

Crescimento da literatura e dos autores sobre a Lei de Lotka

Ruben Urbizagastegui

Doutor na Escola de Ciência da Informação da Universidade Federal de Minas Gerais (Belo Horizonte, Brasil). Bibliotecário na Universidade de Califórnia, Riverside, nos Estados Unidos.

E-mail: ruben@ucr.edu

Resumo

Analisa o crescimento da literatura produzida sobre a Lei de Lotka, assim como o crescimento dos autores produtores dessa literatura, de 1922 a 2003. Nesse período o crescimento da literatura segue o modelo exponencial, com taxa média de crescimento de 7,5% ao ano e taxa de duplicação a cada 9,6 anos. Também o número de autores produtores da literatura cresce à proporção de 7,3% ao ano, com período de duplicação a cada 10 anos. A estatística indica bom ajuste ao modelo exponencial, com R^2 de 0,985 para a literatura e 0,992 para os autores, ambos a um nível de significância de 0,01.

Palavras-chave

Crescimento exponencial. Lei de Lotka. Produtividade dos autores. Cienciometria. Bibliometria. Informetria.

The growth of literature and authors on Lotka's Law

Abstract

The growth of literature produced on Lotka's law from 1922 to 2003, as well as the growth of authors producing that literature, is analyzed. The growth of literature follows an exponential model with an annual average of 7.5% growth and doubling its size every 9.6 years. The number of authors producing this literature grows at a rate of 7.3% per year and duplicating its volume every 10 years. The statistics indicate a good adjustment to the exponential model, with R^2 of 0.985 for literature and 0.992 for authors, both at a level of 0.01 significance.

Keywords

Exponential growth. Lotka's law. Authors' productivity. Scientometrics. Bibliometrics. Informetrics. Scietometrics.

INTRODUÇÃO

O objetivo desta pesquisa é analisar o crescimento da literatura produzida sobre a produtividade dos autores, conhecida também como a “Lei de Lotka”, assim como o crescimento dos autores produtores dessa literatura, no período que vai de 1922 a 2003. Como se sabe, a produtividade científica dos autores, medida através da publicação de artigos em determinada disciplina, começou a ser estudada por Dresden (1922), mas foi Lotka (1926) quem a formulou como um modelo estatístico do *quadrado inverso*, para mais adiante ser questionada e reformulada como o modelo do *poder inverso generalizado*. Como homenagem ao formulador do modelo estatístico, convencionou-se chamá-lo de Lei de Lotka.

Desde essa época, as pesquisas, e por isso a literatura devotada a este assunto, tem crescido e se acumulado ano a ano, de maneira quase constante. É esse crescimento da literatura e seu ajuste a um modelo estatístico, sugerido pelos dados observados, que será analisado nesta pesquisa. Afirma-se que, em qualquer área, a acumulação do conhecimento está crescendo e esse crescimento está diretamente relacionado ao crescimento da literatura publicada nessa área. Entende-se, por crescimento, o aumento paulatino do número de publicações produzida, o que implica o crescimento irreversível do volume dessa literatura e cujo período de duplicação é definido como o tempo necessário para que determinada quantidade, seja esta o número de artigos, seja o número de autores, etc., duplique seu tamanho quando cresce em taxa constante.

Por isso, o crescimento da literatura em determinado período se expressa em termos de uma taxa média de crescimento durante esse período. Isso significa que o aumento da população em determinado período é diretamente proporcional ao total da população que existia no início desse período, ou seja, o aumento da população em determinado período ou intervalo de tempo é diretamente proporcional ao total da população que existia no início desse período ou intervalo de tempo. Então, o que se necessita é conhecer o tamanho do fenômeno estudado no

começo e ao final do período para, dessa forma, estimar a taxa média de crescimento. O interesse reside em conhecer como a quantidade de literatura produzida varia, não só num momento particular, senão também como muda essa quantidade em relação ao tempo. O conhecimento dessas variações permitiria previsões adequadas das tendências observadas no comportamento da literatura.

