

# **BIBLIOMETRIA: UMA FERRAMENTA ESTATÍSTICA PARA A GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO, EM SISTEMAS DE INFORMAÇÃO, DE COMUNICAÇÃO E DE AVALIAÇÃO CIENTÍFICA E TECNOLÓGICA**

VÂNIA L. S.GUEDES\*

([guedes@eq.ufrj.br](mailto:guedes@eq.ufrj.br))

SUZANA BORSCHIVER\*\*

([suzana@eq.ufrj.br](mailto:suzana@eq.ufrj.br))

## **Resumo**

Este artigo tem como objetivo o desenvolvimento de uma revisão da literatura sobre Bibliometria, destacando sua aplicação como ferramenta estatística básica para a gestão da informação e do conhecimento científico e tecnológico. Inicialmente, é apresentada uma descrição do histórico dessa área de assunto e de sua denominação. Em seguida, desenvolve-se uma revisão bibliográfica de autores que se dedicaram aos estudos das leis e princípios que integram a área em questão, enunciando-os e citando suas aplicações como ferramenta empírica objetiva de quantificação dos processos comunicação científica e tecnológica. Finalmente, conclui-se evidenciando sua importância como instrumento de gestão de sistemas de informação, sistemas de comunicação e sistemas de avaliação, em ambientes científicos e tecnológicos.

## **Palavras-Chave**

Bibliometria. Gestão da Informação. Gestão do Conhecimento. Sistemas de Informação Científica e Tecnológica. Sistemas de Comunicação Científica e Tecnológica.

---

\* Mestre em Ciência da Informação e Coordenadora da Biblioteca da Escola de Química da UFRJ

\*\* Doutora em Ciências e Professora Adjunta da Escola de Química da UFRJ na Área de Gestão e Inovação Tecnológica

## 1 INTRODUÇÃO

Bibliometria é um conjunto de leis e princípios empíricos que contribuem para estabelecer os fundamentos teóricos da Ciência da Informação. O termo *statistical bibliography* – hoje Bibliometria – foi usado pela primeira vez em 1922 por E. Wyndham Hulme, antecedendo à data a qual se atribui a formação da área de Ciência da Informação, com a conotação de esclarecimento dos processos científicos e tecnológicos, por meio da contagem de documentos. Após Hulme, pressupõe-se que o termo *statistical bibliography* foi ignorado por 22 (vinte e dois) anos, até ser usado por Gosnell, em 1944, em um artigo sobre obsolescência da literatura. E, o mais curioso, sem reconhecimento de seu uso prévio. Consta, da literatura sobre Bibliometria que existiu um intervalo de cerca de 20 (vinte) anos, até 1962, quando o termo *statistical bibliography* foi mencionado pela terceira vez, por L. M. Raisig, em um estudo sobre análise de citações, intitulado *Statistical bibliography in health sciences*. Menciona-se, ainda, que existia um consenso entre autores dedicados ao assunto, de que o termo *statistical bibliography* não era de todo satisfatório, o que se verificava inclusive pelo seu escasso emprego na literatura. Assim, o termo Bibliometria (em inglês *Bibliometrics*) é sugerido para denominar a área em questão (PRITCHARD, 1969).

Pao (1989), em seu livro *Concepts of Information Retrieval*, refere-se à Bibliometria como um termo introduzido por Allan Pritchard, em seu artigo *Statistical Bibliography or Bibliometrics*, publicado em 1969, para denotar a área de estudo que usa métodos matemáticos e estatísticos para investigar e quantificar os processos de comunicação escrita. Pritchard, segundo Pao (1989), notou que a literatura é o ingrediente chave no processo de comunicação do conhecimento. E, ainda, que o atributo de uma unidade de literatura, que existe em forma publicada, isto é, artigos de periódicos e livros, pode ser estudado em termos estatísticos. Ela acrescenta que publicações, autores, palavras-chave, usuários, citações e periódicos são alguns dos parâmetros observáveis em estudos bibliométricos da literatura. Esses estudos tentam quantificar, descrever e prognosticar o processo de comunicação escrita.

Finalmente, um dos conceitos mais utilizados em Bibliometria é que, segundo Pritchard (1969), significa “todos os estudos que tentam quantificar os processos de comunicação escrita”.

O objetivo deste trabalho é apresentar uma revisão da literatura sobre Bibliometria, incluindo a criação e denominação dessa área de assunto, suas leis e princípios, e a sua utilidade como ferramenta estatística para o tratamento técnico e a gestão da informação e do conhecimento, em sistemas de informação, avaliação e comunicação científica e tecnológica.

