

Bibliometria: evolução histórica e questões atuais

Carlos Alberto Araújo

RESUMO

Promove-se uma análise do campo de conhecimento intitulado bibliometria. Inicialmente, são vistas as três leis fundadoras do campo, a lei de produtividade de autores de Lotka, a lei de dispersão de periódicos de Bradford e a lei de frequência de palavras de Zipf, bem como seus desdobramentos e aplicações em tempos recentes. São também vistos a teoria epidêmica e a análise de citações, em suas aplicações contemporâneas, como por exemplo os estudos de frente de pesquisa e fator de impacto. Por fim, analisa-se o desenvolvimento deste campo de estudos no Brasil e as tendências atuais, com destaque para as abordagens que buscam ampliar o escopo dos estudos realizados integrando os métodos bibliométricos a distintos corpos teóricos.

PALAVRAS-CHAVE: Bibliometria. Lei de Lotka. Lei de Bradford. Lei de Zipf. Análise de citações. Fator de impacto.

1 Introdução

A bibliometria, técnica quantitativa e estatística de medição dos índices de produção e disseminação do conhecimento científico “[...] tal como procede a demografia ao recensear a população” (FONSECA, 1986, p. 10), surge no início do século como sintoma da necessidade do estudo e da avaliação das atividades de produção e comunicação científica. Desenvolve-se inicialmente a partir da elaboração de leis empíricas sobre o comportamento da literatura, sendo que, entre os principais marcos de seu desenvolvimento (TAGUE-SUTCLIFFE, 1992, p. 2, tradução nossa), estão o método de medição da produtividade de cientistas de Lotka (1926), a lei de dispersão do conhecimento científico de Bradford (1934) e o modelo de distribuição e frequência de palavras num texto de Zipf (1949).

Consistindo na aplicação de técnicas estatísticas e matemáticas para descrever aspectos da literatura e de outros meios de comunicação (análise quantitativa da informação), a bibliometria foi originalmente conhecida como “bibliografia estatística” (termo cunhado por Hulme em 1923), sendo o termo “bibliometria” criado por Otlet em 1934 no seu “*Traité de Documentation*”. Contudo, o termo apenas se popularizou em 1969, a partir de um artigo de Pritchard que discutia a polêmica “bibliografia estatística ou bibliometria?” (VANTI, 2002, p. 153). Conforme Nicholas e Ritchie (1978, p. 38, tradução nossa), a diferença essencial entre a tradicional bibliografia e a bibliometria é que esta utiliza mais métodos quantitativos do que discursivos. Assim, a utilização de métodos quantitativos na busca por uma avaliação objetiva da produção científica é o ponto central da bibliometria:

Deixando de lado os julgamentos de valor, parece clara a importância de se dispor de uma distribuição que nos informe sobre o número de autores, trabalhos, países ou revistas que existem em cada categoria de produtividade, utilidade ou o que mais desejarmos saber. (PRICE, 1976, p. 39).

Inicialmente voltada para a medida de livros (quantidade de edições e exemplares, quantidade de palavras contidas nos livros, espaço ocupado pelos

livros nas bibliotecas, estatísticas relativas à indústria do livro), aos poucos foi se voltando para o estudo de outros formatos de produção bibliográfica, tais como artigos de periódicos e outros tipos de documentos, para depois ocupar-se, também, da produtividade de autores e do estudo de citações. Conforme Figueiredo (1977), a bibliometria desde sua origem é marcada por uma dupla preocupação: a análise da produção científica e a busca de benefícios práticos imediatos para bibliotecas (desenvolvimento de coleções, gestão de serviços bibliotecários). Também a promoção do controle bibliográfico (conhecer o tamanho e as características dos acervos, elaborar previsões de crescimento, etc) é apontado como objetivo “mais óbvio” da bibliometria (NICHOLAS; RITCHIE, 1978, p. 12, tradução nossa).

2 As três leis clássicas

A Lei de Lotka, formulada em 1926, foi construída a partir de um estudo sobre a produtividade de cientistas, a partir da contagem de autores presentes no *Chemical Abstracts*, entre 1909 e 1916. Lotka descobriu que uma larga proporção da literatura científica é produzida por um pequeno número de autores, e um grande número de pequenos produtores se iguala, em produção, ao reduzido número de grandes produtores. A partir daí formulou a lei dos quadrados inversos: $y_x = C/p^2x^t$, onde y_x é a frequência de autores publicando número x de trabalhos e a é um valor constante para cada campo científico (2 para físicos e 1,89 para químicos, por exemplo). A lei de Lotka foi, desde então, objeto de larga produção científica:

Desde 1926, época em que Lotka estabeleceu esta lei, muitos estudos têm sido conduzidos para investigar a produtividade dos autores em distintas disciplinas. Até dezembro de 2000, mais de 200 trabalhos, entre artigos, monografias, capítulos de livros, comunicações a congressos e literatura gris (cinzenta) tinham sido produzidos tentando criticar, replicar e/ou reformular esta lei bibliométrica (URBIZAGÁSTEGUI ALVARADO, 2002, p. 14).

Uma grande quantidade desses estudos aponta problemas na lei. Segundo Rao (1986, p. 182), essa lei “[...] é baseada em um conjunto pouco potente de dados e não foi testada estatisticamente”. Entre os aperfeiçoamentos realizados destaca-se o de Price, que a partir de estudos realizados entre 1965 e 1971 concluiu que 1/3 da literatura é produzida por menos de 1/10 dos autores mais produtivos, levando a uma média de 3,5 documentos por autor e 60% dos autores produzindo um único documento. Logo depois foi formulada a lei do elitismo de Price: o número de membros da elite corresponde à raiz quadrada do número total de autores, e a metade do total da produção é considerado o critério para se saber se a elite é produtiva ou não.

