

A FRENTE DE PESQUISA NA LITERATURA SOBRE A PRODUTIVIDADE DOS
AUTORES
RESEARCH FRONT IN THE LITERATURE ABOUT AUTHORS' PRODUCTIVITY

Rubén Urbizagástegui Alvarado
Doutor em Ciência da Informação
Universidade de Califórnia em Riverside
Biblioteca de Ciências
Riverside, CA, USA
ruben@ucr.edu

Resumo

O objetivo deste artigo é identificar, descrever e analisar a frente de pesquisa numa subárea do campo da bibliometria: a produtividade dos autores, conhecida também como a “Lei de Lotka”. Identificaram-se os autores que formam a frente de pesquisa deste campo. Esses autores têm 10 ou mais artigos produzidos, qualificando-os como grandes produtores, e, em alguns casos como produtores moderados. Todos eles tem os trabalhos publicados no idioma inglês, e a grande maioria deles nos periódicos da chamada “corrente principal”, isto é, indexados no Science Citation Index ou no Social Science Citation Index. Verificaram-se a existência de covariação entre o nível de produtividade dos autores, os artigos publicados nos periódicos da chamada “corrente principal” e o idioma da publicação concomitantemente com a frequência de citações. Em outras palavras, para qualificar-se na frente de pesquisa deste campo, o autor tem que ser um grande ou moderado produtor, publicar no idioma inglês e nos periódicos considerados de corrente principal. Fora deste contexto, o uso da informação explicitada na forma de citações fica reduzido a um consumo local.

Palavras-chave: Frente de pesquisa. Produtividade dos autores. Lei de Lotka. Cienciometria. Bibliometria. Informetria.

1 INTRODUÇÃO

O objetivo desta pesquisa é identificar, descrever e analisar a frente de pesquisa numa subárea do campo da bibliometria: a produtividade dos autores, conhecida também como a “lei de Lotka”. É também um intento por desenvolver um modelo que identifique e explique as variáveis que influem na composição das frentes de pesquisa de qualquer área do conhecimento.



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).
DOI 10.5007/1518-2924.2009v14n28p38

Se a publicação de documentos segue um padrão tradicional no processo de comunicação da ciência, essa identificação e descrição da frente de pesquisa podem ser realizadas via análise da literatura produzida sobre o assunto.

Portanto, as redes que se estabelecem por meio das citações num documento publicado poderiam constituir também uma confiável forma de estudar esta subárea científica do campo da bibliometria. Os estudos realizados usando as análises de citações como maneira de compreender melhor a história e estrutura da ciência tem sido descritas por Garfield (1970; 1979), que chama a atenção para o fato de que:

a frequência de citações é uma medida da atividade científica, ou de comunicação sobre a atividade científica. A medida é um desenho sociométrico. Em si mesmo, o número de citações do trabalho de uma pessoa não é uma medida de sua significância ... [e] deve ser usado juntamente com outras escalas para obter qualquer significado útil ou significativo, especialmente se o objeto da avaliação é de alguma forma qualitativo. (GARFIELD, 1977, p. 407).

Também Cawkell (1976) fez observações similares sobre o assunto e não foi gratuitamente que Price (1965) descreveu as redes de conexões na produção de artigos científicos, como obtidas “ligando cada artigo publicado com os outros artigos diretamente associados a eles”. Esta conexão seria estabelecida pela citação de um artigo a outros indicados em suas notas de rodapé ou na bibliografia consultada. Analisava assim a incidência das referências e das citações, ressaltando que:

Apesar de a maioria dos artigos produzidos no ano conterem mais ou menos uma média de referências bibliográficas, a metade dessas referências são referentes à cerca da metade de todos os artigos publicados em anos anteriores. A outra metade das referências liga esses novos artigos a um pequeno número de artigos anteriores, e gera um padrão de múltiplas relações. Assim, cada grupo de novos artigos está ligado a uma seleta pequena parte da literatura científica existente, mas casual e fragilmente conectada à maior parte desta literatura. Como cada ano somente uma pequena parte da literatura mais antiga está entrelaçada a nova produção de artigos, podemos ver esta pequena parte da literatura como uma forma de tecido crescente ou capas epidérmicas, uma ativa frente de pesquisa. (PRICE, 1965, p.149).

Já estudando as colaborações num colégio invisível, voltava a insistir em que

não somente temos verificado que a frente de pesquisa é dominada por um pequeno núcleo de ativos trabalhadores ... senão que também apontamos a possibilidade de que trabalhando juntos, em colaboração, é que ocorre a maior parte da comunicação na frente de pesquisa. (PRICE, 1966, p.1017).

Comunicação que se reforçaria via citações bibliográficas, que também permitiriam identificar os artigos considerados “clássicos”, definidos como aqueles mais frequentemente citados. Este autor insistia que a análise de citações mostrava duas formas de comportamento: a primeira, responsável pela metade das citações mais antigas, às que chamava de arquivísticas, pois evidenciavam ausência de estrutura, eram quase completamente independentes da idade dos artigos mais antigos citados e ocorriam sem um padrão definido. Um segundo tipo, ao qual chamava de frente de pesquisa, como um pequeno número de artigos altamente conectados aos artigos mais recentes e à outra metade da literatura citada por muitas formas e mecanismos. (PRICE, 1969).

Num artigo mais tardio, o autor afirmava que se fossem ordenados 1000 artigos ou cientistas em ordem decrescente de valor, a metade desses valores seriam devidos mais ou menos aos 32 primeiros da lista ordenada; e se tivéssemos um milhão de itens, a metade desses itens seria produzida pelos mil primeiros da lista. No seu entendimento, isto significava que, se era possível cortar aritmeticamente essa população pela metade, então também seria possível contar-se geometricamente tomando-se, por exemplo, a raiz quadrada.

