sociedade, conhecer e controlar as populações, a produção, os fluxos e os usos das mercadorias (CHARTIER apud KURY, p. 110, 2004).

Conforme Dias (1968), a vinculação dos ilustrados luso-brasileiros às correntes utilitaristas do pensamento iluminista era um fato. Nesse contexto, foi privilegiado o lado prático da ciência e da filosofia, em detrimento de temas políticos. Todavia, conforme apontado por Kury (2004) o pragmatismo não é um atributo exclusivo das luzes luso-brasileiras. A defesa da utilidade dos estudos da natureza transformou-se em lugar-comum no Iluminismo.

Como explica Varela (2001), principalmente a partir da segunda metade do século XVIII, a coroa portuguesa passou a incentivar com maior intensidade as expedições para a Amazônia para delimitar o território e coletar espécies vegetais e animais. O referido autor considerou as colônias latino-americanas como agentes históricos plenos, ou seja, sujeitos ativos na aclimação da ciência ilustrada e não meros receptores passivos de modelos científicos e ideologias estrangeiras. O autor analisa a ligação de José Bonifácio de Andrada e Silva com a Academia Real das Ciências em Lisboa e aponta a estreita articulação entre Estado e instituições científicas para difundir as “luzes da razão e do progresso” para o desenvolvimento do Império português. A introdução das atividades científicas no Brasil pode ser

melhor entendida pela análise dos diferentes campos científicos que aqui se desenvolveram.

Os relatórios de José Bonifácio na Intendência das minas portuguesas ajudaram a criar e a sustentar uma rede de informações para que o Estado burocrático conhecesse, de forma mais aprofundada, o seu território e suas potencialidades econômicas. Nesse período de início da difusão das ciências para o Brasil, como aponta Domingues (2001), a Coroa portuguesa controlava o fluxo de informações, conhecimentos e técnicas dentro de seus territórios com uma rede de informações destinada a promover uma maior eficiência no que concerne à exploração econômica do território colonial.

O crescimento da investigação científica ocorrida internacionalmente após a Segunda Guerra Mundial, motivou a criação de políticas científicas e tecnológicas, que repercutiu também no Brasil. Estas fundamentam-se em um tripé básico: definição de áreas prioritárias, financiamento das pesquisas dessas áreas, formação e atuação de cientistas e pesquisadores e, em princípio, a conversão dos produtos da atividade científica em benefícios para a sociedade. Ao lado disso, percebeu-se a importância da comunicação científica para garantir a competitividade econômica. Nas palavras de Costa e Leite, “a comunicação científica pode ser entendida como consequência de uma prática de pesquisa, ao mesmo tempo em que é matéria-prima para novos processos de geração de conhecimento” (2017, p. 21). Além disso, segundo Rozados,

computadores, informática e telecomunicações, utilizados inicialmente a serviço do governo com propósitos de espionagem, defesa e preparo para a guerra, foram sendo incorporados às universidades, à indústria, à pesquisa. O aparecimento e a difusão, ao final do século XX, de redes de informação, inicialmente cobrindo pequenas áreas e comunidades e, mais tarde, expandindo-se para níveis globais, como a Internet, exigiram, e estão exigindo, das unidades de informação uma competência cada vez maior na prestação de seus serviços, conhecidos como serviços de informação (2006, p. 53).

Portanto, gerir e otimizar os fluxos desse tipo de informação passaram a ser fundamentais nas estratégias de governos e organismos internacionais (MARCONDES; SAYÃO, 2009; ROZADOS, 2006).

A valorização da comunicação científica trouxe à tona, também, um novo tipo de negócio: a publicação de revistas científicas. Sua monopolização por grandes

editoras internacionais e seu encarecimento vem, porém, inviabilizando sua assinatura pelas bibliotecas de universidades e instituições de pesquisa. (COSTA; LEITE, 2017; IBICT, 2018; MARCONDES; SAYÃO, 2009). Este é um problema enfrentado também por países desenvolvidos. O movimento mundial pelo acesso aberto à informação nasceu nesse contexto, estimulando a criação e utilização de repositórios de acesso livre e *softwares* de código aberto (COSTA; LEITE, 2017; IBICT, 2018; MARCONDES; SAYÃO, 2009).

Portanto, a comunicação científica, passou a um novo patamar a partir dos anos 1990, sob o impacto das tecnologias de informação e comunicação. Para enfrentar o que Van de Sompel e Lagoze (2000) denominaram “crise dos periódicos”, nascem os periódicos científicos eletrônicos e as bibliotecas digitais, ambos em acesso aberto (COSTA; LEITE, 2017).

O acesso aberto pode ser definido como a disponibilização livre e sem restrições de textos completos dos resultados das pesquisas, na Internet. O crescimento desta ideia impulsionou movimentos em prol do acesso aberto à informação científica (MAA) em todo o mundo. Nesta direção, as declarações de maior impacto internacional foram as realizadas em Budapeste (Hungria, 2002), em Bethesda (EUA, 2003) e em Berlim (Alemanha, 2003), que ficaram conhecidas pela sigla BBB (COSTA E LEITE, 2017).

A Declaração de Budapeste, da *Budapest Open Access Initiative* (BOAI), de 2002, foi a primeira a cunhar o termo *Open Access* e definiu duas estratégias para alcançá-lo, orientando o desenvolvimento de sistemas interoperáveis (COSTA; LEITE, 2017). Dez anos após essa declaração, em 2012, ela reafirmou seus princípios e estratégias através da publicação “Ten years on from the Budapest Open Access Initiative: setting the default to open” (COSTA; LEITE, 2017). Essas estratégias são a via verde (*green open access*) e a via dourada (*gold open access*). A via verde consiste no depósito de cópia das publicações científicas em texto completo em repositórios digitais (institucionais ou temáticos) na internet, para a sua disponibilização livre (COSTA; LEITE, 2017; MARCONDES; SAYÃO, 2009). A via dourada consiste na orientação para que os periódicos científicos publiquem seguindo a lógica do acesso aberto (COSTA; LEITE, 2017; MARCONDES; SAYÃO, 2009). Ambas as vias se referem ao fim, ou à minimização dos custos de acesso à literatura científica. Segundo Costa e Leite, a via verde é a estratégia vista com maior otimismo

pelo MAA porque envolve “diretamente o sistema de informação das próprias instituições produtoras de conhecimento científico” (2017, p. 16).

A Declaração de Bethesda, elaborada em Maryland, nos Estados Unidos, em 2003, propôs o direito ao acesso livre à informação científica a partir da atribuição de licenças de uso e distribuição. A Declaração de Berlim, de 2003, enfatiza a importância da formalização de políticas de informação de acesso aberto (COSTA; LEITE, 2017).

No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) lançou, em setembro de 2005, o “Manifesto Brasileiro de Apoio ao Acesso Livre”, enfatizando o papel dos repositórios institucionais para a concretização de um sistema brasileiro de livre acesso à produção científica (MARCONDES; SAYÃO, 2009).

Costa e Leite (2017, p. 16) destacam três argumentos que mostram a importância do acesso aberto:

O primeiro deles diz respeito à natureza pública do conhecimento científico. O segundo desenvolve-se acerca da promoção da possibilidade de acesso à literatura científica por parte das instituições que não teriam como pagar pelo valor cobrado pelas editoras comerciais. Por fim, o terceiro, vê o acesso aberto como um instrumento que pode tornar visível na Internet a produção científica de regiões historicamente excluídas da comunicação científica internacional.

Ainda segundo Costa e Leite (2017), o Brasil é o país que ocupa a segunda posição, em termos mundiais, na quantidade de periódicos científicos de acesso aberto registrados no *Directory of Open Access Journals* (DOAJ). Dessa forma, nos países da América Latina, as TICs e as iniciativas de acesso aberto promoveram significativa ampliação da disseminação da produção científica. Assim, os Repositórios institucionais da América Latina são dispositivos importantes tanto para gerenciar a produção científica das instituições produtoras quanto para dar maior visibilidade à ciência produzida. Com efeito, o acesso aberto diminuiu os problemas de acesso à informação científica e tecnológica pelos países periféricos. A discussão destas questões é retomada na seção dedicada à parte empírica (estudo de caso) da pesquisa.

2.2 A quantificação de dados sociais

A quantificação da atividade científica decorre da quantificação da vida social. Portanto, nesta pesquisa, que tem como foco os aspectos metodológicos da Bibliometria (área que se fundamenta na Matemática e Estatística), é importante situar, inicialmente, o contexto em que surge a Estatística como instrumento do Estado para intervir e controlar a vida social.

2.2.1 Estatística: determinismo e probabilidade

No Dicionário Brasileiro de Estatística, esta é definida como o “conjunto dos processos que têm por objeto a observação, a classificação formal, e a análise dos fenômenos coletivos ou de massa e, por fim, a indução das leis a que eles obedecem. [...]” (RODRIGUES, 1970, p. 103). O procedimento de observar, classificar, analisar e induzir leis de grande alcance consolidou-se naquilo que o filósofo Ian Hacking (2005, p. 8) denominou “domesticação do acaso.”⁶ Isto é, o modo pelo qual eventos fortuitos ou irregulares foram contados, classificados, analisados e instituídos como fenômenos controlados por leis naturais ou sociais. A argumentação de Hacking convida-nos a refletir sobre a atual organização de conceitos estatísticos em dois domínios, ao menos: o do determinismo e indeterminismo físicos e o da informação estatística desenvolvida para fins de controle social. Na presente pesquisa, foram extensamente utilizadas duas obras de Ian Hacking: *The taming of chance*, de 2005 e *The Emergence of Probability*, de 2006, considerados clássicos na discussão filosófica da lógica indutiva e da probabilidade.

O determinismo é um conceito associado à atividade científica, desde Newton (1643-1726), cujo cerne é a ideia de um universo regido por leis causais absolutas, cabendo à ciência descobri-los. A esta visão contrapõe-se o princípio indeterminista, segundo a qual os fenômenos sociais e naturais são produtos do acaso e do livre arbítrio:

quando se contrapõe a um dado determinismo, as doutrinas indeterministas definem-se geralmente em relação com o que supostamente é a “razão do determinismo”. A razão da determinação dos acontecimentos pode ser um decreto de Deus, ou um princípio, ou a suposição de uma cadeia causal estrita. O indeterminismo em geral nega não só que os acontecimentos (ou alguns acontecimentos)

⁶ “Taming of chance”.

estejam determinados, mas também, e às vezes especialmente, a “razão” da suposta determinação (FERRATER MORA, 2005, p. 1482).

No início do século XIX, as explicações causais e determinísticas entram em crise, mesmo no caso das ciências da natureza. É nesse contexto que a Física relativística e a quântica passam a ter um papel preponderante na explicação dos fenômenos naturais. Segundo Paty (2004) a longa tradição da física determinística é rompida: as teorias sobre a causalidade relativística e os sistemas dinâmicos “não-lineares” procuraram demonstrar que o comportamento da natureza pode ser imprevisível.

Quando consideramos as grandezas mais significativas do ponto de vista das características dos fenômenos físicos, a causalidade se apresenta diferente e o determinismo se mostra antropologicamente centrado e ineficiente. Deixam, desse modo, lugar à consideração dessa outra categoria mais significativa, a saber, a de necessidade, que independe de nossas escolhas conceituais e teóricas e tem, ao mesmo tempo, a capacidade de regulá-las (PATY, 2004, p. 465).

A Física quântica provocou a crise do determinismo ao afirmar que o universo é permeado de ocorrências aleatórias. O matemático Andrei Kolmogorov, por exemplo, ofereceu uma definição formal de acaso:

Considerem-se as duas séries numéricas:

010101010101010101 (1)

01101100110111100010 (2)

(1) repete 01 dez vezes; (2) provém de jogar cara ou coroa vinte vezes.

(1) segue certo padrão; (2) não parece seguir nenhum, já que não parece ser possível prever a sequência de uns e de zeros que serão gerados pelo procedimento. [...]. (1) é uma possibilidade entre 2^{20} . [...]

A fórmula mediante a qual se engendra o programa que dá por resultado (2) é exatamente (2); o programa equivale à enumeração dos elementos da série. Pode-se dizer, pois, em geral, que uma série é completamente fortuita quando não pode ser “comprimida” num algoritmo por meio do qual são dadas instruções que engendram a série. (KOLMOGOROV apud FERRATER MORA, 2004, p. 42)

Sob essa perspectiva, não é fortuita uma série que possa ser “comprimida” em um algoritmo. Ela apresenta uma regularidade explícita que pode ser representada pelo algoritmo 2^{20} . O mesmo não ocorre na série (2), sendo ela incerta, irregular, resultante do acaso.

O deslocamento da ideia de causalidade introduz conceitos probabilísticos na descrição de eventos ou sistemas físicos individuais; assim, o mundo apresenta regularidades, sem estar, no entanto, sujeito a leis universais (PATY, 2004). Estas teorias geraram debates intensos, ultrapassaram o mundo das micropartículas e alcançaram os estudos sobre os fenômenos sociais humanos.

2.2.2 Acaso, Probabilidade e Leis estatísticas

A concepção atual de probabilidade surgiu por volta de 1660, porém, só se generalizou no século XIX. De fato, a estatística se concretiza quando as pessoas se propuseram a se contar e dispunham de meios para isso. Até os primórdios do século XIX, não eram rigorosos os conceitos, como o de população e amostra representativa, sendo também precários os instrumentos de coleta de dados. Três fatores, que ocorreram durante a Revolução Industrial, foram fundamentais para essa virada: a avalanche de números, a erosão do determinismo e a invenção da normalidade (HACKING, 2005).

O acaso estatístico passou, então, a fazer parte dos métodos das ciências naturais, tanto quanto das ciências humanas, para prever eventos fortuitos (RODRIGUES, 1970). A contagem de pessoas e de suas características e hábitos, por exemplo, passou a ser uma atividade realizada pelo Estado.

Os censos coloniais eram constituídos de registros quantitativos das posses das metrópoles (extensões de terra, edifícios, pessoas, gado). Esses procedimentos foram mantidos pelos Estados independentes, tal como os Estados Unidos, que realizava censos cada vez mais complexos e estratificados, a cada 10 anos.

Para Hacking (2006), a necessidade das estatísticas sociais era fundamentada em 3 premissas: 1) o Estado prussiano deveria ser trazido à existência; 2) a verdadeira medida do poder de um Estado é sua população; 3) o Estado deveria ter um escritório estatístico central para conhecer seu poder.

A representação do Estado por meio de dados estatísticos já era feita na Prússia, desde o século XVII. Nasceram, nesse período, os escritórios estatísticos, propostos por um protagonista importante: o filósofo Gottfried Wilhelm Leibniz (1646-1716), tarefa anteriormente executada por estatísticos amadores.

Antes da era napoleônica (1799-1815), as contagens oficiais eram sigilosas, acessíveis apenas aos administradores públicos. A Europa passou a alardear seus números a partir do século XIX para mostrar poder (HACKING, 2005).

Os levantamentos abrangentes, as avaliações estatísticas e as intervenções destinadas a disciplinar todo o corpo social são precursores da Biopolítica (HACKING, 2005), um fenômeno dos séculos XIX e XX de intervenção do Estado na vida da população (FOUCAULT, 2008).

Das contagens surgiram as leis sociais de classificação das pessoas em normais e patológicas. O antissemitismo moderno, embrionário do que viria a ser a concepção antissemita alemã do III Reich, teve origem justamente na Prússia do século XIX e se desenvolveu no processo de formação do Estado-nação que levou à unificação da Alemanha:

a mudança da estrutura política levou a nobreza à perda de seus privilégios e a classe média conquistou o direito à ascensão. Essa reforma, uma 'revolução de cima', transformou a estrutura semifeudal do despotismo esclarecido prussiano num Estado-nação mais ou menos moderno, cujo estágio final foi o Reich alemão de 1871 (ARENDR, 2012, p. 59).

A ideia de controlar o futuro, prever perigos e ameaças, bem como a manipulação de dados por meios estatísticos, têm origem nas práticas do Estado. O historiador Georges Minois (2016), ao analisar os caminhos da predição, no século XVIII, mostra como esses métodos evoluíram. As explicações religiosas ou místicas foram abandonadas e substituídas por novos métodos especulativos, visando ao controle social.

2.3 Leis estatísticas

As estatísticas foram sendo progressivamente usadas para quantificar fenômenos humanos. Dados sobre doenças e morte, por exemplo, foram utilizadas para calcular prêmios de seguro de vida e aposentadorias. O mais célebre autor de leis sobre doenças foi o epidemiologista britânico William Farr que criou, no século XIX, um método para medir a duração relativa de doenças, sua frequência e impactos na morbidade.

O inventor, matemático e economista Charles Babbage (1791-1871) propôs, em 1830, na Royal Society de Londres, uma lista com dezenove categorias de constantes que deveriam ser atualizadas a cada dois anos, como, por exemplo, as constantes do sistema solar; os pesos atômicos de elementos, o poder de condução de eletricidade, os índices de refração e polarização, as espécies da fauna e da flora, as tabelas de dados vitais, a proporção de doenças entre as classes trabalhadoras, entre outras, constantes que foram usadas para estabelecer “leis” (HACKING, 2005).

No século XIX, o suicídio foi amplamente estudado em vista da grande quantidade de dados disponíveis. Durkheim realizou um estudo seminal de correlação entre crime, suicídio, moral e qualidade de vida. Esse episódio é importante porque a noção médica de patologia foi estendida à esfera política. Assim, as narrativas sobre o suicídio, no século XIX, são o preâmbulo da tomada de posição sobre a o uso de leis estatísticas para explicar a sociedade, em perspectiva determinística. As tabelas de suicídio tinham estruturas semelhantes às das tabelas sobre a insanidade, categorizadas por sexo, idade, estado civil. O método para observar o suicídio e suas causas baseavam-se no cruzamento de categorias de dados considerados invariáveis, fato que contribuiu para que as pessoas pudessem "entender" ou aceitar as leis estatísticas (HACKING, 2005).

Com efeito, junto aos dados sobre o suicídio surgiu uma nova categoria de problemas – que hoje chamamos de "problemas sociais" – criada e compartilhada por médicos e juristas. As estatísticas sanitárias e a medicina legal passaram a ser parte do mesmo aparato de ordem moral:

todos os fatos demonstram essa notável proposição, já observada por certo número de escritores, de que os fatos morais, tomados em massa e considerados de maneira geral, obedecem, em sua reprodutibilidade, a leis tão positivas quanto as que reinam no mundo físico⁷ (LISLE apud HACKING, 2005, p.73).

Os dados sobre os julgamentos e condenações publicados pelo Ministério da Justiça francês, foram utilizadas por Poisson para formular a Lei dos Grandes números (HACKING, 2005). Esta lei estatística autoriza a generalização ao afirmar

⁷ “All the facts demonstrate this remarkable proposition, already noticed by a certain number of writers, that moral facts, taken en masse, and considered in a general manner, obey, in their reproducibility, laws as positive as those that reign in the physical world.”

que “na medida em que o número de componentes de um sistema aumenta, a sua dispersão relativa tende a diminuir, indicando que, se um sistema for suficientemente grande, as flutuações do seu estado podem ser consideradas negligenciáveis” (SILVA NETO; BASSO, 2010, p. 319).

O movimento que buscava identificar a estabilidade dos fenômenos de massa teve um importante aliado, o astrônomo Adolphe Quetelet (1796-1874). Lakatos e Marconi, identificam Quetelet como o idealizador do método estatístico de “redução de fenômenos sociológicos, políticos, econômicos, etc. a termos quantitativos e à manipulação estatística, que permite comprovar as relações dos fenômenos entre si, e obter generalizações sobre sua natureza, ocorrência ou significado” (2016, p. 90). Todavia, esse método apresenta um percurso curioso. Assim que as estatísticas judiciais de Poisson foram publicadas, Quetelet considerou possível prever “quantos indivíduos sujarão as mãos com o sangue de outros, quantos serão falsificadores, quantos envenenadores, tão bem quanto se pode enumerar antecipadamente os nascimentos e mortes que devem ocorrer”⁸ (QUETELET apud HACKING, 2005, p. 105).

No início do século XIX, foi desenvolvida a Teoria dos erros para determinar o valor mais provável das grandezas físicas. Essa teoria foi importante para consolidar os dados obtidos em medições experimentais. As variações foram consideradas atribuíveis às características dos equipamentos, aos erros do operador, às condições ambientais, dentre outros. Observou-se, no estudo, que os erros tinham distribuição de frequência muito próxima da distribuição normal, também chamada de distribuição de Gauss ou distribuição de Laplace–Gauss, usada para estimar o valor da grandeza medida e o seu erro.

Quetelet anunciou, em 1844, que muitos atributos humanos apresentavam o mesmo gráfico ou distribuição da curva de Gauss. Com base em recursos retóricos, Quetelet transformou a média em uma quantidade real. Todavia, segundo Hacking (2005), a média, na curva de Gauss, não expressa uma quantidade real. Como afirma Hacking, Quetelet definiu o “homem comum”, no *Treatise on Man* (1835) como segue:

⁸ “how many individuals will dirty their hands with the blood of others, how many will be forgers, how many poisoners, nearly as well as one can enumerate in advance the births and deaths that must take place.”

Primeiro, vemos na epígrafe, que Quetelet não estava se referindo a uma média da espécie humana. Ele estava se referindo ao tipo racial como característico de um povo ou nação. O conceito de povo, anteriormente relacionado à cultura, à geografia, língua, governantes e religião, transforma-se, com Quetelet, em uma nova concepção objetiva mensurável de povo. Uma raça poderia ser caracterizada pela mensuração de suas qualidades físicas e morais, sumarizadas no homem médio daquela raça. Isso constitui a metade inicial da eugenia, a outra metade é a reflexão de que podem ser introduzidas políticas sociais que preservem ou alterem as qualidades médias de uma raça. Em resumo, o conceito de homem comum levou a um novo tipo de informação sobre as populações e a uma nova concepção de como controlá-las⁹ (HACKING, 2005, p. 107-108).

Quetelet estabeleceu a média e a curva em forma de sino como indicadores fundamentais das características antropométricas do homem. Assim, as descrições de regularidades em grande escala se transformaram em leis da natureza. Em uma de suas modelagens, Quetelet aplicou a distribuição normal à altura de recrutas franceses. Não funcionou satisfatoriamente, porém, o fracasso foi atribuído a fraudes. Em outra modelagem, utilizou como amostra dados sobre a circunferência do tórax de soldados escoceses ajustadas à distribuição normal. A consequência imediata desse trabalho foi que todo tipo de atributo físico humano, como também dos demais seres vivos, foi investigado e representado de acordo com a curva de Gauss. Portanto, a retórica de Quetelet, que absorvemos desde o momento em que aprendemos estatística, nos faz acreditar que a interação de uma multiplicidade de causas independentes tende a se comportar como uma curva no formato gaussiano. A matemática da probabilidade e a metafísica da causa subjacente foram reunidas por meio de argumentos disjuntos para apresentar uma "compreensão" sobre a estabilidade estatística de todos os fenômenos. Nesse mesmo contexto, Charles Babbage propôs constantes para avaliar a eficiência da produção de bens, fato que favoreceu o setor industrial, mas não o trabalhador.

⁹ "First, we see from the epigraph that Quetelet was not talking about an average for the human species. He was talking about the characteristics of a people or a nation, as a racial type. Where before one thought of a people in terms of its culture or its geography or its language or its rulers or its religion, Quetelet introduced a new objective measurable conception of a people. A race would be characterized by its measurements of physical and moral qualities, summed up in the average man of that race. This is half of the beginning of eugenics, the other half being the reflection that one can introduce social policies that will either preserve or alter the average qualities of a race. In short, the average man led to both a new kind of information about populations and a new conception of how to control them."

As conexões entre informação, controle e lei estatística criaram um dilema metafísico, o fatalismo estatístico. Este foi analisado por Leibniz, no século XVII, que distinguiu três tipos de fenômenos:

a) o efeito ocorre mesmo quando se evita a causa, como se houvesse necessidade absoluta; b) o homem deve aceitar o fado por ser impossível resistir ao curso dos acontecimentos e c) há um certo destino sobre cada coisa regulado pela presciência e pela providência de Deus (FERRATER MORA, 2005, p. 999-1000).

Na concepção fatalista, o mundo é governado por acontecimentos que não se submetem à interferência humana. O fatalismo estatístico é um utilizado como uma ferramenta de controle das populações, com base em informações estatísticas. Um caso clássico é Francis Galton, antropólogo, meteorologista, matemático e estatístico inglês do século XIX, primo de Charles Darwin, que baseou sua obra na ideologia eugenista: pretendia “melhorar a espécie humana através da genética, esterilização e reprodução controlada”, e “eliminar os defeitos hereditários e, portanto, atacar os problemas morais da decadência, crime e doenças sociais, como o alcoolismo e as doenças venéreas” (GOLISZEK, 2004, p. 102).

Goliszek (2004) afirma que a matemática foi utilizada por Galton para dar legitimidade à eugenia, já que recorria a dados estatísticos que supostamente prediziam o comportamento humano. O movimento eugenista foi olhado com desconfiança, no entanto, muitas vezes esquece-se que ele foi motivado pelas mesmas concepções utilitaristas filantrópicas que dão sustentação às tentativas "liberais" de modificar uma população. Hacking (2005) acredita, inclusive, que as raízes da eugenia galtoniana são antigas. Estão presentes na ideia queteletiana de lei estatística para determinar as características de uma população.

Georg Friedrich Knapp (1842-1926), economista alemão, afirmou que Quetelet era vítima de sua formação: as leis sociais, se existissem, deveriam ser como as leis da física. O astrônomo Quetelet criou uma "concepção astronômica de sociedade", na qual as forças que atuam sobre as pessoas eram semelhantes às forças cósmicas ou da gravidade. As "concepções astronômicas da sociedade", em que são invocadas "forças cósmicas" que atuam sobre uma população, se assemelham aos estudos sobre o suicídio de, Durkheim, como pondera Hacking (2005).

Frédéric Le Play, físico francês que viveu no século XIX, afirmou que a estatística é "a observação e a coordenação de fatos que interessam ao corpo social do ponto de vista do governo [...] A política deve usar incessantemente a estatística como meio de regular suas atividades administrativas"¹⁰ (LE PLAY apud HACKING, 2005, p. 136).

O estudo de Le Play sobre os orçamentos domésticos era uma ferramenta fundamental da economia que poderia ser usada como uma medida objetiva da prosperidade de uma classe ou de uma nação. Esse processo exige a classificação rigorosa dos tipos de gastos para comparação, ou seja, era necessário dispor de uma unidade padrão de 'necessidade de consumo'. Ernst Engel fez, em 1857, um importante complemento ao pensamento de Le Play, que resultou na Lei de Engel, segundo a qual "quanto mais pobre o indivíduo, a família ou um povo, maior deve ser a porcentagem da renda necessária para a manutenção do sustento físico e, portanto, uma maior proporção deve ser concedida para o alimento"¹¹ (HACKING, 2005, p. 140). A Lei de Engel foi adotada nos Estados Unidos em 1875 e permanece integrada às técnicas estatísticas estadunidense.

Os orçamentos domésticos de Le Play eram descrições de famílias individuais representativas dos trabalhadores de uma região, que revelam como vivia a família, quais eram suas necessidades, as possibilidades de serem supridas, como também o uso do tempo livre. Os orçamentos elaborados por Engel, ao contrário, eram medidas para a população. No entanto, Hacking (2005) faz uma importante reflexão sobre o evento: "Deixe um homem propor uma ideia antiestatística para refletir a individualidade e resistir à probabilização do universo; a próxima geração a coopta facilmente para que ela se torne parte da maquinaria estatística padrão de informação e controle"¹² (p. 140).

¹⁰ "the observation and coordination of facts that interest the social body from the point of view of government... Politics must unceasingly use statistics as the means by which to regulate its administrative activities"

¹¹ "the poorer the individual, the family or a people, the greater must be the percentage of the income needed for the maintenance of physical sustenance, and of this a greater proportion must be allowed for food."

¹² "Let a man propose an antistatistical idea to reflect individuality and to resist the probabilification of the universe; the next generation effortlessly coopts it so that it becomes part of the standard statistical machinery of information and control"

2.4 O conceito de normalidade

Os três fatores fundamentais da domesticação do acaso foram: a avalanche de números, a erosão do determinismo e a invenção da normalidade (HACKING, 2005), como discutido acima.

A palavra "normal" é antiga, mas adquiriu seu significado mais comum na década de 1820. Os termos patológico e normal foram adotados, inicialmente, no campo da medicina. Hoje, são conceitos que atravessam todas as esferas da vida e da natureza: pessoas, comportamentos, estados de coisas, relações, moléculas, tudo pode ser normal ou anormal. Hacking (2005) acredita que o conceito de normalidade é muito mais importante do que o de determinismo, embora eles sejam interdependentes.

"Normal" e sua concepção relacionada ao progresso é uma marca do século XIX. Ser 'normal' é ser 'comum, típico, usual' (FERREIRA, 2010). Hacking, todavia, pontua que se pode usar a palavra "normal" para caracterizar como as coisas são, mas também para dizer como elas deveriam ser: "A magia da palavra é que podemos usá-la para fazer as duas coisas simultaneamente"¹³ (HACKING, 2005, p. 163).

Durante o século XIX, o normal era associado à média. Na medicina, a saúde pode ser relacionada às ideias de excesso e deficiência, de desnutrição e obesidade, por exemplo. O patológico é o desvio da condição normal, caracterizado tanto pelo excesso quanto pela deficiência, isto é, qualquer diferença em relação ao estado normal.

August Comte, adepto do conceito de normal (oposto a patológico,) levou essa noção para a esfera política, fomentando a ideia de que o normal é o estado ao qual devemos chegar. Comte criou uma certa inquietação em relação à ideia do normal - o normal como a média existente, e o normal como a perfeição que pode ser alcançada. Hacking (2005) afirma que a inquietação pode ser compreendida de diferentes maneiras se consideramos os trabalhos de Durkheim e Galton. Para o primeiro, o normal se associa ao que é o certo, contribui para preservar o *status quo*. Porém, para Galton, o normal é apenas o que fica na média, portanto, algo a ser melhorado (HACKING, 2005). Para Durkheim, há conceitos comuns entre as ciências

¹³ "The magic of the word is that we can use it to do both things at once."

sociais e as ciências biológicas, principalmente porque, em ambas, a noção de função tem papel importante. A função, ou explicação funcional, permite entender como se relacionam as instituições, as atividades e crenças que compõem a sociedade, em analogia às funções dos órgãos do corpo e suas relações com outros órgãos. Na introdução à primeira edição de *Da divisão do trabalho social*, Durkheim afirma que o normal designa o que é encontrado na média das espécies. Segundo Hacking, a média é entendida como "a massa densa central" da curva de Gauss e "O mesmo método deve ser seguido na ética. Um fato moral é normal para um determinado tipo social quando é observado na média dessa espécie; é patológico em circunstâncias antitéticas"¹⁴ (HACKING, 2005, p. 172).

Galton, por sua vez, foi responsável pela autonomização das leis estatísticas. Criticou as ideias sobre o homem médio, proposto por Quetelet e outros, que afirmavam que a Curva Normal era o produto de uma miríade de pequenas causas independentes (HACKING, 2005). Ainda segundo Hacking (2005), Para Durkheim o patológico era anormal, enquanto, para Galton, o anormal poderia ser o excepcional, o mais saudável da população. O normal, portanto, não era bom, mas medíocre, de modo que o certo e o bom estão concentrados no lado direito da curva de distribuição.

Galton concebeu também, a teoria da correlação e da regressão. Atribui-se a ele a expressão: "a regressão em direção à média" e não a reversão em direção à mediocridade (Hacking, 2005). Sua terminologia revela o fascínio pelo excepcional, em contraposição à preocupação de Quetelet com as médias.

Houve repetidas tentativas para explicar como a homogeneidade estatística, em larga escala, pode resultar de processos determinísticos. No entanto, as leis estatísticas tornaram-se autônomas ao serem usadas não apenas para a previsão de fenômenos, mas também para sua explicação. Galton foi o precursor das leis estatísticas autônomas quando tentou explicar a distribuição da genialidade hereditária em famílias superdotadas. Esse momento é importante porque a estatística saiu do campo da previsão e entrou no campo da explicação, da inferência (HACKING, 2005).