A segunda lei bibliométrica incide sobre conjuntos de periódicos. Com o objetivo de descobrir a extensão na qual artigos de um assunto científico especí- fico apareciam em periódicos destinados a outros assuntos, estudando a distri- buição dos artigos em termos de variáveis de proximidade ou de afastamento, Bradford realiza uma série de estudos que culminam, em 1934, com a formula- ção da lei da dispersão. O autor percebe que, numa coleção de periódicos sobre geofísica, existe sempre um núcleo menor de periódicos relacionados de maneira próxima ao assunto e um núcleo maior de periódicos relacionados de maneira estreita, sendo que o número de periódicos em cada zona aumenta, enquanto a produtividade diminui. Analisando 326 periódicos, ele descobriu que 9 periódicos continham 429 artigos, 59 continham 499 e 258 continham 404 artigos.

Assim, ordenando uma grande coleção de periódicos em ordem de pro- dutividade decrescente relevante a um dado assunto, três zonas aparecem, cada uma contendo 1/3 do total de artigos relevantes (a primeira zona con- têm um pequeno número de periódicos altamente produtivos, a segunda con- têm um número maior de periódicos menos produtivos, e a terceira inclui mais periódicos ainda, mas cada um com menos produtividade).

Bradford viu que era por essa razão que os índices tinham dificuldade para atingir a cobertura completa de assuntos. Havendo grande número de

periódicos na zona exterior, Bradford constatou que mais da metade do total de artigos úteis não estavam sendo cobertos pelos serviços de indexação e resumos.

A Lei de Bradford pode ser enunciada da seguinte forma: se dispormos periódicos em ordem decrescente de produtividade de artigos sobre um determinado tema, pode-se distinguir um núcleo de periódicos mais particularmente devotados ao tema e vários grupos ou zonas que incluem o mesmo número de artigos que o núcleo, sempre que o número de periódicos existentes no núcleo e nas zonas sucessivas seja de ordem de 1: n : n^2 : n^3 Assim, os periódicos devem ser listados com o número de artigos de cada um, em ordem decrescente, com soma parcial. O total de artigos deve ser somado e dividido por três; o grupo que tiver mais artigos, até o total de 1/3 dos artigos, é o “core” daquele assunto. O segundo e o terceiro grupo são as extensões. A razão do número de periódicos em qualquer zona pelo número de periódicos na zona precedente é chamada “multiplicador de Bradford” (B_m): à medida que o número de zonas for aumentando, o B_m diminuirá.

O gráfico formado a partir destes cálculos é uma função em linha reta em que, no eixo x , se tem a soma parcial dos periódicos e, no eixo y , se tem a soma parcial de artigos contidos em x periódicos mais importantes. Essa lei também foi sendo constantemente reformulada e aperfeiçoada, como por exemplo por Vickery, em 1948, que propôs que o número de zonas não precisa ser três mas qualquer número. Outras reformulações foram feitas por outros autores para corresponder à realidade da literatura científica, tais como Leimkuhler, em 1967, Fairthorne, em 1969, e Goffman e Warren em também em 1969 (DONOHUE, 1973, p. 17, tradução nossa).

Essa lei foi muito utilizada para aplicações práticas em bibliotecas, tais como o estudo do uso de coleções para auxiliar na decisão quanto à aquisição, descartes, encadernação, depósito, utilização de verba, planejamento de sistema. Como exemplos (FIGUEIREDO, 1977, p. 21-23) destacam-se Brookes

(estudos sobre descartes; proposta de bibliotecas locais, regionais, nacionais; conceito de percentagem específica de excelência para a coleção de periódicos), Bourne (meta de 90% de satisfação do usuário), Leimkuhler (sistema de dois depósitos, um com 20%, os mais pertinentes, e outro com o resto) e Trueswell (distribuição de Bradford como instrumento para predição da circulação de livros).

No Brasil, destacam-se dois estudos de grande importância. Um deles buscou analisar as várias interpretações sobre a lei que dão a entender existirem, na verdade, duas leis, sendo que na verdade há apenas uma única lei de Bradford (MAIA, 1980). O outro apresenta uma reformulação da lei com a introdução do conceito de produtividade relativa, a partir do argumento de que o núcleo de periódicos de uma área não é formado pelos periódicos mais devotados ao tema e sim pelos mais produtivos num determinado período de tempo (PINHEIRO, 1982).

Estudos atuais têm sido realizados (COOPER; BLAIR; PAO, 1993) buscando identificar *core lists*, isto é, núcleos de periódicos mais produtivos, de uma determinada área, em revisões que confirmam ou reformulam a Lei de Bradford.

A terceira das leis bibliométricas clássicas é a Lei de Zipf, formulada em 1949 e que descreve a relação entre palavras num determinado texto suficientemente grande e a ordem de série destas palavras (contagem de palavras em largas amostragens). Zipf, analisando a obra *Ulisses* de James Joyce, encontrou uma correlação entre o número de palavras diferentes e a frequência de seu uso e concluiu que existe uma regularidade fundamental na seleção e uso das palavras e que um pequeno número de palavras é usado muito mais frequentemente. Ele descobriu que a palavra mais utilizada aparecia 2653 vezes, a centésima palavra mais utilizada ocorria 256 vezes e a duocentésima palavra ocorria 133 vezes. Zipf viu então que a posição de uma palavra multiplicada pela sua frequência era igual a uma constante de aproximadamente 26500.

Sua proposta, assim, é de que, se listarmos as palavras que ocorrem num texto em ordem decrescente de frequência, a posição de uma palavra na lista multiplicada por sua frequência é igual a uma constante. A equação para esse relacionamento é: $r \times f = k$, onde r é a posição da palavra, f é a sua frequência e k é a constante.