Desta forma, entendia que qualquer população de tamanho N tinha uma elite efetiva de \sqrt{n} . Isto significava que o número de autores prolíficos parecia equivaler à raiz quadrada do número total de autores, que em qualquer campo ou subcampo da ciência seriam os que compõem a chamada “Frente de pesquisa” (PRICE, 1971). Também Garvey (1979: 94) enfatiza que muitos acadêmicos parecem colocar a frente de pesquisa como localizada nos processos de comunicação formais, usando as análises de citação para lhe dar maior especificidade e distinguindo a frente de pesquisa em termos da percentagem de literatura agrupada nos últimos cinco anos, de modo que quanto maior é a percentagem encontrada, mais ativa seria a frente de pesquisa.

Seguindo essas propostas, poucos trabalhos têm sido desenvolvidos no campo da ciência da informação para identificar uma frente de pesquisa. Estes trabalhos estão reduzidos aos realizados por Velke (1970), que estudou os artigos publicados em periódicos como resultados de pesquisas patrocinadas pela Air Force Office of Scientific Research, cobrindo um período de dez anos. Para a contagem das citações a autora ignorou as autocitações e múltiplas citações a um mesmo autor, num mesmo artigo. Identificou que de 462 autores, 15 poderiam ser considerados como constituindo a frente de pesquisa da Air Force Office of Scientific Research. Não obstante, a autora considerou seus resultados um exagero. DeHart e Scott (1991) dão conta dos esforços do ISI em identificar frentes de pesquisa ou áreas de ativa

pesquisa on-line correntes e incluí-las na base de dados SCISEARCH. No entanto, Urbizagástegui (1993) postula que um campo científico é uma matriz de articulações sistemáticas de paradigmas, coexistindo no tempo mostrando uma diacronia e uma sincronia, uma relação de hegemonia e dependência ocorrendo num espaço social histórico; portanto, em qualquer campo científico não existiria apenas uma frente de pesquisa, senão uma multiplicidade de frentes de pesquisa, tantas quantas sejam as subáreas ou linhas de pesquisa que estruturam uma área de estudo. Nessa área - na realidade em qualquer área - algumas frentes de pesquisa seriam dominantes e outras emergentes. O autor constatou essas frentes de pesquisa nas diversas subáreas que compunham as linhas de pesquisas existentes na bibliometria brasileira.

Não obstante, nenhum trabalho tem sido desenvolvido estudando um campo homogêneo e sem divisões internas como a que oferece a produtividade dos autores. Existiria esta frente de pesquisa nos estudos da produtividade de autores ou lei de Lotka? E se existisse, que autores a constituiriam? Será que o número de artigos publicados pelos autores está correlacionado à quantidade de citações recebidas? Será possível que a categoria da revista onde o artigo é publicado esteja correlacionada às citações recebidas? Supondo que essa frente de pesquisa exista, será que o idioma predominantemente usado para comunicar os trabalhos influi na integração dos autores nessa frente de pesquisa? Esta pesquisa pretende responder a estas interrogações.

2 METODOLOGIA

Como unidades de análises foram tomadas cada uma das citações bibliográficas contidas em cada um dos documentos publicados sobre a produtividade dos autores, conhecida também como a “Lei de Lotka”, cobrindo o período de 1922 a 2003.

Na primeira etapa, identificaram-se os autores contribuindo com documentos publicados nesta área. Para lograr essa identificação foi realizada uma busca usando a estratégia Dial Index em todas as bases de dados de DIALOG, com os termos Lotka?(5n) Law?, encontrando-se 50 bases de dados que pelo menos continham um artigo sobre o assunto pesquisado. As referências bibliográficas foram depois trasladadas ao PROCITE 5.0 para a elaboração de

uma base de dados específica sobre o assunto. Também foram realizadas buscas no Information Science Abstract (ISA), Library Literature (LL) e Library and Information Science Abstract (LISA). Com esta busca minuciosa foi produzida uma bibliografia analítica sobre a Lei de Lotka listando um total de 390 referências bibliográficas contendo artigos de periódicos, monografias, capítulos de livros, comunicações em congressos, literatura gris, cartas dirigidas a editores de revistas especializadas em biblioteconomia e ciência da informação. (URBIZAGÁSTEGUI, 2005).

Na segunda etapa, cada uma das citas de cada um dos 390 documentos identificados foram isoladas, cortadas e sistematicamente coladas em fichas de 3.5 x 4.5 cm. Em seguida essas fichas foram organizadas alfabeticamente segundo os autores citados e contabilizados para análise e avaliação.

Para a mensuração dos dados foi usada a contagem da frequência das citas, seguindo as propostas de Price (1971) e executadas por Braga (1973), Christovão (1978), Rodrigues (1981), Velke (1990) e Urbizagástegui (1993). Nesta parte do trabalho foram atribuídos pontos a cada um dos autores citados, isto é, a opção foi pelo sistema de contagem completa. Isto significa, ainda, que os múltiplos autores de um único artigo receberam um ponto por cada citação, de forma que o número de pontos obtidos por um autor representa o número de vezes que este foi citado em todos os documentos estudados, independentemente de que no trabalho analisado tenha aparecido como primeiro, segundo, terceiro autor e assim por diante. Na contagem das citas foram eliminados e não considerados para a atribuição de pontos as autocitações e as múltiplas citações a um mesmo autor e artigo feitas num mesmo documento analisado e indicadas na clássica forma de *idem*, *op. cit.*, ou *ibidem*.