## **2 LEIS E PRINCÍPIOS BIBLIOMÉTRICOS**

As principais leis bibliométricas são: Lei de Bradford, (produtividade de periódicos), Lei de Lotka (produtividade científica de autores) e Leis de Zipf (frequência de palavras).

O padrão de distribuição das leis e princípios bibliométricos segue a máxima: “poucos com muito e muito com poucos”. Essa máxima é conhecida como o Efeito Mateus na Ciência, que diz: “aos que mais têm será dado em abundância e , aos que menos têm, até o que têm lhes será tirado” (MERTON, 1968). Trata-se de uma abordagem ao efeito Mateus mediante a análise de processos psicossociais, que afetam o sistema de avaliação e distribuição de recompensas científicas. Por exemplo: cientistas altamente produtivos, de universidades mais conceituadas, obtêm frequentemente mais reconhecimento que cientistas igualmente produtivos, de outras universidades.

Verifica-se na Lei de Bradford, que permite estimar o grau de relevância de periódicos em dada área do conhecimento, que os periódicos que produzem o maior número de artigos sobre dado assunto formam um núcleo de periódicos, supostamente de maior qualidade ou relevância para aquela área. A Lei de Lotka considera que alguns pesquisadores, supostamente de maior prestígio em uma determinada área do conhecimento, produzem muito e muitos pesquisadores, supostamente de menor prestígio, produzem pouco. Nas Leis de Zipf, que permitem

estimar as freqüências de ocorrência das palavras de um determinado texto científico e tecnológico e a região de concentração de termos de indexação, ou palavras-chave, que um pequeno grupo de palavras ocorre muitas vezes e um grande número de palavras é de pequena freqüência de ocorrência.

A seguir serão analisadas e comentadas as leis, acima mencionadas.

## 2.1 LEI DE BRADFORD

A Lei de Bradford, relacionada à dispersão da literatura periódica científica, enuncia que “se periódicos científicos forem ordenados em ordem decrescente de produtividade de artigos sobre determinado assunto, poderão ser divididos em um núcleo de periódicos mais particularmente dedicados ao assunto e em vários grupos ou zonas, contendo o mesmo número de artigos que o núcleo. O número de periódicos ( $n$ ), no núcleo e zonas subseqüentes, variará na proporção  $1:n:n^2$  [...]” (BROOKES, 1969).<sup>1</sup>

A Lei de Bradford sugere que na medida em que os primeiros artigos sobre um novo assunto são escritos, eles são submetidos a uma pequena seleção, por periódicos apropriados, e se aceitos, esses periódicos atraem mais e mais artigos, no decorrer do desenvolvimento da área de assunto. Ao mesmo tempo, outros periódicos publicam seus primeiros artigos sobre o assunto. Se o assunto continua a se desenvolver, emerge eventualmente um núcleo de periódicos, que corresponde aos periódicos mais produtivos em termos de artigos, sobre o tal assunto. Brookes (apud BROOKES, 1969), refere-se a esse fenômeno como o “mecanismo do sucesso gerando o sucesso”.

A Lei de Bradford é um instrumento útil para o desenvolvimento de políticas de aquisição e de descarte de periódicos, em nível de gestão de sistemas de recuperação da informação, gestão da informação e do conhecimento científico e tecnológico. É possível estimar a magnitude de determinada área bibliográfica e o custo de toda e qualquer fração específica da bibliografia, no todo.

---

<sup>1</sup> Onde  $n$  é igual a um número  $x$  de periódicos (CHEN; CHONG; TONG, 1994, p. 536).

## 2.2 LEI DE LOTKA

A Lei de Lotka, relacionada à produtividade de autores e fundamentada na premissa básica de que “alguns pesquisadores publicam muito e muitos publicam pouco” (VOOS, 1974), enuncia que “a relação entre o número de autores e o número de artigos publicados por esses, em qualquer área científica, segue a Lei do Inverso do Quadrado  $1/n^2$ . Isto é, em um dado período de tempo, analisando um número  $n$  de artigos, o número de cientistas que escrevem dois artigos seria igual a  $1/4$  do número de cientistas que escreveram um. O número de cientistas que escreveram três artigos seria igual a  $1/9$  do número de cientistas que escreveram um, e assim sucessivamente. O mesmo autor revela que Price, em “Little Science, Big Science”<sup>2</sup>, observou que, para as ciências em geral, o número de autores decresce mais rapidamente que o inverso do quadrado, mais aproximadamente à Lei do Inverso do Cubo  $1/n^3$ .