A partir daí Zipf formulou o princípio do menor esforço: existe uma economia do uso de palavras, e se a tendência é usar o mínimo significa que elas não vão se dispersar, pelo contrário, uma mesma palavra vai ser usada muitas vezes; as palavras mais usadas indicam o assunto do documento. Se a tendência dos autores dos documentos fosse de variar muito, usar palavras diferentes, a lei não serviria. Também essa lei foi bastante reformulada, por Kendall (que faz um paralelo entre Zipf e Bradford) e Brookes (criador da distribuição Bradford/Zipf, que postula que um largo número de “fontes” contribuem com “itens” casuais para um “campo” determinado). Outros autores que acrescentaram contribuições à Lei de Zipf foram Booth, Donohue e Mandelbrot (RAO, 1986, p. 181). O método foi sendo aperfeiçoado, ainda, principalmente com estudos de frequência e co-ocorrência de descritores.

3 Outras teorias bibliométricas

Uma variação de enfoques bibliométricos é a teoria epidêmica da transmissão de idéias, desenvolvida por Goffman e Newill, em 1967, que explica a propagação de idéias dentro de uma determinada comunidade como um fenômeno similar à transmissão das doenças infecciosas (ou seja, pelo processo epidêmico). Os autores realizam seu estudo por comparação do ciclo da esquistossomose e da informação, fazendo uma analogia entre os dois sistemas. Onde o processo epidêmico na área de saúde identifica a infecção, por exemplo (a pessoa com uma doença), Goffman e Newill identificam o autor com uma idéia. A susceptibilidade, isto é, a pessoa que pode contrair a doença, é vista como o leitor que recebe a idéia. E o material infectante, que no

caso da saúde são os germes da doença, na teoria epidêmica da informação consiste nas próprias idéias contidas na literatura. Há quem avalie que “a principal abordagem usada no estudo da dinâmica da literatura é a teoria epidêmica do crescimento da literatura” (DONOHUE, 1973, p. 28, tradução nossa).

Contudo, a área mais importante da bibliometria é a análise de citações. Citação é um “[...] conjunto de uma ou mais referências bibliográficas que, incluídas em uma publicação, evidenciam elos entre indivíduos, instituições e áreas de pesquisa, visto que mostram o relacionamento de uma publicação com outra”. A análise de citação pode ser definida como “[...] a parte da bibliometria que investiga as relações entre os documentos citantes e os documentos citados considerados como unidades de análise, no todo ou em suas diversas partes: autor, título, origem geográfica, ano e idioma de publicação, etc” (FORESTI, 1989, p. 3). Enfim, para a autora, as citações

[...] contribuem para o desenvolvimento da ciência, provêem o necessário reconhecimento de um cientista por seus colegas, estabelecem os direitos de propriedade e prioridade da contribuição científica de um autor, constituem importantes fontes de informação, ajudam a julgar os hábitos de uso da informação e mostram a literatura que é indispensável para o trabalho dos cientistas. (FORESTI, 1989, p. 2).

De acordo com Moravcsik e Murugesan, as citações podem ser de quatro tipos:

- a) conceitual ou operacional (aquela que se relaciona uma teoria com um método);
- b) orgânica ou perfunctória (a citação é necessária para a compreensão);
- c) evolutiva ou justaposicional (quando o artigo é construído sobre as bases da citação);
- d) confirmativa ou negativa (MORAVCSIK; MURUGESAN apud RODRIGUES, 1981, p.8)

Dentro da bibliometria, particularmente a análise de citações permite a identificação e descrição de uma série de padrões na produção do conhecimento científico. Com os dados retirados das citações pode-se descobrir: autores mais citados, autores mais produtivos, elite de pesquisa, frente de pes-

quisa, fator de impacto dos autores, procedência geográfica e/ou institucional dos autores mais influentes em um determinado campo de pesquisa; tipo de documento mais utilizado, idade média da literatura utilizada, obsolescência da literatura, procedência geográfica e/ou institucional da bibliografia utilizada; periódicos mais citados, “core” de periódicos que compõem um campo.

A análise de citações tem origem no século XVII, com a função primária de “[...] promover uma relação entre dois documentos” (FORESTI, 1989, p. 16). A técnica básica de contar referências é utilizada pela primeira vez em 1927, por P. Gross e E. Gross, depois por Allan, em 1929 e, a seguir, por Gross e Woodford, em 1931. Na década de 1960, com a utilização do computador, a técnica ganha novo fôlego, sendo o ano de 1963 considerado um grande marco, com o surgimento do primeiro índice de citações, o Science Citation Index (SCI) por Eugene Garfield, fundador do *Institute of Scientific Information* – ISI. O autor explica que esse índice é fruto de uma idéia surgida em 1955, e que foram necessários oito anos para se formular as bases teóricas e conceituais da análise de citações, buscando fundamentações nos trabalhos de sociologia da ciência de Merton e de comunicação científica de Crawford, Griffith e Crane (GARFIELD, 1978, p. 98, tradução nossa).

Um conceito extremamente relevante na análise de citações é o de fator de impacto. Formulado por Garfield para constar das análises realizadas pelo ISI, esse conceito consiste em “[...] dividir o número total de citações obtidas por um periódico em um ano qualquer pelo número de artigos publicados naquele ano” (RODRIGUES, 1981, p. 10). A bibliometria, ao longo de sua evolução, contudo, acabou se apropriando do conceito para a análise de autores, correlacionando índices absolutos de citação à quantidade de trabalhos citados. Assim, o fator de impacto é a divisão do número de citações recebidas por um autor dividido pelo número de trabalhos de receberam pelo menos uma citação. Com esse índice, se quer identificar autores que, apesar de terem tido pouca produção, produziram um material muito significativo, isto é, que

receberam muitas citações, em oposição a autores que podem ter tido muitas citações porque publicaram muitos trabalhos, mas cada um desses trabalhos isoladamente com pouca relevância no campo científico. O uso do fator de impacto para a avaliação da produção científica continua a ser atual, algumas vezes relacionado a outros como obsolescência e a idade das referências (MEADOWS, 1999, p. 85-99; STREHL, 2005).