Uma vez identificados os autores que conformam a frente de pesquisa, a mesma foi submetida à análise multivariada para encontrar uma explicação lógica da sua constituição. Como se esperava que existisse associação entre as variáveis dependentes e independentes envolvidas na pesquisa, as relações entre essas variáveis foram exploradas usando a técnica de análise multivariada de contingência. Para avaliar o ajuste dos dados observados e esperados, usou-se o teste qui-quadrado ao 0.001 nível de significância. Os dados foram controlados e analisados usando-se o pacote estatístico SSPS 15.0 na versão para Windows.

3 MODELO TEÓRICO E HIPÓTESES

Qualquer campo científico é um campo dinâmico que se forma como consequência da competência existente no nível intelectual, onde não todos os autores têm a mesma oportunidade de publicar um artigo. Aqueles autores munidos de maior capital cultural terão maiores possibilidades de publicar e, quanto mais artigos publicarem também terão maiores possibilidades de serem citados, isto é, sua visibilidade será maior. Desta forma é possível que exista uma correlação natural entre o número de artigos publicados por um autor no campo, a frequência de citações feitas a esses documentos publicados e a inclusão desses autores numa frente de pesquisa.

Não obstante, é evidente que um grupo desses autores e documentos escapará a esta tendência. Estes documentos serão os chamados “clássicos”, porque acumularão maior frequência de citações. Possivelmente esta acumulação de citações se deva ao fato de que esses autores foram os que chamaram a atenção para o fenômeno estudado ou propuseram novas formas de ver esse fenômeno, ou ainda, revisaram o estado-da-arte do mesmo.

Por outro lado, sabemos que a produção de artigos é veiculada através de periódicos especializados da área, portanto, quando esses são categorizados como de “corrente principal”, maiores serão as frequências de citações dos autores que publicaram nesses periódicos e, vice-versa, quando esses periódicos não são considerados como de “corrente principal”, menores serão as possibilidades de citações dos autores que publicaram nos mesmos.

Por último, como parece haver evidências de que “o inglês é a língua principal da comunicação científica ... [e que] não sendo um escritor de fala inglesa ou familiarizado com o inglês podem ter efeitos negativos sobre a publicação e disseminação dos trabalhos”, como afirma Baldauf (1986), aquele documento que não é publicado em inglês dificilmente será consultado por uma comunidade que não tem familiaridade com um idioma diferente ao seu e dificilmente será citado por uma comunidade que não lê nem entende outra língua que não seja a sua. Daqui se depreende que os documentos mais freqüentemente citados serão aqueles escritos em inglês, idioma em que se publica a maioria dos trabalhos do tipo da Lei de Lotka, portanto, haverá uma correlação positiva entre o idioma e a frequência de citações dos autores, sendo estes hegemônizados pelo idioma hegemônico (o inglês), relegando da frente de pesquisa os autores e trabalhos publicados em outros idiomas. O esquema do modelo probabilístico recursivo proposto nesta pesquisa está resumido na Figura 1.

Para atingir os objetivos desta pesquisa, as variáveis envolvidas no modelo teórico são definidas da seguinte maneira:

A frente de pesquisa é uma variável dependente. Refere-se a todos os autores identificados mediante a contagem de citações (menções) nos artigos publicados sobre a produtividade dos autores, ou lei de Lotka, no período estudado. O clássico \sqrt{n} proposto por Price (1963).

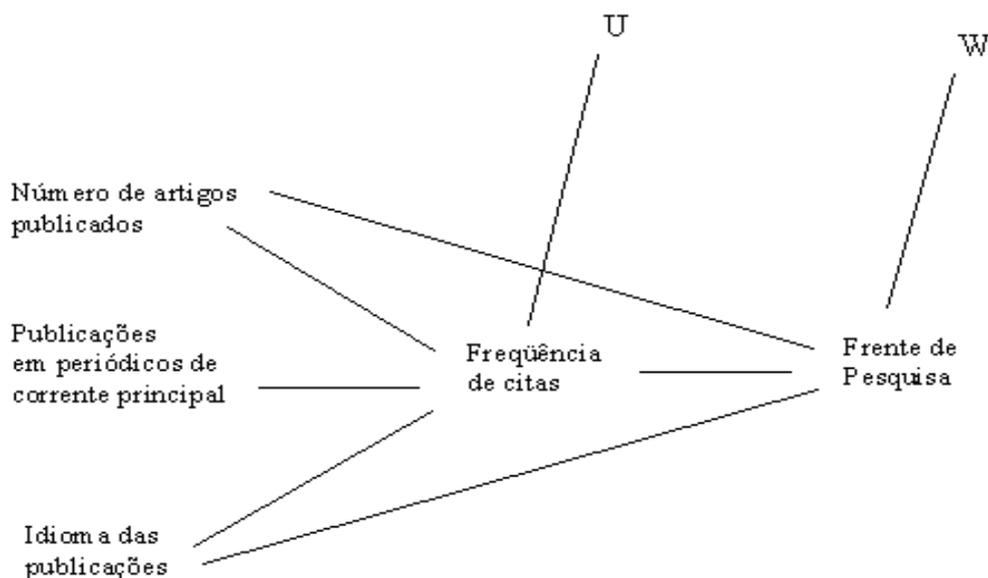


Figura 1: Modelo probabilístico da formação de frentes de pesquisa

A frequência de citações é uma variável interveniente endógena e se refere ao número de menções recebidas por cada autor em cada artigo publicado, que integra a base de dados estudada no período da pesquisa, podendo-se elaborar a primeira hipótese da seguinte forma:

Ho: Não há associação significativa entre a frequência de citações e os autores colocados na frente de pesquisa.