Na gestão da informação, do conhecimento e planejamento científico e tecnológico, sua aplicabilidade se verifica na avaliação da produtividade de pesquisadores, na identificação dos centros de pesquisa mais desenvolvidos, em dada área de assunto, e no reconhecimento da “solidez” de uma área científica. Ou seja, quanto mais solidificada estiver uma ciência, maior probabilidade de seus autores produzirem múltiplos artigos, em dado período de tempo.

## 2.3 LEIS DE ZIPF

As Leis de Zipf, relacionadas à frequência de ocorrência de palavras em um dado texto, enriquecida pelo Ponto de Transição (T) de Goffman<sup>3</sup> relacionam-se diretamente com a representação da informação, isto é, a indexação temática automática. Hans Peter Luhn (1957) propôs que a indexação poderia ser derivada da análise de uma amostra representativa de documentos sobre determinado assunto.

---

<sup>2</sup> *Little Science, Big Science*, refere-se à denominação do Seminário de Estudos em Ciência da Informação, onde pela primeira vez, 1963, Price apresentou dados estatísticos sobre o fenômeno do crescimento exponencial da literatura (PAO, 1989).

<sup>3</sup> Será detalhado na página 10, item c.1.

Zipf observou que, num texto suficientemente longo, existia uma relação entre a frequência que uma dada palavra ocorria e sua posição<sup>4</sup> na lista de palavras ordenadas segundo sua frequência de ocorrência. Essa lista era confeccionada, levando-se em conta a frequência decrescente de ocorrências. À posição nesta lista dá-se o nome de ordem de série (*rank*). Assim, a palavra de maior frequência de ocorrência tem ordem de série 1, a de segunda maior frequência de ocorrência, ordem de série 2 e, assim, sucessivamente.

Zipf observou, também, que o produto da ordem de série (*r*) de uma palavra, pela sua frequência de ocorrência (*f*) era aproximadamente constante (*c*). Enunciou assim que

$r \cdot f = c$ , o que ficou conhecido como Primeira Lei de Zipf.

Fairthorne (1969) ressaltou que Zipf considera essa relação observada como uma consequência do Princípio Geral do Menor Esforço. Pao (1978) reconhece que esta lei é elegante em sua simplicidade. Entretanto, ela se aplica somente a palavras de alta frequência de ocorrência, em um texto. Para palavras de baixa frequência de ocorrência, Zipf propôs uma segunda lei, revisada e modificada por Booth (1967).

A Segunda Lei de Zipf enuncia que, em um determinado texto, várias palavras de baixa frequência de ocorrência (alta ordem de série) têm a mesma frequência. Booth (1967), ao modificá-la, a representa matematicamente da seguinte forma:

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

Onde  $I_1$  é o número de palavras que têm frequência 1,  $I_n$  é o número de palavras que têm frequência *n*, 2 sendo a constante válida para a língua inglesa. Parte da literatura, dedicada a esse tema, tem se referido a essa segunda lei como a Lei de Zipf-Booth.

---

<sup>4</sup> Localização em uma lista, confeccionada segundo a ordem decrescente de frequência de ocorrência das palavras, que compõem um determinado texto.

Guedes e Valois (1988) observam que a expressão  $n(n+1)/2$  corresponde à soma dos  $n$  primeiros números naturais. Por exemplo: a soma de  $1 + 2 + 3 + 4 + 5 = 15$ , pode ser calculada, levando em conta que  $n = 5$ , por  $\frac{5(5+1)}{2} = \frac{5 \times 6}{2} = \frac{30}{2} = 15$ .<sup>5</sup>

Segundo Pao (1978), Goffman observou que a Primeira Lei de Zipf era válida apenas para a região de palavras de alta frequência de ocorrência. Na maioria das vezes, essas palavras têm a propriedade de ocupar *ranking* único na lista de distribuição de palavras, isto é, das palavras de alta frequência de ocorrência, dificilmente, existem duas palavras com a mesma frequência de ocorrência.

A Segunda Lei de Zipf, modificada por Booth, descreve o comportamento das palavras de baixa frequência de ocorrência. Nessa região, observa-se que existem muitas palavras com a mesma frequência.

Esses dois comportamentos, inteiramente distintos, definem as duas extremidades da lista de distribuição de palavras de um dado texto. Assim, é razoável esperar uma região crítica, na qual há a transição do comportamento das palavras de alta frequência para as de baixa frequência. Goffman, segundo Pao (1978), admitiu como hipótese que nessa região de transição estariam as palavras de maior conteúdo semântico, de um dado texto.