Outro conceito relevante é o de frente de pesquisa, que correlaciona os índices absolutos de citação obtidos por cada autor com a data dos trabalhos publicados por cada autor. Assim, para a contagem da frente de pesquisa, só são contabilizados os trabalhos mais recentes desse autor. O período de tempo considerado relevante para se determinar se o trabalho é recente ou não varia conforme o objetivo que se quer atingir. Por exemplo, pode-se definir que se quer identificar a frente de pesquisa para cinco anos. Neste caso, contabiliza-se apenas as citações recebidas referentes a trabalhos publicados nos últimos cinco anos. Entram para a frente de pesquisa os autores que tiverem pelo menos cinco citações, pois a condição para um autor fazer parte da frente de pesquisa é receber pelo menos uma citação por ano.

Também atualmente são realizados estudos desse tipo. Um exemplo são os trabalhos de Mostafa que consistem em estudos de citação com o objetivo de se identificar a frente de pesquisa e, com isso, traçar as tendências epistemológicas de um determinado campo de estudos (MOSTAFA; MÁXIMO, 2003; MOSTAFA, 2002).

Uma outra abordagem buscou ir além da contagem de palavras e frequência das publicações e de citações. Trata-se da contribuição de Kessler, que se preocupou também com a relação entre os artigos que geram e recebem citações, numa ordenação multidimensional. Em 1963, o autor desenvolveu as teorias do acoplamento bibliográfico, que dizem respeito à força de associação entre dois ou mais documentos (FORESTI, 1989, p. 15). Um trabalho mais atual voltado para a mesma temática é o de Spinak (1998).

Um último conjunto de leis bibliométricas são aquelas relacionadas com a obsolescência da literatura – Line e Sandison, em 1974 – e vida média da literatura (por Burton e Kleber, em 1960), que foram desenvolvidas para descrever a queda da validade ou utilidade de informações com o decorrer do tempo. Há dois tipos de envelhecimento da literatura científica: clássico ou efêmero (vida média longa ou curta). Nos estudos realizados dentro dessa perspectiva percebeu-se que existem disciplinas com forte componente de literatura clássica (matemática, geologia, botânica), enquanto outras são quase exclusivamente compostas por literatura efêmera (física, engenharia), sendo algumas de caráter intermediário (fisiologia, química). Uma importante linha de estudos se desenvolveu nessa área a partir dos conceitos de sincronia e diacronia (STINTON; LANCASTER, 1987, tradução nossa).

4 Desenvolvimento dos estudos

No Brasil, os estudos bibliométricos proliferaram na década de 1970, principalmente com os estudos realizados no Instituto Brasileiro de Bibliografia e Documentação – IBBD, hoje Instituto Brasileiro de Informação Científica e Tecnológica, IBICT. Os estudos bibliométricos realizados nesse período incidiram sobre a literatura científica de vários campos científicos, desde algumas áreas por inteiro como, por exemplo, Química (CARVALHO, 1975) e Geologia (FIGUEIREDO, 1972), mas também tendo por objeto assuntos bastante específicos dentro de uma área, como a doença de Chagas (CALDEIRA, 1974), a esquistossomose (OLIVEIRA, 1975) ou o cacau (RIBEIRO, 1978). Alguns estudos incidiram sobre a produtividade de autores de alguma instituição de ensino e pesquisa, como, por exemplo, os pesquisadores do Instituto de Ciências Biológicas da UFMG (CARVALHO, 1976). Há, ainda, estudos realizados sobre um aspecto específico da bibliometria, como a obsolescência (BARBOZA, 1978), a frente de pesquisa (BRAGA, 1972) ou a otimização de coleções (FOLLY, 1976). Conforme Bufrem (1996),

neste período, das 62 dissertações defendidas no IBICT, 18 consistem em estudos bibliométricos, o que corresponde a 29,03% dos estudos realizados.

Ao longo da década de 1980 houve uma queda no interesse pela bibliometria, tanto no Brasil como no exterior. No início dos anos 1990, com as possibilidades do uso do computador, voltou a haver um grande interesse na exploração das metodologias quantitativas. Na verdade desde a primeira *International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval*, na Bélgica, em 1987, essa disposição já havia se manifestado.

Entre os aspectos relacionados ao crescimento da bibliometria na agenda de pesquisa está o surgimento de algumas subdisciplinas ou subcampos da bibliometria (TAGUE-SUTCLIFFE, 1992, p. 1, tradução nossa). Um deles é a informetria, termo utilizado pela primeira vez na Alemanha, por Nacke, em 1979. Outro é a cientometria, também conhecida no Brasil como cienciometria, termo popularizado pelo periódico húngaro de mesmo nome, fundado em 1977 por Braun. Vários autores se preocuparam com a caracterização de cada um deles (TAGUE-SUTCLIFFE, 1992, tradução nossa; LE COADIC, 1996, p. 52). Uma caracterização bem consistente foi proposta por McGrath em 1989, para quem a bibliometria, a cienciometria e a informetria são subdisciplinas que se assemelham por serem métodos quantitativos, mas se diferenciam quanto ao objeto de estudo, as variáveis, os métodos específicos e os objetivos. Apenas para se ter uma idéia dessas diferenças, o autor argumenta que o objeto de estudo da primeira são livros, documentos, revistas, artigos, autores e usuários; da segunda, disciplinas, assuntos, áreas e campos; e da terceira, palavras, documentos e bases de dados (MACIAS-CHAPULA, 1998, p. 135). Mais recentemente, um outro subcampo surgiu, a webometria. Um trabalho que se utiliza da mesma estrutura de McGrath propõe que o objeto de estudo da webometria são os sítios na *world wide web* (VANTI, 2002, p. 160).