Ha: Há associação significativa entre a frequência de citações e os autores colocados na frente de pesquisa.

O número de artigos publicados é uma variável exógena independente e se refere à quantidade total de artigos publicados sobre o assunto por cada autor, no período estudado, daí podendo-se elaborar a segunda hipótese da seguinte forma:

Ho: Não há associação significativa entre o número de artigos publicados por cada autor e a frequência de citações.

Ha: Há associação significativa entre o número de artigos publicados por cada autor e a frequência de citações.

A publicação em periódicos da corrente principal é uma variável exógena independente e se refere à publicação periódica específica onde os artigos foram publicados. Supõe-se que se um artigo é publicado numa revista da corrente principal da área, ou seja, de grande impacto, suas probabilidades de ser visto, recuperado e incorporado na forma de citações em novos artigos será maior do que um artigo não publicado em periódicos dessa corrente principal. Entende-se por periódicos da corrente principal aqueles periódicos que são indexados no Science Citation Index, Social Science Citation Index, e Humanities Citation Index e que se supõe são consumidos (lidos) por uma “comunidade internacional”. A partir daí pode-se elaborar a terceira hipótese:

Ho: Não há associação significativa entre os documentos publicados em periódicos da corrente principal e as frequências das citações.

Ha: Há associação significativa entre os documentos publicados em periódicos da corrente principal e as frequências das citações.

O idioma de publicação é uma variável exógena independente e se refere à língua em que foi publicado o artigo. Supondo-se que se o inglês é a língua da ciência, este será o idioma monopolizador nas citações e nos autores que integrem a frente de pesquisa, podendo-se elaborar a quarta hipótese:

Ho: Não há associação significativa entre o idioma de publicação do artigo e as frequências das citações.

Ha: Há associação significativa entre o idioma de publicação do artigo e as frequências das citações.

U e W são variáveis exógenas residuais que têm efeitos inexplicáveis sobre a frequência de citações e a frente de pesquisa, mas seus efeitos não serão analisados nesta pesquisa.

4 RESULTADOS

No período estudado foram encontrados 390 documentos produzidos por 377 autores diferentes. Esses 377 autores realizaram um total de 6.530 citações aos trabalhos de 2.221 autores diferentes. Isso representa uma média de 16.7 citações por documento. Tomando-se a raiz quadrada do total de 2.221 autores diferentes identificados nesta pesquisa, observou-se que a frente de pesquisa sobre a produtividade dos autores ou Lei de Lotka estaria conformada

por 47 autores, os quais acumularam 23 ou mais citações cada um. Porém, dentre esses autores encontrou-se alguns que nada tinham a ver com a produção de trabalhos sobre este assunto.

Certamente autores como Allan Pritchard (23 citações) foram citados por introduzir o termo bibliometria no campo da ciência da informação, mas seu interesse não vem sendo dirigido especificamente à Lei de Lotka nem à produtividade dos autores. Igualmente, D. K. Simonton (25 citações) tem sido citado por suas preocupações com a “inteligência” como um sistema de capacidades cognitivas que permite a qualquer indivíduo adaptar-se a seu ambiente e produzir numa gama de possibilidades, inclusive a produção de artigos. R. K. Merton (30 citações) e J. R. Cole (42 citações), por suas preocupações com a sociologia da ciência e a característica estratificada da produção científica. S. D. Haitun (35 citações) e N. L. Johnson & Kotz (41 citações), por seus textos de estatística que serviram de embasamento para a aplicação dos modelos estatísticos. William Goffman (46 citações), por suas aplicações da Lei de Bradford e à teoria epidêmica da informação. G. K. Zipf (58 citações), como um dos iniciadores da lei que leva seu nome, bem como S. C. Bradford (73 citações) e Brookes (56 citações), pelas discussões relacionadas à Lei de Bradford. Eugene Garfield (98 citações), pelo uso extensivo do Science Citation Index e o uso da análise de citações nas práticas bibliométricas.

Pode-se notar que esses autores, embora sejam importantes nos estudos quantitativos da ciência e da sociologia da ciência, e de terem aportado escritos ou modelos matemáticos para o estudo da bibliometria, cienciometria e informetria, não tem contribuído com a produção de trabalhos sobre a produtividade científica dos autores, ou Lei de Lotka. Os escritos desses autores tem sido usados basicamente para a construção do referencial teórico e para a revisão da literatura sobre o assunto em pauta.

Como nesta pesquisa o interesse é identificar uma ativa frente de pesquisa sobre a Lei de Lotka, considerou-se prudente separar desse total de 6.530 citações, aquelas referidas unicamente a esses 390 documentos produzidos pelos 377 autores, e a esses autores diferentes identificados como produtores da literatura sobre a Lei de Lotka, no período de 1922 a 2003. Isto é, o que importa é medir a intensidade com que esses documentos foram citados e identificar nessas citações uma ativa frente de pesquisa.

Feita essa separação, encontrou-se 2.186 citações aos documentos produzidos por somente 203 autores, portanto, 174 autores (46%) não foram citados no período pesquisado. Essas quantidades indicam que houve uma média de 10.8 citações por documento. Tomando-se a raiz quadrada desses 203 autores verificou-se que a frente de pesquisa sobre a Lei de Lotka

estaria conformada pelos 14.25 autores, contudo, um pouco mais ousados, pode-se tomar os 15 autores que acumularam 41 ou mais citações cada um no período pesquisado. A Tabela 1 mostra os autores que compõem essa frente de pesquisa e o número correspondente de citações recebidas.