Por essa razão, uma das características mais óbvias da prática científica tem sido a taxa de crescimento da literatura publicada. A taxa de crescimento da literatura publicada será estudada nesta pesquisa, tomando como estudo de caso a produção de trabalhos sobre a Lei de Lotka. Não obstante, é preciso estar conscientes de que se a literatura cresce, não é simplesmente porque os cientistas aumentam sua produtividade, senão também porque mais cientistas entram no campo e contribuem com uma taxa média de produtividade nesse campo. Portanto, será explorado também o crescimento dos autores produtores dessa literatura sobre a Lei de Lotka.

REFERENCIAL TEÓRICO

O crescimento da ciência tem significado não somente um aumento no seu volume, mas também o aumento da mão de obra comprometida com a atividade científica e dos recursos financeiros dedicados à sua produção. Por isso, o crescimento da literatura representa um dos múltiplos aspectos do crescimento geral da comunicação científica. Desde o começo do século XX, muitos pesquisadores têm estudado o crescimento da literatura em diversas áreas do conhecimento, porém foi o trabalho de Price (1963) que popularizou.

Esse pesquisador da história da ciência observou o crescimento da literatura do campo da física, contando o número de resumos no *Physical Abstracts*, de 1900 a 1950. Ele constatou que a literatura de física vinha crescendo *exponencialmente*, com um período de duplicação de aproximadamente doze anos. Não obstante, foram Cole & Eales (1917) que iniciaram o estudo do desenvolvimento histórico do campo da anatomia comparada, de 1550 em diante, cobrindo um período de três séculos. Esses

autores foram os primeiros a utilizar a literatura publicada para construir um perfil quantitativo de uma disciplina científica.

A partir dessa experiência, os esforços por medir o crescimento da literatura se concentraram nas ciências *hard* e foram prosseguidos por Tamiya (1931), que analisou uma bibliografia sobre “*aspergillus*”, espécie de fungo que pode ser encontrado em ambientes fechados, elaborada por ele mesmo, cobrindo o período de 1729 a 1928 e contendo o total de 2.424 publicações. O autor sugere que a literatura sobre esse assunto cresceu da forma *logística* quase semelhante a uma reação catalítica de primeira ordem. Posteriormente, Wilson & Fred (1935) estudaram o crescimento das atividades de pesquisa através das publicações na área da fixação de nitrogênio pelas plantas, especialmente por leguminosas (*Leguminosae*). Os autores usaram uma bibliografia elaborada por Baldwin & McCoy (1932), que listava o total de 2.000 referências bibliográficas, de 1850 a 1915, e observaram que o modelo *logístico* era aquele que melhor descrevia o crescimento da literatura em sentido geral, porém não fizeram nenhum esforço por estimar a melhor curva do modelo por nenhum método estatístico ou matemático convencional.

Paradoxalmente, depois desses dois estudos o crescimento da literatura nas ciências *hard* teve uma mudança para se expressar de forma exponencial. Por exemplo, Conrad (1957), estudando o crescimento da literatura sobre biologia e sua cobertura pelo *Biological Abstracts*, sugeria crescimento *exponencial* desta literatura e estimava que em 2010 teríamos a produção de 348.000 artigos nesse campo. Também May (1966) estudou o crescimento da literatura de matemática usando duas bases de dados: o *Jahrbuch über die Fortschritte der Mathematik*, desde 1868 até 1940, e o *Mathematical Reviews*, de 1941 até 1965.

Esse autor verificou que o número de publicações por ano tinha crescido de 800 a 13.000, com média composta de aproximadamente 2,5% por ano, porém duplicando-se aproximadamente a cada 28 anos, isto é, quatro vezes num século. Igualmente, constatou que o crescimento acumulado da literatura de matemática cresceu de 41.000, em 1867, a 419.000, no final de

1965. O crescimento seguia o modelo *exponencial*, mas com pequenos desvios nos períodos das duas guerras mundiais. Também mostrou que as taxas de crescimento e os períodos de duplicação encontrados nos diferentes campos estudados, com anterioridade à sua pesquisa, não tinham levado em consideração a literatura produzida e acumulada previamente. Os valores estimados estavam baseados em pontos arbitrários dos períodos estudados, ignorando toda a literatura anterior produzida e, conseqüentemente, produzindo taxas de crescimento maiores do que as taxas verdadeiras.