Tal como vários pensadores da época, Galton acreditava que a "raça" humana poderia ser melhorada caso fossem evitados "cruzamentos indesejáveis". O objetivo

¹⁴ "The same method must be followed in ethics. A moral fact is normal for a determined social type when it is observed in the average of that species; it is pathological in antithetical circumstances"

de Galton era incentivar o nascimento de indivíduos mais notáveis ou mais aptos e desencorajar o nascimento dos inaptos. Propôs o desenvolvimento de testes de inteligência para selecionar homens e mulheres brilhantes, destinados à reprodução seletiva. Galton procurou comprovar que traços morais, mentais ou físicos eram transmitidos por pedaços de material hereditário e não por uma série de pequenas influências independentes. Foi assim que chegou à ideia de correlação: **A** não é a única causa de **B**, mas contribui para a produção de **B**, embora possam ocorrer outras causas. A medida de causalidade, portanto, substitui a ideia simples de causa. Nesse momento, a visão filosófica do universo passou a ser a de um sistema correlacionado de variáveis, aproximando-se, mas de modo algum alcançando correlação perfeita, isto é, causalidade absoluta (HACKING, 2005).

2.5 O contraponto

A restauração do acaso é um fenômeno dos séculos XIX e XX. Charles Sanders Peirce e Friedrich Nietzsche tiveram papel preponderante na retomada dessa ideia. Como afirma Dias (2009):

O acaso é assim para Nietzsche um conjunto de coisas humanas, as mais cotidianas. Vem sempre a nosso favor, pois traz o presente. Presente no sentido temporal e presente como dádiva: “Ouve-se, não se busca; toma-se, não se pergunta quem dá”. A vontade de crescer, de dar forma, de devir, de intensificar a potência quer o acaso. Ele não é um incidente que devemos afugentar, mas o elemento essencial que determina a plasticidade da vontade criadora. Se o que vem até nós, o inesperado que, de algum modo, se espera que venha, surge como absolutamente necessário: se o queremos, se o afirmamos com toda a nossa vontade, ele nos traz o presente e impulsiona à ação, a uma ação criadora (DIAS, 2009, p. 11).

Peirce põe em dúvida as crenças determinísticas, tendo como ponto de partida a crítica à Doutrina da necessidade. Segundo Hacking (2005, p. 11), para Peirce, esta doutrina é nula, pois os fenômenos da natureza decorrem de uma miríade de acasos. Assim, as regularidades observadas na natureza não autorizam a formulação de leis determinísticas. A crítica ao determinismo fundamenta a caracterização de raciocínios ou inferências: dedução, indução e abdução. Nessa tríade, do raciocínio dedutivo decorrem inferências necessárias, na indução as

inferências nascem da experimentação de uma teoria, enquanto a abdução é o raciocínio que permite formular novas hipóteses explicativas (PEIRCE, 1977, p.33-35).

Peirce afirma que "A lógica é a ciência necessária para testar o argumento"¹⁵ (PEIRCE, 1982, p.294). Assim, no raciocínio dedutivo, em um argumento válido, quando as premissas são verdadeiras, a conclusão é verdadeira. Os raciocínios dedutivos são seguros pois não lidam com o acaso. O raciocínio indutivo, ao contrário, é permeado de riscos, pois, as premissas podem ser verdadeiras, mas, a conclusão, falsa. É neste contexto que a probabilidade opera como recurso para minimizar os riscos dos raciocínios indutivos (HACKING, 2005). A inferência estatística, portanto, aproxima a estatística e lógica, em particular quando os raciocínios partem de um conjunto representativo (amostra) do universo (população). Para realizar a inferência, é necessário dispor de dados observados experimentalmente. Quando as premissas da inferência são quantitativas, pode-se substituir o 'normalmente' por uma probabilidade numérica.

Segundo Hacking, para Peirce, as leis da natureza são comumente apresentadas como equações com alguns parâmetros fixos, em oposição à ideia de constantes da natureza de Babbage. Afirmava, assim, que era necessário parar de tentar modelar o mundo, pois as "constantes" são apenas variáveis aleatórias que se estabeleceram no curso da evolução das leis (HACKING, 2005).

As ideias de Peirce são apresentadas por meio de tricotomias: "O acaso é o primeiro, a lei é a segunda, a tendência a adotar hábitos é a terceira"¹⁶ (PEIRCE apud HACKING, 2005, p. 215). Porém, o acaso não é anulado por leis estatísticas porque "o acaso flui em todas as direções dos sentidos"¹⁷ (HACKING, 2005, p. 215). Não é algo que possa ser controlado e modelado; ao contrário, pode ser inferido pela combinação de lógica indutiva e probabilidade.

Por que é importante aproximar a lógica indutiva e a probabilidade da Bibliometria? De nosso ponto de vista, muitos estudos bibliométricos partem da ideia de que a estatística é uma ferramenta dedutiva, da qual são derivados resultados verdadeiros, inequívocos. A pulsão quantitativista que perpassa a produção de indicadores bibliométricos com a pretensão de substituir a subjetividade das

¹⁵ "Logic is the Science needed in order to test argument"

¹⁶ "Chance is First, Law is Second, the tendency to take habits is Third."

¹⁷ "chance pours in at every avenue of sense"

avaliações “por pares” por critérios “objetivos”, expressos por números, é perceptível desde a primeira menção ao termo “bibliometria”. Em 1934, no *Tratado de Documentação*, Paul Otlet orienta a “procurar, cada vez mais, revestir os dados com a forma exata dos números, passar do estágio qualitativo ou descritivo para o estágio quantitativo” (OTLET, 2018, p. 18).

Os processos estatísticos são indutivos, seu uso implica riscos que podem ser minimizados pela probabilidade, porém seguem sendo raciocínios arriscados. Como afirmado por Hacking “há uma importante tradição representada igualmente pelo filósofo C.S. Peirce e pelo estatístico Jerzy Neyman: a indução é aquele modo de argumentar que preserva a verdade na maior parte do tempo” (HACKING, 2009, p.188). Podemos afirmar, nessa perspectiva, que um raciocínio indutivo, pode estar contaminado pelos valores científicos e sociais vigentes em cada época.

2.6 Lógica da inferência estatística

Uma das hipóteses levantadas no presente trabalho é a de que a aproximação da lógica da inferência estatística, na abordagem da Filosofia da Ciência, pode contribuir para aprofundar a discussão teórica e metodológica dos estudos bibliométricos e cientométricos.

A Estatística, na Filosofia matemática, tem como objeto a matematização da lógica. Subdivide-se em estatística descritiva, probabilidade e inferência estatística. A estatística descritiva é uma subárea da Estatística que analisa um conjunto de dados e descreve as características importantes desses dados. A probabilidade é umas das teorias matemáticas utilizadas para estudar a incerteza dos fenômenos aleatórios. Por fim, a inferência estatística é a subárea destinada a entender como ocorrem as extrapolações dos dados de uma amostra para uma população, como observado, por exemplo, nos testes de hipóteses ou testes de significância. Os estudos bibliométricos, em sua maioria, inserem-se no campo da estatística descritiva.

Afirma-se, atualmente, que quando o elemento de ligação entre as premissas e a conclusão é de natureza probabilística, estamos diante da inferência estatística. A Estatística floresceu com a conexão indução-probabilidade, proposta por Peirce. Ronald Fisher criou o teste de significância ou teste de hipóteses com base no raciocínio da lógica condicional. J. Neyman e E. Pearson, em meados do século XX,

criaram o que chamaram de comportamento indutivo através da criação de testes de hipóteses. Outro exemplo importante é o Teorema de Bayes, originalmente proposto no século XVII pelo reverendo Bayes, foi retomado e ampliado com o uso de tecnologias computacionais. Esse teorema afirma que uma nova convicção deve depender de duas coisas — uma convicção prévia (e todo o conhecimento que a informou) multiplicada pelo “valor diagnóstico” da nova informação. Porém, o cerne do teorema é a percepção central de Bayes de que é possível chegar gradualmente mais perto da verdade por meio de constantes atualizações proporcionais ao peso da evidência (HACKING, 2015).

Alain Desroisières (2011), um discípulo de Bourdieu, desenvolveu as ideias de Hacking na obra *The politics of large numbers*, na qual analisou o desenvolvimento do pensamento estatístico e os desdobramentos que levaram à probabilidade. Seu ponto de partida é epistemológico, firmemente situado nos debates sobre a apreensão da realidade. O desenvolvimento da estatística tem como pano de fundo a tensão filosófica sobre a percepção da realidade que antecede a produção do conhecimento: de um lado, os objetos a serem descritos são reais; de outro, o objeto resulta de convenções (DÉSROISIÈRES, 2011, p. 1).

Para Desroisières (2011), parece mais frutífero compreender a ciência, e a estatística em particular, como visões que se conectam a um contexto mais amplo, em um sistema de relações. Na tentativa de contornar a tensão entre objetividade e subjetividade estatística, propõe discutir a probabilidade epistêmica (grau de crença) e a probabilidade frequentista.

A conexão complexa entre pontos de vista prescritivos e descritivos é particularmente acentuada na história do cálculo de probabilidade pela oposição recorrente entre probabilidade subjetiva e objetiva; ou, de acordo com uma outra terminologia, entre probabilidade epistêmica e probabilidade frequentista (Hacking, 1975). Na perspectiva epistêmica, a probabilidade é um grau de crença, enquanto na visão frequentista, diversidade e risco são fenômenos naturais, não resultam de conhecimentos incompletos (DESROISIÈRES, 2011, p. 7).

Desroisières (2011, p. 14) afirma que há importantes distinções entre a estatística derivada da matemática (científica abstrata) e a estatística burocrática, ou a face científica e administrativa da estatística, que ganhou força na virada do século XX. É importante ressaltar que Hacking (2009) e Desroisières (2011) procuram

reconstruir historicamente o pensamento quantificador, em especial o dos séculos XVIII e XIX. Enquanto Hacking (2005) investiga a crise de um modelo estatístico baseado em leis, que acaba por promover o ressurgimento e desenvolvimento da probabilidade, em específico a erosão do pensamento determinista e causal e a ascensão da lógica probabilística, Desroisières (2011) aborda o debate epistemológico referente às formas de apreensão da realidade.

Estas concepções sobre a estatística, aqui expostas de forma breve, são importantes para analisar a Bibliometria, e em particular as leis bibliométricas, tema do próximo capítulo.

3 A BIBLIOMETRIA NA QUANTIFICAÇÃO DAS ATIVIDADES CIENTÍFICAS

A Bibliometria é definida, em uma primeira aproximação, como a “análise estatística dos processos de comunicação escrita, tratamento quantitativo (matemático e estatístico) das propriedades e do comportamento da informação registrada” (FIGUEIREDO, 1998, p. 79). Pritchard é considerado um dos primeiros pesquisadores a caracterizar essa disciplina: “a Bibliometria é a aplicação de métodos matemáticos e estatísticos para analisar o curso da comunicação escrita e o curso de uma disciplina”¹⁸ (PRITCHARD, 1969, p. 348-349).

Spinak, por sua vez, acrescenta que a Bibliometria é a:

Aplicação da análise estatística para estudar as características de uso e criação de documentos. Estudo quantitativo da produção de documentos conforme refletido nas bibliografias. Aplicação de métodos matemáticos e estatísticos ao estudo do uso de livros e outras mídias dentro e entre os sistemas de bibliotecas. Estudo quantitativo das unidades físicas publicadas, ou das unidades bibliográficas, ou seus substitutos.¹⁹ (SPINAK, 1996, p. 34).

Spinak conclui esse verbete afirmando que “a bibliometria estuda a organização de setores científicos e tecnológicos a partir de fontes bibliográficas e patentes para identificar os atores, suas relações e tendências”²⁰ (SPINAK, 1996, p. 35). Além de apresentar uma série de definições que associam a bibliometria à aplicação de técnicas de análise quantitativa, o autor amplia a definição ao posicioná-la como um campo de estudos que analisa as atividades científico-tecnológicas, de forma geral, com base em fontes bibliográficas registradas.

As definições acima mostram a centralidade da estatística nos métodos bibliométricos. Na presente pesquisa dirigimos o foco para a aplicação da Bibliometria à avaliação da atividade científica desenvolvida em universidades e institutos de pesquisa. Atualmente, há uma avalanche de indicadores que classificam

¹⁸ “defino bibliometría como la aplicación de las matemáticas y los métodos estadísticos para analizar el curso de la comunicación escrita y el curso de una disciplina.”

¹⁹ “Aplicacion de analisis estadisticos para estudiar las características del uso y creacion de documentos. Estudio cuantitativo de la produccion de documentos como se refleja en las bibliografias. Aplicacion de metodos matematicos y estadisticos al estudio del uso que se hace de los libros y otros medios dentro y entre los sistemas de bibliotecas. Estudio cuantitativo de las unidades físicas publicadas, o de las unidades bibliograficas, o de sus substitutos.”

²⁰ “La bibliometria estudia la organizacion de los sectores científicos y tecnologicos a partir de las fuentes bibliográficas y patentes para identificar a los actores, a sus relaciones y tendencias.”

universidades, laboratórios, pesquisadores. Tais indicadores foram promovidos ao status de evidências objetivas que justificam a atribuição de valor (positivo ou negativo) à pesquisa científica e tecnológica (GINGRAS, 2016). Vale lembrar que um indicador científico é uma medida que fornece informações sobre os resultados da atividade científica de uma instituição, país ou região do mundo[...]”²¹ (SPINAK, 1996, p. 114)

3.1 Constituição da Bibliometria: breve histórico

A exposição a seguir aborda os aspectos mais significativos do contexto de adoção de métodos quantitativos para mensurar fatos sociais, inclusive informações hoje conhecidas como estatísticas sociais, o que motivou esse processo e as concepções presentes nessas práticas até se chegar à consolidação da Bibliometria.

Em 1832, na Royal Society de Londres, o economista inglês Charles Babbage propôs dezenove categorias de constantes como parâmetros de avaliação quantitativa. Uma das categorias referia-se aos números de livros em grandes bibliotecas públicas (Annalen der Physik und Chimie, 1832 apud HACKING, 2005). Essas iniciativas do século XIX eram parte do processo de mensurar e quantificar o mundo. Pode ser interpretado como uma das primeiras evidências da criação de indicadores quantitativos de fenômenos sociais. No entanto, na atual "era da informação", Charles Babbage é lembrado como o criador dos princípios gerais do computador digital, não sendo, portanto, reconhecido como aquele que aplicou métodos matemáticos e estatísticos para contabilizar livros e outros tipos de documentos.

A literatura especializada da Ciência da Informação aponta outros marcos da criação de métodos quantitativos de natureza bibliométrica. Andrés (2009) considera que as primeiras alusões à bibliometria datam de 1873, quando o botânico suíço Augustin Pyrame de Candolle investigou o que influenciava o sucesso científico dos países. Pritchard (1969) informa que E. Wyndham Hulme, foi o precursor da “bibliografia estatística” quando analisava, na Universidade de Cambridge (Reino Unido), por volta de 1922, o crescimento das patentes e sua relação com os processos sociais. Hulme, por sua vez, utilizou o termo bibliografia estatística para designar “os

²¹ “medida que provee informacion sobre los resultados de la actividad científica en una institucion, pais o region del mundo.”

processos da ciência e da tecnologia por meio da contagem de documentos” (PRITCHARD, 1969, p. 348).

Porém, a Bibliometria ganha real impulso nos anos 1920-1930 quando os bibliotecários começaram a utilizar métodos estatísticos para selecionar, dentre as revistas especializadas, as que fossem de maior interesse para os pesquisadores. “É nesse contexto de gestão de coleções de revistas que a análise sistemática, não mais das próprias publicações, mas das referências que elas continham (as citações), surgiu” (GINGRAS, 2016, p. 24). Em um estudo publicado na Revista *Science*, em 1927, os químicos Gross e Gross demonstraram que em vez de selecionar revistas com base no interesse dos pesquisadores, seria mais adequado encontrar uma medida objetiva para conhecer o uso real do periódico e os de maior impacto em seu campo de atuação.

Fonseca (1973), por outro lado, credita a Paul Otlet a primeira definição do termo bibliometria (bibliométrie), em 1934, no *Traité de documentation*.

As ciências do livro também devem procurar hoje em dia introduzir a medição em suas pesquisas. Uma vez que o livro é objeto da psicologia, da sociologia e da tecnologia, seus fenômenos são suscetíveis de ser medidos. A ‘bibliometria’ será a parte definida da bibliologia que se ocupa da medida ou quantidade aplicada aos livros (aritmética ou matemática bibliológica). Todos os elementos estudados pela bibliologia são, em princípio, suscetíveis de medição e é preciso procurar cada vez mais revestir seus dados com a forma exata dos números, passar do estágio qualitativo ou descritivo para o estágio quantitativo (OTLET, 2018, p. 18).

O fetichismo numérico que tomou conta da Inglaterra e da França está igualmente registrado no *Tratado de Documentação* de Otlet:

A sociologia tende também a se tornar quantitativa. Ela trabalha com grupos e estes são suscetíveis de ser recenseados, e para isso a estatística estabelece os métodos e registra os resultados. As coisas do livro não são de modo algum mensuradas, nem em sua realidade objetiva e material, nem em sua realidade subjetiva e intelectual. É, portanto, desejável que se evitem esforços nesse sentido (OTLET, 2018, p. 18).

Na passagem a seguir, Otlet apresenta sua visão sobre como a coleta e análise de dados foi importante para influenciar os hábitos literários da população, na URSS.

A URSS informa as seguintes cifras: 500 milhões de exemplares em 1930, comparados com 120 milhões antes da guerra. Atualmente, 50% de obras sobre assuntos econômicos e sociais e 30% sobre assuntos técnicos ao invés de 5 e 14. Editaram-se 16 milhões de exemplares de obras de Lenin e 50 mil por ano de O capital de Marx, e 30 milhões de exemplares de obras clássicas. No ano passado foram 52 mil títulos de obras das 58 nacionalidades [da URSS], em lugar de 24 em 1913 (OTLET, 2018, p. 19).

Ainda na primeira metade do século XX, Alfred Lotka, estatístico e matemático estadunidense, analisou a frequência de distribuição da produção científica, dando origem à chamada Lei de Lotka, segundo a qual o número de autores que publicam n artigos é inversamente proporcional ao quadrado deste número (LOTKA, 1926). Na mesma época, segundo Godin (2006), a Psicologia passou a avaliar o crescimento da área por meio da contagem da literatura produzida.

Um estudo importante deste período é o do bibliotecário britânico Samuel C. Bradford que, na década de 1930, investigou a distribuição de frequência de artigos em revistas. Seu trabalho deu origem à Lei de Bradford, usada para analisar a produtividade de periódicos (BRADFORD, 1934). Na mesma época, George Kingsley Zipf, linguista da Universidade de Harvard, estudou a frequência das palavras em um texto, dando origem às Leis de Zipf (ZIPF, 1935, 1949). Estas últimas são consideradas uma generalização das Leis de Lotka e de Bradford.

Após a Segunda Guerra Mundial, importantes avanços ocorreram na Bibliometria. O crescimento exponencial do número de artigos publicados tornava difícil a atualização de conhecimentos dos pesquisadores. Surge daí a ideia de criar um sistema de indexação de artigos citados. Eugene Garfield, linguista e empresário estadunidense, desenvolveu, em 1955, o Science Citation Index (SCI), uma base de dados de citações contidas em artigos de revistas científicas (GARFIELD, 1955). O objetivo era, e ainda é, facilitar a pesquisa bibliográfica, ou seja, identificar nos artigos as citações de outros artigos que tratam de um mesmo assunto. “A intuição básica repousa no fato de que a citação de um texto dado indica forçosamente um elo conceitual entre o artigo citante e o artigo citado” (GINGRAS, 2016, p. 27).

Em 1963, o historiador das ciências e físico, Derek de Solla Price publicou a obra *Little Science, Big Science* na qual apresentou as principais características do crescimento científico. O termo “big science” refere-se aos instrumentos e instalações

de grande escala, apoiados por recursos do governo ou de agências internacionais. Porém, no livro, Price afirma que a ciência progrediu a partir da “little science”, que é tradicionalmente realizada por um grupo pequeno de pesquisadores que se tornam eminentes em seus campos de estudo. Para identificar como ocorre a transição da pequena para a grande ciência é necessário analisar a produtividade ao longo do tempo, que cresce por processos multiplicativos. Com base nessa hipótese, Price (1963) formula a lei de crescimento exponencial da ciência. De acordo com a função exponencial, a taxa de crescimento será proporcional ao tamanho da população, de modo que, quanto maior a população, mais rapidamente ela crescerá. A segunda característica dessa função exponencial diz respeito às ciências de rápido crescimento, que tendem a duplicar num período de 10 a 15 anos (PRICE, 1963).

Para Gingras (2016), o verdadeiro impulso da Bibliometria ocorreu durante os anos 1970, em razão do fortalecimento de políticas científicas. O autor apoia sua argumentação em dois fatores: de um lado, no lançamento do *Manual de Frascati* pela Organização de Cooperação e Desenvolvimento Econômico (OCDE), que propõe práticas de medida padrão da pesquisa e desenvolvimento; de outro, na publicação de um artigo de Price, em 1965, na revista *Nature*, que abordou os fundamentos científicos da política científica. É nesse contexto que o Congresso dos Estados Unidos confia à National Science Foundation a tarefa de produzir os indicadores para acompanhar e medir a evolução do estado da ciência e da tecnologia.

Nesse contexto, Narin, quando atuava em um projeto desenvolvido na National Science Foundation (NSF) cunhou o termo Evaluative bibliometrics (EB), definida como segue: “O uso de publicações e análise de citações na avaliação da atividade científica”. (NARIN, 1987, p. 99).

Em 1987 começam a ser organizados os congressos internacionais para discutir as métricas da ciência. Em 1978, Garfield, Malin e Small publicaram o capítulo Citation data as Science indicators na obra coletiva “Para uma métrica da ciência: o advento dos indicadores da ciência” (ELKANA et al., 1978). Em 1981 foi lançada a revista *Scientometrics* e, em 1993, foi criada a International Society for Scientometrics and Informetrics, uma sociedade científica dedicada especificamente ao estudo da Cientometria.

Atualmente, os termos, bibliometria e cientometria, são empregados como sinônimos (GLÄNZEL, 2003). Bufren e Prates (2005), ao investigarem os significados

dos termos relacionados aos métodos quantitativos na literatura especializada, concluíram que o termo Bibliometria é geralmente empregado para referir-se à mensuração da informação nos mais diversos suportes, sendo considerada como ciência de fato, e não uma simples técnica de mensuração.

Com a informatização crescente das atividades relacionadas à produção de indicadores bibliométricos e, particularmente com o advento das redes sociais digitais, surgem novos conceitos. O termo Infometria, de acordo com Glänzel, foi cunhado por pesquisadores do VINITI – *All-Union Institut for Scientific and Technical Information* (subsidiária da *Russian Academy of Sciences*). Glänzel aponta como fonte Gorkova (1988). O termo designa um subcampo da Ciência da Informação que lida com as análises estatísticas do processo de comunicação da ciência, incluindo as mídias eletrônicas (GLÄNZEL, 2003). Porém, Bufren e Prates (2005) afirmam que o termo infometria foi criado em 1979 pelo professor alemão Otto Nacke, um termo guarda-chuva que abrange a bibliometria e a cientometria.

No universo da World Wide Web (WWW), “dentre as medições que podem ser realizadas no campo da webometria, destacam-se a frequência de distribuição e as classificações que compreendem categorias tais como homepages pessoais, institucionais ou organizacionais” (BUFREM; PRATES, 2005, p. 15). Semelhantes redes sociais incrementaram a visibilidade dos produtos das pesquisas. Para acompanhar o fenômeno foram criadas novas métricas, como a Almetria (MARICATO; LIMA, 2017) que objetiva analisar o impacto dos textos e afirmações científicas publicadas em portais de notícias, blogs, mídias sociais e gestores de referências. No entanto, muitas críticas são feitas à Almetria por lidar com dados muitas vezes efêmeros, interpretados de forma descontextualizada.

Santos e Kobashi (2009), relacionam os termos bibliometria, cientometria e infometria:

Enquanto a bibliometria tem como objetos de estudo os livros ou as revistas científicas, cujas análises se vinculam à gestão de bibliotecas e bases de dados e a cientometria preocupa-se com a dinâmica da ciência, como atividade social, tendo como objetos de análise a produção, a circulação e o consumo da produção científica, a infometria abarca as duas primeiras, tendo desenvolvido métodos e ferramentas para mensurar e analisar os aspectos cognitivos da ciência. (SANTOS; KOBASHI, 2009, p.159)

Assim, por motivos pragmáticos, na presente pesquisa adotamos Bibliometria como termo genérico para denominar os diferentes procedimentos de mensuração da produção científica e tecnológica.

3.2 Os métodos bibliométricos

Os principais métodos utilizados nos estudos bibliométricos podem ser sistematizados em duas categorias: 1) Métodos estatísticos descritivos, que utilizam a contagem, classificação e cálculo de frequência, práticas comuns nos estudos relativos às estatísticas primárias; 2) Métodos bibliométricos, que utilizam leis bibliométricas e análises de citação, criados especificamente para lidar com o campo da bibliometria; 3) Métodos relacionais de primeira e segunda gerações, que se dedicam à análise de coautoria, redes de citações, cocitações, coocorrência de palavras. Nesta dissertação, serão abordados somente os métodos 1 e 2, pertinentes ao uso que serão feitos no Estudo de caso – Capítulo 4.

3.2.1 Métodos estatísticos descritivos

As análises descritivas são a maneira mais simples de iniciar um estudo bibliométrico (ANDRÉS, 2009). Apresentam as características do objeto de estudo com base na coleta, classificação e cálculo da frequência dos dados. Essas análises são vantajosas porque oferecem uma visualização rápida e uniforme de certos aspectos do objeto de estudo.

Uma vez coletados os documentos que farão parte do estudo, é possível analisar sua evolução temporal. Para isso, é necessário identificar o ano de publicação de cada documento e calcular a frequência anual das publicações. Glänzel (2003) e Andrés (2009) afirmam que a evolução temporal permite identificar a produtividade científica de áreas de conhecimento. Todas as informações significativas sobre a evolução temporal, como ano, número de publicações (frequência), porcentagem e porcentagem cumulativa são sistematizados em tabelas. Estes dados, juntamente com sua representação gráfica, mostram, diacronicamente, a tendência de evolução do campo estudado.

Apresentamos abaixo uma situação hipotética que exemplifica como a evolução temporal da produção científica pode ser identificada. A tabela 1 mostra, o número de publicações recuperadas no repositório digital da instituição ao longo dos anos 2000 a 2009.

Tabela 1 - Publicações recuperadas no repositório da instituição

Ano	Número de publicações	Frequência relativa (%)	Frequência relativa acumulada
2000	20	2,9	2,9
2001	50	7,3	10,3
2002	80	11,7	22,0
2003	60	8,8	30,9
2004	90	13,2	44,1
2005	50	7,3	51,4
2006	70	10,2	61,7
2007	80	11,7	73,5
2008	90	13,2	86,7
2009	90	13,2	100
TOTAL	680	100	

Fonte: Elaboração do autor

As informações coletadas são, geralmente, organizadas em tabelas para que se possa ter uma visualização global dos dados. Conforme Dantas (2004), na estatística, a amostra coletada (N) é representada pela frequência absoluta que, neste exemplo é o número de publicações. A frequência absoluta (n_i) de um evento i é o número n_i , de vezes que o evento ocorreu. Porém, a frequência absoluta é limitada no que concerne à comparação dos dados, por isso é praxe calcular a frequência relativa (f_i), obtida pela razão entre a frequência absoluta (n_i) e o tamanho da amostra (N), isto é, $f_i = \frac{n_i}{N}$.

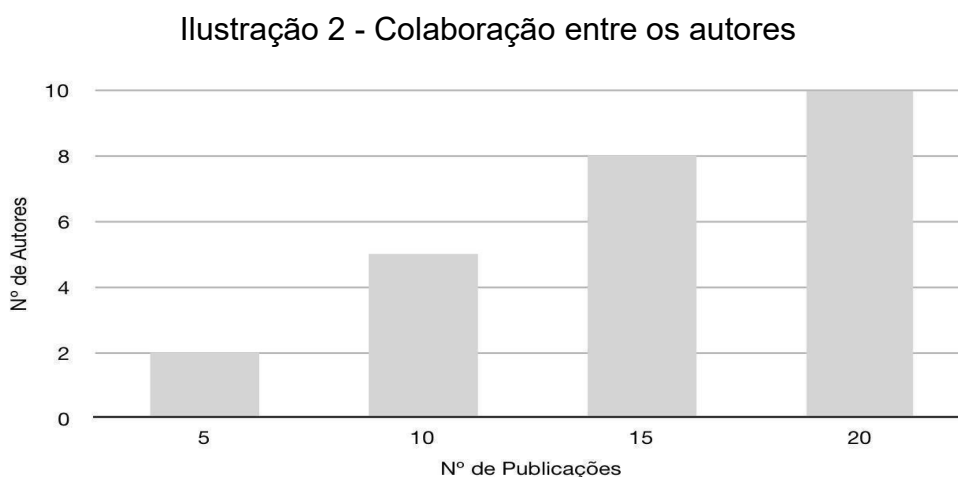
A Tabela 1 indica que a produção científica cresceu ao longo dos anos de 2000 a 2009, tendo havido uma abrupta queda no ano de 2005. A coleta, classificação, cálculos das frequências e apresentação dos dados faz parte do método estatístico, porém, o método estatístico é parte de um universo maior, o universo da lógica indutiva. Qualquer explicação sobre os eventos observados são inferências, ou seja, argumentos de risco que poderiam ser minimizados com a devida combinação entre

o cálculo de probabilidade e informações retiradas da literatura, combinação que precisaria ser avaliada pela lógica indutiva.

As análises descritivas de autores, mais comumente utilizadas são: 1) Número de autor por publicação e 2) Autores mais produtivos. A primeira diz respeito à análise do número de autores de cada publicação. Essa análise permite identificar o grau de colaboração entre autores. A segunda é, basicamente, uma análise classificatória para elaborar *rankings* dos autores mais produtivos de uma instituição, de uma área do conhecimento, laboratório, localidade. Estas análises são complementares e a decisão sobre seu uso dependerá de sua adequação para descrever os dados disponíveis e as características de cada estudo bibliométrico.

Os indicadores de autores têm unidades de agregação em nível micro. É necessário ressaltar que a correta identificação das pessoas que mais publicam em uma determinada área de pesquisa requer depuração para que homônimos, erros ortográficos e variantes não interfiram nos resultados.

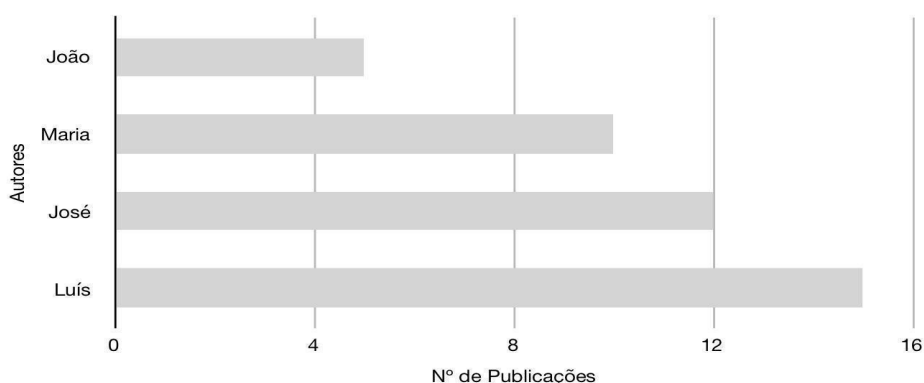
Ainda em relação aos dados do exemplo da Tabela 1, duas situações hipotéticas foram criadas de modo a explicitar as análises descritivas. A Ilustração 2, abaixo, mostra a colaboração entre os autores, pois relaciona o número de autores em uma determinada amostra de publicações.



Fonte: Elaboração do autor.

A Ilustração 3 apresenta a classificação dos autores mais produtivos de uma dada instituição.

Ilustração 3 - Classificação dos autores mais produtivos



Fonte: Elaboração do autor.

Listas classificatórias contendo as instituições ou países mais produtivos de um determinado campo, comum em estudos bibliométricos, são usualmente publicados em *rankings*. Neste caso, o procedimento utilizado consiste em identificar a afiliação de todos os autores dos documentos que compõem a amostra (agregação de dados em nível meso). Andrés (2009) afirma que, embora laboriosa, esta tarefa proporciona uma melhor descrição da participação de autores, bem como a colaboração entre instituições e países. Uma classificação recorrente em estudos bibliométricos é a apresentação de listas das instituições que mais produzem, em termos percentuais.