Autores	No de citações
Lotka, A. J.	231
Price, D. J. de S.	190
Pao, M. L.	98
Bookstein, A.	76
Nicholls, P. T.	68
Schorr, A. E.	64
Simon, H. A.	60
Coile, R. C.	56
Vlachy, J.	56
Sichel, H. S.	54
Potter, W. G.	47
Egghe, Leo	46
Allison, P. D.	46
Murphy, L.	43
Voos, Henry	41

Tabela 1: Frente de Pesquisa da Produtividade dos autores (Lei de Lotka)

Pode-se observar que estes 15 autores têm devotado seus esforços à produção de literatura sobre a produtividade dos autores, por conseguinte, a frente de pesquisa é muito mais claramente estabelecida do que na forma anterior de contar meramente o volume de citações, mas sem relacioná-las com os produtores correspondentes.

Nesta frente de pesquisa existem dois autores que têm se convertido em clássicos da literatura sobre o assunto: Alfred Lotka, o formulador da Lei que leva o seu nome. Este autor foi citado 231 vezes, e quase que regularmente, desde 1932, quando Carr (1932) o citou pela primeira vez. Seu único artigo produzido e publicado em 1926, embora questionado muitas vezes e reformulado para produzir modelos mais adequados para medir a produtividade dos autores, ainda vem sendo citado continuamente. Também a distância, com respeito ao segundo citado, é quase o dobro das citações; isto é, existe uma relação de quase 1 para 2.

John Derek de Solla Price, com muitos textos escritos e várias reedições de seu clássico livro *Big Science Little Science* - inclusive as traduções para o Espanhol e Português, no qual foi inserido um capítulo com o nome de *Galton Revisited*, onde lida especificamente com a Lei de Lotka, foi citado 190 vezes. A distância deste autor, com respeito ao terceiro mais citado, é

quase o dobro das citações; ou seja, novamente existe uma relação de 1 para 2. A partir do terceiro autor mais citado, a relação de proporcionalidade se transforma em insignificante.

Logo, esses dois textos podem ser considerados como os clássicos dos estudos sobre a produtividade dos cientistas. Identificados e separados estes dois textos/autores clássicos, pode-se afirmar que os 13 autores restantes formam a frente de pesquisa deste campo. Curiosamente, a maioria desses autores são grandes produtores (tem produzido 10 ou mais artigos) ou moderados produtores (tem produzidos entre 5 e 9 artigos). Todos publicaram seus trabalhos no idioma inglês, e a grande maioria deles em periódicos da chamada “corrente principal”, isto é, indexados no Science Citation Index ou no Social Science Citation Index.

Por outro lado, note-se que no modelo probabilístico da formação de frentes de pesquisa (Figura 1) sugerido, esta é apenas a expressão da frequência com que um autor é citado num período determinado. Nesse período, existem outras variáveis que podem influir nas razões do porque um autor é citado e, por isso influir indiretamente na formação da frente de pesquisa.

Duas dessas variáveis a serem analisadas nesta pesquisa estão relacionadas ao número de trabalhos publicados pelos autores e a frequência com que elas são citadas. Para analisar essas relações, elaborou-se a Tabela 2 de contingência do nível de produtividade dos autores, segundo a frequência das citações dirigidas aos documentos publicados.

		Nível de produtividade dos autores				
		Grandes	Moderados	Aspirantes	Transeuntes	Total
Nível de citação	Alto	2 66.7%	11 78.6%	13 52.0%	20 6.0%	46 12.2%
	Médio	0	1 7.1%	7 28.0%	21 6.3%	29 7.7%
	Baixo	0	1 7.1%	3 12.0%	124 37.0%	128 34.0%
	Nulo	1 33.3%	1 7.1%	2 8.0%	170 50.7%	174 46.2%
Total		3 100.0%	14 100.0%	25 100.0%	335 100.0%	377 100.0%

Tabela 2: Nível de produtividade dos autores segundo a frequência das citações

Foram considerados como grandes produtores aqueles autores que no período da pesquisa produziram 10 ou mais trabalhos; produtores moderados, aqueles que no mesmo período produziram entre 5 e 9 documentos; aspirantes, aqueles que produziram entre 3 e 4 trabalhos; e transeuntes aqueles que produziram entre 1 e 2 documentos.

Para ser coerente com as medidas, a frequência das citações foi seccionada em quatro estratos: com alta frequência de citação foram considerados os autores cujos trabalhos receberam 10 ou

mais citações; com frequência de citação média foram considerados os autores cujos trabalhos receberam entre 5 e 9 citações; com baixa frequência de citação foram considerados os autores cujos trabalhos receberam entre 1 e 4 citações; e com nula citação os autores cujos trabalhos não receberam nenhuma citação no período da pesquisa.

Nesta Tabela 2 pode-se observar que ocorre uma gradação das citações, segundo o nível de produtividade dos autores. Nos grandes e moderados produtores concentram-se a mais alta percentagem de citações, caindo à medida que as gradações passam de produtores moderados para aspirantes e de aspirantes para transeuntes. As chances dos autores classificados como grandes produtores serem citados são 15% mais elevadas do que dos autores classificados como aspirantes e 61% mais elevadas do que dos autores classificados como transeuntes. Similarmente, as chances de um autor classificado como produtor moderado ser citado é 27% mais elevada do que um classificado como aspirante, e 73% mais elevada do que de um autor transeunte.