Existe atualmente uma rede de pesquisadores envolvidos com o desenvolvimento e a aplicação de métodos e técnicas bibliométricas. Entre os prin-

cipais grupos de pesquisa destacam-se o BIRG (de Sydney, Austrália), o CEST (de Berna, Suíça), o CINDOC (de Madrid, Espanha), o CIS (de Copenhagen, Dinamarca), o CRRM (de Marselha, França), o CWTS (de Leiden, Holanda), o FhG-ISI (de Karlsruhe, Alemanha), o Inforsk (de Umeå, Suécia), o OST (de Montreal, Canadá), a OST (de Paris, França), o SPRU (de Sussex, Inglaterra), e o ISI *Research Service Group* (de Filadélfia, EUA). A cada dois anos realiza-se a *International Conference on Scientometrics and Informetrics*, merecendo destaque ainda os periódicos *Bibliometric Notes*, *Cybermetrics* e *Scientometrics*, especializados no assunto, além de outros 14 que publicam trabalhos relacionados ao assunto (MACIAS-CHAPULA, 1998, p. 138). Entre os principais indicadores de ciência e tecnologia produzidos atualmente destacam-se *Science and Technology Indicators* (EUA), o *Science & Technology Indicators* (Ásia), o *S&T Indicators for de European Research Area* (Europa) e o *Main Science and Technology Indicators* (dos países da OECD). A história da elaboração de indicadores registra ainda, contudo, a existência dos manuais Frascati, de Oslo e de Canberra como importantes momentos de busca de consensos internacionais sobre os índices de inovação científica e tecnológica (SPINAK, 1998, p. 141).

As preocupações atuais dos pesquisadores em bibliometria caminham para direções distintas. Há um grupo particularmente preocupado com o aperfeiçoamento das fórmulas que expressam as leis. É o caso, por exemplo, dos estudos de Burrell (1992) que analisam a possibilidade de uso da curva de Leimkuhler, do índice de Gini e da distribuição de Pareto na análise das distribuições dos valores bibliométricos. Ou, então, de Urbizagástegui Alvarado (2004), que postula o uso da distribuição de Poisson, formulada ainda no século XIX para estimar a probabilidade de ocorrência de um evento durante um período determinado, para a análise da produtividade de autores. Existem ainda vários outros casos (EGGHE; ROUSSEAU, 1990; SICHEL, 1992; PAO, 1992).

Um outro grupo de pesquisadores passou a questionar a validade de estudos exclusivamente quantitativos, colocando em questão a existência da bibliometria (e suas subdisciplinas) como disciplina científica e propondo que ela deveria passar a ser uma técnica de suporte a ser somada a outras na realização de estudos concretos (WHITE; WELLMAN; NAZER, 2004; TRZESNIAK, 1998). A consequência imediata deste tipo de questionamento foi a utilização que se passou a fazer, cada vez com maior intensidade, do instrumental das ciências sociais (MATTOS, 2004)¹. A evolução dos estudos em produção científica, assim, assistiu à conversão da bibliometria, de um campo de pesquisa, em técnica – uma técnica útil, que deve ser adotada em conjunto com métodos qualitativos fornecidos pelas ciências sociais. Essa parece ter sido, conforme discussão empreendida por CARDOSO (1996), a evolução que marcou, como um todo, a Ciência da Informação:

Lidar com o grande volume e a diversificação de informações registradas em variadas formas, com vistas à sua mais ampla difusão, foi o imperativo condicionante da ciência da informação. Fortemente influenciadas pelas ciências empíricas, as primeiras manifestações desse campo embrionário pretendiam estabelecer leis universais que representassem o fenômeno informacional, daí a recorrência a modelos matemáticos (teoria da informação), físicos (entropia) ou biológicos (teoria epidemiológica). [...] Na década de setenta entra em cena um personagem que redireciona o enfoque da ciência da informação: o *usuário*. [...] Com a presença dos usuários, as ciências humanas e sociais passam a contribuir também, com seus métodos e práticas para a composição dessa ciência emergente. (CARDOSO, 1996, p. 73-74).

Assim, atualmente, muitos trabalhos têm se servido de técnicas bibliométricas aliadas a outros referenciais e métodos. São exemplos dessa nova produção os trabalhos que estudam a historicidade da produção científica, a partir de conceitos da arqueologia do saber de Foucault (ALVARENGA, 1996); a região geográfica como fator interveniente na produção científica (TARGINO, 1998); a identidade dos pesquisadores, em aspectos relacionados à carreira, motivações, produtividade, qualidade da produção, colaboração (MEADOWS, 1999; LACEY, 1998; WITTER, 1997); os conceitos de

centro e periferia como base teórica para a compreensão da comunicação científica (MUELLER; OLIVEIRA, 2003). São trabalhos que se utilizam de dados bibliométricos mas que realizam uma leitura desses dados à luz de elementos do contexto sócio-histórico em que a atividade científica é produzida. É o caso, por exemplo, do Centro de Estudos Informétricos de Copenhague que busca

Uma nova abordagem para a área, qual seja, a combinação de teorias e metodologias avançadas de recuperação da informação com o estudo científico dos fluxos de informação. O Centro objetiva aplicar métodos bibliométricos não somente em estudos cienciométricos e em avaliações de pesquisa científica e tecnológica, mas também na análise de suas relações sociais, econômicas etc., ampliando as análises bibliométricas tradicionais [...] (WORMELL, 1998, p. 210).