Mais tarde, Mernard (1971) examinou a literatura em diversos subcampos da geologia, encontrando que, em paleontologia dos vertebrados, disciplina que começou no início do século XVI, a literatura produzida cresceu lentamente até finais do século XVIII e depois começou a crescer *exponencialmente*, com período de duplicação a cada 15 anos. Em meados do século XVIII, a taxa de crescimento começou a declinar e, de 1870 em diante, o período de duplicação aumentou para 35 anos. No caso de geologia glacial, depois de um período inicial muito curto, verificou que houve rápido crescimento *exponencial* da literatura até meados do século XIX, com período de duplicação de 10 anos. Similarmente, Brookes (1973) afirmava que, julgada pelo número de artigos publicados a cada ano, a literatura produzida na maioria dos campos científicos continuava a crescer *exponencialmente*, com taxa de duplicação de aproximadamente dez anos.

As pesquisas sobre o crescimento da literatura continuaram em outros campos, tanto que, revisando a literatura indexada no *Nuclear Science Abstracts*, de 1950 a 1972, Sullivan, White & Barboni (1977) compilaram exaustiva bibliografia de artigos de periódicos publicados no campo da física das interações fracas (*weak interactions*) e identificaram o total de 4.691 artigos produzidos por 3.949 autores de todas as nacionalidades, que tinham publicado pelo menos um artigo nesse período. Os autores analisaram e compararam o crescimento da literatura dos experimentalistas e dos teóricos do campo e verificaram que essa literatura não mostrava um crescimento *exponencial*, mas sim *linear*.

No mesmo ano, Braun, Lyon & Bujdosó (1977) observaram crescimento *exponencial* na literatura sobre análise da ativação nuclear até 1970, com período de duplicação a cada três anos. Notaram, também, que certos subcampos mostravam períodos de duplicação diferentes: ativação forense (2,1 anos) e detectores do estado sólido (1,1 ano).

Tague; Beheshti & Rees-Poter (1981) analisaram o crescimento *exponencial* da literatura indexada no *Chemical Abstracts* (1907-1979 e 1960-1979), *Science Abstracts* (1960-1979), e *Biological Abstracts* (1960-1979). Observaram que o crescimento linear se ajustava melhor à maioria da literatura coberta pelos serviços de indexação da ciência, no período de 1960 a 1979, mas o crescimento *exponencial* se ajustava melhor à literatura indexada pelo *Chemical Abstracts*, no período de 1907 a 1979. No mesmo ano, Baker (1981) elaborou um documento especial sobre as tendências mais recentes do crescimento da literatura no campo da química, com ênfase na desaceleração do crescimento da literatura no período 1975 a 1980. Porém, uma simples extrapolação da curva de crescimento nesse período indicaria a diminuição do crescimento contínuo observado nas décadas anteriores, levando à paralisação gradual da taxa de crescimento anual.

Ainda nesse ano, Rangarajan & Bhatnagar (1981), estudando o crescimento da literatura dos estudos sobre o “efeito Mössbauer”, de 1960 a 1977, no campo da física, observaram flutuações erráticas para diferentes períodos e diferentes subcampos de aplicação do “efeito Mössbauer”. Já Chakraborty (1983) estudou a literatura sobre petrologia citada em três periódicos da área, cobrindo o período de 1980-1981, e observou que nos três periódicos a literatura citada foi de 9.847 itens citados, sendo a citação mais antiga uma publicação de 1804. Constatou também que a taxa de crescimento da literatura era muito baixa, notando crescimento regular a partir de 1930. Nas duas últimas décadas, porém, a taxa de citação tinha crescido muito rapidamente. Apesar dessas afirmações, o autor não oferece parâmetros que possam servir de comparação.

Como a forma de crescimento da literatura passava de logística a exponencial e de exponencial a linear, era necessário procurar maiores precisões

nos modelos de crescimento observados nas pesquisas; por isso Fleischer (1984), utilizando o número de artigos indexados e resumidos na Seção 53 (química mineralógica e geológica) do *Chemical Abstracts*, da qual o autor era editor desde 1941, estudou o crescimento da literatura de química geológica, de 1946 a 1980 e, embora seus dados mostrem certas flutuações, tinham crescimento constante. Para estabilizar as flutuações, usou o número médio por ano num período de cinco anos, verificando que o número de resumos aumentou de 1.150 para 11.135 no período estudado, mostrando com isto uma taxa de duplicação a cada nove anos.