---

<sup>5</sup> Carl Friedrich Gauss (1777-1855) calculou a soma aritmética  $1+2+3+4+5+6 [\dots] +99+100$ , substituindo-a pela equação matemática  $n(n+1)/2$  (KIYUKAWA; SHIGEKIYO; YAMAMOTO, 1993).

## 2.4 PONTO DE TRANSIÇÃO (T) DE GOFFMAN

De acordo com Goffman, para chegar a essa região de transição, onde estariam as palavras de alto conteúdo semântico (termos de indexação, descritores ou palavras-chave), a expressão da 2ª Lei de Zipf modificada por Booth teria que fornecer o comportamento típico das palavras de alta frequência, isto é, o número de palavras que têm frequência  $n$  tenderia a 1. Substituindo-se, na expressão da 2ª Lei de Zipf-Booth,  $I_n$  por 1 obtém-se:

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

ou ainda, rearranjando:  $n^2 + n - 2I_1 = 0$ ,

cujas raízes são:  $n = \frac{-1 \pm \sqrt{1 + 8I_1}}{2}$ .

Da expressão acima, interessa somente determinar a raiz positiva, assim

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

Ao valor de  $n$  assim determinado dá-se o nome de Ponto de Transição (T) de Goffman. 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 de um texto em questão. Goffman apresenta, com o Ponto de Transição, a primeira oportunidade de se decompor um texto sintaticamente, visando sua indexação.

Braga (1996) explica que “[...] a aplicação das Leis de Zipf, Zipf/Booth, Ponto T de Goffman e similares produz uma listagem de palavras acompanhadas de respectivas frequências de aparecimento no texto. Há uma total desconstrução do contexto das palavras, transformando-as em seqüências de caracteres entre espaços e pontuações.” Ela acrescenta que, apesar de existirem poucos estudos, eles indicam



que é possível determinar o conteúdo semântico dos textos, por meio de tais procedimentos.

Rouault (1987), em sua abordagem a métodos estatísticos de indexação temática automática, apesar de questionar a validade das leis de Zipf, destaca que podemos aproximadamente delimitar três zonas de ocorrência de palavras, em qualquer índice, e que as fronteiras, entre elas, não são claras. Acrescenta, entretanto, que essas observações são válidas para todos os textos analisados.

A primeira zona de ocorrência de palavras é composta de formas com um número elevado de ocorrências. Ela contém, acima de tudo, palavras cuja presença deve-se às razões de sintaxe (estrutura lingüística do idioma no qual é escrito o texto). No idioma francês, encontramos, essencialmente, determinantes, tais como: *le* (o), *la* (a); preposições: *des* (dos, das), representando cerca de 10% do texto; conjunções, auxiliares; etc. Essa zona contém poucos representantes das categorias de substantivos, adjetivo, verbos (ROUAULT, 1987).

Já a segunda zona, na opinião de Rouault (1987), se caracteriza, sobretudo, pelo fato de conter uma quantidade maior que a primeira zona de representantes de categorias morfológicas “informativas”, tais como: substantivo, adjetivo, verbo [...] entretanto, as formas gramaticais citadas na primeira zona ocorrem, ainda, sobretudo no topo dessa segunda zona. O autor enfatiza que o limite entre a primeira, a segunda e a terceira zona é muito tênue. Segundo ele, este limite deve ser resultado de uma decisão arbitrária de fixação de um limite entre as efetivas formas de cada zona e depende do texto utilizado.

A terceira zona de ocorrência contém as formas que jamais serão utilizadas como descritores (termos de indexação ou palavras-chave). A parte mais importante da terceira zona é composta por formas que ocorrem uma única vez e que representam freqüentemente 50% das formas distintas de um texto (ROUAULT, 1987).

Esta linha de raciocínio representa um passo importante, na busca de um critério para o processo de indexação automática; principalmente, em meios eletrônicos. As Leis de Zipf vêm sendo aplicadas, inclusive, para identificar estilos distintos de autores, na redação de artigos científicos e tecnológicos. Além disso, ela

vem sendo utilizada com sucesso como ferramenta estatística, em diferentes áreas do conhecimento, tais como: lingüística, urbanismo, física, medicina, economia, engenharia, química, entre outras.

Os estudos que utilizam a freqüência de ocorrência de palavras, como ferramenta de representação temática da informação, têm evoluído para tentativas de desenvolvimento de algoritmos, visando contribuir para automatização, em parte ou no todo, da indexação temática da informação.