Diversas frentes de estudo são levadas a termo na atualidade com essa proposta. Há, por exemplo, estudos de usuários feitos com o auxílio de técnicas bibliométricas. É o caso do estudo de Oliveira (2004), que analisa a possibilidade de aquisição de itens para uma biblioteca universitária a partir de indicativos de necessidades de usuários obtidos com o estudo bibliométrico das referências bibliográficas de teses e dissertações. Ou, então, de estudos de uso de coleções também a partir da aplicação da bibliometria às referências de teses e dissertações defendidas em uma instituição específica (COITO et al, 2002; PAULA et al., 2002). E, ainda, estudos que evidenciam os processos de comunicação científica e formação de colégios invisíveis a partir da análise de citação (MELLO, 1996; NORONHA, 1998).

Também o uso de dados bibliométricos como indicadores da produção científica passou a ser cada vez mais freqüente, diante do conjunto de ações que vêm sendo desenvolvidas no sentido de dispor desses indicadores para o planejamento nacional das atividades de pesquisa científica (MUGNANI; JANNUZZI; QUONIAM, 2004, p. 123; GUEDES; BORSCHIVER, 2005; KRZYZANOWSKI; FERREIRA, 1998). Ou então para a análise do desenvolvimento da pesquisa científica e tecnológica dentro de uma instituição

específica (PENTEADO FILHO et al., 2002; ROUSSEAU, 1998), da análise dos periódicos de uma área específica (CAMPOS, 2003; ELKIS, 1999) ou da produtividade de pesquisadores (POBLACIÓN; NORONHA, 2002). Há ainda estudos bibliométricos para determinar o “léxico básico” de um campo, com possibilidades de aplicação inclusive para a construção de tesouros e linguagens documentárias desse campo (ROBREDO; CUNHA, 1998).

Percebe-se, dessa forma, que a bibliometria vem se consolidando como método de estudo dentro de uma preocupação com leituras mais ricas da realidade, mais atentas às reivindicações contemporâneas do pensamento complexo (MORIN, 1987; 2001) bem como às críticas ao projeto da *mathesis universalis* (a consideração exclusiva daquilo que é mensurável como cientificamente relevante) do projeto positivista de ciência (SANTOS, 1986). A imensa popularidade que a bibliometria passou a ter após as possibilidades digitais foi acompanhada, portanto, de uma série de avanços relativos ao aperfeiçoamento das leis bibliométricas mas, sobretudo, pela busca de fundamentação teórica e conceitos oriundos dos contextos concretos em que os fenômenos informacionais ocorrem (BORGMAN; FURNER, 2002, tradução nossa).

Bibliometrics: historical development and current issues

ABSTRACT

In this study, the field of knowledge known as bibliometrics is analyzed. First, the three basic laws in the field are reviewed: Lotka's Law of productivity of authors; Bradford's Law of journal dispersion; and Zipf's Law of word frequency, along with their developments and applications in recent times. The epidemic theory and citations analysis are also reviewed, as well as their contemporary applications, like, for example, advanced research studies and the impact factor. Finally, it analyzes the development of this field of studies in Brazil and current trends, emphasizing the approaches that seek to broaden the scope of studies that have integrated bibliometric methods and distinct theories.

KEYWORDS: Bibliometrics. Lotka's Law. Bradford's Law. Zipf's Law. Citation analysis. Impact factor.

Bibliometria: evolução histórica y cuestiones actuales

RESUMEN

Se promueve un análisis del campo del conocimiento intitulado bibliometria. Inicialmente, son vistas las tres leyes fundadoras del campo, la ley de productividad de autores de Lotka, la ley de dispersión de periódicos de Bradford y la ley de frecuencia de palabras de Zipf, bien como sus desdoblamientos y aplicaciones contemporáneas, como por ejemplo los estudios de frente de pesquisa y factor de impacto. Por fin, se analiza el desenvolvimiento de este campo de estudios en el Brasil y las tendencias actuales, con destaque para los abordajes que buscan ampliar el escopo de los estudios realizados integrando los métodos bibliométricos a distintos cuerpos teóricos.

PALABRAS-CLAVE: Bibliometria. Ley de Lotka. Ley de Bradford. Ley de Zipf. Análisis de citaciones. Factor de impacto.

Referências

ALVARENGA, Lídia. **A institucionalização da pesquisa educacional no Brasil.** 1996. Tese (Doutorado) – Faculdade de Educação da Universidade Federal de Minas Gerais, Belo Horizonte, 1996.

BARBOZA, Elza Maria Ferraz. **Estudo comparativo dos métodos usados para medir a obsolescência da literatura científica.** Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1978.

BORGMAN, Christine; FURNER, Jonathan. Scholarly communication and bibliometrics. **Annual Review of Information Science and Technology**, New Jersey, v. 36, p. 3-72, 2002.

BRAGA, Gilda Maria. **Relações bibliométricas entre a frente de pesquisa e revisões de literatura: estudo aplicado à ciência da informação.** 1972. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Biblioteconomia e Documentação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1972.

BUFREM, Leilah Santiago. **Linhas e tendências metodológicas na produção acadêmica discente do mestrado em Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia – Universidade Federal do Rio de Janeiro (1972-1995).** 1996. Tese (Concurso de professor titular) – Universidade Federal do Paraná, Curitiba, 1996.

BURRELL, Quentin. The Gini index and the Leimkuhler curve for bibliometric processes. **Information processing & management**, Oxford, v. 28, n. 1, p. 19-33, 1992.