Mahapatra (1985) sugeriu um novo método para medir o crescimento da literatura e introduziu o conceito de taxa média relativa de crescimento. Este método está baseado na taxa média relativa de crescimento entre períodos específicos de intervalo e a tendência do crescimento total ou a curva absoluta de crescimento durante o período estudado. Usando este método, o autor analisou o crescimento da literatura em fisiologia das plantas, publicados em inglês na Índia, na forma de artigos, desde 1951 até 1980, que incluíam 2.703 artigos. Seu resultado indicava que a taxa de crescimento nessa área levava a um ponto de saturação que seria atingida no ano 2000.

Contudo os resultados observados nas pesquisas pareciam não ser ainda concluintes e as procuras continuaram com Stephenson (1985) examinando o crescimento do conhecimento e da literatura de geologia, propondo, testando e analisando um algoritmo baseado no método de pesquisa como a variável independente e o crescimento da literatura publicada como a variável dependente. Ele selecionou dois subcampos da geologia para testar suas hipóteses: geoquímica e paleontologia dos vertebrados. A população foi definida como todas as publicações nessas duas subáreas publicadas e indexadas, entre 1965 e 1966, na *Bibliography of North American Geology*. Foi observado que o subcampo de paleontologia dos vertebrados cresceu a uma taxa muito lenta, contrária à taxa de rápido crescimento da área de geoquímica.

Essa taxa de crescimento foi confirmada pelo teste de Kendall tau c (coeficiente de correlação de 0.57182 com $p < 0.00001$). Similarmente, Akhibge (1988) estudou o crescimento da literatura sobre cola (cola acuminata, cola nítida, e cola verticilat), comumente chamada de Kolanut, na Nigéria. O autor observou o total de 156 referências bibliográficas produzidas de 1901 a 1978, porém não realizou uma análise aprofundada de seu crescimento. Igualmente, Qinxian (1990), baseando-se no *Catalogue of Chinese Bamboo Literature from 1975-1986*, compilado em 1988 pelo Instituto de Pesquisas Florestais da China, estudou o crescimento da literatura sobre bambu (Bambusoideae, subfamília: Gramineae), observando o total de 1.200 itens. Embora o traçado dos seus dados descreva crescimento exponencial, não foi estimada a taxa de crescimento desta literatura, limitando-se a descrever alguns aspectos interessantes desta literatura, como no período da chamada “revolução cultural”, quando a pesquisa e produção sobre bambu na China foi forçada a parar, retomando o crescimento a partir daí.

Nijagunappa; Gunjal & Parvathamma (1996) realizaram um estudo para mapear as características nacionais da literatura de geologia indiana, especialmente o seu crescimento. Os autores coletaram o total de 7.250 itens bibliográficos publicados durante o período de 1978 a 1988 e indexados no *Bibliography and Index of Geology (USA)*, *Mineralogical Abstracts (UK)*, *Indian Geoscience Abstracts (New Deli)*, além de outras fontes. No que se refere ao crescimento da literatura desta área, observaram taxa de crescimento decrescente, de 0,65, em 1979, para 0,07, em 1988. Não obstante, o período de duplicação da literatura cresceu de 1,07 anos, em 1979, para 9,9 anos em 1988, com a média de duplicação para esta literatura de 4,8 anos.

Ravi (2001) fez estudo exploratório do crescimento *exponencial* da literatura de ciências nucleares, em língua inglesa, produzida pelos pesquisadores indianos. Coletou as referências bibliográficas indexadas na base de dados do INIS e no *Indian Science Abstracts*, num longo período de 15 anos, de 1980 a 1994. Nesse período observou taxa de

crescimento de 0,21, com período de duplicação de 5,12 anos. Não obstante, observou que este período variava de 2,33 anos para 1981-1985, 5,26 anos para 1986-1990, e 8,45 anos para 1991-1994.

No mesmo ano, Saxena, Gupta & Jauhari (2001) discutiram as metodologias seguidas para ajustar uma curva não linear e mostraram um método de linearização e estimação do modelo não linear. Para demonstrar seu método, coletaram literatura do campo de química, indexada no *Chemical Abstracts*, no período de 1901 a 1994, tendo observado que o número de publicações atingiu seu mínimo aproximadamente ao redor de 1918. O próximo valor mínimo foi verificado em 1945. De 1918 a 1945 a série se comporta como uma curva inclinada e de 1945 a 1970 mostra crescimento *exponencial*, com pequena queda em 1972, e daí em diante volta a crescer exponencialmente.