A análise da produtividade científica por instituição enfrenta um problema técnico relacionado ao modo como a informação sobre a afiliação dos autores é obtida e organizada. Primeiramente, como apontado por van Raan (2005), uma instituição pode ser referida de diferentes formas. Além disso, pode ser difícil distinguir os muitos institutos dentro de uma organização principal. Outra dificuldade refere-se aos problemas linguísticos, pois, a mesma instituição pode ser mencionada em mais de um idioma. Glänzel (2003) e Andrés (2009) afirmam que a escolha de bases de dados adequadas pode contornar esses problemas.

Outros indicadores passíveis de serem gerados, a partir de análises descritivas, são: o idioma do documento e seu assunto. O primeiro é obtido pela contagem do idioma das publicações. No entanto, como a maioria das revistas listadas nas grandes bases de dados referenciais - Web of Science, Scopus, PubMed, etc - são publicadas em inglês, este é o principal idioma aferido por indicadores. O segundo

indicador refere-se aos assuntos dos documentos. Como os resultados são dependentes dos procedimentos internos de indexação de cada base de dados, a quantificação de assuntos é uma tarefa bastante complexa.

3.2.2 Métodos bibliométricos

Os métodos bibliométricos propriamente ditos procuram medir a produtividade de autores, a produtividade de periódicos, a ocorrência e frequência de palavras em um dado texto e a análise de citação.

3.2.2.1 Produtividade de autores: a Lei de Lotka

A produtividade de autores pode ser analisada por meio da Lei de Lotka. Esta lei postula que o número de autores que fazem x contribuições em um determinado campo científico é de aproximadamente $\frac{1}{x^2}$ daqueles que fazem apenas uma contribuição (LOTKA, 1926). A lei de Lotka, assim como a Lei de Coulomb e a Lei da Gravitação de Newton, são exemplos de aplicação da lei do inverso do quadrado (ALVARADO, 2002; ANDRÉS, 2009; LOTKA, 1926). Dito de outro modo: o número de autores que produzem x estudos é proporcional a $1/x^2$. Supondo-se que, em um determinado campo, 100 autores publicaram um único estudo, pode-se prever quantos autores terão publicado dois estudos, de acordo com a seguinte fórmula: $y_x = c \cdot 1/x^2$ ou $y_x = c \cdot x^{-2}$.

Sendo,

y_x = número de autores com n publicações

c = número de autores com uma única publicação

x = número de publicações

Portanto, para identificar quantos autores contribuíram com 2 artigos, basta calcular $y_x = 100/2^2 = 25$. Porém, Lotka afirma que o número de autores prolíficos diminui de forma progressiva, de modo que apenas 11 autores teriam publicado três artigos, de acordo com a fórmula $y_3 = 100/3^2 = 11$. À medida que o número de publicações aumenta, o número de autores diminui. Segundo esses cálculos, 25%

das contribuições corresponderão a 75% dos autores menos produtivos, enquanto os mais produtivos, em torno de dez autores, responderão por 50% das contribuições.

Andrés (2009) afirma que esse padrão de produtividade não depende do campo em que a lei de Lotka é aplicada; depende, apenas, do período de tempo considerado. Desse modo, se a intenção é encontrar um grupo de autores prolíficos, é necessário considerar um período de tempo suficientemente longo para se obter resultados consistentes. Este período é fixado em intervalos de dez anos, aproximadamente.

No entanto, há controvérsias. Oppenheimer (1986), por exemplo, afirma que a Lei de Lotka foi testada em vários conjuntos de amostras e os resultados nem sempre seguiram o padrão postulado. De acordo com a distribuição descrita acima, o perfil de produção de um autor não é distribuído aleatoriamente. Isso significa que, até certo ponto, quanto mais artigos um autor produz, maior será a probabilidade de produzir outros. Portanto, a produtividade não está relacionada ao número de artigos publicados por um autor, mas ao seu logaritmo. Assim, parece tornar-se progressivamente mais fácil publicar uma vez que um autor tenha publicado um grande número de artigos. Esse fato pode ser descrito como uma vantagem perversa, denominado por Merton (1968) de efeito Mateus, em que os ricos ficam cada vez mais ricos enquanto os pobres ficam cada vez mais pobres. Conseqüentemente, a produtividade de um autor prolífico tende a crescer de modo acelerado.

Andrés (2009) explica como se deve aplicar a Lei de Lotka na avaliação da produtividade de um autor sem relação ao padrão postulado. O primeiro passo é decidir quais dados serão considerados para análise. E essa é uma questão controvertida. Wolfram (2003) chama a atenção para os impasses encontrados na análise de publicações com múltiplos autores, que pode ser feita de várias maneiras:

- a) Contagem completa: todos os autores (principal e secundário(s)) recebem crédito igual, independentemente do número de autores associados ao artigo.
- b) Contagem direta: somente o primeiro autor recebe crédito, com base no pressuposto de que o primeiro autor é o principal contribuinte da publicação.
- c) Contagem ajustada: os autores recebem crédito fracionário, em que cada autor de uma publicação com dois autores, por exemplo, receberá 0,5.

Embora a formulação geral da lei de Lotka atribua o valor 2, *a priori*, ao expoente n , ($1/n^2 = 1/x^n$), é possível calcular o valor de n de acordo com o critério adotado para calcular o crédito recebido pelo autor. Essa possibilidade resulta da variação do valor de n , que pode ser maior ou menor do que 2. De fato, quando a distribuição inclui autores altamente prolíficos, a diferença entre o número de autores de alta e baixa produtividade será maior. Neste caso, o número de autores altamente prolíficos diminuirá mais rapidamente do que o quadrado inverso, e o expoente n estará mais próximo de 3 (ANDRÉS, 2009).

Vlachy (1974), em um conjunto de estudos analisados, aponta uma série de desacordos entre os dados empíricos e o ajuste de n . Pao (1985) alerta para o fato de que atribuir 2 a n sem estimar o valor de n , ou seja, caso não seja realizado o teste apropriado do grau de ajustamento dos dados observados, os resultados poderão ser discrepantes. Nicholls (1989) definiu os métodos de cálculo dos coeficientes específicos na aplicação da Lei da Lotka a uma dada distribuição, porém, considera-se aconselhável utilizar a contagem completa. Um exemplo da aplicação da lei de Lotka é apresentado por Andrés (2009), em uma amostra de 1205 autores com trabalhos publicados em um período de 10 anos. No exemplo, são considerados todos os autores que contribuíram para a elaboração dos trabalhos (contagem completa), mostrado abaixo.

Tabela 2 - Número de publicações por número de autores

Número de publicações	Número de autores
1	1005
2	130
3	32
4	15
5	7
6	6
7	4
8	3
9	1
10	2

Fonte: Andrés (2009, p.26).

É importante observar, inicialmente, que o exemplo acima foi moldado pela autora em conformidade com os cânones tradicionais: a maioria dos autores contribuiu com um pequeno número de publicações, enquanto um pequeno grupo de autores foi bastante produtivo. Para testar a Lei de Lotka, os dados devem ser reunidos em uma nova tabela que inclui as informações adicionais necessárias para o cálculo.

Tabela 3 – Exemplo de aplicação da Lei de Lotka

X	y_x	$X = \lg x$	$Y = \lg y$	X^2	XY	$\frac{y_x}{\sum y_x}$	$\sum(\frac{y_x}{\sum y_x})$	f_e	$\sum f_e$	D
1	1.005	0.000	3.002	0.000	0.000	0.834	0.834	0.812	0.812	0.022
2	130	0.301	2.114	0.091	0.636	0.108	0.942	0.114	0.926	0.016
3	32	0.477	1.505	0.228	0.718	0.027	0.968	0.036	0.962	0.016
4	15	0.602	1.176	0.362	0.708	0.012	0.981	0.016	0.978	0.003
5	7	0.699	0.845	0.489	0.591	0.006	0.987	0.009	0.987	0.000
6	6	0.778	0.778	0.606	0.606	0.005	0.992	0.005	0.992	0.000
7	4	0.845	0.602	0.714	0.509	0.003	0.995	0.003	0.995	0.000
8	3	0.903	0.477	0.816	0.431	0.002	0.998	0.002	0.997	0,001
9	1	0.954	0.000	0.911	0.000	0.001	0.998	0.002	0.999	0.001
10	2	1.000	0.301	1.000	0.301	0.002	1.000	0.001	1.000	0.000
	1205	6560	10801	5215	4500					

Fonte: Andrés (2009, p.27).

As duas primeiras colunas correspondem às colunas da tabela 2, que relacionam número de autores e quantidade de publicações. A coluna x corresponde ao número de trabalhos publicados. A coluna seguinte, y_x , relaciona quantidade de autores e quantidade de artigos publicados. De acordo com a tabela, 1005 pesquisadores publicaram um único trabalho, enquanto dois autores publicaram 10 trabalhos cada um.

A amostra total corresponde a 1205 autores ($\sum y_x$). As duas colunas seguintes (X e Y) correspondem ao logaritmo da frequência dos artigos (x) e autores (y_x). O produto de X e Y é mostrado nas duas próximas colunas. Na coluna (y_x/\sum), é apresentada a frequência de autores com um trabalho, e aqueles com dois, três, e assim sucessivamente. O total desses dados está consolidado na coluna ao final da tabela.

Uma vez obtida esta frequência (a frequência observada), a Lei de Lotka pode ser aplicada para obter a frequência esperada de autores que publicaram x artigos. Assim, é necessário calcular o expoente n desse exemplo em particular.

A Lei de Lotka propõe o cálculo do crescimento da produção de acordo com a fórmula $y_x = c \cdot x^{-n}$, onde n é igual a 2. O cálculo resultará no valor do expoente n , que se refere à distribuição da produtividade atual do autor, além de informar sobre a correspondência desses dados com a lei. O primeiro passo do teste é obter o valor do expoente n de acordo com a seguinte fórmula:

$$\frac{N\sum XY - \sum X\sum Y}{N\sum X^2 - (\sum X)^2} = n$$

Todos os dados necessários para realizar o cálculo estão na Tabela 3. O índice que requer mais trabalho para ser obtido é N , que representa o número de pares considerados. Neste exemplo, são considerados os autores que publicaram entre 1 e 10 artigos, o que representará 10 pares de dados ($N = 10$).

Há um caso específico em que nem todos os pares de dados estão incluídos na análise da Lei de Lotka. Isso é, quando $y_x = 1$ é encontrado o final da distribuição, que corresponde ao maior valor de x (número de publicações). Nesse caso, esse pequeno grupo de autores mais prolíficos é excluído da análise para não superestimar os resultados.

No entanto, esses pares de dados devem ser excluídos somente após análise cuidadosa. Portanto, se $y_x = 1$ está localizado em outros pares de dados, mas não no final da distribuição, ele deve ser incluído no cálculo da Lei de Lotka. No exemplo de Andrés (2009), os valores de n foram obtidos a partir da tabela e assumiu-se que $N = 10$. Inserindo os valores correspondentes na fórmula de n , temos:

$$\frac{10 \cdot 4,5 - (6,56 \cdot 10,801)}{10 \cdot 5,215 - (6,56)^2} = -2,83$$

Neste caso particular, o valor de n (valor absoluto) é 2,83, o valor específico do coeficiente na fórmula de Lotka para explicar a produtividade do autor. Como a fórmula da lei de Lotka é $y_x = c \cdot x^{-n}$, e conhecido o valor n , o único índice que falta calcular é c . Esse valor é obtido da seguinte forma:

$$c = \frac{1}{\sum \frac{1}{x^n}}$$

Para os dados do exemplo, c terá o valor 0,81, e assim, a fórmula de Lotka será $y_x = 0.81 \cdot x^{2,83}$. Aqui, y_x , representa a frequência esperada dos autores que publicaram x documentos. A fim de evitar confusões na nomenclatura, Andrés (2009) define a frequência esperada (y_x). Ao introduzir os valores tomados pela variável “número de artigos” (x), será possível obter as frequências esperadas correspondentes. Os valores para f_e são demonstrados na tabela assim como a respectiva frequência acumulada ($\sum f_e$).

Como o objetivo deste exemplo é verificar se os dados se enquadram na Lei de Lotka, é necessário conhecer a magnitude das diferenças entre frequências observadas e esperadas, diferença obtida subtraindo-se o acumulado da frequência esperada com o acumulado da frequência observada: $\sum(y_x/\sum y_x) - \sum f_e$.

A última etapa do processo refere-se à aplicação do teste de Kolmogorov-Smirnov para verificar se os dados observados correspondem à distribuição teórica, de acordo com a lei de Lotka. O valor mais alto na coluna (D_{MAX}) é tomado como referência para a comparação com o valor crítico (v.c.) cuja formulação geral é:

$$c. v. = \frac{1,63}{(\sum y + \left(\frac{\sum y}{10}\right)^2)^{1/2}}$$

Nesse exemplo, o valor crítico é 0,047, obtido pela fórmula:

$$c. v. = \frac{1,63}{(1,205 + \left(\frac{1,205}{10}\right)^2)^{1/2}} = 0,047$$

Uma vez que a diferença máxima obtida da tabela é 0,022, que é menor do que o valor crítico (0,047), a hipótese nula deve de ser aceita. Pode-se concluir,

portanto, que a produtividade do autor nesta hipotética área de pesquisa enquadra-se na Lei de Lotka.

A validade da Lei de Lotka foi testada por inúmeros pesquisadores. Os resultados nem sempre confirmaram a lei. Alvarado (2002) identificou críticas e propostas de reformulação da lei em mais de 200 publicações. Nicholls (1989) atribuiu os problemas de ajuste a falhas na medição, aos parâmetros de estimação, às formas de testagem ou interpretação do modelo. Alguns autores criticaram a forma de contagem de autoria. Nath e Jackson (1991) afirmaram que a contagem direta e a contagem ajustada produzem os mesmos resultados, portanto, “não seria necessário considerar a contagem ajustada, sendo necessário prestar maior atenção à contagem direta” (1991, p.207). Outros autores argumentam que se o período de cobertura respeitar o parâmetro de no mínimo 10 anos e a amostra for ampla, a produtividade se enquadra no postulado de Lotka (Potter, 1981). Pao (1985), defensora da lei, afirma que os trabalhos que não se enquadram na lei apresentam diferentes tipos de erros: inexistência de um método uniforme de organização dos dados, atribuição errônea do valor 2 a n , sem a devida estimativa do valor dos dados observados.

Glänzel (2003) acredita que Lotka faz parte de um grupo de jovens “*bibliometricians*” que adotavam modelos de outros campos para descrever regularidades, porém, “[...] na maioria dos casos, esses modelos provaram ter validade limitada” (2003, p.17). Segundo ainda este autor, a Lei de Lotka é um exemplo de lei formulada como lei natural, universal, sem levar em conta outros fatores que influenciam a produtividade.

A Lei de Lotka pode ser analisada com base em leis estatísticas, apresentadas no subcapítulo 2.3, acima. O ponto importante a ser observado refere-se a duas concepções formuladas no século XIX: que a lei estatística deriva da ideia de que a lei social é um fenômeno fixo, produto da vontade dos indivíduos. Para a segunda corrente, as leis sociais não são fixas, estão em constante evolução e transformação, assim como a sociedade, de modo que as leis não poderiam ser lidas nas tabelas de dados numéricos ou obtidas a partir de resumos de fatos sobre indivíduos. Assim, as leis estatísticas são apenas regularidades numéricas e não leis fundamentais.

Nessa ótica, com base nas considerações de Hacking (2005), é possível desenvolver alguns raciocínios sobre a Lei de Lotka, que foi formulada com base em certas regularidades. Lotka afirma que no padrão de publicação da ciência mundial, a

maior parte da produção é de autoria de uma pequena parte dos pesquisadores. A partir da constatação dessa regularidade, a Lei de Lotka é considerada adequada para quantificar o comportamento científico em escala mundial.

Pode-se considerar que a lei de Lotka faz parte do movimento estatístico que procura identificar estabilidades nos fenômenos de massa. Ao generalizar o comportamento de pesquisadores, Lotka agiu como Quetelet (subcapítulo 2.3), que aplicou a curva de distribuição normal à antropometria de soldados. Há inclusive semelhanças entre as críticas à Lei de Lotka e as dirigidas a Quetelet. Em ambos os casos, quando os resultados não se encaixavam na lei ou na distribuição esperada, os problemas são atribuídos a falhas na medição, na amostra, ou na modelagem dos dados.

Porém, o fato questionável da Lei de Lotka é o enquadramento do comportamento científico mundial de publicação dentro da lei do quadrado inverso. Isso implica ignorar as particularidades culturais, sociais e econômicas dos pesquisadores, variáveis que afetam as formas de publicação.

O percurso de uniformização dos fenômenos sociais, para contá-los e generalizá-los, como ocorreu na França e na Inglaterra, no século XIX, é elucidado por Hacking (2005). Interessante notar também que contemporaneamente, Potter (1981) propôs, justamente, a uniformização dos dados para contornar as inconsistências observadas na aplicação da Lei de Lotka. Em 1835, Poisson afirmou que “nos assuntos estatísticos ... o primeiro cuidado antes de tudo é perder de vista o homem tomado isoladamente para considerá-lo apenas como uma fração da espécie”²² (POISSON apud HACKING, 2005, p.81). Nath e Jackson (1991), por sua vez, afirmaram que a contagem direta – quando somente o primeiro autor recebe crédito pela publicação - é a que cria a melhor modelagem para obter resultados compatíveis com a Lei de Lotka. Trata-se de um ajuste perverso para confirmar uma lei controvertida. De fato, creditar arbitrariamente a produção a um único autor diminui o número de autores que publicaram e, por outro lado, aumenta o número de autores que não publicaram. Este recurso, claramente questionável, torna mais fácil “modelar” um evento para justificar a validade uma lei.

²² “In statistical affairs ... the first care before all else is to lose sight of the man taken in order to consider him only a fraction of the species.”

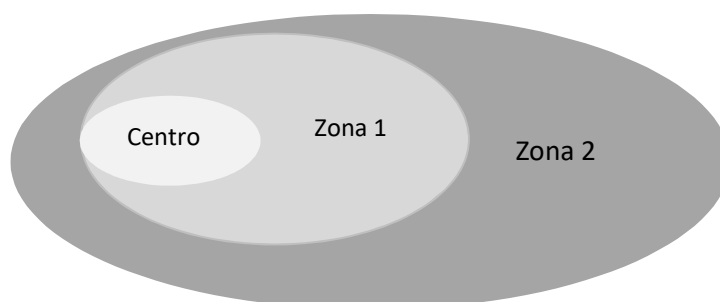
Outra discussão deve ser feita a respeito da Lei de Lotka. Ao se postular que a desigualdade no campo da pesquisa científica é uma lei, acaba-se fomentando a normalização da desigualdade. E, como foi discutido anteriormente, a ideia de normal é, de acordo com Hacking (2005), uma poderosa ferramenta ideológica. A ideia de que uma lei determina o comportamento de cientistas e pesquisadores de forma que alguns poucos são responsáveis pela maioria das publicações e outros muitos são responsáveis por uma menor quantidade, reforça duas concepções equivocadas: a) um pequeno grupo de seres humanos é abençoado com a dádiva da inteligência e/ou b) um grande grupo de seres humanos não está se esforçando o suficiente. E, sendo uma lei natural, não há nada que possa ser feito para mudá-la.

3.2.2.2 Produtividade de periódicos: a Lei de Bradford

Samuel C. Bradford (1934) investigou, na década de 30, o padrão de publicação de artigos científicos no campo da Geologia. Em uma amostra de 326 periódicos, notou que 9 deles continham 429 artigos sobre o campo, 59 continham 499 e 258 continham 404 artigos. Esses resultados mostravam a existência de um núcleo de 9 periódicos que abordava os temas da área de maneira mais extensiva, e uma região periférica, dividida em 2 zonas, composta por periódicos que publicavam quantidades menores de artigos. Identificou, portanto, uma relação inversa entre o número de artigos publicados em uma área de assunto e o número de periódicos em que os artigos aparecem.

As investigações de Bradford levaram-no à conclusão de que seria possível medir a produtividade dos periódicos e estabelecer o núcleo e as zonas de dispersão de um campo científico. Essa constatação permitiria estimar o grau de relevância de cada periódico para uma determinada área do conhecimento. Portanto, segundo a Lei de Bradford, se os periódicos forem classificados em três grupos segundo o número de artigos sobre um determinado campo, então o número de periódicos em cada grupo será proporcional a $1: n: n^2$.

Ilustração 4 – Diagrama de Bradford



Fonte: Elaboração do autor.

A lei de Bradford pode ser usada como uma ferramenta para o gerenciamento de coleções de bibliotecas. Ao identificar as principais revistas, por áreas temáticas são obtidas evidências para a tomada de decisões sobre a formação de coleções de periódicos.

Um exemplo da aplicação da Lei de Bradford é apresentado por Andrés (2009). Com dados de Pulgarín e Gil-Leiva (2004), que coletaram 527 artigos sobre o tema indexação automática, nos anos de 1956 a 2000, Andrés criou a seguinte tabela de frequências.

Tabela 4 – Dispersão na literatura científica que trata de indexação automática

Número de periódicos	Número de artigos	Periódicos - Acumulativo	Artigos - Acumulativo	Ln (Periódicos - Acumulativo)
1	66	1	66	0
1	40	2	106	0.6931
1	36	3	142	1.0986
1	28	4	170	1.3862
1	21	5	191	1.6094
1	18	6	209	1.7917
1	10	7	219	1.9459
1	8	8	227	2.0794
1	7	10	241	2.3025
1	6	14	265	2.6390
5	5	19	290	2.9444
6	4	25	314	3.2188
14	3	39	356	3.6635
27	2	66	410	4.1896
117	1	183	527	5.2094

Fonte: Pulgarín e Gil-Leiva (2004) apud Andrés (2009, p. 35).

A partir destes dados é necessário calcular o valor da constante k de Bradford. A constante k é o multiplicador que explica como o número de periódicos cresce de uma zona para a próxima. Egghe (1986, 1990) desenvolveu a seguinte fórmula para o cálculo de k :

$$k = (e^\gamma \cdot Y_m)^{1/P}$$

Onde,

γ = número de Euler ($\gamma = 0,5772$)

Y_m = produtividade máxima da revista de primeiro grau

P = número de zonas

No exemplo de Andrés (2009), o periódico mais produtivo apresentou 66 artigos: então $Y_m = 66$. P é definido como 4, um dado que merece discussão. De acordo com Andrés, foi atribuído o valor 4 a P porque os autores, Pulgarín e Gil-Leiva (2004), afirmam que seus dados se encaixam numa distribuição em quatro zonas. A análise do trabalho desses autores mostra que, de fato, seria possível distribuir com certa regularidade os artigos em 4 zonas. Porém, também seria factível a divisão em um número maior de zonas, preservando-se a regularidade. Infere-se, portanto, que a escolha de 4 zonas diz respeito à necessidade de adequação à Lei de Bradford. Uma observação deve ser feita sobre a amostra utilizada neste exemplo: na coleta foi utilizado um indexador automático. Os indexadores automáticos são ferramentas criticadas por não serem capazes de identificar significados. Portanto, é uma amostra questionável que pode gerar resultados igualmente questionáveis.

Andrés (2009) adverte que em cada caso individual é necessário testar o número de zonas que se enquadram nos dados, de acordo com a formulação da Lei de Bradford. Assim, k será obtido da seguinte forma:

$$k = (1.781 \times 66)^{1/2} = 3.29$$

A autora também afirma ser possível calcular o número de periódicos que pertencerão ao grupo central de Bradford, representado por r_0 , calculado com a seguinte fórmula:

$$r_0 = \frac{T(k-1)}{(k^P - 1)}$$

Onde,

T = número total de periódicos

k = constante de Bradford

P = número de grupos de Bradford

De acordo com o exemplo:

$$r_0 = \frac{183(3.29-1)}{(3.29^4-1)} = 3.60$$

O resultado indica que 4 revistas são as mais produtivas (zona central). Conhecendo os valores de k e r_0 , é possível, teoricamente, obter a distribuição de revistas nas zonas propostas por Bradford (correspondentes a r_0, r_1, r_2 e r_3). O número esperado de periódicos, por zonas, neste exemplo, será: $r_0 = r_0 \cdot 1 = 3.6$, $r_1 = r_0 \cdot k = 11.8$, $r_2 = r_0 \cdot k^2 = 38.9$ e $r_3 = r_0 \cdot k^3 = 128.2$.

A partir dessa distribuição teórica, é possível testar o ajuste exato da lei de Bradford aos dados empíricos. O número de artigos em cada zona pode ser contado, conforme demonstrado na Tabela 5.

Tabela 5 – Zonas de Bradford na área de indexação automática

Zonas	Nº de periódicos	Nº de artigos	K
Zona central	4	170	-
Zona 1	12	105	3.00
Zona 2	39	113	3.25
Zona 3	128	139	3.28

Fonte: Pulgarín e Gil-Leiva (2004) apud Andrés (2009, p. 36).

Nesta tabela, é apresentado o número exato de zonas de k , levando em consideração o número real de periódicos incluídos em cada zona. Como os valores de k são muito semelhantes, e também semelhantes ao calculado pela fórmula ($k = 3,29$), pode-se concluir que os dados se encaixam numa distribuição de 4 zonas.

Tendo confirmado que os dados se enquadram na lei de Bradford, calcula-se a equação para a curva de Bradford, com base na derivação de Leimkuhler (1967), cuja fórmula é:

$$R(r) = a \log_e(1 + br)$$

Sendo que,

$R(r)$ = número acumulado de artigos publicados pela revista de classificação 1, 2, 3... r

a = uma constante calculada por: $a = y_0 / \log_e k$

y_0 = número de artigos encontrados em cada grupo (considerando que cada zona incluirá o mesmo número de artigos), calculado por meio da fórmula: $y_0 = A/P$, onde A é o número total de artigos encontrados na literatura e P é o número de zonas de Bradford. No exemplo acima, isto é obtido como segue: $y_0 = \frac{527}{4} = 131,75$ ou 132. Conhecendo-se este valor, é possível obter o valor de a . Nesse caso, $a = 132 / \log_e 3.29 = 110.85$.

b = que é obtido por $b = (k - 1) / r_0$. Como os valores k e r_0 são conhecidos, $b = (3.29 - 1) / 3.6 = 0.636$.

Finalmente, com base na formulação de Leimkuhler (1967),

$$R(r) = 110.8 \log_e(1 + 0,64r)$$

Esta fórmula mostra o número cumulativo de artigos $R(r)$ através de diferentes classificações (do 1 ao r). Substituindo os diferentes valores do número cumulativo de revistas r , é possível obter o valor do número esperado de artigos $R(r)$. Por exemplo, para $r = 25$ (quando o número cumulativo de revistas toma o valor 25), pode-se calcular o número acumulado esperado de artigos $R(r)$. Então, $R(r) = 110.8 \log_e(1 + 0,64 \times 25) = 313.9$, bastante semelhante ao resultado encontrado na Tabela 4.

A Lei de Bradford recebeu variadas críticas. Gilchrist (1966) e Saracevic (1971) aplicaram a lei e apontaram ambiguidades, disparidades e incoerências entre a formulação e sua aplicação prática. Drott (1981) afirma que nenhuma das variáveis próprias da situação empírica foi relacionada ao modelo teórico, demonstrando,

assim, a necessidade de maior quantidade de experimentos para validá-la. Em relação às variáveis, Braga (1978) alerta que os periódicos prolíficos em determinada área não necessariamente publicam trabalhos inovadores.

Pinheiro (1983), por sua vez, aponta as seguintes variáveis que escapam à lei: 1) as variáveis referentes às práticas de indexação e resumos - termos novos, barreiras linguísticas, entre outros – podem influenciar a coleta de artigos; 2) a subjetividade da noção de relevância – com efeito, definir o que é importante para uma área do conhecimento não é uma tarefa trivial; 3) a política editorial e o sistema de avaliação de periódicos são fatores determinantes na seleção dos artigos para publicação; 4) a periodicidade, o número de páginas, o formato, a extensão dos artigos, etc., são variáveis que interferem no número de artigos. Por exemplo, uma revista mensal estaria muito mais propensa a se localizar no centro do diagrama de Bradford do que uma revista anual (PINHEIRO,1983)

A exclusão de variáveis próprias da composição de uma área do conhecimento é a principal crítica à Lei de Bradford. De fato, se pensarmos em áreas interdisciplinares como, por exemplo, a Ciência da Informação, definir quais são as publicações de interesse é uma atividade complexa. Para elaborar a presente revisão bibliográfica, por exemplo, foi necessário considerar uma série de variáveis e acasos que nos levaram a publicações importantes das mais diversas áreas do conhecimento, para construir o referencial teórico desta pesquisa.

Sendo assim, questiona-se: como coletar uma amostra que seja interessante ou importante para conhecer uma área do conhecimento? Interessante e importante são conceitos subjetivos, de modo que se pode perguntar que tipos de especialistas estariam capacitados a realizar esses tipos de estudos.

É importante lembrar as críticas de Peirce às tentativas de moldar a 'realidade' com base em leis invariáveis. Segundo Peirce (1967), seria absolutamente necessário deixar de modelar o mundo por meio de "constantes", que são apenas variáveis aleatórias que se estabeleceram no curso da evolução das leis. Certos autores tentaram reformular a Lei de Bradford, como Pinheiro (1983) e Braga (1978). No entanto, esses estudos ficam comprometidos porque foram incorporadas miríades de acasos a variáveis cuja validade se esgota no tempo e no espaço.

3.2.2.3 Contagem de palavras: as leis de Zipf

As Leis de Zipf possibilitam estimar as frequências de ocorrência de palavras em um determinado texto, bem como a região de concentração de palavras-chave. A primeira Lei de Zipf foi formulada por George Kingsley Zipf na década de 1940, com base na proposta de indexação automática de Luhn (1957). Guedes e Borschiver (2005) afirmam que as leis de Zipf, juntamente com o Ponto de Transição (T) de Goffman, têm relação direta com a indexação automática. Na obra *Human Behaviour and the Principle of Least-Effort*, Zipf (1949) explica que em textos ou conjuntos de textos suficientemente longos um pequeno grupo de palavras ocorre muitas vezes e um grande número de palavras tem baixa frequência de ocorrência:

o número de vezes que cada palavra aparece em um texto é contado. As palavras são, então, ordenadas em uma tabela, na qual a primeira palavra é a mais frequente, a segunda palavra mais frequente e assim sucessivamente. A ordem de qualquer palavra na tabela é chamada ordem de série (r), e o número de vezes que ela aparece, frequência (f), de tal forma que $r \cdot f = C$, onde C é uma constante para qualquer texto (BOOTH, 1967).

Porém, essa equação aplica-se apenas a palavras de alta frequência, de modo que Zipf formulou uma segunda lei para as palavras de baixa frequência. A Segunda Lei de Zipf enuncia que, em um determinado texto, várias palavras de baixa frequência de ocorrência têm a mesma frequência. Em 1967, Booth reformulou a Segunda Lei de Zipf cuja representação matemática é expressa em,

$$\frac{I_1}{I_n} = \frac{n(n+1)}{2}$$

Sendo,

I_1 = número de palavras que têm frequência 1

I_n = o número de palavras que têm frequência n

2 = é a constante válida para a língua inglesa.

Entretanto esses dois comportamentos precisam ser analisados. Na maioria das vezes, palavras de alta frequência têm a propriedade de ocupar classificação única em uma lista de distribuição de palavras, ou seja, em listas de palavras de alta

frequência dificilmente serão encontradas duas palavras que ocupem a mesma posição na classificação. Supondo que em um determinado texto existam duas palavras de alta frequência, dificilmente elas ocupam o primeiro lugar na classificação. Porém, em listas de palavras de baixa frequência é comum observar muitas palavras com a mesma frequência, isto é, que ocupam a mesma posição na classificação, pois é comum um grande número de palavras diferentes serem citadas apenas uma única vez no texto. Esses dois comportamentos definem as duas extremidades de uma lista de distribuição de palavras em um texto. Goffman (1966) inferiu que entre essas duas extremidades, existiria uma região de transição do comportamento dessas palavras, região na qual estariam as palavras de maior conteúdo semântico (PAO,1978).