Um fenômeno inverso ocorre nos níveis mais baixos das frequências de citação, que é a tendência que as percentagens das citações mais baixas e/ou nulas apresentam para aumentarem, segundo a gradação dos produtores. Isto é, as chances de não serem citados são maiores para os autores aspirantes e moderados. Novamente, há uma indicação de que existe uma relação de dependência entre ambas as variáveis. Essa relação de dependência pode ser testada com o qui-quadrado, mas este teste exige que os valores observados em cada uma das celas da tabela de contingência não sejam menores do que 5.

Como na Tabela 2 observaram-se valores menores de 5, para se realizar o teste qui-quadrado foi necessário agrupar os valores observados com valores menores do que 5. Essa agrupação pode ser observada na Tabela 3, onde os grandes e moderados produtores foram agrupados conjuntamente, assim como os aspirantes e transeuntes. Também as frequências altas e médias das citações foram agrupadas conjuntamente, bem como ambos os níveis baixos e nulos. Desta forma, minimizou-se a presença de celas com valores menores de 5, as quais podiam distorcer o teste qui-quadrado, ainda que, apesar da agrupação conjunta dos níveis de citação baixas e nulas, observe-se uma única cela com valor menor do que 5.

		Nível de produtividade dos autores		
		Grandes e Moderados Produtores	Produtores Aspirantes e Transeuntes	Total
Nível de citação	Alto e Médio	14 (82.4)	61 (16.9)	75 (19.9)
	Baixo e Nulo	3	299	302

	(17.6)	(83.1)	(80.1)
Total	17	360	377
	(100.0)	(100.0)	(100.0)

Tabela 3: Nível de produtividade agrupado segundo a frequência das citações

O padrão de citação continua sendo o mesmo, no entanto agora mostra mais claramente que as chances dos autores classificados conjuntamente como grandes e moderados produtores serem citados nos níveis altos e médios são 65.5% mais elevadas do que daqueles autores classificados como aspirantes e transeuntes juntos. No sentido inverso, as chances dos autores classificados como grandes e moderados produtores serem citados nos níveis baixos ou nulos da frequência de citação são 65.5% menos elevadas do que daqueles autores classificados conjuntamente como aspirantes e transeuntes.

O teste qui-quadrado dos valores observados e esperados desta tabela de contingência, a um nível de significância de 0.001 e com 1 grau de liberdade, produziu um valor igual a 43.581, muito maior que o valor crítico de 10.828, rejeitando-se a hipótese nula de não existência de covariação entre a frequência de citações e o nível de produtividade dos autores dos documentos publicados sobre a Lei de Lotka; conclui-se, portanto, que ambas as variáveis são dependentes. Em outras palavras, quanto maior a produtividade de um autor, maior a sua chance de ser citado.

Outras relações entre variáveis a serem analisadas nesta pesquisa são: a publicação dos trabalhos sobre a produtividade dos autores ou lei de Lotka nos periódicos da chamada “corrente principal”; isto é, a condição de documentos serem publicados nos periódicos indexados pelo Science Citation Index ou pelo Social Science Citation Index e a frequência com que essas publicações são citadas. Para analisar essa relação elaborou-se a Tabela 4 de contingência dos periódicos da corrente principal, segundo a frequência das citações recebidas. O nível da frequência de citação continua sendo a mesma que no caso anterior, e já foi explicada.

		Documentos publicados nos periódicos da corrente principal		
		Sim	Não	Total
Nível de citação	Alto	31 (23.8)	15 (6.1)	46 (12.2)
	Médio	15 (11.5)	14 (5.7)	29 (7.7)

(Continua)

Baixo	34 (26.2)	94 (38.1)	128 (34.0)
Nulo	50 (38.5)	124 (50.2)	174 (46.1)
Total	130 (100.0)	247 (100.0)	377 (100.0)

Tabela 4: Periódicos da corrente principal segundo a frequência de citações

Pode-se notar que ocorre uma gradação das citações segundo a condição dos documentos serem publicados nos periódicos da corrente principal ou naquelas não pertencentes a esse grupo. Nos periódicos da corrente principal concentra-se a mais alta percentagem das citações de nível alto e médio, caindo estas citações nos estratos de baixas e nulas. As chances dos documentos publicados em periódicos da corrente principal serem altamente citados são 18% mais elevadas do que daqueles documentos publicados em periódicos não pertencentes a essa corrente.

Similarmente, as chances de esses documentos serem medianamente citados são 6% mais elevadas do que daqueles documentos publicados em periódicos não pertencentes à corrente principal. Igualmente, as chances dos documentos publicados em periódicos da corrente principal obterem baixa citação ou nenhuma citação são 12% menores do que daqueles documentos publicados em periódicos não pertencentes à corrente principal.

Essa relação monotônica negativa indica que existe uma relação de dependência entre ambas as variáveis. Essa relação de dependência foi testada usando-se o teste qui-quadrado a um nível de significância de 0.001 e com 3 graus de liberdade, que produziu um valor igual a 31.964, muito maior do que o valor crítico de 16.266, rejeitando-se a hipótese nula de não existência de covariação entre a frequência de citações e a publicação dos documentos em periódicos da corrente principal; conclui-se, portanto que ambas as variáveis são dependentes. Em outras palavras, se um documento sobre a Lei de Lotka é publicado nos periódicos da chamada “corrente principal”, maiores são as suas possibilidades de ser citado do que se fosse publicado em periódicos fora ou marginais àqueles da corrente principal. Como se sabe, somente 55 periódicos do campo da biblioteconomia e da ciência da informação são indexados pelo SSCI, e todos eles procedem do mundo anglo-saxão.