A figura 1, a seguir, representa espacialmente as três principais leis bibliométricas e seus respectivos focos de estudo; considerando-as inseridas em um sistema de informação científica e tecnológica e este, num sistema de comunicação científica e tecnológica.

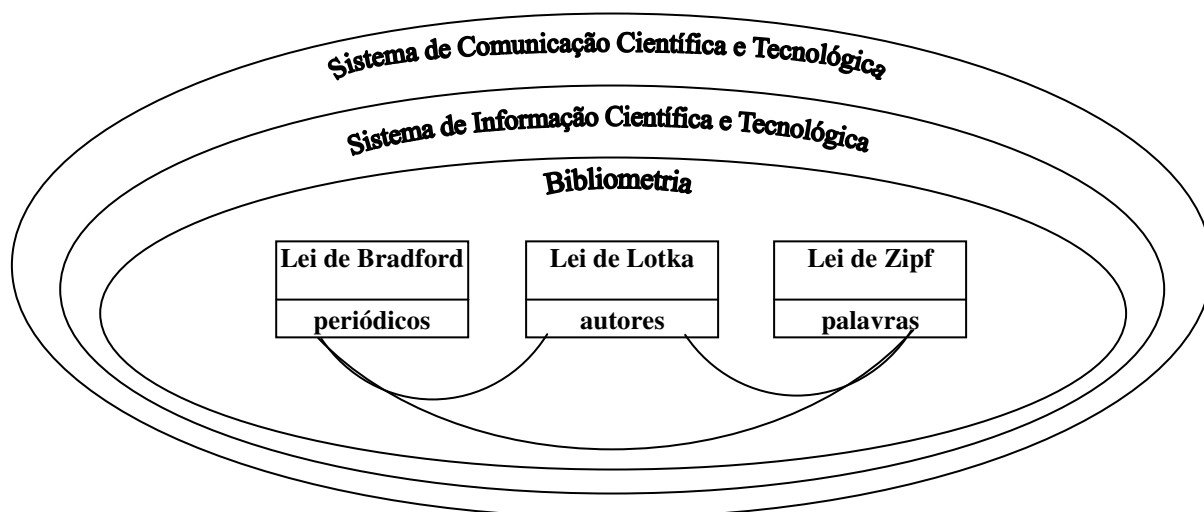


Figura 1. Principais leis da Bibliometria, seus focos de estudo e suas relações com os sistemas de comunicação e de informação científica e tecnológica (Fonte - Elaboração Própria).

Vale ressaltar aqui o conceito de sobrecedade<sup>6</sup>, na Ciência da Informação, que se faz absolutamente indispensável à aplicação das Leis de Bradford, Zipf e Ponto T de Goffman. A Lei de Bradford tem como objetivo estimar o grau em que um periódico é sobre dado assunto, da mesma forma que as de Zipf e Ponto T de Goffman têm como objetivo determinar sobre que assunto é um determinado texto científico e tecnológico.

<sup>6</sup> Sobrecedade é um conceito da área de Ciência da Informação, que significa tematicidade, “sobre o que trata” um documento.

### **3 OUTROS ESTUDOS E CONCEITOS APLICADOS À BIBLIOMETRIA**

Neste item, serão detalhados outros estudos e conceitos relacionados à Bibliometria . Dentre esses, valem ser destacados os estudos baseados na Análise de Citações<sup>7</sup>, que parte da hipótese de que citação é um indicador válido de influência de um determinado trabalho sobre outro(s), evidenciando conexões intelectuais.

#### **3.1 FRENTE DE PESQUISA E COLÉGIOS INVISÍVEIS**

A análise de citações permite identificar a Frente de Pesquisa, de uma determinada área científica, por meio de um conjunto de autores, que se citam na literatura recente, revelando um estreito padrão de relações múltiplas, na literatura sobre o assunto. Ela permite, também, identificar, nesse pequeno grupo de artigos entrelaçados, o trabalho de algumas centenas de colaboradores que formam os Colégios Invisíveis. Price (apud PAO, 1989) observou que um grupo pequeno de autores e de publicações exerce maior influência, em uma dada área de assunto. Esses autores são líderes, na área, e seus trabalhos são mais citados. Outras citações são distribuídas, regularmente e uniformemente, por toda a literatura passada, com frequência decrescente.