Estudando as duas leis de Zipf, Goffman (1966) derivou uma fórmula de transição, cujo resultado final é uma Ordem de série T, que indica, aproximadamente, as ordens de séries nas quais se encontram as palavras significativas de um texto em língua inglesa:

$$T = \frac{-1 + \sqrt{1 + 8 I_1}}{2}$$

Onde:

T = ordem de série qualquer

8 = constante fixa para qualquer idioma

I = frequência 1, correspondente à ordem de I série mais alta

2 = constante válida para a língua inglesa

Guedes e Borschiver (2005, p.8) afirmam que:

O Ponto T de Goffman determina graficamente a localização onde ocorre a transição das palavras de baixa frequência para as de alta frequência. Existe uma determinada região, ao redor deste ponto, com probabilidade de concentrar as palavras de alto conteúdo semântico e, portanto, aquelas que seriam usadas para indexação do texto em questão. Goffman apresenta, com o Ponto de Transição, a primeira proposta de decomposição sintática de um texto, visando à sua indexação.

Maia (1973) fez uso das Leis de Zipf e do Ponto de Transição de Goffman em um estudo sobre o comportamento da Língua portuguesa em uma amostra de dois textos:

Texto A: Edson Nery da Fonseca, "Desenvolvimento da Biblioteconomia e da Bibliografia no Brasil", publicado na Revista do Livro, em 1957, p. 95-124. Contém 6395 palavras, 1971 palavras diferentes.

Texto B: Hagar Espanha Gomes, "Ensino de Técnicas Bibliográficas nos Cursos de Ciências Agrícolas", apresentado no Seminário sobre Bibliotecas Agrícolas, organizado pelo IBBD, no Rio de Janeiro, 1963, 10 f. Contém 2210 palavras, sendo 839 palavras diferentes.

Primeiramente, foi estabelecido um critério para realizar a contagem de palavras: a) palavras-compostas, nomes de pessoas e instituições e títulos de publicações foram consideradas como um todo; b) não foram incluídas na contagem datas e inscrições numéricas; c) o processo de contagem foi idêntico para os dois textos; d) todas as palavras de cada texto foram transcritas em fichas; e) os termos foram agrupados em conjuntos segundo o número de ocorrências; f) as palavras foram ordenadas em séries, em uma tabela, em ordem decrescente; g) foi obtida a ordem de duas ou mais palavras diferentes com mesma frequência, dividindo-se a soma de suas ordens de série, pelo número de palavras a elas correspondentes. Ex.: A ordem de série 10,5 foi obtida somando-se os números 10 e 11, e dividindo-se o resultado por 2 (número de palavras). O número 12, da ordem de série seguinte, foi aquele não envolvido na soma; h) foi calculado o produto da ordem de série versus frequência de palavras, a fim de se obter a constante c e verificar se os resultados estavam em conformidade com a primeira lei de Zipf. O resultado deste processo é apresentado na tabela abaixo:

Tabela 6 – Ordem de série versus frequência das palavras nos textos A e B.

Texto A					Texto B				
Ordem de série	Nº de palavras	Freq.	Produto Total	$r . f = c$	Ordem de série	Nº de palavras	Freq.	Produto Total	$r . f = c$
1	1	320	320	320	1	1	128	128	128
2	1	303	303	606	2	1	72	72	144
3	1	246	246	738	3	1	68	68	204
4	1	177	177	708	4	1	53	53	212

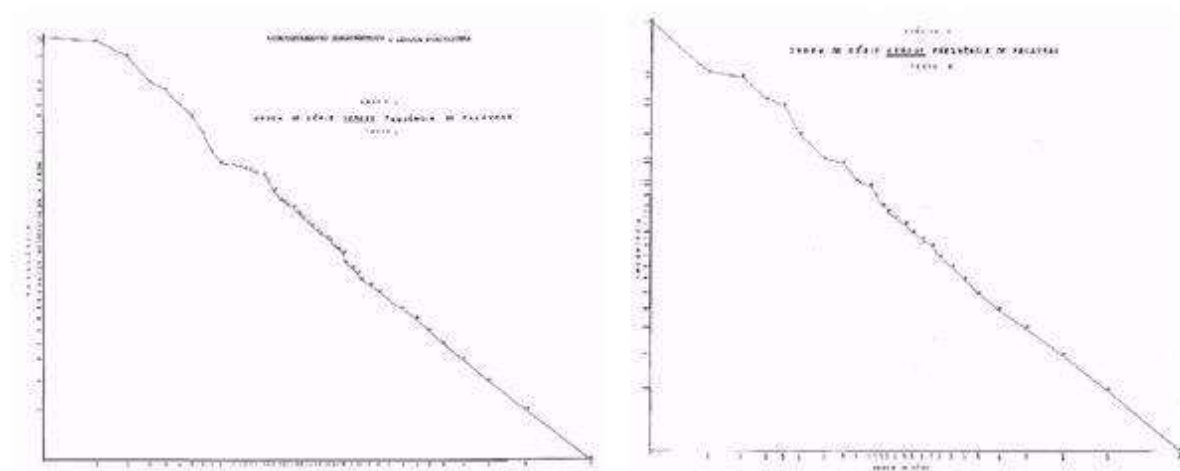
5	1	156	156	780	5	1	50	50	250
6	1	126	126	756	6,5	2	36	72	234
7	1	112	112	784	8,5	2	27	54	229,5
8	1	88	88	704	10,5	2	26	52	273
9	1	69	69	621	12,5	2	21	42	262,5
10,5	2	58	116	609	14	1	20	20	280
12	1	56	56	672	15	1	18	18	270
13	1	55	55	715	16	1	16	16	256
14	1	54	54	756	17	1	15	15	255
15	1	53	53	795	19	3	14	42	266
16,5	2	50	100	825	21,5	2	13	26	279,5
18	1	49	49	882	23	1	12	12	276
19	1	46	46	874	26	5	11	55	286
20,5	2	40	80	820	29	1	10	10	290
22	1	35	35	770	32	5	9	45	288
23	1	34	34	782	37,5	6	8	48	300
24,5	1	33	66	808,5	43,5	6	7	42	304,5
26	1	32	32	832	51	9	6	54	306
27	1	30	30	810	65	19	5	95	325
28	1	29	29	812	90,5	32	4	128	362
29	1	28	28	812	141	69	3	207	423
31	3	26	78	806	236,5	122	2	244	473
33,5	2	25	50	837,5	568,5	542	1	542	568,5
33,5	2	23	46	816,5		839		2210	
37,5	2	22	44	825					
40	3	21	63	840					
42,5	2	20	40	850					
44	1	19	19	836					
46	3	18	54	828					
49	3	17	51	833					
51,5	2	15	30	772,5					
55	5	14	70	770					
59	3	13	39	767					
63,5	6	12	72	762					
70,5	8	11	88	775,5					
79,5	10	10	100	795					
91	13	9	117	819					
109	23	8	184	872					
130	19	7	133	910					
152	25	6	150	912					
185,5	42	5	210	927,15					
236	59	4	236	944					
329	127	3	381	987					
533	281	2	562	1066					

1232,5	1118	1	1118	1232,5					
	1791		6395						

Fonte: Maia (1973, p.105-108).

Maia (1973) listou, em seguida, as palavras dos textos com suas respectivas ordens de série e calculou os pontos de encontro dos logaritmos das frequências com as Ordens de série das palavras, gerando os gráficos a seguir:

Ilustração 5 - Ordem de série versus frequência de palavras nos textos A e B



Fonte: Maia (1973, p.129-131).

A segunda Lei de Zipf, foi aplicada na última etapa, conforme reformulação proposta por Booth (1967) para identificar as variações da constante 2 e garantir resultados mais precisos. Além disso, Maia (1973) calculou a transição de Goffman, para obter variações da constante 2.

Para Goffman (1966), o ponto de transição é a região que concentra as palavras de maior conteúdo semântico e deverão ser utilizadas para indexar um texto ou conjunto de textos para representar seu conteúdo temático. Os termos selecionados tornar-se-ão pontos de acesso na recuperação da informação.

No experimento, as palavras significativas encontradas para o Texto A (Desenvolvimento da Biblioteconomia e da Bibliografia no Brasil), foram: "brasileiras", "bibliografias", "corrente", "Cursos de Biblioteconomia da Biblioteca Nacional",

No caso do Texto B (Ensino de Técnicas Bibliográficas nos Cursos de Ciências Agrícolas) foram identificadas como significativas as palavras "informação" e "especialista".

Os resultados obtidos nesse experimento podem ser questionados. Os termos "corrente" e "Cursos de Biblioteconomia da Biblioteca Nacional" são de fato os de maior peso semântico para representar o Texto A (Desenvolvimento da Biblioteconomia e da Bibliografia no Brasil)? Deve-se lembrar que o termo "corrente", isoladamente, pode ter inúmeros significados.

No caso do Texto B (Ensino de Técnicas Bibliográficas nos Cursos de Ciências Agrícolas), foram identificados os termos "informação" e "especialista". Esses dois termos são efetivamente pertinentes para representar o conteúdo do texto analisado?

O questionamento fundamenta-se na literatura do campo da Ciência da informação que mostra os limites da indexação automática baseada na palavra como unidade de observação. A validade das leis de Zipf deve, de nosso ponto de vista, ser discutida no contexto da indexação, tal como se compreende hoje a significação e a representação de conteúdos de textos. De fato, a indexação baseada na contagem de palavras foi abandonada após os diversos experimentos de indexação automática realizados nos anos 1970 (BORKO, 1968).

A conclusão da autora sobre a pertinência da aplicação da fórmula de transição de Goffman, com constante 2, à língua portuguesa é também questionável. A língua evolui segundo os seus usos sociais, portanto, para afirmar que a constante ocorre na língua portuguesa é necessário realizar testes massivos diacrônicos para verificar sua validade.

3.2.2.4 Análise de citação

A Análise de citação é um procedimento utilizado para identificar os padrões e a frequência das citações feitas e recebidas por autores, periódicos, disciplinas de pesquisa, bem como as relações entre os documentos citados (SPINAK, 1996). A Norma Brasileira 10.520:2002 define citação como a "menção de uma informação extraída de outra fonte" (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2002, p.1). No entanto, citação é um termo polissêmico:

em espanhol, a citação é a reprodução textual de uma passagem de um documento, geralmente entre aspas. A palavra correspondente em inglês seria "*quotation*". Em vez disso, a palavra referência é usada para indicar a referência de um documento para outro que indica a relação entre eles, e isso também pode ser chamado de citação²³ (SPINAK, 1996, p. 50-51).

Na norma brasileira, o termo referência diz respeito ao “conjunto padronizado de elementos descritivos, retirados de um documento, que permite sua identificação individual” (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018, p. 2). Com base nas definições propostas pelas NBRs, pode-se inferir que para estudar as menções presentes em um determinado texto é necessário considerar tanto a citação quanto a referência. Porém, para efeito de análises bibliométricas, Price distingue “referência” de “citação”:

Acho lamentável desperdiçar um bom termo técnico usando as palavras citação e referência de forma intercambiável. Eu, portanto, proponho e adoto a convenção de que se o artigo R contém uma nota bibliográfica usando e descrevendo o artigo C, então R contém uma referência a C, e C tem uma citação de R. O número de referências de um artigo é medido pelo número de itens da bibliografia apresentada, seja no final ou como notas de rodapé, etc., enquanto que o número de citações de um artigo é informado por algum tipo de índice de citação que informa quantos artigos o mencionam²⁴(PRICE, 1970 apud SPINAK, 1996, p.51).

Embora os dois termos sejam usados indistintamente, ou como no caso brasileiro, complementarmente, cada um deles representa uma visão diferente na perspectiva citante ou citado. Nas orientações de Price (1970), uma referência é realizada quando se cita um documento. Essa ação representa o reconhecimento de

²³ “En idioma español cita es la reproducción textual de un pasaje de un documento, normalmente entrecomillado. En inglés se correspondería la palabra "quotation". En cambio se usa la palabra referencia para señalar la remisión de un documento a otro que indica la relación entre éstos, y esto también puede llamarse citación.”

²⁴ “Me parece lamentable desperdiciar un buen término técnico al usar las palabras citación y referencia de manera intercambiable. Yo, por eso, propongo y adopto la convención de que si el artículo R contiene una nota bibliográfica usando y describiendo al artículo C, entonces R contiene una referencia a C, y C tiene una citación desde R. La cantidad de referencias que tiene un artículo se mide por la cantidad de ítems que tiene su bibliografía, sea al final o como notas al pie de página, etc., mientras que la cantidad de citas que tiene un artículo se halla buscándolas en algún tipo de índice de citas y viendo cuántos artículos lo mencionan.”

outro estudo. Já a citação representa a confirmação recebida pelo documento citado e, é justamente essa confirmação, o objeto de estudo da análise de citação.

Spinak afirma que as análises de citação são realizadas porque “um artigo científico não é uma entidade solitária, mas está imerso na literatura sobre o assunto. O trabalho de um autor é construído sobre as obras dos antecessores²⁵” (SPINAK, 1996, p.8). Quando um autor publica um estudo, incluirá referências de trabalhos anteriores de autores que abordaram a mesma temática. Estas citações revelam ligações entre autores, grupos de pesquisadores, tópicos de estudo ou países.

Latour (2000) discute um outro aspecto: a motivação para citar. De acordo com o sociólogo, a citação é utilizada em caráter persuasivo, retórico, e não dentro de uma dinâmica de reconhecimento. Na Bibliometria, no entanto, considera-se a citação como a expressão do relacionamento entre dois (ou mais) participantes do circuito científico, de modo que as citações revelam a dinâmica dessa cooperação e reconhecimento. Estes padrões dão origem a diferentes tipos de relações entre autores, instituições, países, assuntos, áreas, de modo a estabelecer redes de pesquisas. Portanto, considera-se que a análise de citação revela características que podem ser fundamentais para predizer os caminhos da ciência e os ajustes que serão necessários para alcançar os objetivos desejados.

A criação do *Institute for Scientific Information* (ISI), em 1958, por Garfield, é considerada um marco histórico na prática da análise de citação. O contexto de criação da ISI relaciona-se ao protagonismo assumido pela avaliação científica após a diminuição dos fundos públicos para a pesquisa, a partir dos anos de 1960-70, nos Estados Unidos (ROMANCINI, 2010). A análise de citação pareceu ser um caminho adequado para realizar avaliações científicas e definir no que e onde investir. Porém, segundo Yancey (2005), a contribuição verdadeiramente inovadora de Garfield foi reunir estudos de diferentes áreas em uma única base de dados. Cria-se, assim, um índice de citações que poderia promover a identificação de artigos em abordagem interdisciplinar. Portanto, a ISI conciliava dois interesses fundamentais: a avaliação da atividade científica e a ampliação das possibilidades de pesquisa, segundo a política científica norte-americana da guerra fria.

²⁵ “un articulo científico no es una entidad solitaria sino que esta inmerso dentro de la literatura sobre el tema. La obra de un autor se construye sobre las obras de los predecesores.”

A análise de citação é também utilizada para avaliar a obsolescência da literatura científica. Para Price (1963), a obsolescência é o declínio do uso de documentos ao longo do tempo e pode ser medida por meio da análise de citação. Como o uso de um documento é medido pelo número de citações que recebe, quando não é mais citado, é considerado obsoleto.

No entanto, há pelo menos dois entraves à adoção desse princípio: as práticas científicas. Meadows (1999) chama a atenção para as distintas práticas de publicação e citação nas diferentes áreas do conhecimento. É comum uma grande proporção de documentos receber citações nos anos imediatamente após a publicação, bem como documentos que permanecem não citados ou raramente citados nos anos seguintes à sua publicação e se tornam reconhecidos posteriormente. Isso ocorre porque, com o tempo, maior número de pesquisadores pode tomar consciência do valor de um dado documento e passa a receber número maior de citações.

A efemeridade de boa parte dos artigos científicos é também outro fator a ser considerado. Nos estudos bibliométricos, livros, monografias, relatórios, teses e artigos publicados em periódicos são unidades que podem compor um *corpus* de análise. No entanto, autores reconhecidos (PRICE, 1963; GLÄNZEL, 2003) afirmam ser fundamental que as unidades de análise sejam padronizadas. Esses padrões seriam os sistemas de revisão, critério de originalidade dos resultados da pesquisa, disponibilização da literatura e regras transparentes de publicação. Nesse sentido, considera-se que os artigos científicos publicados em revistas científicas de referência são as unidades mais adequadas para realizar estudos bibliométricos. O artigo científico pode ser mais rapidamente acessado, seus argumentos, hipóteses, metodologias e conclusões podem ser rapidamente ratificadas ou rejeitadas. Essa mesma propriedade não é observada em livros, cujo processo de produção é mais lento. Assim, a definição de obsolescência da literatura científica é subjetiva, constituída por variáveis que vão muito além das práticas de citações utilizadas em artigos.

A compreensão de que a citação é um indicador de impacto, influência ou desempenho de um pesquisador é, de acordo com Romancini (2010), um parâmetro adotado pela vertente da Sociologia da Ciência encabeçada por de Solla Price e Robert Merton, posição compartilhada por muitos pesquisadores. Embora Leydesdorff (1998) defenda a importância de definir primeiramente “qualidade”, “impacto” e

“influência”, conceitos que significam de acordo com os campos do conhecimento considerados, a análise de citação é suportada pela ideia de que a ciência funciona como uma instituição com normas e valores comuns respeitados por todos os membros que a compõem. Por conseguinte, neste modelo, a citação faz parte de um sistema de reconhecimento de pesquisadores. Está implícita, nesse modelo, a noção de que existe correlação entre quantidade e qualidade de citações. Uma citação indica um reconhecimento, duas citações, dois reconhecimentos e assim sucessivamente. Os índices de citação são, nessa medida, considerados representações objetivas da ciência, fato que, de acordo com Wouters (1999), fortaleceu a “cultura da citação”. Deve-se considerar, no entanto, com base em Hacking (2005, 2006, 2009) e Peirce (1967), que não existem modelos fechados capazes de representar a sociedade a longo prazo. Em algum momento, essa concepção se esfacela, pois, o acaso, o livre-arbítrio, contextos e variáveis diversas agem sobre os modelos.

Determinados usos da citação são considerados polêmicos. A autocitação é um deles. Essa prática não é perniciosa em si porque o autor pode recorrer às suas ideias anteriores para compor seus argumentos. Porém, há situações em que são usadas para inflar o número de citações de seus trabalhos. Deschacht (2017) estudou a autocitação e concluiu que o excesso é um fenômeno mais comum entre autores que residem em países onde o individualismo e a competitividade predominam como traços culturais, como nos Estados Unidos e Reino Unido, sendo menos frequente, segundo o estudo, em países onde predomina uma cultura coletivista, como a China e a Coreia do Sul.

Não há consenso sobre o papel da autocitação na literatura. Há os que consideram as autocitações como um aspecto natural da comunicação científica e argumentam que uma completa falta de autocitações seria tão patológica quanto a quantidade exagerada (GLÄNZEL et al, 2014). Ao comparar a evolução temporal de autocitações e citações externas (feitas por outros autores), Glänzel et al (2014) verificaram que as autocitações crescem rapidamente nos anos seguintes à publicação e diminuem rapidamente com o passar do tempo. Em contraste, as citações estrangeiras levam mais tempo para atingirem seu pico, porém, mantêm uma taxa de citação alta durante mais tempo do que as autocitações.

Uma explicação coerente sobre o papel das autocitações é atribuída a van Raan (2008), que considera que as autocitações têm papéis e funções diferentes,

variando de acordo com os grupos de pesquisa. Para entender sua explicação é necessário esclarecer primeiros dois termos por ele utilizados: o tamanho de um grupo e a vantagem cumulativa. O tamanho de um grupo é definido pelo número de publicações produzidas. O termo vantagem cumulativa está relacionado ao efeito Mateus discutido por Merton (1968). Aplicado ao presente caso, este conceito implica que grupos de pesquisa maiores não simplesmente recebem mais citações, mas angariam cada vez mais vantagens. Conseqüentemente, os grupos de pesquisa de alto desempenho não precisam de promoção interna, já que os autores de outros grupos são encorajados a tomar conhecimento de suas publicações e as citarão. Como os grupos de baixo desempenho tendem a receber menor quantidade de citações externas (por autores de outros grupos de pesquisa), uma boa maneira de promover seus trabalhos é começar com autocitações e, em seguida, incentivar as citações externas. Desta forma, o número de publicações não citadas por grupos de baixo desempenho será reduzido. Parece, portanto, que a dimensão quantitativa reforça um mecanismo interno de promoção.

O fenômeno da autocitação se estende aos periódicos. A este respeito, van Raan (2008) observou semelhanças com as autocitações dos grupos: a proporção de autocitações tende a diminuir com o impacto e desempenho do veículo de publicação, sendo menor nas revistas com maior fator de impacto.

Um outro problema apontado na literatura diz respeito às citações que criticam um trabalho citado. Não se trata de um reconhecimento e sim a negação do que está exposto no trabalho citado. No entanto, o termo reconhecimento não é restrito à validação ou ao elogio. A crítica é importante para a evolução da ciência e pode ser considerada uma forma de cooperação científica.

As duas variáveis aqui expostas, autocitação e citação crítica, podem alterar os resultados da avaliação de citações no modelo fechado exposto acima. Neste modelo, a ciência é vista como uma instituição com normas e valores comuns respeitados por todos os seus membros. Estes são passíveis de serem avaliados por meio de menções sobre a cooperação realizada entre eles. Porém, entendemos que essas variáveis não são as principais responsáveis pelas falhas no funcionamento deste modelo; ao contrário, são consequência do contexto onde este modelo está inscrito.

Romancini (2010) alerta ser necessário pensar a análise de citação no contexto mais amplo dos estudos da ciência, pois o nível micro (a citação) conecta-se com o macro (o sistema científico). Nesse sentido, o ato de citar implica o desejo do autor de incluir uma referência em seu trabalho e oferecer informações importantes sobre o desempenho e a influência científica do artigo citado. O fato de citar um estudo também afeta a relação entre o documento citado e o citante, por exemplo, quando se argumenta invalidando as descobertas do trabalho citado. Essa dinâmica, em princípio baseada na solidariedade e na cooperação, converte-se em competição. Dito de outra maneira, quando a atividade científica e seus atores (o pesquisador, a instituição, o país) são majoritariamente avaliados quantitativamente, por meio de artigos científicos ditos certificados, a produção científica converte-se em mercadoria, e na lógica do mercado, a competição supera a cooperação.

3.2.2.4.1 Índice de Hirsch e Fator de Impacto

O índice de Hirsch, mais conhecido como índice h, foi proposto por Hirsch (2005) para quantificar a produção científica de um pesquisador. Especificamente, o índice mede o impacto cumulativo da produção científica de um indivíduo. Esse índice postula que um cientista tem o índice h se h de seus N artigos têm pelo menos h citações cada e os outros (N-h) têm $\leq h$ citações cada. Assim, se um pesquisador tem um índice $h = 12$, significa que ele escreveu 12 trabalhos, cada um dos quais recebeu pelo menos 12 citações.

Os dados necessários para calcular o índice h são o número de artigos publicados por um autor durante a sua carreira e o número de vezes que esses artigos foram citados. Ao classificar os trabalhos em ordem decrescente de citações recebidas, pode-se identificar o índice h como a classificação mais alta, de modo que as primeiras h publicações receberam pelo menos h citações.

A título de exemplo, Andrés (2009) selecionou um autor que publicou um dos artigos mais citados no campo bibliométrico. A autora realizou a pesquisa bibliográfica na *Web of Science* (WOS) com a palavra-chave “bibliometrics” no tópico, sem filtrar um limite de anos de publicação. Stephen P. Harter foi o autor identificado. Na etapa seguinte, a autora realizou outra busca bibliográfica para identificar todos os artigos que Stephen P. Harter havia publicado durante sua carreira. Essa busca foi

novamente feita na WoS, desta vez com a palavra-chave “Harter SP” no campo de autor, sem delimitação de ano de publicação. Como o índice h considera tanto sua produtividade como o impacto de suas citações, foram reunidos todos os artigos publicados por esse autor entre 1971 e 2000. Os 77 itens identificados foram classificados de acordo com o número de citações recebidas, conforme tabela abaixo.

Tabela 7 – Ranking de citações nos artigos publicados por Stephen P. Harter

			Continuação:		
Posição do artigo no ranking	Número de citações	Ano de publicação	Posição do artigo no ranking	Número de citações	Ano de publicação
1	173	1992	40	0	1979
2	64	1996	41	0	1979
3	56	1997	42	0	1979
4	47	1975	43	0	1980
5	46	1998	44	0	1980
6	43	2000	45	0	1981
7	36	1985	46	0	1981
8	33	1993	47	0	1981
9	27	1975	48	0	1981
10	26	1996	49	0	1981
11	22	1971	50	0	1982
12	21	1997	51	0	1982
13	17	1984	52	0	1984
14	16	1990	53	0	1985
15	15	1992	54	0	1985
16	14	1990	55	0	1985
17	13	1979	56	0	1986
18	13	1984	57	0	1987
19	11	1978	58	0	1987
20	10	1982	59	0	1987
21	8	1998	60	0	1987
22	8	2000	61	0	1987
23	7	1988	62	0	1987
24	6	1990	63	0	1988
25	5	1978	64	0	1989
26	5	1996	65	0	1990
27	4	1996	66	0	1990
28	2	1988	67	0	1991

29	2	1990		68	0	1991
30	2	1992		69	0	1992
31	1	1981		70	0	1993
32	1	1986		71	0	1995
33	1	1993		72	0	1996
34	1	1998		73	0	1997
35	0	1971		74	0	1998
36	0	1971		75	0	1998
37	0	1975		76	0	1998
38	0	1978		77	0	1998
39	0	1979				

Fonte: Andrés (2009, p.64-65).

A primeira coluna mostra o número de classificação atribuído a cada artigo, de acordo com o número de citações recebidas. O primeiro item da tabela é o artigo que recebeu o maior número de citações, neste caso 173. Em contraste, os últimos itens da tabela são aqueles que não foram citados desde que foram publicados.

A Tabela 7 fornece as informações necessárias para obter o índice h . Neste caso, $h = 15$, porque o autor tem 14 artigos com 16 ou mais citações cada, enquanto 15 têm 15 ou mais citações cada. No entanto, o índice h deste autor não pode ser superior a 15, pois não há 16 documentos com 16 ou mais citações cada. Especificamente, o artigo classificado na décima sexta posição tem 14 citações e, portanto, o valor mais alto possível para h é 15. Por conseguinte, pode-se afirmar que este autor tem 15 artigos com pelo menos 15 citações cada. Entre estes 15 artigos, há alguns que têm um grande número de citações, como no caso do artigo classificado em primeiro lugar, com 173 citações. Esses 15 documentos formam o núcleo de Hirsch, constituído pelo grupo de documentos incluídos no índice h (do primeiro classificado em relação ao classificado na posição h).

Andrés (2009) afirma que, no índice h , o importante não é o número de citações recebidas, mas sim os documentos h que obtiveram pelo menos h citações cada. No entanto, o que significa ter um índice h igual a 15? Deborah Dias, especialista em treinamento de clientes da Clarivate Analytics, afirma que esse valor nada indica se não for comparado com os valores de distintos pesquisadores que tratam do mesmo tema (DIAS, 2019). Portanto, o índice h de Stephen P. Harter deve ser comparado com os índices h de seus pares da área de bibliometria. De acordo com a especialista, o valor de referência não é a média de todos os índices h de todos os

pesquisadores da área de bibliometria, porém, o maior índice h alcançado por alguém dentro do conjunto. Deve-se acrescentar, ainda, que a validade de um índice comparativo deve levar em conta um mesmo período de tempo para todos os autores.

No subcapítulo 2.4, discutimos a invenção da normalidade, na perspectiva de Ian Hacking, enquanto fenômeno fundamental da consolidação da estatística. O autor distingue o trabalho de dois atores fundamentais do processo de invenção da normalidade: Durkheim e Galton. Durkheim defendia que o normal era a média. Galton, por outro lado, afirmava que a média indicava mediocridade e, portanto, era algo a ser melhorado. Hacking (2005) afirma que a concepção de normal de Galton é uma história de sucesso, mas não a de Durkheim. Na sociedade contemporânea, o normal de Durkheim é o medíocre, de modo que se busca o ideal galtoniano. O índice h parece ser, a partir da perspectiva apresentada por Dias (2019), um exemplo da vitória do pensamento de Galton. Os índices h dos pesquisadores não têm como referência, para efeito de comparação, a média de todos os pesquisadores da área. A referência é o mais produtivo, aquele que provavelmente estará mais à direita da curva de Gauss. Portanto, o índice h é uma medida que fomenta a competição entre pesquisadores, instituições, universidades. Não promove, portanto, um aspecto fundamental da atividade científica: a cooperação, o trabalho coletivo.

Os autores adeptos da lógica de mercado podem contra-argumentar afirmando que não há qualquer problema em premiar os melhores. De fato, não seria um problema se a competição fosse justa. Glänzel (2003) afirma que, para descrever a produtividade de uma determinada área, deve-se ter certeza de que os documentos incluídos no estudo são verdadeiramente representativos do campo de pesquisa. De acordo com Glänzel, embora existam problemas na cobertura de periódicos e na política de processamento de dados, no *Science Citation Index* (SCI), suas características únicas - multidisciplinaridade, seletividade, cobertura ampla, completude de endereços, referências bibliográficas e disponibilidade - tornaram-na a fonte básica mais aceita de análises bibliométricas. Andrés (2009), por sua vez, afirma que a primeira e mais importante característica das bases de dados é a cobertura da literatura publicada em periódicos. Nas avaliações, deve-se assegurar, portanto, que os dados relativos a país, língua e publicações é exaustiva.

Por outro lado, há pesquisadores, dentre os quais Van Raan, que discordam de Glänzel. Esses pesquisadores afirmam que os critérios adotados pela WoS

privilegiam a inclusão de periódicos dos Estados Unidos e de países de língua inglesa, o conjunto de periódicos indexados, por ser restrito, produz representação desigual de áreas do conhecimento, é evidente a falta de padronização de dados (nomes de autores e instituições) (VAN RAAN, 2005).

Uma premissa fundamental em um estudo comparativo é o uso de critérios justos. Segundo Hacking (2009b), algo é injusto se favorece uma parte em detrimento da outra. O autor afirma que, no caso de jogos de azar, podemos utilizar a palavra “tendencioso”. O mesmo pode ser afirmado em relação à base de dados Web of Science. De fato, se uma base de dados privilegia as publicações de uma determinada região geográfica, os resultados de uma pesquisa sobre os países que mais se destacaram em uma determinada área do conhecimento, não são representativos. Os resultados serão parciais porque os dados de partida foram coletados em uma base de dados não exaustiva, cuja configuração atende a políticas específicas.

Uma resposta possível para o problema aqui exposto seria a utilização de repositórios digitais institucionais. De acordo com Marcondes, “repositórios institucionais trazem para universidades e instituições de pesquisa a oportunidade de se fortalecerem institucionalmente a partir da visibilidade de sua produção acadêmica organizada e disponível” (2009, p. 21).

No início deste item, citamos alguns autores que assinalaram a importância das citações para a avaliação da cooperação científica; a análise de citação permitiria revelar a arquitetura das redes de pesquisa, seus padrões e características. No entanto, dois dos principais indicadores de análise de citação – Índice h e Fator de Impacto – medem produtividade, classificam e comparam os mais produtivos. Esses índices não são adequados para avaliar as atividades científicas porque se limitam a quantificar a produção de artigos. A atividade científica é, sem dúvida, maior do que a produção de artigos.

Em artigo recente publicado na revista *Nature*, Wouters e outros (2019) alertam para a necessidade de criar um conjunto de métricas mais amplo e transparente para melhorar a publicação científica. Wouters afirma que a criação de novos indicadores não é sinônimo de melhoria do cenário atual, porém, associado à utilização responsável, poderia ser um caminho para avaliar as qualidades de um periódico. O autor pontua ser esta uma tarefa desafiadora, pois os interesses variam e, muitas vezes, entram em conflito. Talvez este seja o momento de os pesquisadores,

grupos de pesquisas, instituições e países analisarem os métodos de avaliação utilizados, a fim de propor modelos mais justos. Caberia perguntar, ainda, se são necessários novos indicadores, ou, ao contrário, se o mais importante é desenvolver estudos rigorosos, como propõe o Manifesto de Leiden.