Outra relação entre variáveis a ser analisada nesta pesquisa refere-se à frequência de citação segundo os idiomas nos quais os documentos foram publicados. Observou-se uma gradação das citações segundo os idiomas, sendo o inglês o idioma que monopoliza as citações, com alta (17.7%) e média (8.5%) frequência. Também as percentagens das frequências baixas

(28.2%) e nulas (45.6%) são menores para este idioma em relação aos outros citados. Neste nível, a percentagem de baixa frequência para os idiomas diferentes do inglês corre, no geral, de 42% até 100% e no nível de citação nula essas percentagens correm de 46% a 100%. Isso indica que existe uma relação de dependência entre ambas as variáveis. Essa relação de dependência pode ser provada com o teste qui-quadrado, mas este teste exige que os valores observados em cada célula da tabela não sejam menores do que 5. Como se observaram muitos valores menores do que 5, para realizar o teste qui-quadrado foi necessário agrupar os idiomas encontrados com valores menores do que 5. Essa agrupação pode ser verificada na Tabela 5, onde os idiomas diferentes do inglês foram agrupados apenas numa coluna, segundo os níveis (alto, médio, baixo e nulo) de citação.

		Idiomas		
		Idioma Inglês	Outros idiomas	Total
Nível de citação	Alto	44 (17.7)	2 (1.6)	46 (12.2)
	Médio	21 (8.5)	8 (6.2)	29 (7.7)
	Baixo	70 (28.2)	58 (45.0)	128 (34.0)
	Nulo	113 (45.6)	61 (47.2)	174 (46.1)
Total		248 (100.0)	129 (100.0)	377 (100.0)

Tabela 5: Idiomas dos documentos publicados, segundo a frequência de citações.

Observa-se mais claramente que a percentagem com alta frequência de citações é 16.1% mais elevada para o inglês do que para todos os outros idiomas citados. Similarmente, a percentagem com média frequência de citações do inglês é 2.3% mais alta do que a dos outros idiomas juntos. Um padrão oposto ocorre nos níveis mais baixos das frequências de citações, revelando que a baixa citação é 16.8% mais alta para os outros idiomas em relação ao inglês. Também a citação nula dos outros idiomas juntos é 1.6% maior do que a do inglês. Esses valores demonstrariam que existe uma relação de dependência entre os idiomas e a frequência de citação, embora as possibilidades de citação dos documentos publicados em outros idiomas sejam muito mais baixas do que dos documentos publicados no idioma inglês.

Para testar esta possibilidade usou-se o teste qui-quadrado que, a um nível de significância de 0.001 e com 3 graus de liberdade, produziu um valor igual a 25.854, muito maior do que o valor crítico de 16.266, rejeitando a hipótese nula de não existência de covariação entre a frequência de citações e os idiomas dos documentos publicados sobre a Lei de Lotka; logo,

conclui-se que ambas as variáveis são dependentes. Em outras palavras, os documentos publicados no idioma inglês tem maiores chances de serem citados do que os documentos publicados em outros idiomas.

Esta tendência já tinha sido observada em 1955 pelos psicólogos anglo-saxões no uso da literatura, tanto que “os escritores ingleses quase completamente se limitam a citar sua própria língua” (LOUTTIT, 1955, p.486). Esse mesmo comportamento foi observado pelo mesmo autor em 1957, nos campos da química e da física, quando afirmou que os autores que escrevem em inglês citam sua própria língua em excesso face às outras línguas (LOUTTIT, 1957, p.315). Por outro lado, já é conhecida a hegemonia do idioma Inglês na comunicação científica, tanto que Price (1971: 257) afirmava que neste idioma se publica pouco mais da metade da produção filosófica e científica do mundo, embora os científicos não leiam muito e virtualmente nada em idiomas estrangeiros. Para ser mais preciso, “... olham a literatura estrangeira através de [seus] óculos obscuros que lhes deixam ver somente uma décima parte do que realmente existe”. (PRICE, 1971:258).

5 CONCLUSÕES

No período pesquisado identificaram-se dois textos e dois autores clássicos. Os autores são Alfred Lotka, com seu texto “The Frequency Distribution of Scientific Productivity” e John Derek de Solla Price, com seu texto “Little Science, Big Science”, especialmente o capítulo “Galton revisited”. Identificou-se também 13 autores que formam a frente de pesquisa deste campo. Basicamente, esses 13 autores tem 10 ou mais artigos produzidos, qualificando-os como grandes produtores, ainda em alguns casos como produtores moderados, aqueles que tem entre 5 e 9 artigos produzidos. Todos eles possuem os trabalhos publicados no idioma inglês, e a grande maioria deles nos periódicos da chamada “corrente principal”, isto é, indexados no Science Citation Index ou no Social Science Citation Index.

Verificou-se também a existência de covariação entre a frequência de citações e o nível de produtividade dos autores, concluindo-se que ambas as variáveis são dependentes. Isso significa que quanto maior a produtividade de um autor maior são as suas chances de ser citado, e no sentido inverso, quanto menor a produtividade dos autores menores são as suas chances de serem citados. Igualmente, as chances dos documentos publicados em periódicos

da corrente principal obterem baixa citação ou nenhuma citação são menores do que daqueles documentos publicados em periódicos não pertencentes à corrente principal. Isto significa que, se um artigo é publicado nos periódicos da chamada “corrente principal”, maiores são as suas possibilidades de ser citado do que se fosse publicado em periódicos não indexados nos Science Citation Index ou no Social Science Citation Index.