#### **3.2 FATOR DE IMEDIATISMO OU DE IMPACTO**

A análise de citações é utilizada, ainda, para estimar o Fator de Imediatismo de um artigo publicado, pelo estudo da concentração de citações a esse artigo, em documentos publicados nos últimos quinze anos (PRICE, 1965). A hipótese é de que, em determinada área científica, artigos de periódicos citados mais freqüentemente são mais relevantes do que artigos menos citados. Jones (2003), ao externar sua preocupação com o uso e abuso do fator de impacto, na avaliação da importância e do prestígio de periódicos científicos e de cientistas, defende que o fator de impacto de um determinado periódico é calculado pela divisão do número de citações correntes a um dado artigo, publicado nos últimos dois anos, pelo total de artigos publicados, no mesmo período de tempo.

---

<sup>7</sup> Citação é a “menção no texto de uma informação extraída de outra fonte” . (ABNT,2002).

### **3.3 ACOPLAMENTO BIBLIOGRÁFICO E CO-CITAÇÃO**

O Acoplamento Bibliográfico (retrospectivo) consiste na união de artigos, citando o(s) mesmo(s) documento(s), e a Co-citação (prospectivo), na união de artigos, citados pelos mesmos documentos. Na opinião de Marshakova (1981), o Acoplamento Bibliográfico mede o grau de ligação entre dois ou mais artigos, segundo o número de documentos idênticos citados por esses artigos, e Co-citação mede o grau de ligação de dois ou mais artigos, pelo número de documentos onde esses artigos são citados, simultaneamente.

### **3.4 OBSOLESCÊNCIA DA LITERATURA E VIDA-MÉDIA**

A Obsolescência da Literatura consiste na análise do declínio do uso da literatura, no decorrer do tempo, e a Vida-Média é estimada a partir da razão de obsolescência e da razão de crescimento, de um determinado corpo de literatura (LINE, 1970). Em termos de uso da literatura, a Vida-Média tem sido interpretada mediante o estudo do número de citações feitas a um determinado item. Pao (1989) acrescenta que, na área de química, metade das referências citadas na literatura relaciona-se a artigos com menos de oito anos de publicação; enquanto que, na matemática, a Vida-Média de uma unidade da literatura é estimada em cerca de vinte anos.

### **3.5 LEI DO ELITISMO**

A Lei do Elitismo enuncia que toda população de tamanho  $N$  tem uma elite efetiva tamanho  $\sqrt{N}$  (PRICE, 1965). Price (1963), Crawford (1971), Crane (1972), Cole and Cole (1972) and Griffith and Mullins (1972) encontraram, em seus estudos, evidências de elites e elitismo na ciência (PAO, 1989).

### **3.6 TEORIA EPIDÊMICA DE GOFFMAN**

A Teoria Epidêmica de Goffman fundamenta-se na analogia entre a transmissão de uma doença infecciosa, segundo um processo epidêmico, e a transmissão e desenvolvimento de idéias, informações registradas, em uma

comunidade científica (GOFFMAN & NEWILL, 1964). Segundo esse modelo, as idéias científicas são materiais infecciosos, no curso de uma epidemia intelectual; transmitidas, por exemplo, por comunicações diretas, entre um conferencista e o público, ou através de conversações. Essas idéias podem também ser expostas por um autor, em artigos de periódicos, para um determinado público. A análise matemática de Goffman foi capaz de prognosticar as condições de controle da epidemia, a razão de crescimento e de declínio, de uma dada área do conhecimento, e permitiu definir as condições sob as quais a epidemia declinaria e se tornaria estável (PAO, 1989). Segundo Goffman (1966), sua teoria possibilita estimar os níveis de importância de linhas de pesquisa, em uma determinada área de assunto, e prognosticar o comportamento dessas linhas de pesquisas.

Citações visam, primordialmente, acessibilidade ao material citado. Citação indica uso do citado pelo citante. Pelas citações posso contar artigos, periódicos, autores, departamentos acadêmicos, universidades, institutos de pesquisa, entre outros. Além disso, um artigo de periódico muito citado representa a aceitação da comunidade que o cita. Smith (apud PAO, 1989), destaca a citação como meio de quantificação, aceitando como premissa que citação normalmente indica uma certa relação de assunto, ainda que essa relação não seja especificada

Basicamente, a análise de citações tem três grandes aplicações: bibliotecas (gestão de coleções), ciência (mapeamento do desempenho dos autores), administração (financiamento de pesquisa, auxílio, bolsas, orçamento de sistemas de informação/bibliotecas). Além disso, ela é uma ferramenta para a recuperação da informação, avaliação de periódicos, produtividade de autores, medida de qualidade de uma dada informação, medida do fluxo de informação em uma unidade, sociologia da ciência, indicador de estruturas e tendências científicas, entre outras.