4 ESTUDO DE CASO: PRODUÇÃO CIENTÍFICA DO INSTITUTO BUTANTAN

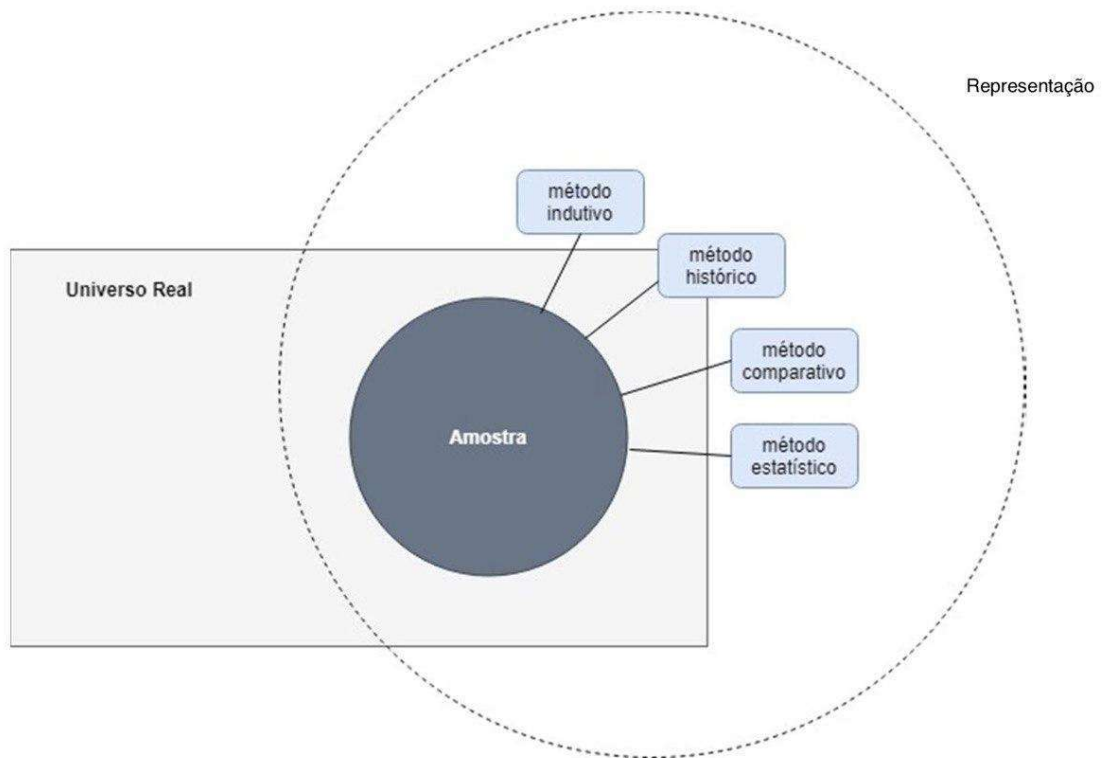
Este estudo de caso tem como objeto a produção científica do Instituto Butantan. A escolha não é aleatória. De 2016 a 2017, fui estagiária da Biblioteca do Instituto Butantan tendo realizado atividades de classificação, catalogação e inserção da produção científica no repositório institucional. No estágio, tive a oportunidade de levantar dados sobre a produção científica da instituição, uma lista preliminar de autores responsáveis pelas pesquisas, os projetos desenvolvidos em colaboração com outras instituições, estudar a terminologia e o vocabulário controlado utilizado no repositório, e outras informações úteis para compor o corpus da pesquisa.

No item a seguir, apresenta-se, inicialmente, o enquadramento metodológico do estudo de caso, sendo também justificados, de forma circunstanciada, os parâmetros adotados em seu desenvolvimento.

4.1 Percurso metodológico do estudo de caso

A questão metodológica, deve-se salientar, inclui o quadro de referência utilizado nas pesquisas, definido como uma totalidade que abrange a teoria ou teorias de base de um estudo e os métodos mais adequados propostos por essas teorias. No caso da Bibliometria, é fundamental compreender que se trata de aplicar o método indutivo para construir hipóteses e realizar inferências. Este se associa aos métodos histórico, comparativo e estatístico, juntamente com as técnicas específicas de coleta de dados. O quadro abaixo procura explicitar os aspectos metodológicos gerais da presente pesquisa. Na Ilustração 6, procura-se evidenciar que o estudo de caso parte de uma amostra da população de um universo concreto e é analisado segundo a combinação de quatro métodos: indutivo, histórico, comparativo e estatístico descritivo. Os resultados da aplicação estão organizados em representações discursivas e/ou gráficas e, ao final, interpretados.

Ilustração 6 - Quadro de referência dos estudos de construção de indicadores bibliométricos.



Fonte: Elaboração do autor.

O cuidado com a descrição do método e as possibilidades de replicação contribuem para assegurar a qualidade de um estudo bibliométrico. É esse rigor que poderá conferir confiabilidade aos resultados alcançados. Assim, a apresentação de hipóteses e explicações deve estar fundamentada em um discurso lógico, devidamente encadeado, com base em proposições cientificamente sólidas.

O método indutivo, como dito acima, é fundamental nos estudos empíricos. A indução é um processo de abstração em que, a partir de uma amostra restrita de dados da realidade, são inferidas conclusões gerais ou universais que vão além da amostra analisada. No caso dos estudos bibliométricos, como o próprio nome indica, trata-se de mensurar determinados fenômenos da realidade por procedimentos estatísticos e probabilísticos. Deve-se lembrar que a argumentação indutiva difere da dedutiva porque, nesta, as premissas verdadeiras levam, necessariamente a conclusões verdadeiras. Nos processos indutivos, contudo, a conclusão extrapola as premissas, de tal forma que as conclusões são consideradas prováveis. Cervo e Bervian (1978, p. 25) afirmam que “as premissas de um argumento indutivo correto sustentam ou atribuem certa verossimilhança à sua conclusão. Assim, quando as

premissas são verdadeiras, o melhor que se pode dizer é que a sua conclusão é provavelmente verdadeira”.

Entendemos que um estudo bibliométrico consistente requer estar atento às leis e regras do método indutivo. Nesse sentido, é importante considerar as três etapas fundamentais da indução: observação dos fenômenos, descoberta das relações entre eles e generalização da relação.

Ilustração 7 - Percurso da atividade de análise bibliométrica

Etapas	Descrição
Prenúncio (Algo aconteceu)	Algum fenômeno desperta o interesse do analista.
Primeiras Percepções (O que está acontecendo?)	O analista passa a indagar sobre as causas desses problemas e a construir hipóteses iniciais a fim de explicar o fenômeno que chamou a sua atenção. Neste estágio, é cedo para se apontar as causas precisas do fenômeno. Os esforços visam identificar o contexto e elaborar uma estrutura de trabalho constituída de hipóteses alternativas que permitam observar o fenômeno.
Explicação (Por que está acontecendo?)	Para isso, parte-se de um processo de abstração em que se relaciona o fenômeno observado a outros. Recorre-se, neste caso, a fontes de dados diversas para fundamentar as explicações sobre o fenômeno que chamou a sua atenção.
Implicações (O que se pode concluir?)	Nesta etapa, o resultado obtido é expresso por indicadores. É neste estágio que ocorre a generalização dos resultados.

Fonte: Faria (2017, adaptado).

No quadro acima, estão apresentadas as etapas que orientam o trabalho baseado no método indutivo. As etapas e regras desse método repousam nas seguintes leis:

- 1) “nas mesmas circunstâncias, as mesmas causas produzem os mesmos efeitos”;
- 2) “o que é verdade em muitas partes suficientemente enumeradas de um sujeito, é uma verdade para todo esse sujeito universal” (Nérici, 1978, p. 72).

Em estudos bibliométricos, as etapas e regras são orientadas também por uma visão global do contexto de produção científica, que se refere aos aspectos o histórico, o comparativo e o estatístico.

O método histórico, de acordo com Lakatos, parte da premissa que “as atuais formas de vida social, as instituições e os costumes têm origem no passado, é importante pesquisar suas raízes, para compreender sua natureza e função” (2016, p.88). Nesse sentido, o método histórico investiga acontecimentos, processos e instituições do passado para verificar sua influência no fenômeno observado. Em um estudo bibliométrico, a contextualização do objeto de análise, quer seja de um instituto de pesquisa, ou de uma rede de pesquisadores, é fundamental para a compreensão da dinâmica da atividade científica. Em outras palavras, os indicadores de produção científica gerados a partir de estudos bibliométricos devem levar em conta os aspectos históricos, econômicos e sociais no qual o objeto de estudo está inserido.

No que concerne ao método comparativo: um estudo bibliométrico deve realizar comparações com a finalidade de observar similaridades e divergências. De fato, “o método comparativo permite analisar o dado concreto, deduzindo do mesmo os elementos constantes, abstratos e gerais. Constitui uma verdadeira “experimentação indireta” (LAKATOS, 2016, p. 89).

Os indicadores de produção científica podem apontar algo que já aconteceu ou uma tendência futura. O que se busca é estabelecer comparações: pode-se comparar a produção científica de um período com outro, as citações de um autor em relação a outro, de uma instituição com outra, de um país com outro, etc. Compara-se a fim de posicionar aquilo que queremos estudar em relação aos casos semelhantes. Através de indicadores também procura-se entender relações de causa e efeito, ou seja, as implicações que os indicadores apontam sobre uma determinada situação (FARIA, 2017).

O método estatístico descritivo é o mais utilizado em estudos bibliométricos. Esses procedimentos “permitem obter, de conjuntos complexos, representações simples e constatar se essas verificações simplificadas têm relações entre si” (QUETELET apud LAKATOS, 2016, p.90). Nesse sentido, o método estatístico descritivo propõe a redução dos fenômenos a termos quantitativos para identificar e objetivar as relações dos fenômenos entre si e obter generalizações. No entanto, salienta-se que a estatística, por ser um método de experimentação e prova, não é isenta de aspectos subjetivos. Sua aplicação deve, portanto, levar em conta as limitações do método e ser utilizado com o devido cuidado.

Como abordado anteriormente, a Bibliometria possui procedimentos e ferramentas próprios de análise - os métodos bibliométricos - que compreendem as leis bibliométricas (Bradford, Lotka e Zipf), as análises de citações e os indicadores de desempenho da atividade científica. Deve salientar que os métodos bibliométricos e estatísticos são utilizados para diminuir os riscos próprios de uma argumentação indutiva.

4.2 O Instituto Butantan (IB): contextualização do objeto de pesquisa

Em um estudo bibliométrico, a contextualização do objeto é fundamental para a compreensão da dinâmica da atividade científica. A contextualização, via de regra, leva em conta os aspectos históricos, econômicos e sociais no qual o objeto de estudo está inserido. Nesta pesquisa, de análise da produção científica de uma instituição dedicada à saúde pública, apresentamos, primeiramente, uma breve contextualização das políticas de saúde pública latino-americanas a fim de compreender como e por que foram criados os institutos de pesquisa voltados para a área de saúde e, em um segundo momento, contextualizaremos o objeto deste estudo de caso: o Instituto Butantan (IB).

Cueto e Palmer (2016) afirmam que, na época da independência dos países latino-americanos, houve um retrocesso na ordem institucional, inclusive no campo da medicina. Esse quadro se alterou somente no final do século XIX, com a criação das agências nacionais de saúde para implementar as práticas de saúde pública e controle.

As elites médicas da América Latina, atores importantes do campo da saúde pública, organizados em associações nacionais, conjugaram interesses profissionais e científicos e classificaram as doenças da região, como, por exemplo, a doença de Chagas no Brasil. Isso conferiu grande prestígio a essa comunidade, bem como permitiu que se tornassem líderes mundiais em pesquisas nas áreas de bacteriologia médica e parasitologia (CUETO; PALMER, 2016).

No Brasil, no final do século XIX e início do século XX, as agências nacionais de saúde, nas quais as elites médicas desempenhavam papéis políticos e profissionais importantes, passaram a centralizar as ações de saúde do Estado, integrando-se a diferentes ministérios: educação pública, interior, agricultura e

comércio. Com a anexação, a saúde estatal obteve capacidade coercitiva, pedagógica e proteção de áreas econômicas consideradas prioritárias. Em função da crescente dependência econômica e política crescente do país aos EUA, as instituições de saúde passaram a ser apoiadas pelo Estado ou por iniciativas filantrópicas bilaterais (como as patrocinadas pela Fundação Rockefeller). Porém, as instituições brasileiras ficavam muito aquém das instituições de pesquisa em saúde das potências internacionais.

Gilberto Hochman (1998), ao procurar evidenciar o porquê, quando e como ocorreu a estatização e nacionalização das políticas sanitárias no Brasil, durante a Primeira República, assevera que a saúde passou a ser pública, coletivizada, em decorrência do desenvolvimento capitalista e seus efeitos (industrialização, imigração, aumento populacional, intensa urbanização), que geraram problemas como as epidemias. A criação de organizações e políticas permanentes, pactuada entre os estados e a Federação, foi centralizada no poder central, para assegurar as funções coercitivas desse aparato. Embora o autor atribua a estatização à tomada de consciência das elites na defesa de seus interesses, um aspecto interessante se destaca: o custo-benefício da coletivização. A opção era adotada quando os custos externos eram maiores do que os custos da estatização e estes menores do que os seus benefícios.

Os programas de saúde latino-americanos foram implementados de forma verticalizada. Os modelos, altamente tecnocráticos, estavam concentrados em doenças específicas, tendo sido deixado de lado o princípio do atendimento à saúde primária. Nas décadas de 1980 e 1990, as propostas neoliberais tecnocráticas tornaram ainda mais restritiva a atenção primária à saúde, tanto em nível nacional quanto internacional (CUETO; PALMER, 2016).

O acesso à saúde pública integral na América Latina foi marcado por dois processos. O primeiro, de apoio à proposta de atenção primária à saúde (APS), definida pela OMS em 1978, documento universalizante, na mesma linha das declarações de Direitos humanos. O segundo, de crítica à Atenção Primária à Saúde (APS), defende sua reformulação em APS seletiva. A APS abrangente, que tem Cuba como modelo, foi bastante criticada pela lógica neoliberal, no sentido de que os Estados deveriam abster-se ao máximo de expandir os serviços sociais: "...a lógica desse consenso era que o Estado deveria deixar de operar muitos programas,

passando a gerenciar a área da saúde de forma a abrir espaço a prestadores privados” (CUETO; PALMER, 2016, p. 228).

A América Latina tornou-se um campo de batalha entre os que defendem um atendimento global e aqueles que defendem a seletividade, bem como um laboratório das reformas neoliberais dos sistemas de saúde e previdência social. Essa lógica se respalda nas atitudes conservadoras enraizadas na maioria dos profissionais de saúde. Os médicos latino-americanos, oriundos das classes alta e média-alta, são formados em escolas que se assemelham muito às universidades estadunidenses. Outro obstáculo para a implementação da APS no subcontinente foi a vontade política real, uma vez que as ditaduras latino-americanas, favorecendo a APS seletiva, criaram programas de atenção primária subfinanciados. Cueto e Palmer (2016) evidenciam que o modelo neoliberal, consistente em medidas que enfatizam o papel do mercado sobre o Estado, a desregulamentação, privatização de empresas públicas, sistemas de taxas para serviços públicos, remoção de subsídios, foi aplicado na América Latina com maior ou menor intensidade. O termo globalização passou a ser utilizado para explicar a rápida disseminação dessas medidas como o modelo ideal de governança para qualquer sociedade, o que justificava a redução de empregos públicos, o fim da intervenção do Estado na economia e privatização dos serviços públicos em nome da eficiência.

4.2.1 Contexto de criação do Instituto Butantan.

Para entender o processo de criação e consolidação do Instituto Butantan (IB), importante instituição de pesquisa pública nacional, bem como para conhecer a situação atual de sua produção científica, objetivo deste estudo de caso, é indispensável traçar o histórico dessa instituição.

O IB é o principal e mais destacado centro de excelência em pesquisa biomédica e produção de imunobiológicos do Brasil, com reputação internacional incontestável. Nosso estudo detém-se na análise de sua constituição e consolidação, ao longo dos séculos XX e XXI, dentro do contexto de “subdesenvolvimento” latino-americano.

Uma característica comum às instituições de pesquisa em saúde no Brasil e em toda América Latina, periferia dos centros capitalistas europeus e estadunidense,

é o “imediatismo” com que são implementadas as políticas de saúde e a destinação dos respectivos recursos (CAMARGO, 2002). Essas instituições, desde os primórdios, por algumas razões que serão apresentadas a seguir, se pautaram pelo enfrentamento pontual de demandas de saúde pública e não para responder a políticas públicas concretas, consistentes e pensadas a longo prazo. De fato, o IB também não foge a esta lógica.

O histórico do IB é marcado por crises persistentes, com oscilações periódicas entre a priorização ou a destinação de recursos financeiros diretamente para a produção e comercialização de soros e vacinas, diante de alguma urgência pontual, ocasionada por surtos epidêmicos, infestações e doenças tropicais e períodos de priorização da pesquisa para o desenvolvimento de trabalhos em medicina experimental, química, farmacologia, toxilogia e outras áreas do conhecimento científico (INSTITUTO BUTANTAN, 2013). As oscilações das prioridades dentro do instituto estão relacionadas a fatores políticos tanto internos quanto externos. A instituição tem sua história marcada por denúncias de corrupção, afastamento de diretores por irregularidades, descaso com a ciência, com o erário público, com os trabalhadores e pesquisadores da instituição, ainda que sua imagem, perante a grande mídia, seja constantemente associada à excelência em pesquisa (TAVARES, 2019 ; ALVES, WATANABE, 2017; AMORIM, 2009).

As origens do IB remetem ao contexto internacional de criação de institutos de pesquisa dessa natureza, como a do Instituto Pasteur de Paris, em 1888. No cenário brasileiro, a abolição da escravidão, a expansão da cultura cafeeira, a urbanização desordenada e o início da industrialização levaram à criação do Instituto Bacteriológico de São Paulo em 1892, sob a Direção de Adolfo Lutz. Contudo, esses centros, nos moldes do Instituto Pasteur de Paris, tiveram conformações específicas, de acordo com a realidade histórico-social latino-americana (GOMES, 2013). O próprio IB foi inicialmente concebido, em 1899, como um laboratório dentro do Instituto Bacteriológico, em meio à reformulação do Serviço Sanitário do Estado de São Paulo, à época comandado por Emílio Ribas. Sua atividade inicial foi prioritariamente orientada à produção de soro antipestoso para combater um surto epidemiológico de peste bubônica no município de Santos/SP, importante porto de escoamento de *commodities*, principalmente de café. Posteriormente, o modelo francês, assentado

no tripé produção, pesquisa e ensino, foi replicado pelo IB, porém, aqui se traduziu ora na priorização da pesquisa, ora na produção de soros e vacinas (DIAS, 2010).

O Instituto Butantan, enquanto instituição de pesquisa independente, foi fundado oficialmente em 23 de fevereiro de 1901, por meio do Decreto Estadual nº 878-A, com a denominação de “Instituto Serumtherápico”. Aproximadamente na mesma época, foi criado em Manguinhos, no Rio de Janeiro, o “Instituto Soroterápico Federal”, atualmente Instituto Oswaldo Cruz ou Fiocruz. Por intermédio de Adolfo Lutz, foi nomeado como diretor da instituição paulista o médico sanitarista Vital Brazil e, para o instituto carioca, o também médico sanitarista Oswaldo Cruz. O IB começou a produzir soros antipestosos em instalações precárias e cocheiras improvisadas na Fazenda Butantan, em São Paulo, local à época bastante afastado do centro urbano da capital do estado. Portanto, as demandas sociais eram atendidas dentro de suas poucas possibilidades de atuação (FUNDAÇÃO BUTANTAN, 2015). Porém, uma vez sob controle o surto epidêmico na Baixada Santista, Vital Brazil passou a desenvolver pesquisas sobre soros antiofídicos, diante da incidência cada vez maior de acidentes envolvendo animais peçonhentos nas fazendas de café do interior do estado de São Paulo. Esta passou a ser umas das características singulares da instituição ao longo de mais de um século de existência (IBAÑEZ, WEN, FERNANDES, 2005).

Um outro fator que impulsionou as atividades do IB, no início do século XX, foi a vinda da missão da Fundação Rockefeller (FR) ao Brasil. A organização estadunidense firmou acordos de cooperação direta com o Estado para combater doenças como a malária. Carlos Chagas e Vital Brazil integraram a comissão escolhida pela FR, de modo a direcionar recursos para as atividades de combate a determinadas doenças tropicais (SANTOS, 1985). Esse tipo de cooperação seletiva, como já evidenciado acima, não tinha como foco os problemas da saúde pública em geral; ao contrário, era direcionada para resolver questões pontuais, de interesse das elites econômicas e não propriamente para atender aos problemas de saúde da sociedade como um todo.

Esse imediatismo é apontado como um dos principais fatores da descontinuidade da produção científica. Segundo Antonio Carlos de Camargo (2002), a pressão política pela solução de problemas urgentes de saúde pública, baseada somente em cálculos de custo-benefício, acabou determinando a oscilação de suas prioridades, ora voltada para a pesquisa, ora para a produção de vacinas, oscilação

que persiste até hoje. A título de ilustração, sob a direção de Vital Brazil, seu primeiro diretor, foram desenvolvidas pesquisas nas áreas de Biologia, Bioquímica e Farmacologia, importantes para o preparo da vacina BCG, de prevenção à tuberculose, bem como de vacinas para prevenção da febre tifoide. No entanto, em 1917, foi firmado um acordo entre o IB e a Casa Ambrust, em que essa empresa detinha o direito de ser a única depositária e distribuidora dos produtos produzidos pelo IB durante um período de 5 anos. No ano seguinte, evidenciou-se o balanço desfavorável desse acordo, pois, sem assumir qualquer risco advindo da produção, a Casa Ambrust ficava com a maior parte dos dividendos (IBAÑEZ, WEN, FERNANDES, 2005). Esse fato, em tese, teria levado à saída de Vital Brazil do IB, em 1919, porém, ele retornou ao cargo de diretor de 1924 a 1927. O episódio, entretanto, evidencia os tipos de crises institucionais e a falta de autonomia da instituição desde a sua criação.

Após a saída de Vital Brazil, assumiu a direção do IB o herpetólogo Afrânio do Amaral, que promoveu uma grande reforma administrativa. O IB passou a ser subordinado à Secretaria de Educação e Saúde Pública, estabelecido como um Centro de Medicina Experimental destinado ao estudo de questões relativas à patologia humana. Nesse período, foram contratados renomados pesquisadores nacionais e estrangeiros, tais como Karl Heinrich Slotta e José Lemos Monteiro para auxiliarem na estruturação das novas áreas de pesquisa. Porém, uma nova crise institucional, que ficou conhecida como o “Caso Butantan”, se instalou, em decorrência de acusações de improbidade administrativa e disputas políticas envolvendo o diretor Afrânio do Amaral, o então deputado Ademar de Barros e empresas privadas concorrentes do IB, como o Instituto Pinheiros (posteriormente adquirido pela empresa estadunidense “Laboratórios Sintex do Brasil”) (IBAÑEZ, WEN, FERNANDES, 2005). Em 1937, foi firmado um contrato de terceirização, com exclusividade de compra e distribuição de soros e vacinas com a empresa Delpech e Cia., muito similar ao acordo firmado com a Casa Ambrust, em 1917. Mais uma vez, constatou-se o caráter desfavorável do acordo para o IB, pois a empresa privada colhia todos os lucros da produção e a instituição pública produtora arcava com todos os prejuízos.

Com a saída de Afrânio do Amaral, em 1938, e após o término da II Guerra Mundial, a diretoria do IB foi assumida, em 1947, pelo médico Eduardo Vaz,

proveniente do próprio Instituto Pinheiros. Nesse período (1947-1951), a instituição paralisou a pesquisa, tendo sido rebaixada à condição de um entreposto de soros e vacinas (IBAÑEZ, WEN, FERNANDES, 2005). O diretor, além de suas estreitas relações com o então governador do estado de São Paulo, Ademar de Barros (1947-1951), também afastou pesquisadores ligados às administrações anteriores e deu ênfase à modernização da produção em escala industrial, dando início, por exemplo, à confecção de vacina antirrábica. Eduardo Vaz deixou o IB na mesma época em que terminou o mandato de Ademar de Barros, também sob acusações de improbidade e enfrentamento de sindicâncias internas (IBAÑEZ, WEN, FERNANDES, 2005). Depois de 1951, bem como no decorrer das décadas de sessenta e setenta, sob a direção de lideranças internas, o IB passou por um novo processo de reestruturação e de retomada da prioridade em pesquisa, sem prejuízos ou interrupções na produção de imunobiológicos.

Durante a Ditadura Civil-Militar brasileira, o Instituto esteve vinculado a programas de vacinação em âmbito nacional, tais como as campanhas de vacinação contra a poliomielite, difteria, varíola, meningite, cólera, sarampo e o Programa Nacional de Imunização (PNI) de 1973. Vale destacar que, em 1978, ainda no período ditatorial, os funcionários do IB fundaram a Associação de Servidores do Instituto Butantan (ASIB). Essa organização sindical atuava na luta por direitos trabalhistas e contra os abusos administrativos, passando a se configurar como uma das forças políticas internas, bem como um marco significativo na história da instituição (DIAS, 2012).

Na década de 1980, período de “abertura” política, assumiu a diretoria do IB, o geneticista Willy Beçak (1983-1991). Esse fato coincidiu com a promulgação da Lei Complementar estadual nº 335/1983, que regulamentou a carreira de pesquisador científico, que permitiu a contratação de lideranças científicas e renovação do corpo técnico-científico do instituto. Outro fato importante que mudou a capacidade de atuação do Instituto Butantan foi a crise nacional no abastecimento de soros antiofídicos, na qual se descobriu que o IB produzia soros inócuos, ou seja, impróprios para uso humano. Como resposta, foi lançado o Programa Nacional de Autossuficiência em Imunobiológicos (PASNI), pelo Ministério da Saúde, para financiar o desenvolvimento científico-tecnológico e a modernização dos laboratórios públicos. O IB também se tornou, no período, um dos maiores fornecedores de

imunoderivados do Sistema Único de Saúde (SUS), arrecadando, assim, somas vultosas.

Os recursos recebidos permitiram criar o Centro de Biotecnologia, área considerada prioritária no desenvolvimento tecnológico nacional, que significou a retomada da priorização da produção de soros e vacinas, com a implementação do programa de Boas Práticas de Fabricação (BPF). Além disso, em 1989, foi criada a Fundação Butantan, com a finalidade de atuar como intermediária, agilizar e flexibilizar os processos de produção e comercialização de soros e vacinas e, também, de contratação de pessoal. Uma medida controversa porque estranha à natureza da instituição, uma entidade de direito público que deveria seguir os princípios éticos e legais próprios da administração pública, como por exemplo os processos de licitação e de concursos públicos para provimento de cargos.

Como resultado dessas circunstâncias, chegou à direção do IB o médico Isaias Raw (1991-1997). Em 1997, Raw deixou o cargo de diretor do instituto e passou à direção da Fundação Butantan, onde permaneceu até 2009, momento em que foi afastado cautelarmente, devido a mais um escândalo de corrupção envolvendo o Instituto e a Fundação. Isso se deu, de início, com a reformulação do estatuto da Fundação Butantan, em 1998, ocasião em que praticamente o instituto público de pesquisa científica se tornou subordinado à Fundação, a qual não atuava mais como mera intermediária do IB. Entre 2006 e 2007, o Tribunal de Contas estadual constatou irregularidades orçamentárias na Fundação Butantan. Em 2009, o Ministério Público identificou desvios de cerca de R\$ 35 milhões. A autoria de tais desvios recaiu sobre alguns funcionários administrativos e fornecedores do setor privado (JERONYMO, 2010).

Essa situação trouxe à tona discussões sobre a privatização do IB por parte do governo estadual. O cenário de privatização se acirrou ainda mais com o grande incêndio de 2010 no prédio das coleções, no qual se perdeu a maior parte do acervo zoológico do instituto, formado desde 1901, constituído por animais peçonhentos armazenados em recipientes com líquidos de alta inflamabilidade, tais como álcool e formol. Esta coleção zoológica possuía um valor incalculável, tendo em vista tratar-se de uma das maiores coleções dessa natureza no mundo, inclusive com espécimes já extintos.

Diante de mais uma grave crise institucional, o médico Jorge Kalil foi indicado para assumir a direção do instituto, em 2011, e no ano seguinte, assumiu também a presidência da Fundação, unificando o poder dentro do IB. Entretanto, por pressão do governo estadual, Kalil foi obrigado a deixar o cargo de diretor da Fundação Butantan, em favor do economista André Franco Montoro Filho. Os desdobramentos das investigações para apuração dos desvios de verbas e afins, levaram o governo do estado de São Paulo a contratar uma auditoria privada das contas da instituição, realizada pela empresa “Colorado Consultoria Empresarial Ltda.” Em fevereiro de 2017, André Franco Montoro Filho entregou o seu cargo na Fundação, alegando graves irregularidades, como por exemplo relacionadas à construção da fábrica de hemoderivados, pelo custo de mais de R\$ 200 milhões, mas que permanecia fechada e inativa, mesmo tendo sido inaugurada por diversas vezes para fins de campanhas eleitorais de governadores. Em razão disso, Jorge Kalil foi afastado do cargo e um dos membros da auditoria privada, o médico Dimas Tadeu Covas, foi nomeado diretor, no qual permanece atualmente (CRUZ, 2017). Isso denota a crescente relação de interesses políticos e do capital privado na privatização do IB, uma instituição pública de pesquisa, que deveria servir aos interesses da sociedade para a solução de problemas de saúde pública.

Atualmente, o Instituto expressa oficialmente a sua missão como a “de contribuir com a saúde pública por meio de pesquisas, inovação, produção e desenvolvimento de produtos biológicos, compartilhando conhecimento com a sociedade” (FUNDAÇÃO BUTANTAN, 2014, p.5). Para consolidar sua missão, o IB deve produzir imunobiológicos e outros produtos afins, o que implica a realização de pesquisas que mobilizam várias áreas de conhecimento como: Bacteriologia, Imunoquímica, Immunopatologia, Farmacologia, Bioquímica, Biofísica, Parasitologia, Biologia Celular e Molecular, Genética, Imunologia Viral, Zoologia, Ecologia e Evolução, Imunogenética, Biotecnologia, dentre outras áreas correlatas. A Toxinologia é uma das áreas do conhecimento associada ao desenvolvimento de produtos biotecnológicos, farmacêuticos e de ferramentas para estudos de processos biológicos. O IB é comumente mencionado como uma referência mundial na pesquisa toxinológica e conta, desde 2010, com o único Programa de Pós-Graduação em Toxinologia do Brasil, o PPGTox. Porém, como evidenciado acima, em determinados períodos, os trabalhos de medicina experimental, química, farmacologia, toxinologia e

outras áreas do conhecimento científico se expandiam, mas em outros, a prioridade era dada à produção e comercialização de soros e vacinas. Esse conflito de interesses mostra a complexidade das relações que se estabeleceram no campo da produção científica (INSTITUTO BUTANTAN, 2013).

4.3 Produção científica do Instituto Butantan: metodologia do estudo bibliométrico

Na estatística, o conjunto de dados que contém as características que se pretende analisar é chamado de população, que pode ser constituída de diferentes maneiras. No presente estudo bibliométrico, a população é composta de artigos científicos publicados em periódicos por pesquisadores do Instituto Butantan nos anos de 2007 a 2016.

4.3.1 Moldura analítica

O histórico do Instituto Butantan indica que ele enfrentou problemas administrativos associados a conflitos entre os interesses públicos e privados. Trata-se de uma prática comum observada nos institutos de pesquisa em saúde latino-americanos que, desde os primórdios, tiveram sua atuação pautada no enfrentamento pontual de demandas de saúde pública e não para responder a políticas públicas concretas, consistentes e pensadas a longo prazo. No caso do Instituto Butantan, essa oscilação se manifesta na descontinuidade de seu foco de atuação que oscila entre produção de soros e vacinas e pesquisa científica. Além das descontinuidades acima referidas, o histórico do Instituto mostra os episódios de denúncias de corrupção e descaso com o erário público. Somente na última década, ocorreram duas intervenções do Ministério Público no Instituto a fim de investigar desvios financeiros. No entanto, a bibliografia analisada não menciona se esses eventos interferiram na produção científica do Instituto. Sendo assim, pode-se perguntar: as crises institucionais e o padrão de atuação do Instituto que ora privilegia a pesquisa e ora privilegia a produção de imunobiológicos, afetaram a produção científica do Instituto Butantan? Como?

As respostas para essas perguntas foram construídas com base em dois conjuntos de dados: as evidências qualitativas enunciadas no quadro histórico, exposto no item 5.2 e na análise e quantificação dos seguintes dados do *corpus*:

- a) Evolução cronológica das publicações;
- b) Produtividade dos pesquisadores vinculados ao Instituto e à Fundação Butantan;
- c) Cooperação científica nacional e internacional.