Similarmente, Sharma, Gupta & Kumar (2002) selecionaram três áreas de pesquisa: física, tomando a literatura indexada no *Physical Abstracts*, de 1907 a 1994; química, tomando a literatura indexada no *Chemical Abstracts*, de 1907 a 1994; e, engenharia elétrica e eletrônica, refletida na literatura indexada no *Electrical and Electronical Abstracts*, no período de 1907 a 1994. Nestas áreas os autores estudaram o crescimento da literatura através do modelo de crescimento exponencial, crescimento logístico e o modelo do poder, e constataram que o modelo do poder se ajusta melhor aos dados das três áreas. Os três modelos estudados (crescimento exponencial, crescimento logístico e modelo do poder) descreveram bem os dados coletados no campo da engenharia elétrica e eletrônica. O modelo do poder e o modelo exponencial se ajustaram melhor à literatura de química.

Embora a concentração das pesquisas estivesse nas ciências *hard*, elas não estiveram ausentes das ciências *soft*. Ainda com limitações, as ciências sociais e humanas foram objeto de pesquisa; por exemplo, Bottle & Efthimiadis (1984), tomando uma amostra de cinco bases de dados que indexavam a literatura no campo da ciência da informação e biblioteconomia,

abrindo o período de 1860 a 1983, estudaram o crescimento de autores e da literatura. Observaram que a literatura mostrava crescimento *exponencial* com um período de duplicação de 13,8 anos.

Usando a literatura referenciada na “*A classified bibliography on OPACs*”, elaborada em parceria com Mitev (1987), Efthomiadis (1990) analisou o crescimento da literatura dos catálogos de acesso público em linha (OPACs), de 1970 a 1985. O método do mínimo quadrado foi empregado para obter o melhor ajuste dos modelos linear, exponencial e logístico. No geral, os três modelos produziram ajustes não satisfatórios, embora o modelo logístico descrevesse melhor o comportamento dos dados, porém conceitualmente uma curva logística não poderia ser extrapolada indefinidamente.

No mesmo ano, Hu (1987) reportava os resultados observados em estudo acerca do crescimento da literatura sobre microcomputadores aplicados aos serviços bibliotecários. Os dados foram coletados usando as palavras-chaves relacionadas à automação dos processos bibliotecários, especificados nas formas impressas das bases de dados bibliográficas do Library Literature (LL), Library & Information Science Abstracts (LISA) e Information Science Abstracts (ISA), abrindo o período de 1973 a 1986, com 2.661 itens sobre o assunto. Embora o traçado dos seus dados descreva crescimento exponencial, ele não estimou a taxa de crescimento, limitando-se a descrever alguns aspectos interessantes desta literatura, tais como o crescimento rápido desde 1973, duplicando-se a cada ano até 1981, e reduzindo o crescimento desde então.

Wolfram e colaboradores (1990), estudando a literatura indexada em diversas bases de dados do campo da ciência e tecnologia, ciências sociais e humanidades, observaram que o modelo de crescimento linear e do poder tem melhor desempenho na explicação do crescimento da literatura do que o modelo exponencial. Os autores concluíram que o modelo do crescimento exponencial não seria adequado para representar o crescimento da literatura nos últimos 20 anos. Gupta, Kumar, Sangam & e Karisiddappa

(2002) aplicaram seletivo número de modelos de crescimento da literatura aos dados acumulados de publicações em antropologia (1963-1997), economia (1969-1997), história (1970-1997), ciências políticas (1970-1997), psicologia (1974-1998) e sociologia (1963-1998). Os modelos escolhidos foram avaliados em termos dos seus parâmetros, ajustes estatísticos e gráficos dos dados observados e esperados. Para as publicações dessas áreas, os melhores ajustes foram dados pelos modelos do poder e logístico. Não obstante, esta pesquisa contradiz os resultados obtidos por Egghe & Ravichandra Rao (1992), que afirmam que o modelo de Gompertz explica melhor o crescimento da literatura nas ciências sociais.