CALDEIRA, Paulo da Terra. **Crescimento da literatura brasileira de doença de Chagas**. 1974. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Biblioteconomia e Documentação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1974.

CAMPOS, Mauro. Conceitos atuais em bibliometria. **Arquivos Brasileiros de Oftalmologia**, São Paulo, v. 66, n. 1, p. 18-21, 2003.

CARDOSO, Ana Maria Pereira. Pós-Modernidade e informação: conceitos complementares? **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 63-79, jan./jul. 1996.

CARVALHO, Maria de Lourdes. **Análise de citações dos artigos de periódicos publicados pelos professores do Instituto de Ciências Biológicas da UFMG, no período de 1968 a 1973**. 1976. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1976.

CARVALHO, Maria Martha. **Análises bibliométricas da literatura de química no Brasil**. 1975. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Biblioteconomia e Documentação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1975.

CAVALCANTI, Ilce G. M. **Padrões de citação em comunicação: análise das dissertações apresentadas à ECO/UFRJ**. 1989. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1989.

COITO, M.I. et al. Avaliação do uso da coleção de periódicos por docentes e alunos da Faculdade de Ciências Farmacêuticas – Unesp. In: **SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS**, 12., 2002, Recife. **Anais...** Recife: UFPE, 2002.

COOPER, Randolph; BLAIR, David; PAO, Miranda. Communicating MIS research: a citation study of journal influence. **Information processing & management**, New York, v. 29, n. 1, p. 113-127, 1993.

DONOHUE, Joseph. **Understanding scientific literatures: a bibliometric approach**. Cambridge: The MIT Press, 1973.

EGGHE, L.; ROUSSEAU, R. **Introduction to informetrics: quantitative methods in library, documentation and information science**. Amsterdam: Elsevier Science, 1990.

ELKIS, Hélio. Fatores de impacto de publicações psiquiátricas e produtividade científica. **Revista Brasileira de Psiquiatria**, São Paulo, v. 21, n. 4, dez 1999.

FIGUEIREDO, Laura Maia. **Distribuição da literatura geológica brasileira: estudo bibliométrico**. 1972. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Bibliografia e Documentação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1972.

FIGUEIREDO, Nice. **Tópicos modernos em Bibliometria**. Brasília: Associação dos Bibliotecários do Distrito Federal, 1977.

FOLLY, Eneyda. **Otimização da coleção de periódicos da Biblioteca de Veterinária da UFF**. 1976. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1976.

FONSECA, Edson Nery da (Org). **Bibliometria: teoria e prática**. São Paulo: Cultrix, Ed. da USP, 1986.

FORESTI, Nôris. **Estudo da contribuição das revistas brasileiras de biblioteconomia e ciência da informação enquanto fonte de referência para a pesquisa**. 1989. Dissertação (Mestrado) – Departamento de Biblioteconomia da Universidade de Brasília, UnB, Brasília, 1989.

GARFIELD, Eugene. **Citation indexing: its theory and application in science, technology na humanities**. New York: John Wiley and Sons, 1978.

GUEDES, Vânia; BORSCHIVER, Suzana. 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: CINFOM – ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6., 2005, Salvador. **Anais...** Salvador: ICI/UFBA, 2005.

KRZYŻANOWSKI, Rosaly; FERREIRA, Maria Cecília. Avaliação de periódicos científicos e técnicos brasileiros. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 165-175, maio/ago. 1998.

LACEY, Hugh. **Valores e atividade científica**. São Paulo: Discurso Editorial, 1998.

LE COADIC, Yves-François. **A ciência da informação**. Brasília: Briquet de Lemos, 1996.

MACIAS-CHAPULA, Cesar. O papel da informetria e da cienciométrica e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998.

MAIA, Maria José da Fonseca. A unicidade da lei de Bradford. 1980. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, 1980.

MATTOS, Lincoln. Bibliometria: a metodologia acadêmica convencional em questão. **ERA-eletrônica**, São Paulo, v. 3, n. 2, jul./dez. 2004. Disponível em: <www.era.com.br/artigos/2199.pdf>. Acesso em: 17 fev. 2006.

MEADOWS, Arthur Jack. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MELLO, Paula Maria Abrantes. A citação bibliográfica no contexto da comunicação: um estudo exploratório na área de botânica. **Ciência da Informação**, Brasília, v. 25, n. 3, p. 387-395, set./dez. 1996.

MORIN, Edgar. **O método III: o conhecimento do conhecimento**. Lisboa: Europa-America, 1987.

MORIN, Edgar. **O método II: a vida da vida**. Porto Alegre: Sulina, 2001.