Para fins operacionais, no presente estudo de caso, os pesquisadores do Instituto Butantan foram divididos em três grupos: servidores do Instituto Butantan, colaboradores da Fundação Butantan e bolsistas, caracterizados conforme a explicação a seguir:

- a) Primeiro grupo: Servidores do Instituto Butantan - originários de concurso público realizado pela Secretaria de Estado da Saúde de São Paulo. Importante salientar que o último certame ocorreu no ano de 2009 e, desde então, a contratação de recursos humanos está a cargo da Fundação Butantan.
- b) Segundo grupo: funcionários da Fundação Butantan. Desde 1989, o Instituto Butantan sofre intervenções da Fundação Butantan, uma entidade civil, com personalidade jurídica de direito privado. A instituição, credenciada pela Resolução SDECTIC nº 55/2018 como Fundação de Apoio da ICETESP Instituto Butantan, tem como uma de suas atribuições contratar recursos humanos em Regime CLT, bem como estagiários. Sendo assim, a Fundação é responsável por dois tipos de potenciais autores: 1) funcionário da Fundação, denominado colaborador, que possui vínculo empregatício, ou seja, um prestador de serviços pessoal, não-eventual e oneroso subordinado à Fundação Butantan mas que pode atuar tanto na Fundação quanto no Instituto; 2) estagiário da Fundação, comumente estudantes no começo da carreira profissional que possuem contratos firmados com a Fundação Butantan e interveniência da instituição de ensino de origem. Os estágios podem ser voluntários ou remunerados, não possuem vínculo empregatício e são realizados em caráter eventual. O estagiário, assim como o funcionário, pode atuar tanto na Fundação como no Instituto Butantan.

- c) Terceiro grupo: bolsistas, estudantes de pós-graduação ou iniciação científica e pós-doutorandos que recebem bolsa de instituições de fomento à pesquisa.

4.3.1.1 A escolha do *corpus* de análise

As unidades dos estudos bibliométricos podem ser livros, teses, dissertações, artigos de periódicos, trabalhos apresentados em eventos científicos, dentre outros. De acordo com Glänzel (2003), a produção de artigos é submetida a sistemas de revisão, análise de originalidade da pesquisa, disponibilização pública dos resultados e regras transparentes de publicação. Os sistemas de difusão devem, ainda, apresentar as informações referenciais de forma padronizada. Nesse sentido, o autor considera que os artigos científicos publicados em revistas científicas de referência são as unidades mais adequadas para os estudos bibliométricos. A justificativa de Glänzel (2003) é curiosa, pois livros, monografias e teses também costumam respeitar critérios canônicos de avaliação e certificação (bancas, de avaliação de trabalhos acadêmicos, conselhos editoriais, revisores, entre outros) e disponibilização pública desses produtos em formatos padronizados. Pode-se questionar, portanto, se o artigo científico é realmente a melhor unidade a ser analisada para avaliar a atividade científica dos diferentes campos do saber.

Antes de apresentar o *corpus* é interessante retomar a discussão sobre as unidades analisadas nos estudos bibliométricos. De forma geral, a bibliometria é definida como um campo que se propõe a quantificar os produtos da comunicação escrita (CUNHA; CAVALCANTI, 2008), principalmente pela análise de artigos científicos²⁶. Estes processos se materializam nos documentos. Buckland, citando Briet (BUCKLAND, 1997), afirma que são características fundamentais dos documentos a materialidade, a intencionalidade e o seu processamento.

O conceito de materialidade dos documentos foi retomado por Frohmann (2008), que afirma que a materialidade é uma ponte que liga “duas áreas conceituais: informação, por um lado, e práticas públicas e sociais por outro.” (p.18). Para sustentar

²⁶ O artigo científico, ou *paper* ou *scientific article*, segundo o Dicionário de Biblioteconomia e Arquivologia, é o “texto escrito que foi aprovado para publicação ou publicado num periódico científico.

esse ponto de vista, Frohmann recorre ao conceito de *enunciado* de Foucault. Para este último, o *enunciado* é qualquer mensagem: ele surge, transforma-se, amplia-se, conecta-se e desvanece. Neste sentido, os enunciados são materiais, pois estão agrupados em uma sequência linguística e só assim podem ser considerados e analisados. Foucault diferencia materialidade de fisicalidade, pois a materialidade é medida pela massa, inércia e resistência. Considerando que massa tem peso, inércia, estabilidade e permanência no tempo e no espaço, em clara referência à teoria da relatividade de Einstein, Foucault afirma que o *enunciado* é também fonte de energia e de poder.

Segundo Foucault, os enunciados ganham estabilidade e materialidade pela institucionalização (FOUCAULT apud FROHMANN, 2008). Em outros termos: a instituição é responsável pela materialidade porque é ela que decide o que será preservado, como será interpretado, como vai circular. A institucionalização, que podemos caracterizar como processos de documentação, dão massa, inércia e resistência ao documento, materializam a informação.

O raciocínio acima permite aprofundar a discussão sobre o objeto que nos interessa caracterizar nesta pesquisa: a materialidade e a estabilidade dos documentos nos estudos bibliométricos.

Livros, relatórios, teses, artigos publicados em periódicos, dentre outros, são unidades que podem compor um *corpus* de análises bibliométricas. A padronização é fundamental para quantificar e medir. No item 2.2, a revisão bibliográfica, mostra que a quantificação de pessoas, ao longo da história, foi precedida da análise de suas características, dentro de determinados padrões pré-estabelecidos, para posterior classificação e quantificação.

Em estudos bibliométricos, o *corpus* deve ser constituído de unidades padronizadas e estáveis. De fato, o resultado de uma determinada pesquisa científica materializada em um documento, contém uma série de dados identificadores que o representam, tais como autor, título e ano. São esses dados representacionais estáveis que permitem compor o *corpus* de análise.

Deve-se lembrar que os documentos nem sempre estão representados de forma rigorosa. É comum encontrar fichas catalográficas e referências bibliográficas com informações imprecisas e/ou erros. Nessa medida, para compor um *corpus* consistente, sólido, é muitas vezes necessário ter acesso ao documento original para

dirimir dúvidas. A oferta de artigos científicos, na *web*, pelos diferentes sistemas de informação é fator que facilita sobremaneira a consulta a documentos na íntegra. Nesta pesquisa, esse fator foi determinante para compor a amostra.

Outro aspecto a ser considerado: a construção de bases de dados de livros, teses, dissertações, relatórios de pesquisa e demais tipos de monografias são fundamentais para conhecer a atividade científica de campos como as Ciências Sociais e Humanidades. Porém, a construção de bases exaustivas e confiáveis sobre esses tipos de documentos está para ser feita.

4.3.1.2 A escolha da base de dados

Andrés (2009) afirma que a primeira e mais importante característica das bases de dados é a cobertura. Deve-se assegurar que a cobertura não seja parcial em relação à produção científica de países, idiomas e tipos de publicações. Uma outra característica observada por Andrés (2009) é a necessidade de normalização dos nomes de autores e das instituições às quais os primeiros estão vinculados.

É necessário, portanto, conhecer em profundidade as bases para selecionar as mais adequadas à natureza do estudo que se pretende realizar. Para isso, no presente estudo de caso, analisamos inicialmente três bases: Web of Science (WoS), Scopus e Plataforma Lattes.

A Web of Science é uma base de dados multidisciplinar, mundial, composta atualmente de 159 milhões de artigos de periódicos e 1,7 bilhão de referências citadas nos artigos (CLARIVATE, 2019). Estabelece pré-requisitos mínimos para a inclusão de um periódico na coleção, dentre os quais se destacam: 1) regularidade de publicação; 2) diversidade internacional de autores (citados e citantes); 3) qualidade editorial do periódico, que se refere à apresentação de informação bibliográfica e referências bibliográficas completas, afiliação dos autores, título e resumo descritivos; 4) apresentação de título, resumo e palavras-chave em inglês e 5) avaliação por pares. (TESTA, 1997).

Outra base de dados multidisciplinar também utilizada em estudos bibliométricos é a Scopus. Lançada em 2005 pela Elsevier, concorre diretamente com a Web of Science. Os registros da Scopus contêm dados de afiliação de todos os autores dos artigos indexados e as citações feitas e recebidas pelos artigos. Conta

também com ferramenta para cálculo do Índice h. Embora bem mais recente do que a Web of Science, a Scopus afirma ser “o maior banco de dados de resumos e citações da literatura com revisão por pares: revistas científicas, livros, processos de congressos e publicações do setor” (ELSEVIER, 2019).

A Plataforma Lattes é um ambiente virtual nacional, administrado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), composta por três bases de dados: Currículo Lattes, Diretório de Grupos de Pesquisa e Diretório de Instituições. É uma plataforma estratégica para o planejamento da política científica nacional pelo Ministério da Ciência e Tecnologia, tendo também a função de tornar públicas as atividades de fomento do CNPq. A base de dados Currículo Lattes foi criada em 1999 com o objetivo de registrar nacionalmente o ciclo de vida acadêmico de estudantes e pesquisadores do Brasil. Em 2016, estavam registrados 37.640 grupos de pesquisa de 531 instituições e 199.566 currículos de pesquisadores. (CNPQ, 2016).

Cada base de dados apresenta aspectos positivos e negativos. Chavarro, Ráfols & Tang (2018), van Leeuwen, Moed, Tijssen, Visser e Van Raan (2001) alertam para as seguintes limitações de cobertura tanto da WoS quanto da Scopus: baixa presença de publicações em língua não inglesa, problemas relativos à indexação e normalização de nomes de instituições e autores, baixa cobertura de periódicos que fogem do circuito Estados Unidos-Europa, especialmente nas ciências humanas e sociais. Apesar dessas desvantagens, há unanimidade sobre as vantagens: facilidade para identificação, descarga e tratamento dos dados. Estas funcionalidades são mais precárias na base de Currículos Lattes, cuja principal vantagem é a grande extensão de cobertura da produção científica nacional (MUGNAINI, 2013).

Para definir a fonte de dados mais adequada para a presente pesquisa, analisamos as três bases acima mencionadas. O quadro abaixo apresenta dados comparativos sobre os pontos fortes e fracos de cada base:

Ilustração 8 – Pontos fortes e fracos de bases de dados selecionados.

Base de dados	Pontos Fortes	Pontos Fracos
Web of Science	<ul style="list-style-type: none"> • Abrangência mundial; • Liberdade para buscas elaboradas; • Dados sobre citação; 	<ul style="list-style-type: none"> • Conjunto restrito de periódicos; • Baixa cobertura de artigos em língua portuguesa;

	<ul style="list-style-type: none"> • Dados sobre índice h e Fator de impacto; • Facilidade para descarregar e tratar os dados; • Disponibilização de estatísticas prévias. 	<ul style="list-style-type: none"> • Representação de áreas desigual; • Representação de países desigual; • Não há padronização de nomes de autores²⁷, o que praticamente inviabiliza as buscas da produção de autores do Instituto Butantan; • Não há padronização de nomes de instituições; • Divergências entre os dados do registro (metadados) e os dados bibliográficos dos artigos. • Não é gratuita.
SCOPUS	<ul style="list-style-type: none"> • Abrangência mundial; • Liberdade para realizar buscas elaboradas; • Dados sobre índice h; • Facilidade para descarregar e tratar os dados; • Padronização dos nomes de Instituições; • Permite relacionar nomes de autores e vínculo institucional. 	<ul style="list-style-type: none"> • Conjunto restrito de periódicos; • Baixa cobertura de artigos em língua portuguesa; • Representação de áreas desigual; • Representação de países desigual. • Não é gratuita.
Currículo Lattes	<ul style="list-style-type: none"> • Cobertura exaustiva nacional; • Possibilidade de busca pelo nome completo do autor; • Permite relacionar nomes de autores e vínculo institucional; • Informa todos os vínculos institucionais do autor durante sua trajetória acadêmica; • Permite relacionar nomes de autores e vínculo institucional; • Informa todos os vínculos institucionais do autor durante sua trajetória acadêmica; • Gratuita. 	<ul style="list-style-type: none"> • Dificuldade para importar e tratar os dados; • Dificuldade para descarregar e tratar os dados; • Variações nos registros (falta de padronização dos registros) • Dados sobre citação e Índice H nem sempre confiáveis. - Falta de padronização das informações por ser preenchida pelos próprios autores.

Fonte: Elaboração do autor.

Os registros foram analisados para verificar sua potencialidade para elaborar os indicadores previstos na pesquisa, quais sejam:

- a. Evolução cronológica das publicações;
- b. Produtividade dos pesquisadores vinculados ao Instituto e à Fundação Butantan;
- c. Cooperação científica nacional e internacional.

²⁷ Ao final desta pesquisa, a Web of Science disponibilizou para seus clientes uma versão beta que objetiva oferecer um serviço de padronização dos autores.

O indicador a) Evolução cronológica das publicações necessita de bases de dados com cobertura exaustiva em pelo menos duas áreas do conhecimento: Ciências biológicas e ciências sociais e humanas, isso porque o Instituto Butantan atua na área de Saúde pública, como também no campo da educação científica e museologia²⁸.

O indicador b) Produtividade dos pesquisadores vinculados ao Instituto e à Fundação Butantan, requer dados sobre o vínculo institucional do autor. De modo geral, na produção científica, os pesquisadores afirmam estar vinculados ao Instituto Butantan. No entanto, atualmente há, ao menos, cinco tipos de vínculos: 1) servidor público do Instituto Butantan; 2) colaborador²⁹ da Fundação Butantan; 3) estagiário da Fundação Butantan; 4) aluno de pós-graduação e 5) aluno de iniciação científica. Esse tipo de informação só é possível de ser encontrada no Currículo Lattes.

O indicador c) Cooperação científica nacional e internacional também requer dados precisos sobre os pesquisadores e sua produção. Dessa forma, em face dos problemas de cobertura e padronização identificados na WoS e Scopus, optou-se por utilizar os dados do Currículo Lattes para a análise da cooperação científica.

4.3.2 Recuperação dos dados

A escolha do Currículo Lattes para a coleta de dados implicou a execução de alguns procedimentos discutidos a seguir. O Lattes permite fazer pesquisas por autor e por assunto. Além disso, possibilita aplicar os seguintes filtros: título acadêmico (Doutor, Mestre, Graduado, Estudante, Técnicos, etc.), nacionalidade (brasileiro ou estrangeiro, com especificação do país de origem), bolsistas CNPq (de produtividade e demais níveis), campo de atuação profissional, idioma, nota do curso de pós-graduação, orientação de alunos, áreas ou setores de produção em C&T e presença no Diretório de Grupos de pesquisa.

Embora os filtros permitam realizar buscas avançadas, os modos de busca “por autor” e “por assunto” são limitadas. Isso torna necessário pesquisar o tipo de vinculação de cada pesquisador à Instituição. Além disso, a interface para buscas avançadas que permitiria, supostamente, relacionar assunto a vínculo institucional,

²⁸ O Instituto Butantan mantém quatro museus: Museu Biológico, Museu de Microbiologia, Museu Histórico e Museu Emílio Ribas.

²⁹ É o termo que a Fundação Butantan utiliza para denominar seus funcionários contratados sob o regime da CLT.

não funciona na prática. Resultam da consulta à base uma quantidade imensa de currículos que utilizaram a palavra “Butantan” em quaisquer de seus campos. Sendo assim, no caso da presente pesquisa, solicitamos ao Instituto Butantan a lista de funcionários para buscar seus currículos na Plataforma Lattes.

4.3.2.1 Mapeamento dos nomes dos servidores, colaboradores e bolsistas do Instituto e Fundação Butantan

O mapeamento dos nomes dos servidores do Instituto Butantan, bem como dos colaboradores da Fundação Butantan e bolsistas envolveu uma série de procedimentos que explicitados a seguir. Em princípio é preciso expor dois dos principais problemas enfrentados na realização desse mapeamento: 1) O Instituto Butantan não disponibiliza sua lista de funcionário por transparência ativa³⁰, prática comum nas instituições de ensino públicas paulistas; 2) O Portal de Transparência Estadual do Governo de São Paulo (<http://www.transparencia.sp.gov.br/>) disponibiliza dados sobre os servidores do Estado, porém a busca só é possível pelo nome do servidor e não pelo nome da Instituição. Além disso, no portal, não constam dados sobre os funcionários da Fundação Butantan.

A solução para os problemas citados acima foi fazer uso da Lei de Acesso à Informação (Lei nº 12.527/2011) que regulamenta o direito constitucional de acesso às informações públicas. Os nomes dos servidores, colaboradores e bolsistas do Instituto Butantan foram solicitados por meio do Portal de Transparência Estadual do Governo de São Paulo. A Instituição disponibilizou três listas intituladas:

- 1) Nomes – Fundação Butantan (FB)
- 2) Nomes – Instituto Butantan (IB)
- 3) Nomes bolsistas

A primeira lista contém 1392 nomes de funcionários da Fundação Butantan, e respectivos setores de atuação. A segunda lista, além de 650 nomes dos servidores públicos vinculados ao Instituto Butantan, disponibiliza dados sobre função e unidade administrativa à qual o servidor está alocado. A terceira lista traz 515 nomes de alunos e seus respectivos setores de atuação. Porém, ao unir as três listas em um único

³⁰ Transparência ativa é o princípio que exige de órgãos e entidades públicas a divulgação de informações de interesse geral, independentemente de terem sido solicitadas.

arquivo, deparamo-nos com 30 funcionários com vínculo com o Instituto e com a Fundação. Para consolidar a lista, recorremos ao vínculo informado no Currículo Lattes para classificá-lo em apenas uma das categorias. Além disso, 2 servidores do Instituto Butantan e 5 bolsistas estavam com seus nomes duplicados na segunda e terceira listas.

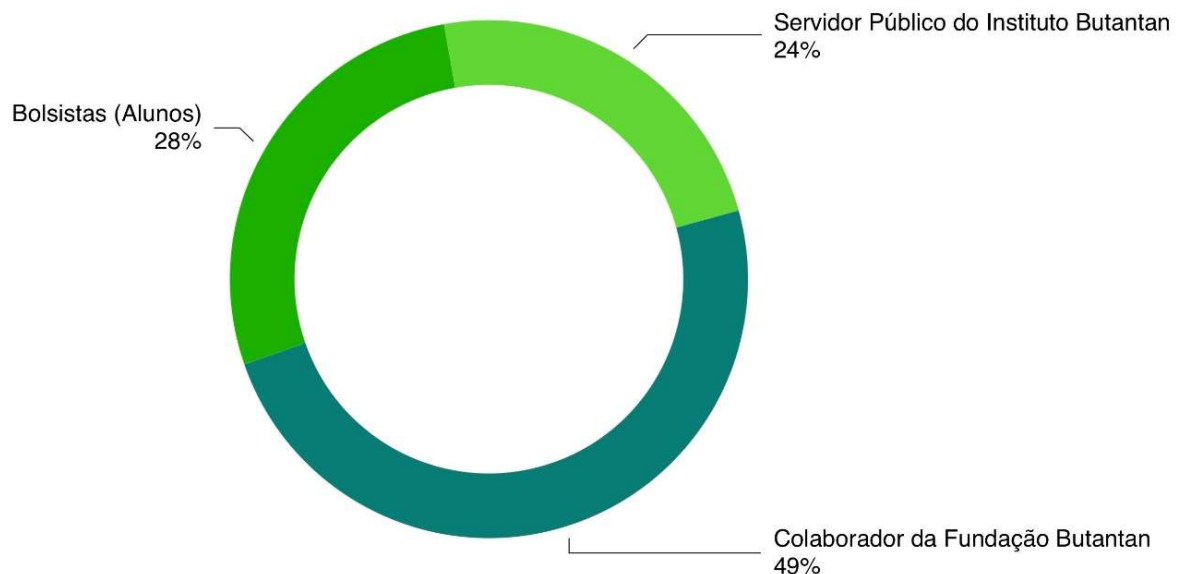
No entanto, é importante pontuar que os dados sobre os servidores, colaboradores e bolsistas fornecidos pela Fundação Butantan (FB) dizem respeito aos anos 2017/2018, porém o recorte temporal do presente estudo de caso refere-se aos anos de 2007 a 2016. Tendo em vista que o último concurso de ingresso no Instituto Butantan foi realizado em 2009, pode-se inferir que não houve muitas mudanças no quadro de servidores do Instituto. Todavia, esse raciocínio não se aplica aos bolsistas e funcionários da Fundação Butantan, pois os mesmos não possuem a estabilidade própria do funcionalismo público. Para solucionar esse problema, duas eram as alternativas possíveis: 1) Solicitar, via Lei de Acesso à Informação, as listas de servidores públicos, colaboradores e bolsistas vinculados ao Instituto e à Fundação nos anos de 2007 a 2016; 2) Fazer a primeira coleta dos artigos científicos produzidos pelos autores constantes das primeiras listas fornecidas pela Fundação e buscar, nos registros, outros autores que registraram vínculo com o Instituto nos anos de 2007 a 2016.

Optamos pela segunda alternativa por conta das dificuldades de se conseguir dados a respeito de funcionários da Fundação Butantan via Lei de Acesso à Informação. Essa decisão baseou-se na morosidade e problemas encontrados no primeiro pedido.³¹

Após os procedimentos citados, a lista consolidada resultou em 3025 nomes. O gráfico a seguir apresenta a distribuição desses nomes de acordo com o vínculo com a instituição:

³¹ A Lei de Acesso à Informação estipula o prazo de 20 dias para a Instituição responder a um pedido de informação. Esse prazo pode ser estendido, desde que justificado, bem como a Instituição pode se recusar a dar a informação com base nas restrições inscritas na Lei de Acesso. O autor do pedido pode apresentar até dois recursos, caso o pedido seja negado, porém isso implica espera de mais 10 a 20 dias para obter uma resposta, lembrando que esse prazo pode ser mais uma vez prorrogado desde que justificado. No primeiro pedido realizado pela presente autora, a Fundação Butantan negou acesso aos nomes dos funcionários da Fundação sob o argumento de que não se tratavam de funcionários públicos. A resposta levou a autora a elaborar recurso explicando o caráter público das Fundações de apoio a órgãos públicos e o recurso foi deferido. No entanto, todo o processo levou quase dois meses, além disso soma-se o tempo que a autora precisou para pesquisar e elaborar o recurso.

Ilustração 9 - Gráfico de distribuição entre servidores públicos do Instituto Butantan, colaboradores da Fundação Butantan e bolsistas após segundo mapeamento de autores.



Fonte: Elaboração do autor.

A etapa seguinte envolveu o mapeamento dos currículos Lattes dos servidores, funcionários e bolsistas. Foram recuperados 740 currículos, cujos endereços estão disponibilizados na amostra em anexo.³²

4.3.2.2 Mapeamento dos artigos publicados pelos servidores, colaboradores e bolsistas do Instituto e Fundação Butantan

a) Coleta:

Após a recuperação dos Currículos Lattes, a etapa seguinte objetivou analisar o vínculo que o pesquisador possuía com o Instituto Butantan, bem como coletar os

³² Anexo I - Lista de autores.

registros dos artigos completos publicados em periódicos no anos de 2007 a 2016. O procedimento, relativamente simples, consiste no acesso ao Currículo Lattes do pesquisador e seleção dos dados referentes a produção dos anos de 2007 a 2016 no campo “Artigos Completos Publicados em Periódicos”. Esta etapa resultou na coleta de 10.074 registros.

b) Deduplicação:

A fim de evitar duplicação na contagem dos artigos, as duplicatas foram identificadas por meio de casamento dos títulos registrados nos artigos. Importante pontuar que essa etapa é bastante trabalhosa em se tratando de dados coletados na Plataforma Lattes, isso porque os autores de um mesmo artigo utilizam inúmeras variações no título. É comum, por exemplo, a utilização de letras gregas e das suas variações nos idiomas inglês e português. Há registros em que o pesquisador informa apenas parte do título. Essas práticas acabaram por inviabilizar a utilização de softwares de deduplicação³³ como única estratégia para eliminação de duplicatas e exigiu uma segunda limpeza manual a fim de se obter a quantidade real de artigos publicados pelos pesquisadores componentes no conjunto de dados.

c) Análise da amostra:

Ao final da deduplicação a amostra ficou composta por 4.585 registros que foram migrados para o editor de planilhas Microsoft Office Excel, para processamento. Cada registro foi segmentado nas seguintes partes: título, autores, ano e título do periódico. Os resumos e afiliação dos autores foram também relevantes para a execução do estudo. Porém, a tarefa de verificar a afiliação dos autores levantou um problema: a produção científica de um determinado pesquisador não está necessariamente vinculada à instituição em que ele está alocado. Dito de outro modo e a título de exemplo, o pesquisador Jorge Elias Kalil Filho atuou no Instituto Butantan nos anos de 2011 a 2017 e em seu Lattes ele informa ter publicado 115 artigos nesse período. No entanto, Kalil também exercia a função de docente da Universidade de

³³ Deduplicação ou Deduplication é o processo que identifica e remove duplicidade nos dados.

São Paulo e o acesso aos artigos publicados permitiu averiguar que em apenas 16 artigos o pesquisador afirmava possuir vínculo com o Instituto Butantan. Sendo assim, a estratégia adotada foi compor a amostra apenas de artigos em que pelo menos um dos autores afirma vínculo com o Instituto Butantan. Essa solução se justifica porque a produção científica de um determinado autor não contou, necessariamente, com o apoio da Instituição que ele está vinculado de modo que cabe a ele informar se o trabalho possui ou não vínculo com a instituição.

A princípio a busca por informações sobre os vínculos dos autores foi realizada nos registros disponibilizados pelas bases de dados Web of Science, Scopus, e PubMed além da rede social Researchgate, porém foram observadas nessas plataformas, divergências entre os dados do registro (metadados) e os dados bibliográficos. Sendo assim, foi necessário consultar os trabalhos originais na web. Os trabalhos publicados em periódicos de livre acesso foram coletados nos sites das revistas e os trabalhos de acesso restrito foram coletados na Plataforma Sci-hub. Não foram encontrados 76 artigos de uma amostra de 4.585. Foram eliminados 2.167 trabalhos, pois os autores não informavam o vínculo com o Instituto Butantan. O resultado desse trabalho, disponibilizado no arquivo anexo³⁴, é constituído de 2342 registros bibliográficos. A Tabela 8 apresenta os dados gerais, de modo a permitir a análise do volume de informações consolidadas, de acordo com o ano de publicação.

Tabela 8 - Número total de artigos recuperados, por período analisado

Período	Número de artigos
2007	196
2008	243
2009	217
2010	236
2011	241
2012	263
2013	252
2014	247
2015	223

³⁴ Anexo II: Lista de Registros Bibliográficos.

2016	224
Total	2342

Fonte: Elaboração do autor.

5 RESULTADOS

A análise dos resultados compreende a discussão dos seguintes aspectos: a) utilidade do indicador para a avaliação do instituto de pesquisa; b) apresentação do contexto em que o indicador está inserido e as primeiras interpretações do indicador; c) explicação para o(s) evento(s) e ocorrência(s) relacionando-os com outras fontes de informações e; d) análise das implicações.

5.1 Evolução cronológica das publicações

O indicador de evolução temporal permite identificar a tendência de produtividade científica do Instituto Butantan. Para isso, construiu-se uma matriz com as frequências relativas e acumuladas dos artigos publicados por ano, no período 2007-2016, coletados na Plataforma Lattes.

Tabela 9 – Frequência relativa e frequência relativa acumulada dos artigos publicados pelo Instituto Butantan nos anos de 2007 a 2016

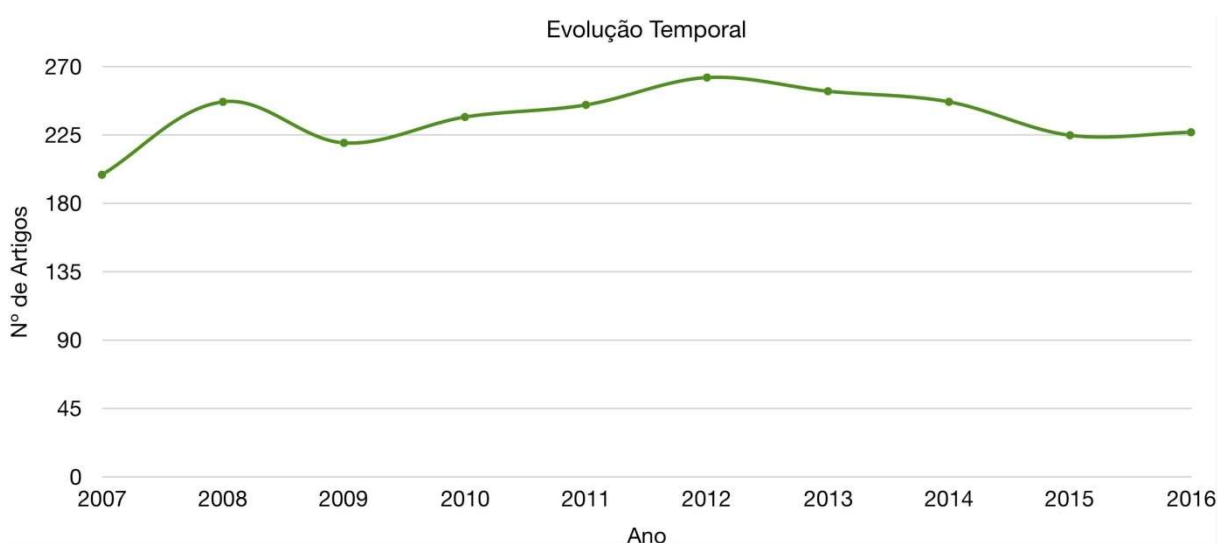
Ano	Número de artigos	Frequência relativa (%)	Frequência relativa acumulada
2007	196	8,3%	8,3
2008	243	10,3%	18,7
2009	217	9,2%	28,0
2010	236	10,0%	38,1
2011	241	10,2%	48,3
2012	263	11,2%	59,6
2013	252	10,7%	70,3
2014	247	10,5%	80,9
2015	223	9,5%	90,4
2016	224	9,5%	100
TOTAL	2342	100%	

Fonte: Elaborado pelo autor.

A tabela acima mostra que o número de artigos dos anos de 2007 a 2016 teve aumento de 14,2% na base de dados Currículo Lattes. Em 2007, foram contabilizados

196 artigos e 224 artigos em 2016. Os dados de frequência apontam a estagnação da produção, considerando que em dez anos a amostra manteve frequência relativa anual próxima a 10%. A Ilustração 10 dá maior visibilidade ao cenário referido.

Ilustração 10 - Evolução do número de artigos publicados pelo Instituto Butantan nos anos de 2007 a 2016



Fonte: Elaboração do autor.

No gráfico acima, é possível notar que houve a retomada da produção científica no ano de 2008, após uma grave crise institucional, que ocorreu entre 2006 e 2007. A instituição foi investigada pelo Tribunal de Contas do Estado de São Paulo (TCE-SP) por suspeita de desvios de verbas públicas. Depois de uma leve queda em 2009, a produção do instituto apresentou um pequeno crescimento entre os anos de 2010 e 2013, seguido de um declínio a partir de 2014. Este aumento pode ter ocorrido em razão do início das atividades do Programa de Pós-Graduação em Toxinologia (PPGTox), dos estudos sobre a vacina da dengue, e dos 38 (trinta e oito) artigos referentes ao vírus *influenza* (H1N1), em decorrência da epidemia de gripe ocorrida nesse período. Em 2016, a produção se estabiliza, devido aos estudos envolvendo o zika vírus, também em face de um surto ocorrido na região Sudeste.