Como se pode observar pela literatura revisada, o modelo exponencial aparece com maior frequência para descrever o crescimento da literatura tanto no campo das ciências puras quanto nas aplicadas e nas humanidades.

METODOLOGIA

Como unidades de análises foram tomados cada um dos autores e dos artigos, capítulos de livros e trabalhos apresentados em congressos sobre a chamada Lei de Lotka, desde 1922 até 2003, isto é, um longo período de 82 anos. Para identificar a literatura e os autores contribuindo com artigos nesta área, foi feita uma busca usando a estratégia Dial Index em todas as bases de dados de DIALOG, com os termos Lotka²(5n)Law², produzindo 50 bases de dados que pelo menos continham um artigo sobre o assunto pesquisado. Esta estratégia de busca produziu o total de 315 registros que, depois da depuração dos duplicados e falsas recuperações, foram acumulados no total de 275 referências bibliográficas. Essas referências foram trasladadas ao PROCITE 5.0 para a elaboração de uma base de dados específica sobre o assunto.

Posteriormente foi realizada minuciosa leitura de cada um dos artigos identificados na busca, dedicando especial atenção a cada citação efetuada no documento lido. Depois, cada citação referente à Lei de Lotka era confrontada com a base de

dados e nela incorporada, se não tivesse sido identificada na busca anterior no DIALOG. Também foram realizadas buscas no *Information Science Abstract* (ISA), *Library Literature* (LL) e *Library and Information Science Abstract* (LISA). Com a leitura minuciosa foi produzida uma bibliografia analítica da Lei de Lotka que lista o total de 381 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 (URBIZAGASTEGUI, 2005). A bibliografia analítica de 381 referências produzidas entre 1922 e 2003 constitui o universo desta pesquisa. O período coberto pela literatura recuperada é suficientemente longo para assegurar a publicação de artigos sobre o assunto nos periódicos do campo da ciência da informação e afins.

O crescimento exponencial representa o aumento da população numa proporção fixa em cada unidade de tempo e expressa em percentagens. Também supõe taxa de crescimento constante com limite não definido de crescimento. O modelo não somente proporciona taxa média de crescimento, senão também oferece uma taxa de duplicação, isto é, uma taxa na qual o tamanho da população estudada se duplica. Geralmente, o crescimento exponencial assume forma côncava em sua representação gráfica inicial, nas palavras de Gupta & Karisiddappa (2000, p.333) “quando os valores da variável y formam uma progressão geométrica, no entanto, se os valores correspondentes da variável x formam uma progressão aritmética, a relação entre ambas variáveis é determinada por uma função exponencial”. Esta função é, matematicamente, representada como:

$$C(t) = C(0)e^{at} \quad (1)$$

Segundo Egghe & Rao (1992), esta função pode ser reescrita como:

$$C(t) = c g^t \quad (2)$$

onde $c > 0$, $g > 1$, e $t \geq 0$.

Certamente, a literatura publicada através do tempo em qualquer campo do conhecimento cresce e

se acumula, mas não conhecemos a forma desse crescimento e acumulação. Por isso, quando se estuda o crescimento da literatura e dos autores produtores dessa literatura, postula-se uma relação entre o tempo (variável independente), medido em anos, e o volume acumulado da literatura produzida (variável dependente), medida em unidades produzidas (para o caso do crescimento da literatura) e o volume de autores acumulados (variável dependente) para o caso do crescimento dos autores. Supõe-se que essa relação possa ser estatisticamente modelável, tanto que para avaliar o ajuste do modelo se constrói um gráfico da “nuvem de pontos” dispersos que estão baseados nos dados observados. Isso permite avaliar se existe certa regularidade na distribuição das frequências observadas e quando essa regularidade se assemelha à curva mostrada no gráfico se tenta ajustar essa curva à nuvem de pontos, via a regressão não linear. Numa distribuição exponencial, o que vai se testar é se o volume acumulado da produção de documentos segundo os anos (t) procedem de uma distribuição exponencial; isto é, a probabilidade de uma frequência na amostra seja igualmente provável para todas as frequências na mesma situação.

A análise da variância, o cálculo da inclinação e da interceptação da distribuição exponencial, foi realizada pelo método da determinação da regressão da curva não linear, usando os *softwares* estatísticos SPSS 14.0 e Mathematica 5.0 para Windows, disponíveis na Universidade da Califórnia em Riverside. Como se espera alta correlação entre as variáveis dependentes e independentes, essa correlação foi explorada usando-se o coeficiente de correlação de Pearson.