Finalmente, embora não baseada na análise de citações, vale destacar a **Lei dos 80/20**, que consiste em um fenômeno, inicialmente observado no comércio e na indústria, segundo o qual em sistemas de informação 80% da demanda de informação se satisfaz com 20% do conjunto de fontes de informação. (TRUESWELL, 1969). Em sistemas de informação, esta lei pode ser usada nas tomadas de decisão relacionadas à composição e redução de acervos.

O Quadro 1, a seguir, relaciona as principais leis e princípios bibliométricos, seus focos de estudo e suas principais aplicações na gestão da informação e do conhecimento, em sistemas de informação e comunicação científica e tecnológica.

<b>Ciência da Informação</b>		
<b>Bibliometria</b>		
<b>Leis e Princípios</b>	<b>Focos de Estudo</b>	<b>Principais Aplicações</b>
Lei de Bradford	periódicos	estimar o grau de relevância de periódicos, em dada área do conhecimento
Lei de Lotka	autores	estimar o grau de relevância de autores, em dada área do conhecimento
Leis de Zipf	palavras	indexação automática de artigos científicos e tecnológicos
Ponto de Transição (T) de Goffman	palavras	indexação automática de artigos científicos e tecnológicos
Colégios Invisíveis	citações	identificação da elite de pesquisadores, em dada área do conhecimento
Fator de Imediatismo ou de Impacto	citações	estimar o grau de relevância de artigos, cientistas e periódicos científicos, em determinada área do conhecimento
Acoplamento Bibliográfico	citações	estimar o grau de ligação de dois ou mais artigos
Co-citação	citações	estimar o grau de ligação de dois ou mais artigos
Obsolescência da Literatura	citações	estimar o declínio da literatura de determinada área do conhecimento
Vida-média	<b>citações</b>	estimar a vida-média de uma unidade da literatura de dada área do conhecimento
Teoria Epidêmica de Goffman	citações	estimar a razão de crescimento e declínio de determinada área do conhecimento
Lei do Elitismo	citações	estimar a o tamanho da elite de determinada população de autores
Frente de Pesquisa	citações	identificação de um padrão de relação múltipla entre autores que se citam
Lei dos 80/20	demanda de informação	composição, ampliação e redução de acervos

Quadro 1. Leis e princípios bibliométricos, seus focos de estudo, principais aplicações e áreas de interesse (Fonte - Elaboração própria).

Há de se admitir as possíveis “injustiças” cometidas em nome da máxima que norteia as leis e princípios bibliométricos. Ocorre que leis e princípios, de qualquer área de assunto, são enunciados a partir da observação de fenômenos que se repetem com frequência, dado um determinado contexto. A generalidade no comportamento, verificada em suas aplicações, é que os caracteriza como leis e princípios, sem, contudo, deixar de se considerar suas limitações. O conceito considerado fundamental à aplicação de tais leis e princípios é a probabilidade.

## 5 Conclusão

A Bibliometria é uma ferramenta estatística que permite mapear e gerar diferentes indicadores de tratamento e gestão da informação e do conhecimento, especialmente em sistemas de informação e de comunicação científicos e tecnológicos, e de produtividade, necessários ao planejamento, avaliação e gestão da ciência e da tecnologia, de uma determinada comunidade científica ou país.

Pode-se observar, ao longo deste trabalho, a diversidade de leis e conceitos que norteiam o tema. As Leis de Bradford, Lotka e Zipf, focando respectivamente a produtividade de periódicos, a produtividade de autores e a frequência de ocorrência de palavras, em sistemas de informação e comunicação científica e tecnológica, são as principais. Os outros conceitos, baseados na prática da análise de citações, dentre de uma visão macro, são mais utilizados como ferramenta na política científica e tecnológica, mediante diagnóstico e prognóstico dos fenômenos que norteiam a comunicação científica e tecnológica, de uma determinada instituição ou país.

Finalmente, a Bibliometria é também um instrumento quantitativo, que permite minimizar a subjetividade inerente à indexação e recuperação das informações, produzindo conhecimento, em determinada área de assunto. Em última análise ela contribui para tomadas de decisão na gestão da informação e do conhecimento, uma vez que auxilia na organização e sistematização de informações científicas e tecnológicas.