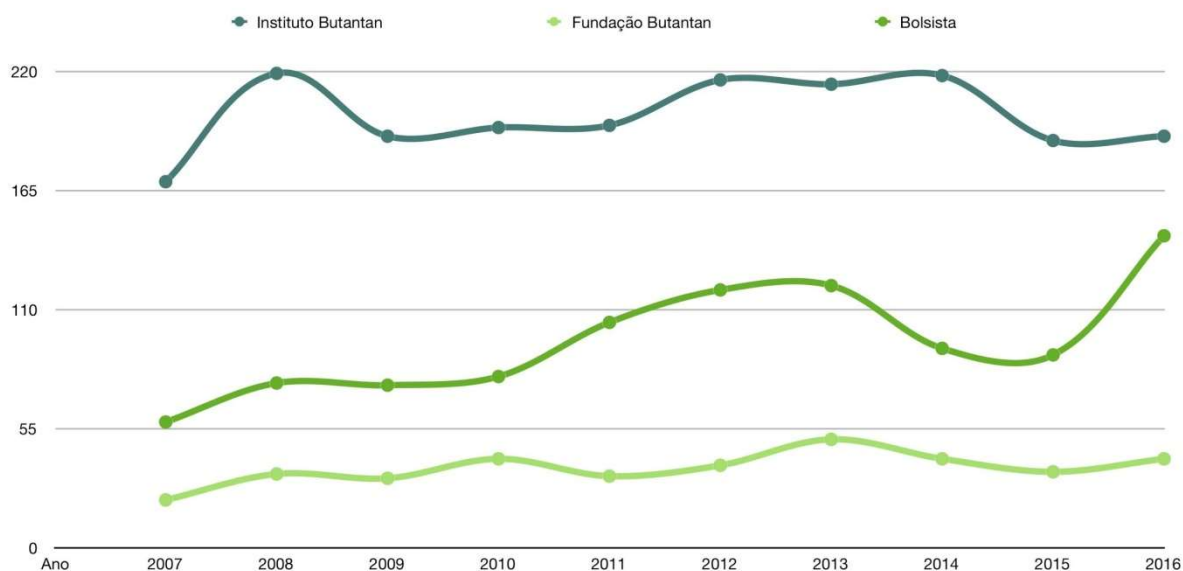
Como foi observado no item 4.2 de contextualização do Instituto Butantan, trata-se de uma característica comum às instituições de pesquisa da América Latina,

o imediatismo com relação às políticas públicas de saúde. Essa prática faz com que os recursos financeiros sejam constantemente realocados para áreas prioritizadas em determinado período, para responder a demandas pontuais de saúde. Assim, ficam prejudicados os programas e políticas de longo prazo. Essa lógica, que caracteriza o Instituto Butantan, desde a sua fundação até o presente, portanto, ao longo de mais de um século, pode explicar as oscilações na produção científica mostrados na Ilustração 10.

Esses indicadores permitem inferir que o Instituto Butantan não apresentou tendência de crescimento no período. As oscilações presentes na Ilustração 10 são de certa forma, o retrato da instabilidade das políticas públicas brasileiras de apoio à pesquisa científica.

A Ilustração 11 ilustra a estratificação da produção científica por grupos de autores, classificados segundo o critério de tipo de vínculo do autor com o Instituto Butantan.

Ilustração 11 - Evolução do número de artigos publicados pelo Instituto Butantan nos anos de 2007 a 2016 por tipo de vínculo na instituição



Fonte: Elaboração do autor.

Segundo o gráfico acima, em 2015 e 2016 houve importante queda na produção de artigos por parte dos servidores do Instituto Butantan. Porém, cresceu a

produção dos bolsistas, fato que compensou a queda de produção dos servidores. Uma das hipóteses para explicar o aumento significativo da produção dos bolsistas é a avaliação dos programas de pós-graduação do quadriênio 2013-2016, pela Capes, fato que pode ter estimulado o aumento da produção científica de alunos de pós-graduação. Esta hipótese pode explicar também o aumento verificado em 2012, considerando que nesse ano ocorreu a avaliação trienal da Capes.

Por outro lado, os dados coletados, não permitem inferir quais fatores foram responsáveis pelo aumento total do número de artigos publicados pelos bolsistas. Isso porque, conforme já exposto, houve dificuldade de acesso aos nomes dos bolsistas do Instituto Butantan nos anos de 2007 a 2016. Quanto à queda na produção dos servidores públicos do Instituto, suas causas serão melhor exploradas no próximo indicador.

5.2 Produtividade dos pesquisadores

A Tabela 10 refere-se ao vínculo dos autores mais produtivos do Instituto Butantan. A avaliação de autoria pode ser feita com base em dois critérios: a) número de publicações e b) número de citações. Em face das limitações da Plataforma Lattes no que concerne às citações, o presente estudo apresenta dados de publicações com base na estratificação dos pesquisadores por vínculo institucional com o Instituto Butantan: servidores públicos vinculados ao Instituto Butantan, colaboradores da Fundação Butantan e bolsistas.

Após a normalização dos nomes dos autores para depurar a lista de modo a evitar ruídos provocados por homônimos, erros ortográficos e outras variantes, a amostra de artigos foi classificada conforme o tipo de vínculo do autor ou autores com o Instituto Butantan. A classificação foi feita através do cruzamento dos dados fornecidos pela Fundação Butantan com os dados dos autores registrados na Plataforma Lattes.

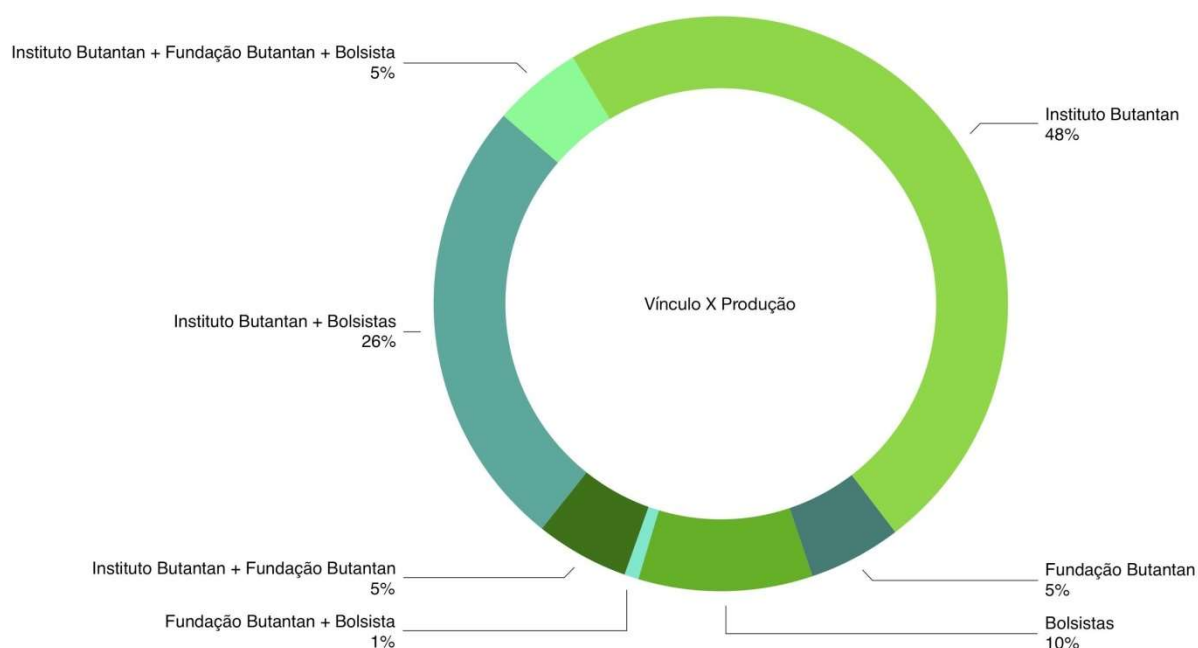
Tabela 10 - Vínculo institucional por número de artigos nos anos 2007-2016

Vínculo	Número de artigos
Instituto Butantan	1081

Instituto Butantan e Fundação Butantan	118
Instituto Butantan e Bolsistas	575
Instituto Butantan, Fundação Butantan e Bolsistas	113
Fundação Butantan	117
Fundação Butantan e Bolsistas	18
Bolsistas	220

Fonte: Elaborado pelo autor.

Ilustração 12 - Artigos publicados nos anos de 2007 a 2016 por vínculo institucional do pesquisador no Instituto Butantan



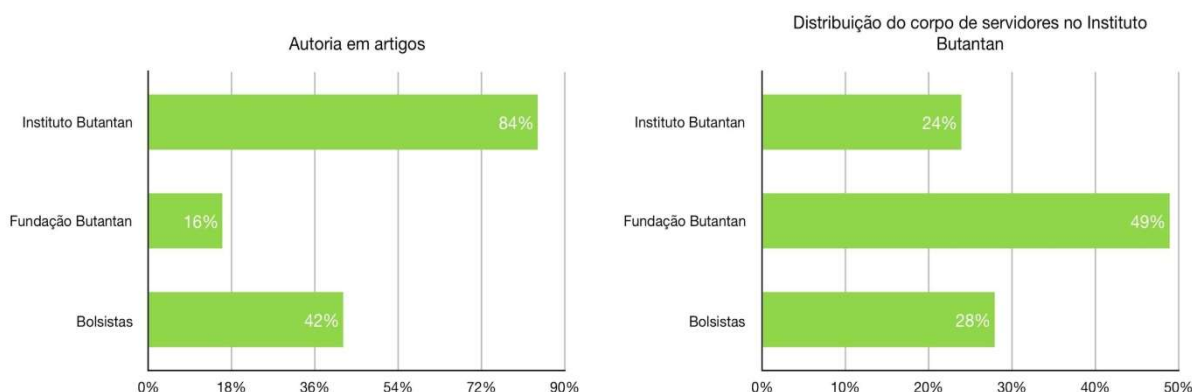
Fonte: Elaboração do autor.

O Gráfico acima foi construído respeitando as práticas de coautoria entre os autores do Instituto Butantan. Os autores foram estratificados em sete categorias: 1) Artigos publicados por servidores públicos do Instituto Butantan e colaboradores da Fundação Butantan; 2) Artigos publicados por servidores públicos do Instituto Butantan e bolsistas; 3) Artigos publicados por servidores públicos do Instituto Butantan, colaboradores da Fundação Butantan e bolsistas; 4) Artigos publicados por

colaboradores da Fundação Butantan e bolsistas; 5) Artigos publicados somente por servidores públicos do Instituto Butantan; 6) Artigos publicados somente por colaboradores da Fundação Butantan e 7) Artigos publicados somente por bolsistas.

A Tabela 10 e a Ilustração 12, acima, mostram que a maioria das publicações é de autoria dos servidores públicos do Instituto Butantan. O gráfico revela que, somados com artigos em coautoria, os servidores participam de 84% dos artigos publicados pelo Instituto nos anos de 2007 a 2016. Deve-se assinalar que os servidores correspondem a 24% dos funcionários do Instituto conforme apresentado no item 4.3.2.1. Trata-se de um dado ainda mais significativo se comparado aos dados de colaboradores da Fundação Butantan, que participam de 16% dos artigos e representam 49% dos funcionários do instituto. Essa diferença pode ser melhor visualizada nos gráficos a seguir:

Ilustrações 13 e 14 - Autoria em artigos e distribuição do corpo de servidores no Instituto Butantan



Fonte: Elaboração do autor.

O indicador da Tabela 11 refere-se à média de artigos publicados. A média foi de aproximadamente 15,61 artigos por pesquisador, considerando que foram mapeados 2342 artigos em 350 Currículos Lattes de autores com o título de doutor. Este valor está acima da média de produção da área de Ciências Biológicas, conforme estudo realizado por Mugnaini et al. (2019). Os autores analisaram a dispersão da produção científica dos 260.663 pesquisadores, do nível de doutorado, registrados na Plataforma Lattes, cuja amostra é composta por artigos completos publicados em

periódicos, entre os anos de 1998 e 2016. Nesse período, o valor médio da área de Ciências Biológicas era de 10,78 artigos por pesquisador.

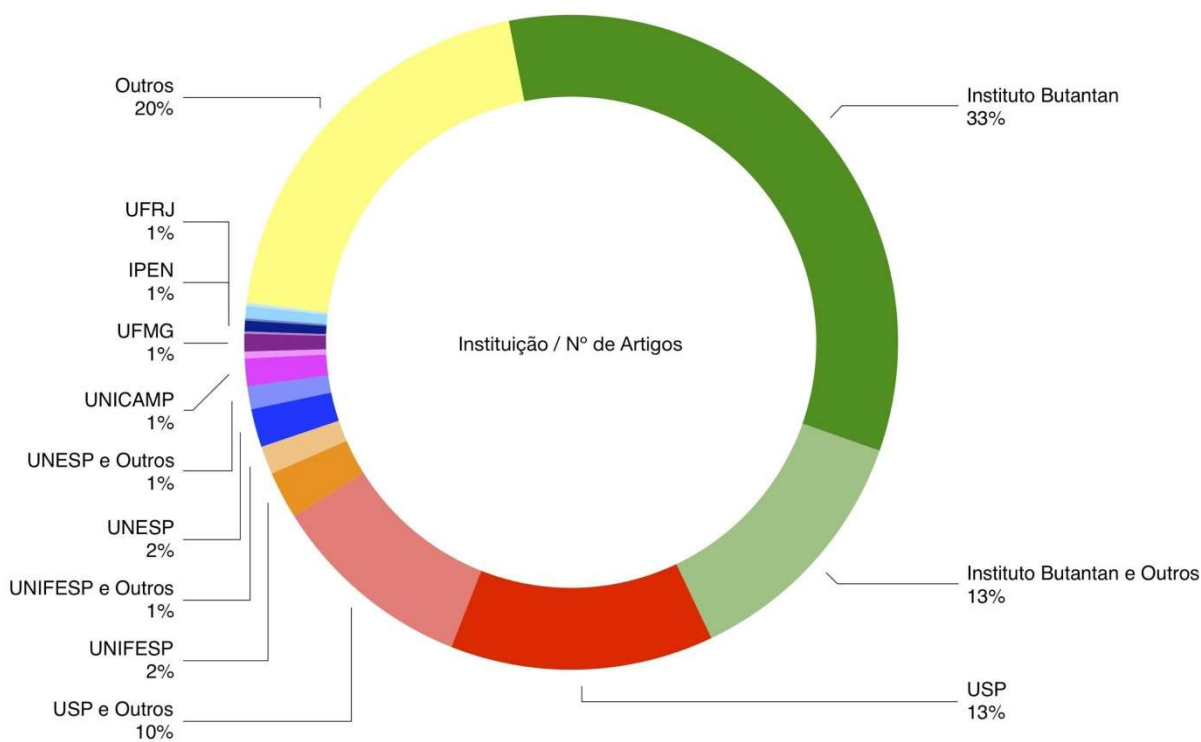
Tabela 11 - Relação número de artigos por pesquisador

Instituição	Número de Pesquisadores	Número de artigos	Artigos / Pesquisador
Instituto Butantan	350	2342	15,61

Fonte: Elaborado pelo autor.

Um outro indicador interessante para avaliar a produtividade dos autores do Instituto Butantan, pode ser obtido pela análise da primeira autoria. O Instituto tem como prática a menção do primeiro autor como o autor que deu a maior contribuição para a realização da pesquisa e também da redação do manuscrito. O vínculo dos primeiros autores da amostra analisada pode ser visualizado no gráfico a seguir:

Ilustração 15 - Afiliação dos primeiros-autores na amostra

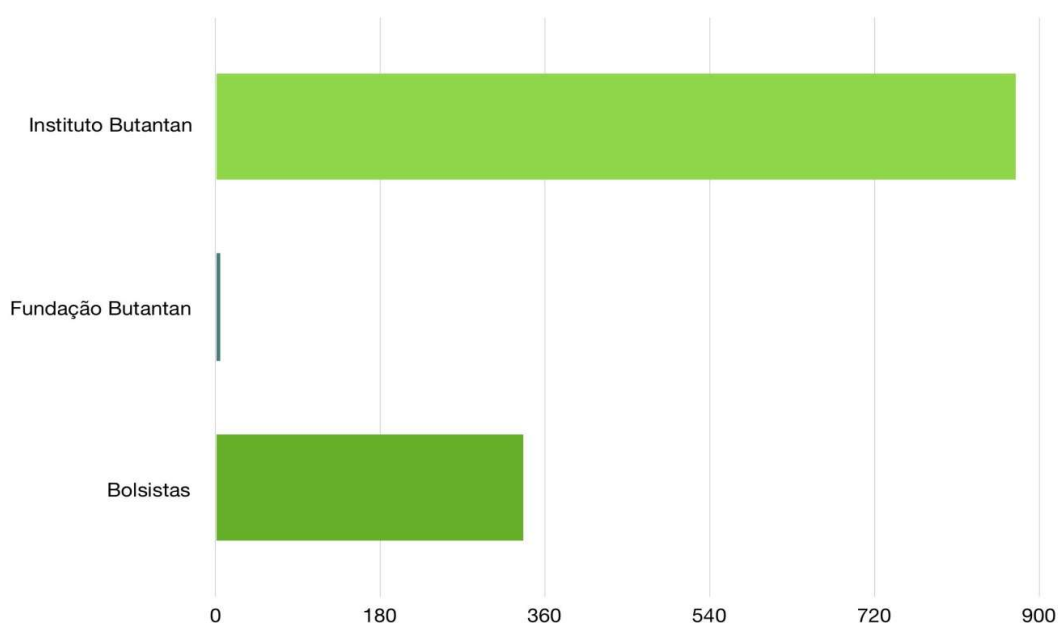


Fonte: Elaboração do autor.

O termo “Outros” foi utilizado para designar outros vínculos que um mesmo autor informa. Assim, o grupo “Instituto Butantan e outros” diz respeito aos autores que informaram ser do Instituto Butantan e também de outra instituição. A Ilustração 15 acima chama a atenção para o protagonismo do Instituto Butantan em sua área de atuação: 46% de sua produção científica tem como primeiro autor membros da instituição. O indicador também aponta para os principais parceiros do Instituto Butantan, instituições de ensino e pesquisa altamente conceituadas, com destaque para a Universidade de São Paulo (23%), dado que pode ser explicado pela localização espacial, a Universidade de São Paulo e o Instituto Butantan são vizinhos. Além disso, boa parte dos pesquisadores do Instituto Butantan fazem suas graduações e/ou pós na USP, de forma que suas conexões profissionais são inicialmente construídas dentro da Universidade.

Há outros dados preocupantes sobre a situação do Instituto Butantan, como por exemplo, a distribuição da primeira autoria no Instituto. A próxima Ilustração retrata essa realidade.

Ilustração 16 - Distribuição da Primeira Autoria de acordo com Vínculo Institucional



Fonte: Elaboração do autor.

O gráfico acima apresenta enorme desigualdade na primeira autoria da produção científica do instituto, sendo que os servidores e bolsistas responsáveis por quase a totalidade da produção. No item 3.3.1 Produtividade de autores: a Lei de Lotka, apresentamos a forma de calcular produtividade de autores. Para relembrar: número de autores que fazem x contribuições em um determinado campo científico é de aproximadamente $\frac{1}{x^2}$ daqueles que fazem apenas uma contribuição (LOTKA, 1926). A relação de desigualdade na produção de trabalhos científicos por pesquisadores do Instituto Butantan pode ser modelada com a Lei de Lotka. No entanto, conforme discutido anteriormente, ao se postular que a desigualdade no campo da pesquisa científica é uma lei, acaba-se fomentando a normalização da desigualdade no campo científico. No caso do Instituto Butantan, pode-se afirmar que não se trata de uma lei social e sim o resultado de políticas mal formuladas. O Instituto não realiza concursos para provimento de cargos desde o ano de 2009, a Fundação não supre a necessidade de recursos humanos altamente especializados e parece não oferecer oportunidades que atraiam novos pesquisadores. A título de exemplo, o autor vinculado a Fundação Butantan e que possui o maior número de artigos publicados na presente amostra, Vidal Haddad Junior, deixou o Instituto em 2014 e integra, atualmente, o corpo de servidores da UNESP.

Um outro dado importante: os servidores públicos aposentados que continuam pesquisando e publicando representam 36% dos servidores do Instituto Butantan.

5.3 Colaboração Internacional

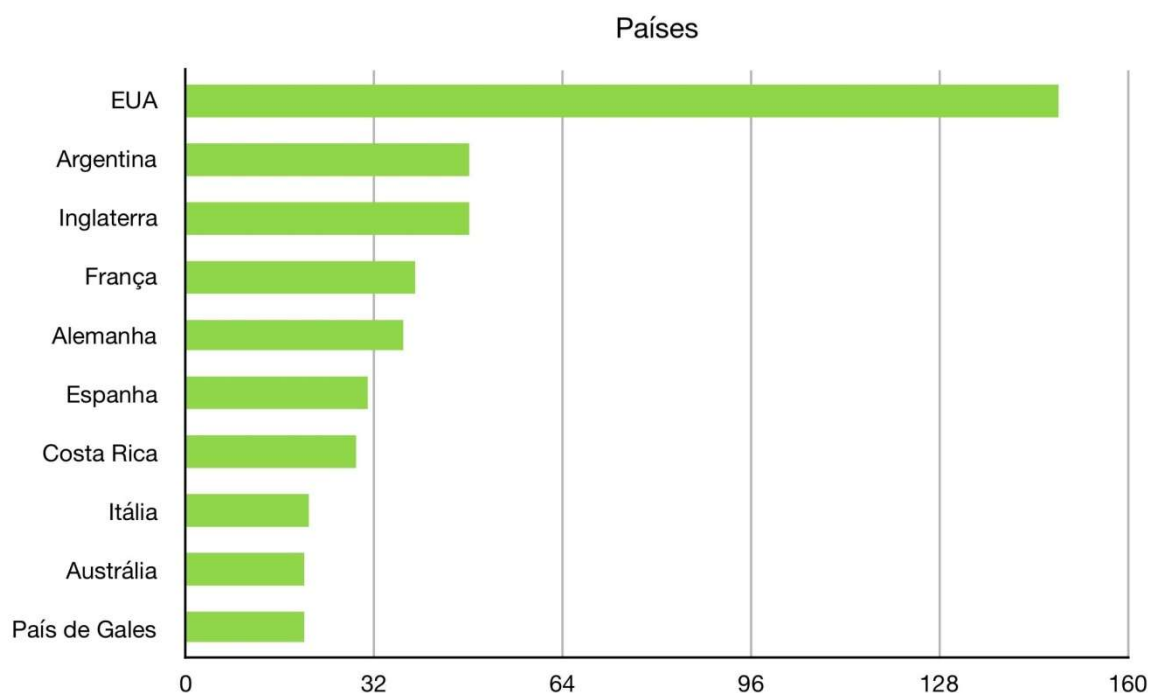
A colaboração internacional ou internacionalização é um importante critério utilizado para avaliar a pesquisa científica. A premissa subjacente é a de que a internacionalização contribui para a melhoria da qualidade das pesquisas e para um maior impacto e visibilidade internacional, fatos que contam pontos na avaliação nos programas de pós-graduação. Tal colaboração é estimulada pelas universidades e órgãos de fomento por ser um dos critérios utilizados para classificar universidades nos rankings internacionais. Por esse motivo, as agências de fomento nacionais vêm investindo na intensificação de pesquisas colaborativas. A FAPESP, por exemplo, destinou R\$ 180 milhões, em 2018, para apoiar pesquisas em colaboração

internacional (ZIEGLER, 2018). O indicador comumente utilizado para avaliar a cooperação internacional é a coautoria em artigos publicados em periódicos de prestígio internacional.

A elaboração desse tipo de indicador implica a normalização dos nomes das instituições internacionais, tarefa laboriosa pois uma instituição pode ser referida de diferentes formas. Além disso, muitos artigos fazem menção a um departamento de uma instituição, fato que torna difícil identificar a qual organização pertence o departamento. O indicador de colaboração foi elaborado após a normalização dos nomes das instituições com base em uma lista de instituições criada para a normalização dos dados.

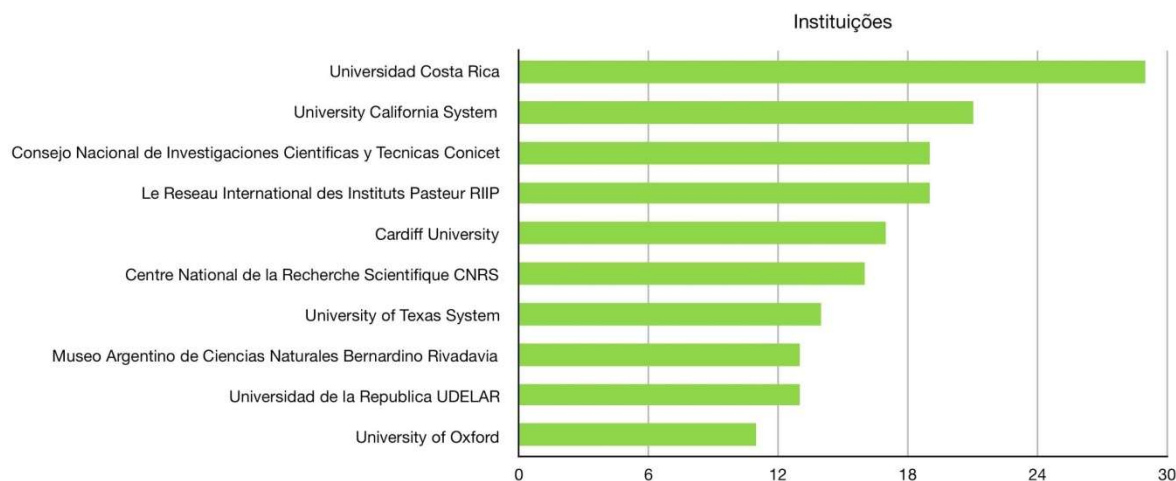
Os gráficos apresentam os principais países que cooperaram com o Instituto Butantan, registrados nos artigos dos anos de 2007 a 2016.

Ilustração 17 - Principais países colaboradores nos artigos publicados pelo IB



Fonte: Elaboração do autor.

Ilustração 18 - Principais instituições do exterior que colaboraram nos artigos publicados pelo Instituto Butantan

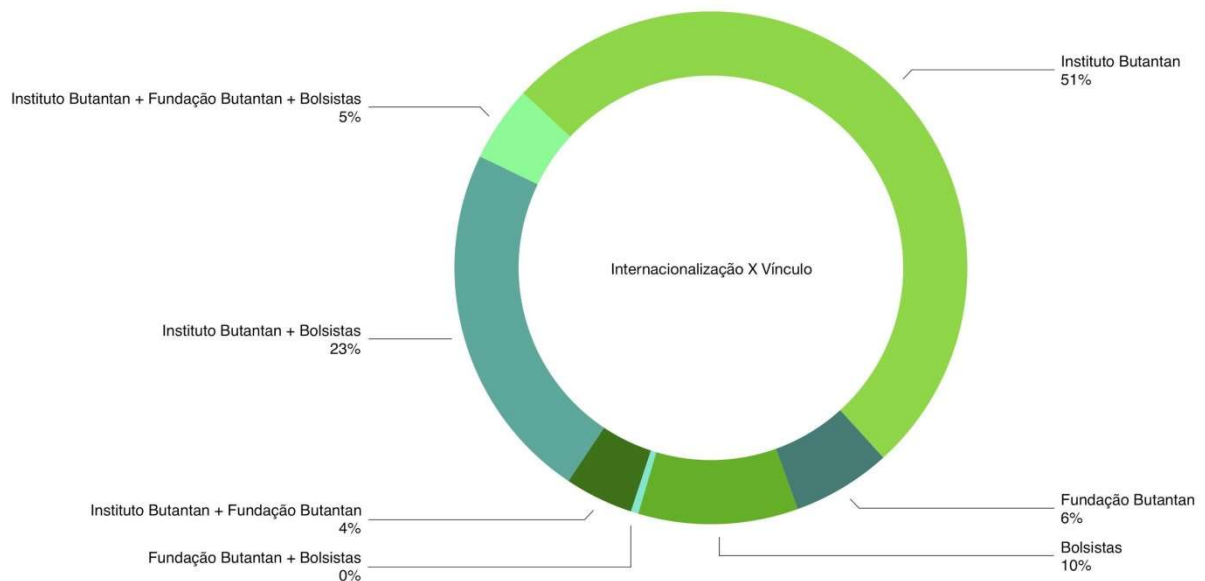


Fonte: Elaboração do autor.

Foram produzidos 417 artigos em colaboração internacional, que corresponde a 17,8% do total da amostra. Esse valor é baixo se utilizarmos como referencial o dado da FAPESP (ZIEGLER, 2018). De acordo com essa agência de fomento, a taxa de artigos resultantes de colaboração internacional, financiados pela FAPESP, foi de 40%. No entanto, uma hipótese para explicar esse dado está no fato de o Instituto Butantan ser pioneiro em pesquisas em um campo bastante específico: a produção de imunobiológicos fazendo uso de substâncias produzidas pela fauna de países tropicais. Essa realidade pode limitar a coautoria com determinados países e instituições.

Ademais, um indicador mais preocupante pode ser observado ao se analisar a distribuição de artigos produzidos em cooperação internacional por pesquisadores do Instituto Butantan.

Ilustração 19 - Vínculo do Pesquisador em relação à internacionalização da pesquisa



Fonte: Elaboração do autor.

A Ilustração acima mostra que 83% dos artigos em cooperação internacional foram produzidos por servidores do Instituto. Como se pode notar através dos indicadores apresentados, a produção científica do Instituto Butantan é dependente de um grupo específico, os servidores públicos da Instituição. Além disso, os pesquisadores bolsistas, são completamente dependentes do primeiro grupo, uma vez que a produtividade da instituição define o número de bolsas concedidas. Muito embora não se trate de uma instituição voltada exclusivamente para o ensino, a produção científica se assenta na relação dos pesquisadores com os respectivos bolsistas. Contudo, a saída de pesquisadores produtivos, o decréscimo da produtividade de pesquisadores já aposentados, somadas à não contratação de novos quadros nesta área, decorre de políticas deliberadas de sucateamento do Instituto e denotam a precarização da atividade de pesquisa na instituição.

A moldura analítica aventada para o presente estudo de caso propôs a apresentação de indicadores que auxiliassem na resposta para a seguinte pergunta: as crises institucionais e o padrão de atuação do Instituto que ora privilegia a pesquisa e ora privilegia a produção de imunobiológicos, afetaram a produção científica do Instituto Butantan? Com base nos resultados expostos, é possível inferir que as crises institucionais e a instabilidade nas políticas científicas do Instituto afetam sua produção científica gerando um quadro interno de desigualdade entre grupos de

pesquisadores, bem como uma produção científica vulnerável a crises e surtos de saúde sazonais.

6 CONSIDERAÇÕES FINAIS

Esta pesquisa exploratória sobre a produção de indicadores bibliométricos, foi realizada com base nos seguintes procedimentos: a) construção do referencial teórico, fundamentado nos estudos históricos, sociais e filosóficos da ciência; b) sistematização da literatura sobre a Bibliometria e seus usos na avaliação da atividade científica; c) realização de um estudo de caso de análise da produção científica do Instituto Butantan por técnicas bibliométricas.

Dois pressupostos sustentaram este trabalho: 1) a atividade científica é uma atividade social e 2) a estatística tem importância central nos métodos bibliométricos. Deve-se admitir que esses pressupostos nem sempre estão claros nas pesquisas de avaliação da atividade científica. Com efeito, estudos bibliométricos cujos dados são coletados e trabalhados de forma mecânica, têm sido criticados por serem pouco úteis, ou mesmo prejudiciais para as finalidades com que são desenvolvidos. Os estudos ingênuos, denominados *Desktop bibliometrics* (MOED, 2018), são hoje fontes de preocupação de pesquisadores renomados da área pelos danos que podem causar à formulação de políticas de ciência.

Nesta pesquisa, elaborada em perspectiva crítica, foram sistematizadas inicialmente, as ideias e eventos que contribuíram para a incorporação da estatística aos métodos das Ciências Sociais e da Ciência da Informação. Este percurso foi fundamental para identificar os aspectos problemáticos, desafiadores, da aplicação da estatística em estudos sobre os fatos sociais.

A pesquisa bibliográfica sobre as abordagens propostas pela Filosofia da Ciência e Sociologia da Ciência permitiu identificar como, quando e porque os fenômenos sociais começaram a ser quantificados para subsidiar as políticas de Estado. As obras do filósofo Ian Hacking permitiram reconstruir o percurso histórico do processo quantificador, em especial do século XVIII até o século XX, quando a estatística adquire *status* de ferramenta estratégica do Estado. Ian Hacking trata justamente do desenvolvimento da quantificação e da matemática aplicadas aos estudos e interpretação de fenômenos populacionais, por ele caracterizado como processos de “domesticação do acaso”, isto é, quando se busca descobrir regularidades e criar leis sobre eventos fortuitos.

Muitos eventos contribuíram para a consolidação das leis estatísticas, no entanto, segundo Hacking, há um divisor de águas: em 1844, o astrônomo belga Quetelet (1796-1874) aplicou a teoria de mensuração de grandezas físicas para medir as propriedades ideais ou abstratas de uma população. Este foi um passo crucial para a domesticação do acaso: Quetelet transformou a descrição de regularidades de grande escala em leis da sociedade.