RESULTADOS

Os modelos estatísticos são utilizados para deduzir a variabilidade de um fenômeno particular do qual se procura compreender os componentes que afetam essa variabilidade. Como já tem sido dito, o crescimento da literatura representa um dos múltiplos aspectos desse fenômeno do crescimento da comunicação científica. Porém quando se modela

estatisticamente tal crescimento, temos que ver esse modelo estatístico como reflexo de um sistema real que tem sido desenhado, observando cada uma das características desse mundo real. O modelo do sistema real pode ser modificado e refinado de várias maneiras, até atingir um retrato mais realístico do sistema representado. No caso do crescimento da literatura, costuma-se representar os dados numa coordenada cartesiana, de tal maneira que a dispersão da nuvem dos pontos na coordenada sugira a forma do crescimento. Uma vez determinada a forma do crescimento, estima-se e testa-se o modelo pertinente para se determinar o seu ajuste ou não, ao modelo sugerido pelos

dados. Nesta parte, serão analisados o crescimento da literatura produzida sobre a produtividade dos autores, conhecida também como a Lei de Lotka, assim como o crescimento dos autores produtores desta literatura.

Crescimento da literatura produzida sobre a Lei de Lotka

A produtividade de artigos ordenados cronologicamente, de 1922 até 2003, bem como seu número acumulado, podem ser observados na tabela 1. O traçado da literatura sobre a Lei de Lotka, de 1922 a 2003, está mostrado na

TABELA 1

Cronologia da produção de publicações sobre a Lei de Lotka, 1922-2003

Anos	Nº de artigos	Artigos acumulados	Anos	Nº de artigos	Artigos acumulados	Anos	Nº de artigos	Artigos acumulados
1922	1	1	1950	0	6	1978	6	73
1923	0	1	1951	0	6	1979	6	79
1924	0	1	1952	0	6	1980	6	85
1925	0	1	1953	1	7	1981	11	96
1926	1	2	1954	3	10	1982	11	107
1927	0	2	1955	2	12	1983	5	112
1928	0	2	1956	0	12	1984	7	119
1929	0	2	1957	1	13	1985	14	133
1930	0	2	1958	0	13	1986	10	143
1931	0	2	1959	0	13	1987	11	154
1932	1	3	1960	1	14	1988	11	165
1933	0	3	1961	1	15	1989	12	177
1934	0	3	1962	0	15	1990	15	192
1935	0	3	1963	1	16	1991	11	203
1936	0	3	1964	0	16	1992	16	219
1937	0	3	1965	1	17	1993	12	231
1938	1	4	1966	1	18	1994	12	243
1939	0	4	1967	1	19	1995	17	260
1940	0	4	1968	1	20	1996	15	275
1941	0	4	1969	2	22	1997	10	285
1942	1	5	1970	4	26	1998	12	297
1943	0	5	1971	3	29	1999	17	314
1944	1	6	1972	2	31	2000	17	331
1945	0	6	1973	4	35	2001	20	351
1946	0	6	1974	5	40	2002	14	365
1947	0	6	1975	6	46	2003	16	381
1948	0	6	1976	12	58			
1949	0	6	1977	9	67			

figura 1. Este traçado ascende lentamente de 1922 até aproximadamente 1970 e, mais rapidamente, conforme o tempo expressado em anos aumenta a partir dessa data. O crescimento da literatura se assemelha a um crescimento exponencial. Nesse período observou-se um rango mínimo de 1 trabalho e um rango máximo de 20 publicações por ano, com média de 4,65 trabalhos e desvio padrão de 5,85 publicações por ano. Claramente, pode-se observar que a quantidade da literatura produzida começa a crescer a partir dos anos 1970.

O traçado da produção acumulada de trabalhos sobre a Lei de Lotka, de 1922 a 2003, é mostrado na figura 2, a seguir. Este traçado mostra uma forma côncava no início da distribuição e ascende lentamente de 1922 até aproximadamente 1970 e, mais rapidamente, após essa data. Porém, no geral a função é crescente e a curva inicial é côncava,

ascendendo conforme o tempo, expressado em anos, aumenta para se transformar numa linha quase reta. O crescimento da literatura se assemelha a um crescimento exponencial, conforme esperado.