Percebeu-se ainda, que existe uma relação de dependência entre os idiomas e a frequência de citação, já que as possibilidades de citação dos documentos publicados em outros idiomas diferentes do inglês são muito mais baixas do que dos artigos publicados nesse idioma. Em outras palavras, para qualificar-se na frente de pesquisa de uma disciplina, o autor tem que ser um grande ou moderado produtor, publicar no idioma inglês e nos periódicos considerados de corrente principal. Fora deste contexto, o uso da informação explicitada na forma de citações fica reduzido a um consumo local.

REFERÊNCIAS

BALDAUF, R. B. Linguistic constrains on participation in psychology. **Psychologist**, New York, v. 41, p. 220-240, 1986.

BRAGA, G. M. **Relações bibliométricas entre a frente de pesquisa (research front) e revisões de literatura**: estudo aplicado à Ciência da Informação. 1973. Dissertação (Mestrado em Ciência da Informação)- IBICT/UFRJ, Rio de Janeiro, 1973.

CARR, L. J. The patenting performance of 1,000 inventors during ten years. **The American Journal of Sociology**, Chicago, v. 37, n. 4, p. 569-580, 1932.

CAWKELL, A. E. Understanding science by analyzing its literature. **The Information Scientist**, v. 10, n. 1, p. 3-10, 1976.

CHRISTÓVÃO, H. T. **Da comunicação informal à comunicação formal**: identificação da frente de pesquisa através de filtros de qualidade. 1978. Dissertação (Mestrado em Ciência da Informação)- IBICT/UFRJ, Rio de Janeiro, 1978.

DeHART, F. E.; SCOTT, L. ISI research fronts and online subject access. **Journal of the American Society for Information Science**, Maryland, v. 42, n. 5, p. 386-388, 1991.

GARFIELD, E. Citation indexing for studying science. **Nature**, n. 227, p. 669-671, 1970.

_____. **Citation frequency as a measure of research activity and performance.**
In: _____. Essays of an information scientist. Philadelphia: ISI Press, 1977.

_____. **Citation indexing: its theory and application in science, technology and humanities.**
New York: Wiley, 1979.

GARVEY, W. D. **Communication, the essence of science: facilitating information exchange among librarians, scientists, engineers and students.** Oxford ; New York: Pergamon Press, 1979.

_____. Identification of important authors in science: a comparison of two methods of identification. **Information Processing & Management**, v. 21, n. 3, p. 177-186, 1985.

LOUTTIT, C. M. The use of foreign languages by psychologists. **The American Journal of Psychology**, Chicago, v. 68, n. 3, p. 484-486, sept. 1955.

_____. The use of foreign languages by psychologists, chemists, and physicists. **The American Journal of Psychology**, Chicago, v. 70. n. 2, p. 314-316, jun. 1957.

PRICE, J. D. de S. Galton revisited. In: _____. **Little science, big science.** New York: Columbia University Press, 1963. p. 33-61.

_____. Networks of scientific papers. **Science**, v. 149, n. 3683, p. 510-515, (1965)

_____. The growth of knowledge. In: KOCHEN, M. **Some problems in information science.** New York : Scarecrow Press, 1965. p. 145-155.

_____. Collaboration in an invisible college. **American Psychologist**, v. 21, p. 1011-1018, 1966.

_____. The structure of publication in science and technology. In: GRUBER, W. H.; MARQUIS, D. G. (Ed.) **M.I.T, conference on the human factor in the transfer of technology.** Cambridge: M.I.T. Press, 1969. p. 91-104.

_____. The expansion of scientific knowledge. **Annals of the New York academy of sciences**, New York, v. 184, p. 257-259, 1971.

_____. Some remarks on elitism in information and the invisible college phenomenon. In _____ . **Science. journal of the american society for information science**, v. 22, n. 2, p. 74-75, 1971.

RODRIGUES, M. da P. L. **Estudo das citações constantes das dissertações de mestrado em Ciência da informação do IBICT/UFRJ**. Dissertação (Mestrado em Ciência da Informação)-IBICT/UFRJ, Rio de Janeiro, 1973.

URBIZAGÁSTEGUI ALVARADO, R. Aparência e realidade da frente de pesquisa na bibliometria brasileira. **Revista Española de Documentación Científica**, v. 16, n. 4, p. 321-340, oct./dic. 1993.

_____. **Lotka's law: an annotated bibliography**. Riverside, Ca: 2005. Unpublished.

VELKE, L. The use of citation patterns in the identification of "research front" and classic papers. In: NORTH, J. B. (Ed.). ANNUAL MEETING OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE, 33, 1970. Philadelphia, **Proceedings ...** Washington, D. C.: The Society, 1970, v. 7, p. 49-51.

Abstract

The research front of a sub-area of bibliometrics: the author's productivity, also known as the "Lotka's law", was analyzed. The authors who make up the research front of this field were identified. These authors have produced 10 or more articles, qualifying them as the major producers, although in some cases as moderate producers. They all have their works published in English, and the vast majority of them in journals of the so-called "principal current", that is journals indexed by the Science Citation Index or the Social Science Citation Index. Covariation between author's levels of productivity, articles published in journals of "principal current", language of publication, and frequency of citations were observed. In other words, to qualify as being in the research front of a discipline the author has to be a great or moderate producer, publish in English and in journals considered in the mainstream. Outside this context, the use of the information in the form of citations is reduced to a local consumption.

Keywords: Research front. Author's productivity. Lotka's law. Bibliometrics. Scientometrics. Informetrics.

Originais recebidos em: 02/12/2008

Texto aprovado em: 10/09/2009