No entanto, o otimismo relativo aos usos da estatística foi abalado em 1848. A restauração do acaso foi fomentada por Nietzsche (1844-1900) e principalmente por Peirce (1839-1914). Para Peirce, o acaso pode ser inferido pela combinação da lógica indutiva e da probabilidade, não sendo, porém, algo que possa ser controlado e modelado como leis. Essas ideias, que contribuíram para o desenvolvimento da estatística moderna, foram importantes para discutir seus impactos nos estudos quantitativos, em especial na aplicação das leis bibliométricas de Lotka (1926), Zipf (1935; 1949) e Bradford (1985).

As reflexões de Hacking foram também fundamentais porque permitiram desenvolver raciocínios críticos sobre a Lei de Lotka, que propõe um padrão internacional relativo à produtividade de pesquisadores. Essa lei postula que a maior parte produção científica mundial é de responsabilidade de uma “elite” de pesquisadores.

A lei de Lotka se configura como concepção vinculada ao movimento estatístico que procurou descobrir estabilidades nos fenômenos de massa. Portanto, ao generalizar o comportamento de cientistas e pesquisadores, Lotka procurou reproduzir Quetelet (item 2.3) que aplicou a curva de distribuição normal à antropometria de soldados. Há inclusive semelhanças entre as críticas à Lei de Lotka e as dirigidas a Quetelet. Em ambos os casos, quando os resultados não se encaixam na lei ou na distribuição esperada, os problemas são atribuídos a falhas na medição, à inadequação da amostra, ou na inconsistência na modelagem dos dados de pesquisa.

Com base no referencial teórico proposto pela Filosofia da Ciência, foi possível, ainda, identificar as semelhanças entre a estatística do século XIX e as leis bibliométricas. A Lei de Lotka, por exemplo, ao tentar enquadrar o comportamento de pesquisadores à lei do quadrado inverso, ignora as particularidades culturais, sociais e econômicas do processo de criação científica. A análise criteriosa das variáveis que

afetam o comportamento de publicação dos autores, não pode ser ignorada. A Lei de Lotka é parte, portanto, da uniformização de fenômenos sociais a fim de possibilitar sua generalização e expressão na forma de leis estatísticas, concepção característica da estatística do século XIX e das leis bibliométricas.

Uma das hipóteses levantadas na pesquisa é a de que a aproximação dos estudos da lógica da inferência estatística, realizada no campo da Filosofia da Ciência, poderia contribuir para discutir a Bibliometria e a Cientometria, precisamente porque estas últimas usam a estatística para produzir indicadores sociais. As concepções sobre o acaso e a necessidade, de Charles Peirce, foram fundamentais para construir a estatística contemporânea. Peirce foi responsável, também pela aproximação da lógica indutiva das teorias probabilísticas para a análise de argumentos do campo da Estatística.

As ideias de Hacking e Peirce autorizam a construção da hipótese de utilização dos métodos da lógica indutiva em estudos bibliométricos. Tal concepção pode ser um caminho para pavimentar a realização de estudos bibliométricos consistentes, uma hipótese a ser testada em pesquisas futuras.

Após o percurso de construção do referencial teórico e metodológico, foi planejado e executado um estudo de caso. Nesse estudo, procurou-se analisar, por meio de métodos estatísticos descritivos e bibliométricos, um conjunto selecionado de artigos publicados por pesquisadores do Instituto Butantan. Os dados foram coletados na plataforma de Currículos Lattes. Os problemas e dificuldades encontrados no estudo de caso permitem inferir que a escolha da fonte de dados, a constituição do *corpus* e a reflexão permanente no processo de aplicação dos métodos e técnicas bibliométricas, são essenciais para assegurar a produção de indicadores consistentes. Os princípios norteadores da pesquisa como um todo, como também do estudo de caso, foram inspirados no Manifesto de Leiden (HICKS et al, 2015). Nessa perspectiva, a pesquisa que realizamos procurou:

- 1) realizar a análise quantitativa para dar suporte à avaliação qualitativa;
- 2) produzir indicadores de desempenho de instituições respeitando a sua missão, qual seja, as particularidades próprias de um instituto de saúde pública latino-americano;
- 3) tratar com respeito a pesquisa realizada por um instituto conceituado nacional e internacionalmente;

- 4) manter a coleta de dados e os processos analíticos abertos, transparentes, anexando a amostra da pesquisa à dissertação. Isso permitirá confrontar os resultados com o corpus utilizado;
- 5) considerar as diferenças das práticas de publicação e citação das áreas de conhecimento;
- 6) basear a avaliação de pesquisadores individuais nos dados da carreira do pesquisador;
- 7) evitar ao máximo as falsas precisões, analisando cuidadosamente os dados de partida para compor a amostra;
- 8) reconhecer os efeitos e impactos sistêmicos da avaliação e dos indicadores.

O décimo princípio, que trata do exame e atualização regular dos indicadores, será realizado à medida em que novos dados estiverem disponíveis. Deve-se ressaltar que pautar o estudo de caso nos princípios do Manifesto de Leiden permitiu compreender também as limitações de um estudo bibliométrico.

Quanto aos resultados do estudo de caso, é importante destacar que o Instituto Butantan possui importante produção científica na área da Saúde pública. No entanto, apesar de um Repositório digital ter sido criado pela Biblioteca do Instituto Butantan, há pouco mais de três anos, não foram feitos, até o momento, os procedimentos para disponibilizar publicamente esse acervo. Como recomendado na literatura crítica da área, a construção e disponibilização pública das pesquisas em um repositório institucional de acesso livre é fundamental para produzir indicadores consistentes de produção científica, já que as coleções das bases de dados do exterior não representam, na exaustividade desejável, a pesquisa produzida em países periféricos.

Outro aspecto revelado no processo de pesquisa se relaciona ao contexto político e econômico em que estão inseridas as instituições de ensino e pesquisa. A capacidade de produzir novos conhecimentos está diretamente relacionada ao financiamento público da pesquisa. As instituições brasileiras, longamente submetidas a processos de sucateamento, ou reféns de políticas científicas equivocadas, ficam, não raro, impossibilitadas de promover avanços. O estudo de caso desta pesquisa procurou analisar os impactos das relações entre política científica e produção científica das instituições.

Finalmente, a presente pesquisa permitiu compreender melhor a Bibliometria, seus usos e limitações. Observar a estatística, sob o olhar da Filosofia da Ciência, promoveu acréscimos aos conhecimentos sobre os métodos e as técnicas bibliométricas. Em diversos momentos, foi necessário recorrer à literatura de outros campos do conhecimento para propor soluções para os problemas encontrados.

Esta pesquisa indicou também temas para estudos futuros. Alguns tópicos relevantes são: a) os modelos de avaliação acadêmica atendem a interesses políticos, econômicos e sociais. É, portanto, um campo de disputas, cuja compreensão requer o aporte dos estudos sociais da ciência; b) para analisar epistemologicamente o campo da Bibliometria é necessário recorrer à Sociologia e Filosofia da Ciência; c) as teorias que propõem a interseção da lógica e da inferência estatística poderão contribuir para aprofundar as discussões sobre o campo da Bibliometria e da Cientometria; d) a Bibliometria adota métodos estatísticos para gerar indicadores. Nessa medida, é importante revisitar as teorias estatísticas para melhor compreender os fundamentos de seus métodos.

Em síntese, esta pesquisa mostra que os métodos da Bibliometria não devem ser utilizados apenas como um conjunto de técnicas estatísticas de descrição. Ao contrário, o diálogo com outras disciplinas, ou seja, a perspectiva interdisciplinar, que inclui a Sociologia da Ciência, a História da Ciência e a Filosofia da Ciência, entre outras, é fundamental para o aprimoramento das pesquisas apoiadas em métodos bibliométricos.

7 REFERÊNCIAS

- AMORIM, A. N. Instituto Butantan é alvo de investigação de desvio de R\$ 35 milhões. **Radioagência NP**. 22 set. 2009.
- ALVARADO, R. U. A Lei de Lotka na bibliometria brasileira. **Ciência da Informação**, v. 31, n. 2, p. 14-20, maio/ago. 2002.
- ALVES, G.; WATANABE, P. Promotora denuncia 11 por suspeita de crimes contra Fundação Butantan. **Folha de São Paulo**. 23 fev. 2017.
- ANDRÉS, A. **Measuring Academic Research: How to undertake a bibliometric study**. Oxford: Chandos Publishing, 2009.
- ARENDT, H. **Origens do totalitarismo: Antissemitismo, imperialismo, totalitarismo**. São Paulo: Companhia das Letras, 2012.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6023: informação e documentação: referências. Rio de Janeiro, 2018. 24 p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 10520: informação e documentação: citações em documentos: apresentação. Rio de Janeiro, 2002. 7 p.
- BASALLA, George. The spread of Western Science. **Science**, v. 156, p. 611- 622, 1967.
- BLOOR, D. Wittgenstein and Mannheim on the Sociology of mathematics. **Studies in the History and Philosophy of Science**, v. 4, n. 2, p. 173-181, 1973.
- BOOTH, A. D. A. A "law" of occurrences for words of low frequency. **Information and Control**, v. 10, n. 4, p. 386-393, abr. 1967.
- BORKO, H. Information Science: What is it? **American Documentation**, v.19, n.1, p.3-5, Jan. 1968.
- BRADFORD, S. C. Sources of information on specific subjects. **Engineering**, v. 37, p. 85-86, 1934. [Reprinted in *Journal of Information Science*, v. 10, p. 176-180, 1985.
- BRAGA, G. M. Some aspects of the Bradford's distribution. **Proceedings of the ASIS**, v. 15, p. 51-54, 1978.
- BUFREM, L.; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. **Ciência da Informação**, Brasília, v. 34, n. 2, p. 9-25, maio/ago. 2005.
- BUCKLAND, Michael K. What Is a "Document"? **Journal of the American Society for Information**, North Carolina, v. 48, n. 9, p. 804-809, Set. 1997
- BUNGE, M. **Epistemologia: curso de atualização**. São Paulo: T.A. Queiroz: EDUSP, 1980.

CAMARGO, A. C. M. de. **As contradições da política de saúde no Brasil o Instituto Butantan**. São Paulo: Perspectiva, 2002.

CERVO, A. L.; BERVIAN, P. A. **Metodologia científica: para uso dos estudantes universitários**. São Paulo: McGraw-Hill do Brasil, 1978.

CHAVARRO, D.; RÀFOLS, I.; TANG, P. To what extent is inclusion in the Web of Science an indicator of journal 'quality'? **Research Evaluation**, v. 27, n. 3, 2018.

CLARIVATE ANALYTICS. **Web of Science: Confident research begins here. 2018**. Disponível em: <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>. Acesso em 04 dez. 2018.

CNPQ. Dados e estatísticas da Plataforma Lattes. Disponível em: <http://lattes.cnpq.br/>. Acesso em 04 dez. 2018.

COSTA, M. P. da; LEITE, F. C. L. **Repositórios institucionais da América Latina e o acesso aberto à informação científica**. Brasília: IBICT, 2017. Disponível em: <http://eprints.rclis.org/31109/1/Repositorio%20America%20Latina_Digital.pdf>. Acesso em: 16 junho 2018.

CRUZ, E. P. Dimas Tadeu Covas é o novo presidente do Instituto Butantan. **Agência Brasil**. 22 fev. 2017.

CUETO, M.; PALMER, S. **Medicina e Saúde Pública na América Latina: uma história**. Rio de Janeiro: Ed Fiocruz, 2016.

CUNHA, M.N. da; CAVALCANTI, C. R. de O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília, DF: Briquet de Lemos, 2008.

DANTAS, C. A. **Probabilidade: um curso introdutório**. São Paulo: Editora da Universidade de São Paulo, 2004.

DESCHACHT, P. Cross-cultural differences in self-promotion: A study of self-citations in management journals. **Journal of Occupational and Organizational Psychology**, v. 90, p. 77–94, 2017.

DESROSIÈRES, A. **The politics of large numbers: a history of statistical reasoning**. Cambridge, Mass.: Harvard University Press, 2011.

DIAS, Deborah. **Workshop Web of Science, JCR e EndNote Web**. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, 2019.

DIAS, C. E. As relações Brasil-França na criação do Instituto Butantan. **Cadernos de História da Ciência**, v. 6, n. 1. Instituto Butantan, 2010.

DIAS, C. E. História social dos trabalhadores do Instituto Butantan: ASIB - Associação dos Servidores do Instituto Butantan - Entrevista de Antônio Carlos Barbosa. **Cadernos de História da Ciência**, v. 8, n. 1. Instituto Butantan, 2012.

DIAS, R. A questão da criação para Nietzsche. **Viso: Cadernos de estética aplicada, Revista eletrônica de estética**. n. 7, p. 3-12, jul-dez. 2009.

- DOMINGUES, Â. Para um melhor conhecimento dos domínios coloniais: a constituição de redes de informação no Império português em finais do setecentos. **História, ciências, saúde. Manguinhos**, v. 8, suplemento, p. 823-838, 2001.
- DOSTOIEVSKI, F. M. **Memórias do subsolo**. São Paulo: Editora 34, 2008.
- DROTT, M. C. Bradford's law: theory, empiricism and the gaps between. **Library trends**, p. 41-52. 1981.
- DURKHEIM, E. **Da divisão do trabalho social**. São Paulo: Edipro, 2016.
- DURKHEIM, E. **O suicídio: estudo de sociologia**. São Paulo: Edipro, 2017.
- EGGHE, L. The dual of Bradford's law. **Journal of the American Society for Information Science**, n. 37, v. 4, p. 246–55, 1986.
- EGGHE, L. A note on different Bradford multipliers. **Journal of the American Society for Information Science**, n. 41, v. 3, p. 204–9, 1990.
- ELKANA, Y.; LEDERBERG, J.; MERTON, R. K.; THACKRAY, A.; ZUCKERMAN, H. (org). **Toward a metric of Science: the advent of scientific indicators**. Nova York: John Wiley & Sons, 1978.
- ELSEVIER. **Scopus**. Disponível: <https://www.elsevier.com/pt-br/solutions/scopus>. Acesso em 04 dez. 2018.
- FARIA, L. I. L. de. **Bibliometria: apostila**. São Carlos: UFSCAR, 2015.
- FERRATER MORA, J. **Dicionário de filosofia: tomo I (A-D)**. 2. ed. São Paulo: Edições Loyola, 2004.
- FERRATER MORA, J. **Dicionário de filosofia: tomo II (E-J)**. 2. ed. São Paulo: Edições Loyola, 2004.
- FERRATER MORA, J. **Dicionário de filosofia: tomo III (K-P)**. 2. ed. São Paulo: Edições Loyola, 2004.
- FERRATER MORA, J. **Dicionário de filosofia: tomo IV (Q-Z)**. 2. ed. São Paulo: Edições Loyola, 2004.
- FERREIRA, A. B. de H. **Dicionário da língua portuguesa**. 5. ed. Curitiba: Positivo, 2010.
- FIGUEIREDO, N. M. de. **Desenvolvimento e avaliação de coleções**. 2. ed. rev. e atual. Brasília: Thesaurus, 1998.
- FONSECA, E. N. da. Bibliografia estatística e bibliometria: uma reivindicação de prioridades. **Ciência da Informação**, v. 2, n. 1, p. 5-7, 1973.
- FOUCAULT, M. **História da sexualidade 1: a vontade de saber**. São Paulo: Paz e Terra, 2015.
- FOUCAULT, M. **Nascimento da biopolítica**. São Paulo: Martins Fontes, 2008.

FROHMANN, B. O caráter social, material e público da informação. In: FUJITA, M. S. .; MARTELETO, R. M.; LARA, M. L. G. A dimensão epistemológica da Ciência da Informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da Informação. São Paulo: Cultura Acadêmica; Marília: Fundepe, 2008. p.17-34.

FUNDAÇÃO BUTANTAN. **Relatório de Atividades 2015**. São Paulo: 2015.

GARFIELD, E. Citation indexes for science – new dimension in documentation through association ideas. **Science**, Washington, DC, v. 122, n. 3159, p.108-111, 1955.

GARFIELD, E.; SHER, I.H. New factors in the evaluation of scientific literature through citation indexing. **American Documentation**, Washington, DC, v. 14, n. 3, p.195-201, 1963.

GILCHRIST, A. Documentation of documentation: a survey of leading abstracts services in documentation and na identification of Key-journals. **Aslib Proceedings**, v. 18, n. 3, p. 62-80, 1966.

GINGRAS, Y. **Os desvios da avaliação da pesquisa: o bom uso da bibliometria**. Rio de Janeiro: Editora UFRJ, 2016.

GLÄNZEL, W. **Bibliometrics as a Research Field: A course on theory and application of bibliometric indicators**. Course handouts, 2003.

GLÄNZEL, W., THUS, B., SCHLEMMER. A bibliometric approach to the role of author self-citations in scientific communication. **Scientometrics**, v. 59, n.1, p. 63–77, 2014.

GODIN, B. On the origins of bibliometrics. **Scientometrics**, n. 68, p. 109-133, 2006.

GOFFMAN, W. Mathematical approach to the spread of scientific ideas: the history of mast cell research. **Nature**, v. 212, p. 449-452, out. 1966.

GOFFMAN, W.; NEWILL, V. A. Generalization of epidemic theory: an application to the transmission of ideas. **Nature**, v. 204, n. 4955, p. 225-228, out. 1964.

GOLISZEK, A. **Cobaias humanas: a história secreta do sofrimento provocado em nome da ciência**. Rio de Janeiro: Ediouro, 2004.

GOMES, A. C. V. **Uma ciência moderna e imperial: a fisiologia brasileira no final do século XIX**.

GORKOVA, V.I. Infometrics. **Informatics**, Moscou, v. 10, 1988.

GROSS, P.L.K.; GROSS, E.M. College libraries and chemical education. **Science**, Washington, DC, v. 66, n. 1713, p. 385-389, 1927.

GUEDES, V. L.; BORSCHIVER, S. Bibliometria: Uma ferramenta estatística para a Gestão da Informação e do Conhecimento, em Sistemas de Informação, de Comunicação e de Avaliação Científica e Tecnológica. In: Encontro Nacional de Ciência da Informação, 6., Salvador. Anais... Salvador: UFBA, jun. 2005.

HACKING, I. **The taming of chance**. Cambridge New York: Cambridge University Press, 2005.

HACKING, I. **The emergence of probability: a philosophical study of early ideas about probability, induction and statistical inference**. Cambridge New York: Cambridge University Press, 2006.

HACKING, I. **Ontologia histórica**. São Leopoldo: Unisinos, 2009.

HACKING, I. **An introduction to Probability and Inductive Logic**. Cambridge: Cambridge Univ. Press, 2009b.

HICKS et al. The Leiden Manifesto for research metrics. **Nature**, v. 520, p. 429-431, 2015. Disponível em: <http://www.nature.com/news/bibliometrics-the-leiden-manifesto-for-research-metrics-1.17351> e <http://www.leidenmanifesto.org/>.

HIRSCH, J.E. An index to quantify an individual's scientific research output. **Proceedings of the National Academy of Sciences of the United States of Sciences**. v. 104, p.19193-19198, 2005.

HOCHMAN, G. **A era do saneamento: as bases da política de Saúde Pública no Brasil**. São Paulo: Hucitec, 1998.

IBAÑEZ, N.; WEN, F. H.; FERNANDES, S. Instituto Butantan: história institucional. **Cadernos de História da Ciência**, v. 1, n. 1. Instituto Butantan, 2005.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Ibict lança manifesto pelo acesso livre à informação científica**. Disponível em: <<http://www.ibict.br/Sala-de-Imprensa/noticias/2005/ibict-lanca-manifesto-pelo-acesso-livre-a>>. Acesso em: 17 junho 2018.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Repositórios Digitais**. Disponível em: <<http://www.ibict.br/informacao-para-ciencia-tecnologia-e-inovacao%20/repositorios-digitais/repositorios-brasileiros>>. Acesso em: 17 junho 2018.

INSTITUTO BUTANTAN. **Relatório Anual 2013**. São Paulo: 2013.

JERONYMO, G. **Instituto Butantan, público, subordina-se a fundação privada**. Revista ADUSP, setembro, 2010.

JOLIVET, R. **Curso de filosofia**. Rio de Janeiro: Agir, 1979.

KATZ, J.S.; HICKS, D. Desktop Scientometrics. **Scientometrics**, n.38, p.141-153, 1997.

KUHN, T. S. The function of measurement in modern physical science. **Isis**, v. 52, n. 2, p. 161-193, 1961.

KUHN, T. S. **A estrutura das revoluções científicas**. São Paulo: Editora Perspectiva, 1997.

- KURY, L. Homens de ciência no Brasil: impérios coloniais e circulação de informações (1780-1810). **História, ciências, saúde. Manguinhos**, v. 11, suplemento 1, p. 109-130, 2004.
- LAKATOS, E. M.; MARCONI, M. de A. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2016.
- LATOUR, B. Joliot: a história e a física misturadas. In: SERRES, Michel (Org.). **Elementos para uma história das ciências: De Pasteur ao computador**. Lisboa: Terramar, 1996. p.131-155.
- LATOUR, B. **Ciência em ação**. São Paulo: Ed. UNESP, 2000.
- LEYDESDORFF, L. Theories of Citation? **Scientometrics**. v. 43, n. 1, p. 5-25, 1998.
- LEIMKUHNER, F.F. The Bradford distribution'. **Journal of Documentation**. v. 23, p. 197-207, 1967.
- LE COADIC, Y. F. **A ciência da informação**. 2. ed. Brasília, DF: Brinquet de Lemos, 2004.
- LOTKA, A. J. The frequency distribution of scientific productivity. **Journal of the Washington Academy Sciences**, Washington, DC, v. 16, n. 12, p.317-323, 1926.
- LUHN, H. P. A statistical approach to mechanized encoding and searching of literary information. **Journal of Research and Development**, New York, v. 1, n. 4, p. 309-317, out. 1957.
- MAIA, E. L. e S. Comportamento bibliométrico da língua portuguesa, como veículo de representação da informação. **Ciência da Informação**, n. 2, v. 2, 1973.
- MARCONDES, C. H.; SAYÃO, L. F. Introdução: repositórios institucionais e livre acesso. In.: SAYÃO, L. F. et al. (org.). **Implantação e gestão 16 de repositórios institucionais: políticas, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009.
- MARCONDES, C. H.; SAYÃO, L. F. À guisa de introdução: repositórios institucionais e livre acesso. In: SAYÃO, L. F.; TOUTAIN, L. B.; ROSA, F. G.;
- MARCONDES, C. H. **Implantação e gestão de repositórios institucionais: política, memória, livre acesso e preservação**. Salvador: EDUFBA, 2009. p. 9-21.
- MARICATO, J. de M.; Lima, E. L. M. Impactos da Altimetria: aspectos observados com análises de perfis no Facebook e Twitter. **Informação & Sociedade: Estudos**. v. 27, n. 1, abr. 2017.
- MARQUES, F. Avaliação em cinco dimensões. **Revista Fapesp**, ed. 286, dez. 2019.
- MARQUES, R. **Dicionário breve de pedagogia**. Lisboa: Presença, 2000.
- MEADOWS, A. J. **A comunicação científica**. Brasília: Brinquet de Lemos, 1999.
- MERTON, R. K. **Social theory and social structure**. New York: Free Press, 1968.

MINOIS, G. **História do futuro**. São Paulo: Ed. UNESP, 2016.

MOED, H. F. Assessment and support of emerging research groups. **FEMS Microbiology Letters**, v.365, n.17, 2018.

MUGNAINI, R. 40 anos de bibliometria no Brasil: da bibliografia estatística à avaliação da produção científica nacional. In: HAYASHI, Maria Cristina Piombato Innocentini; LETA, Jacqueline. (org.) **Bibliometria e cientometria: reflexões teóricas e interfaces**. 1ed. São Carlos: Pedro e João Editores, 2013.

MUGNAINI, R. et al. Panorama da produção científica do Brasil além da indexação: uma análise exploratória da comunicação em periódicos. **Transinformação**, v.31, e190033, 2019.

NALIMOV, V.V.; MULCHENKO, B.M. **Scientometrics**. Moscou: Nauka, 1969.

NARIN, F. **Evaluative bibliometrics: the use of publication and citation analysis in the evaluation of scientific activity**. Alexandria (Virginia): NSF, 1976.

NARIN, F. Bibliometric techniques in the evaluation of scientific activity. **Science and public policy**, 1987, p. 99-106.

NATH, R., JACKSON, W. M. Productivity of management information system researchers: Does Lotka's law apply?. **Information Processing & Management**, v. 27, n. 2/3, p. 203-209, 1991.

NEEDHAM, J. **Science and civilisation in China**. Cambridge, Cambridge University Press, 1954.

NÉRICI I. **Introdução à lógica**. São Paulo: Nobel, 1978.

NICHOLLS, P. T. Bibliometric modeling process and the empirical validity of Lotka's Law. **Journal of the American Society for Information Science**, v. 40, n. 6, p. 379-385, 1989.

OKUBO, Y. **Bibliometric indicators and analysis of research systems: methods and examples**. OECD - Science, Technology and Industry Working Papers 1997/1. Paris: OECD Publishing, 1997.

OPPENHEIMER, C. The use of online database in bibliometric studies. In: INTERNATIONAL ON-LINE INFORMATION MEETING, 9, 1986, Oxford. Oxford : [s. n.], 1986, p. 355-364.

OTLET, Paul. **Tratado de documentação: o livro sobre o livro, teoria e prática**. Brasília: Briquet de Lemos / Livros, 2018.

PAO, M. L. Automatic text analysis based on transition phenomena of word occurrences. **Journal of the American Society for Information Science**, New York, v. 29, n. 3, p. 121-124, maio. 1978.

PAO, M. L. Lotka's law : a testing procedure. **Information Processing & Management**. v. 21, n. 4, p. 305-320, 1985.

- PATY, M. A noção de determinismo na física e seus limites. **Scientiae Studia**. São Paulo, v.2, n.4, p. 465-492, 2004.
- PEIRCE, C. S. **Annotated Catalogue of the Papers of Charles S. Peirce**. Richard S. Robin (org.). Amherst, MA: University of Massachusetts Press, 1967.
- PEIRCE, C.S. **Semiótica**. São Paulo: Perspectiva, 1977.
- PESTRE, D. Por uma nova história social e cultural das ciências: novas definições, novos objetos, novas abordagens. **Cadernos IG-UNICAMP**, v. 6, n. 1, p. 3-56. 1996.
- PINHEIRO, L. V. R. Lei de Bradford: uma reformulação conceitual. **Ciência da Informação**, Brasília, n. 12, v. 2, p. 59-80, jul./dez. 1983.
- POTTER, W. G. Lotka's Law revisited. **Library Trends**, v. 31, p. 21-39, 1981.
- PRICE, D. J. de S. **Little science, big science**. New York: Columbia University Press, 1963.
- PRICE, D. J. de S. Networks of scientific papers. **Science**, n.149, p. 510-515, 1965.
- PRITCHARD, A. Statistical bibliography or bibliometrics? **Journal of Documentation**, v. 24, p. 348-349, 1969.
- PULGARÍN, A.; GIL-LEIVA, I. Bibliometric analysis of the automatic indexing literature: 1956-2000. **Information Processing and Management**, v. 40, p. 365-377. 2004.
- RAJ, K. **Relocating modern Science. Circulation and the construction of knowledge in South Asia and Europe, 1650-1900**. London: Palgrave MacMillan, 2007.
- RODRIGUES, M. da S. **Dicionário brasileiro de estatística: seguido de um vocabulário inglês-português**. Rio de Janeiro: Instituto Brasileiro de Estatística, 1970.
- ROMANCINI, R. O que é uma citação? A análise de citações na ciência. **Intexto**, Porto Alegre: UFRGS, v. 2, n. 23, p. 20-35, jul./dez. 2010.
- ROZADOS, H. B. F. A informação científica e tecnológica e os serviços de informação. **Informação & Sociedade: Estudos**, João Pessoa, v. 16, n. 1, p. 49-62, jan./jun. 2006.
- RUIZ, J. A. **Metodologia científica: guia para eficiência nos estudos**. São Paulo: Atlas, 1979.
- SANTOS, R. N. M. dos; KOBASHI, N. Y. Bibliometria, cientometria, infometria: conceitos e aplicações. **Pesquisa brasileira em Ciência da Informação**, Brasília, v. 2, n. 1, p. 155-172, jan./dez. 2009.
- SARACEVIC, T. Five years, five volumes and 2345 pages of the Annual Review of Information Science and Technology. **Information Storage and Retrieval**, v. 7, p. 127-39, 1971.

SILVA NETO, B.; BASSO, D. A ciência e o desenvolvimento sustentável: para além do positivismo e da pós-modernidade. **Ambiente & Sociedade**. Campinas, SP, v. 13, n. 2, 2010.

SPINAK, E. **Diccionario enciclopédico de bibliometría, cienciometría e informetría**. Caracas: Unesco, 1996.

SPINAK, E. Indicadores cienciometricos. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 141-148, maio/ago. 1998.

TAVARES, B. MP de SP faz operação contra corrupção em Fundações; ex-promotor é um dos alvos. **TV Globo e G1 SP**. 17 dez. 2019.

TESTA, J. A. base de dados ISI e seu processo de seleção de revistas. **Ciência da Informação**, Brasília, DF, v. 27, n. 2, p. 233-235, 1998.

TRUJILLO FERRARI, A. **Metodologia da Ciência**. Rio de Janeiro: Kennedy, 1974.

VAN LEEUWEN T.N.; MOED H.F.; TUSSEN R.J.W.; VISSER M.S.; VAN RANN A.F.J. Language biases in the Science Citation Index and its consequences for International Comparisons of National Research Performance. **Scientometrics**, v. 51, n. 1, p. 335–346, 2001.

VAN RAAN, A.F.J. Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. **Scientometrics**, v. 62, n. 1, p. 133-143, 2005.

VAN RAAN, A.F.J. Self-citation as an impact-reinforcing mechanism in the science system. **Journal of the American Society for Information Science and Technology**. v. 59, n. 10, p. 1631–1643, 2008.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 152-162, maio/ago. 2002.

VARELA, Alex Gonçalves. **Juro-lhe pela honra de bom vassalo e bom português: filósofo natural e homem público – Uma análise das memórias científicas do ilustrado José Bonifácio de Andrada e Silva (1780-1819)**. 2001. Dissertação (Mestrado em História das Ciências Aplicada ao Ensino) – Instituto de Geociências, Universidade de Campinas, São Paulo.

VLACHY, J. Distribution patterns in creative communities. In: WORLD CONGRESS OF SOCIOLOGY, 8., 1974, Toronto. Toronto : [S. n.], 1974.

WOLFRAM, D. **Applied informetrics for information retrieval research**. Westport, CT: Libraries Unlimited, 2003.

WOUTERS, P. F. The Citation Culture. Tese de PhD. Amsterdam: University of Amsterdam, 1999.

WOUTERS, P.; SUGIMOTO, C. R.; LARIVIÈRE, V.; MCVEIGH, M. E.; PULVERER, B.; RIJCKE, S. de; WALTMAN, L. Rethinking impact factors: better ways to judge a

jornal. **Nature**, maio. 2019. Disponível em: <https://www.nature.com/articles/d41586-019-01643-3>. Acesso em: 29 de maio de 2019. Acesso em: 01 jul. 2019.

YANCEY, R. **Fifty Years of Citation Indexing and Analysis**. Newsletter from Thomson Scientific. 2005. Disponível em: <http://ips.clarivate.com/m/pdfs/klnl/2005-08/50-years-citation-indexing.pdf>. Acesso em: 08 nov. 2018.

YIN, R. K. **Estudo de caso: planejamento e métodos**. Porto Alegre: Bookman, 2001.

ZIEGLER, M. F. Internacionalização das universidades aumenta a qualidade da pesquisa, defendem reitores. **Agência FAPESP**. 2018. Disponível em: <http://agencia.fapesp.br/internacionalizacao-das-universidades-aumenta-a-qualidade-da-pesquisa-defendem-reitores/31330/>. Acesso em: 08 nov. 2019.

ZIPF, G.K. **Human Behavior and the Principle of Least Effort**. Cambridge, MA: Addison-Wesley, 1949.

ZIPF, G.K. **The Psycho-biology of Language**. Cambridge, MA: MIT Press, 1935.

ANEXOS

CD-ROM contendo:

Anexo I: Lista de autores;

Anexo II: Lista de registros bibliográficos.