- MOSTAFA, Solange. Citações epistemológicas na educomunicação. **Comunicação & Educação**, São Paulo, v. 8, n.24, p. 15-28, 2002.
- MOSTAFA, Solange; MÁXIMO, Luís. A produção científica da ANPED e da Intercom no GT de Educação e Comunicação. **Ciência da Informação**, Brasília, v. 32, n. 1, p. 96-101, jan./abr. 2003.
- MUELLER, Suzana; OLIVEIRA, Hamilton. Autonomia e dependência na produção da ciência: uma busca conceitual para estudar as relações na comunicação científica. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n.1, p. 58-65, jan./jun. 2003.
- MUGNANI, Rogério; JANNUZZI, Paulo; QUONIAM, Luc. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 123-131, maio/ago. 2004.
- NICHOLAS, David; RITCHIE, Maureen. **Literature and bibliometrics**. London: Clive Bingley, 1978.
- NORONHA, Daisy Pires. Análise das citações das dissertações de mestrado e teses de doutorado em saúde pública (1990-1994): estudo exploratório. **Ciência da Informação**, Brasília, v. 27, n. 1, p. 66-75, jan./abr. 1998.
- OLIVEIRA, Elias. Identificando necessidades de acesso de usuários de um departamento acadêmico: uma abordagem bibliométrica. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 126-131, jan./abr. 2004.
- OLIVEIRA, Margarida Pinto. **Estudo bibliométrico da literatura brasileira de esquistossomose**. 1975. Dissertação (Mestrado) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro 1975.
- PAO, Miranda Lee. Global and local collaborators: a study of scientific collaboration. **Information processing & management**, Oxford, v. 28, n. 1, p. 99-109, 1992.
- PAULA, T.M.B. et al. Avaliação das citações em dissertações e teses do departamento de ciência política da FFLCH/USP defendidas no ano de 2000: um estudo de uso. In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS, 12., 2002, Recife. **Anais...** Recife: UFPE, 2002.
- PENTEADO FILHO, Roberto et al. Aplicação da bibliometria na construção de indicadores sobre a produção científica da Embrapa. In: CONGRESSO ANUAL DA SOCIEDADE BRASILEIRA DE GESTÃO DO CONHECIMENTO, 1., 2002, São Paulo. **Anais...**São Paulo, 2002.
- PINHEIRO, Lena Vania R. **Lei de Bradford**: uma reformulação conceitual. 1982. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1982.
- POBLACIÓN, Dinah; NORONHA, Daisy. Produção das literaturas “branca” e “cinzenta” pelos docentes/doutores dos programas de pós-graduação em ciência da informação no Brasil. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 98-106, maio/ago. 2002.

- PRICE, Derek de Solla. **O desenvolvimento da ciência: análise histórica, filosófica, sociológica e econômica.** Rio de Janeiro: Livros Técnicos e Científicos, 1976.
- RAO, I.K. Ravichandra. **Métodos quantitativos em biblioteconomia e ciência da informação.** Brasília: Associação dos Bibliotecários do Distrito Federal, 1986.
- RIBEIRO, Eliane de Oliveira. Estudo sobre a literatura do cacau no período 1950-1975. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1978.
- ROBREDO, Jaime; CUNHA, Murilo. Aplicação de técnicas infométricas para identificar a abrangência do léxico básico que caracteriza os processos de indexação e recuperação da informação. **Ciência da Informação**, Brasília, v. 27, n. 1, p. 11-27, jan./abr. 1998.
- RODRIGUES, Maria da Paz. **Estudo das citações constantes das dissertações de mestrado em ciência da informação do IBICT/UFRJ.** 1981. Dissertação (Mestrado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1981.
- ROUSSEAU, Ronald. Indicadores bibliométricos e econométricos para a avaliação de instituições científicas. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 149-158, maio/ago. 1998.
- SANTOS, Boaventura. **Um discurso sobre as ciências.** Porto: Afrontamento, 1996.
- SICHEL, H.S. Anatomy of the generalized inverse gaussian-poisson distribution with special applications to bibliometric studies. **Information processing & management**, Oxford, v. 28, n. 1, p. 5-17, 1992.
- STINTON, E.R.; LANCASTER, F.W. Synchronous versus diachronous methods in the measurement of obsolescence by citation studies. **Journal of Information Science**, London, v. 13, p. 65-74, 1987.
- STREHL, Letícia. O fator de impacto do ISI e a avaliação da produção científica: aspectos conceituais e metodológicos. **Ciência da Informação**, Brasília, v. 34, n. 1, p. 19-27, jan./abr. 2005.
- SPINAK, Ernesto. Indicadores cientimétricos. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 141-148, maio/ago. 1998.
- TAGUE-SUTCLIFFE, Jean. An introduction to informetrics. **Information processing & management**, Oxford, v. 28, n. 1, p. 1-3, 1992.
- TARGINO, Maria das Graças. **Comunicação científica: o artigo de periódico nas atividades de ensino e pesquisa do docente universitário brasileiro na pós-graduação.** 1998. Tese (Doutorado em Ciência da Informação) – Curso de Pós-Graduação em Ciência da Informação, Universidade de Brasília, Brasília, 1998.
- TRZESNIAK, Piotr. Indicadores quantitativos: reflexões que antecedem seu estabelecimento. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 159-164, maio/ago. 1998.

URBIZAGÁSTEGUI ALVARADO, Rubén. A bibliometria no Brasil. **Ciência da Informação**, Brasília, v. 13, n. 2, p. 91-105, jul./dez. 1984.

_____. Aplicações da distribuição Poisson zero truncada à produtividade de autores. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n. 1, p. 18-33, jan./jun. 2004.

_____. A lei de Lotka na bibliometria brasileira. **Ciência da Informação**, Brasília, v. 31, n. 2, p. 14-20, maio/ago. 2002.

VANTI, Nadia Aurora. 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.

VANZ, Samile. A bibliometria no Brasil: análise temática das publicações do periódico *Ciência da Informação*. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., 2003, Belo Horizonte. **Anais...** Belo Horizonte: Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação e Biblioteconomia, 2003. 1 CD-ROM.

WHITE, Howard; McCAIN, Katherine. Visualization of literatures. In: WILLIAMS, Martha (Ed). **Annual review of information science and technology**, New Jersey, v.32, 1997.

WHITE, H.; WELLMAN, B.; NAZER, N. Does citation reflect social structure? **Journal of the American Society for Information Science and Technology**, Washington, v. 55, n. 2, p. 111-126, 2004.

WITTER, Geraldina Porto (Org). **Produção científica**. Campinas: Átomo, 1997.

WORMELL, Irene. Informetria: explorando bases de dados como instrumentos de análise. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 210-216, maio/ago. 1998.

Carlos Alberto Araújo

Doutor em Ciências da Informação

Professor Adjunto da ECI/UFMG

E-mail: casalvila@yahoo.com.br