## 6 Referências Bibliográficas

1. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *NBR 10520: informação e documentação. Citações em documentos: apresentação*. Rio de Janeiro, 2002. 7 p.
2. BOOTH, Andrew D. A. A “law” of occurrences for words of low frequency. *Information and Control*, [s.l.], v. 10, n.4, p.386-393, April 1967.
3. BRAGA, Gilda Maria. A representação da informação na reconstrução do texto. In COLÓQUIO: ORGANIZAÇÃO DO CONHECIMENTO E REPRESENTAÇÃO DA INFORMAÇÃO: ENFOQUES E PERSPECTIVAS. *Resumos...* Rio de Janeiro: IBICT-CNPq, 1996.
4. BRADFORD, S. C. Sources of information on specific subjects. *Engineering*, [s.l.], v.137, p. 85-86, 1934.
5. BROOKES, B. C. Bradford’s law and the bibliography of science. *Nature*, [s.l.], v.224, p.953-956, Dec. 1969.
6. CHENG, Y. S. Booth’s Law of word frequency. *Journal of American Society for Information Science*, v. 41, n. 5, p. 387-388. July 1990.
7. CHENG, K. H. Automatic identification for topics of electronic documents. *Bulletin of Library Association of China*, v. 59 , p. 43-58, Dec. 1997.
8. CHEN, Ye-Sho; CHONG, P. Pete; TONG, Morgan Y. The Simon-Yule approach to bibliometric modeling. *Information Processing and Management*, v. 30, n.4, p. 535-556, 1994.
9. FAIRTHONE, R. A. Empirical hyperbolic distribution (Bradford – Zipf – Mandelbrot) for bibliometric description and prediction. *Journal of Documentation*, [s.l.], v. 25, n. 4, p.521-534, Dec. 1969.
10. GOFFMAN, W. Mathematical approach to the spread of scientific ideas: the history of mast cell research. *Nature*, [s.l.], v. 212, p.449-452, Oct. 1966.
11. GOFFMAN, W.; NEWILL, V. A. Generalization of epidemic theory: an application to the transmission of ideas. *Nature*, [s.l.], v. 204, n. 4955, p. 225-228, Oct. 1964.



12. GUEDES, V. L. da S.; VALOIS, E. C. *Adequação das Leis de Zipf (1ª e 2ª) e Ponto T de Goffman à indexação de documentos científicos: uma aplicação em Mecânica dos Solos (engenharia civil)*. Rio de Janeiro, 1988. Trabalho não publicado apresentado á Disciplina de Bibliometria da ECO/UFRJ. P.3.
13. JONES, A W. Impact factors of forensic science and toxicology journals: what do the numbers really mean? *Forensic Science International* , V 133, n.1-2 ,p. 1-8,2003  
Disponível em : [www.sciencedirect.com](http://www.sciencedirect.com) Acesso: 05/04/2005
14. LINE, M. B. The “half-life” of periodical literature: apparent and real absolescence. *Journal of Documentation*, [s.l.], v.26, n.1, p. 48-54, Mar. 1970.
15. LUHN, H. P. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*, New York, v. 1, n.4, p. 309-317, Oct. 1957
16. LOTKA, A. J. The frequency of distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, v. 16, n.12, p. 317-323, 1926.
17. MARSHAKOVA, I. V. Citation networks in information science. *Scientometrics*, [s.l.], v.31, n.1, p.13-16, 1981.
18. MERTON, R. K. The Mathew effect in science. *Science*, [s. l.], v. 159, n. 3810, p. 58, Jan. 1968.
19. 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, May 1978.
20. PAO, M. L. *Concepts of information retrieval*. Englewood, Colorado:Libraries Unlimited, Inc., 1989. 285 p.
21. PRICE, Derek J. De Solla. Networks of scientific papers. *Science*, [s.l.], v. 149, n.3683, p. 56-64, July 1965.
22. PRITCHARD, A. Statistical bibliography or bibliometrics? *Journal of Documentation*, [s. l.], v. 25, n.4, p. 348-349, Dec. 1969.
23. ROUAULT , J. *Linguistique automatique: applications documentaries*. Berne:

- PETER LANG, 1987. p.
24. TRUESWELL, R. W. Some behavioral patterns of library users: the 80/20 rules. *Wilson Library Bulletin*, [s.l.], v. 43, n. 5, p. 458-461, 1969.
- 25 VOOS, H. Lotka and information science. *Journal of the American Society of Information Science*, New York, v. 25, p. 270-272, July/Aug. 1974.
- 26-IPF, G. K. *Human behavior and the principle of least effort*. Cambridge, Ma: Addison Wesley, 1949.