A regressão não linear do número de trabalhos acumulados produzidos de acordo com os anos, executada usando SPSS 14.0 para Windows, produziu um R^2 igual a 0,989, significando que 98,9% da variação é explicada pelos anos de produção dos artigos. Isso significa que somente 1,1% da variação não é explicada pelos anos de produção dos trabalhos. O valor estimado de c foi igual a 1,738, com erro padrão de 0,170. O valor estimado de g foi igual a 1,070, com erro padrão de 0,001. Com esses valores pode-se construir a equação seguinte, como os valores que podem predizer o crescimento exponencial da literatura

FIGURA 1
Número de publicações sobre a lei de Lotka segundo os anos

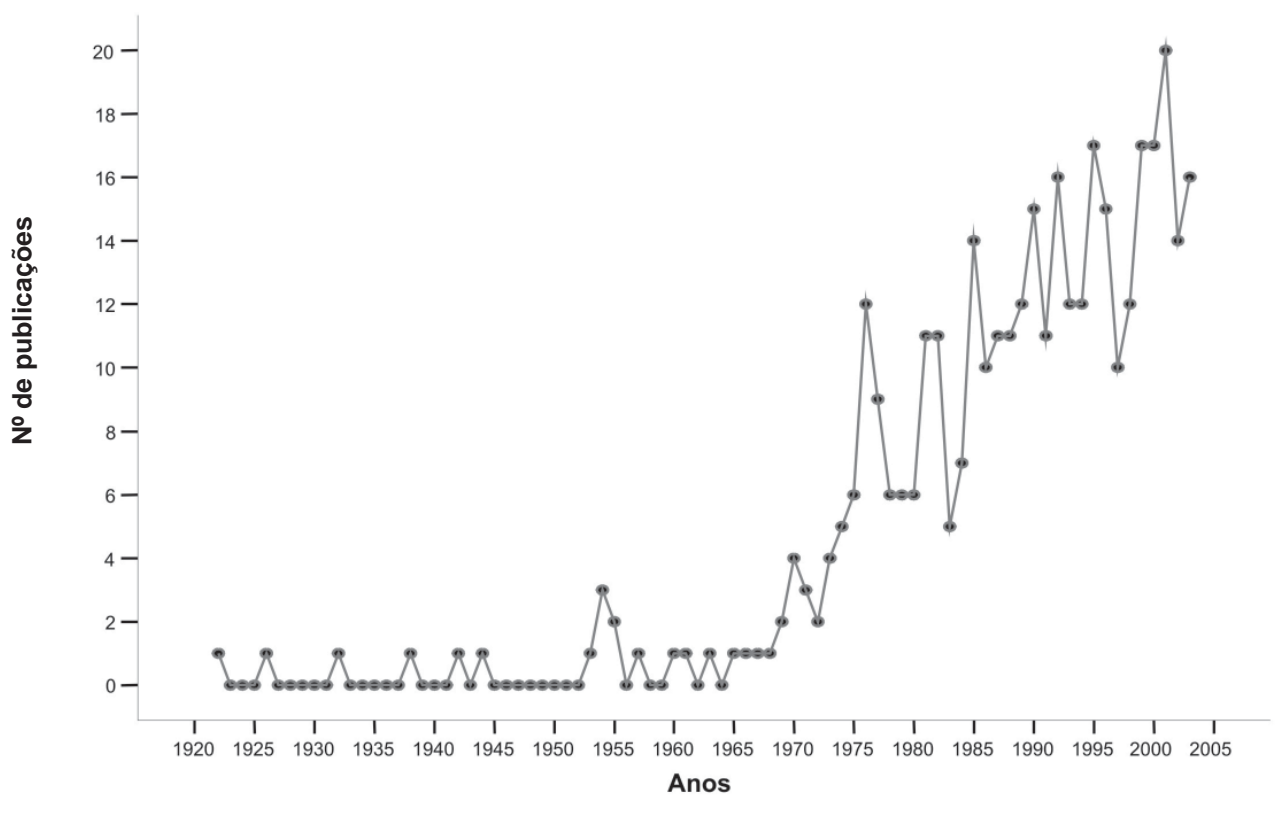
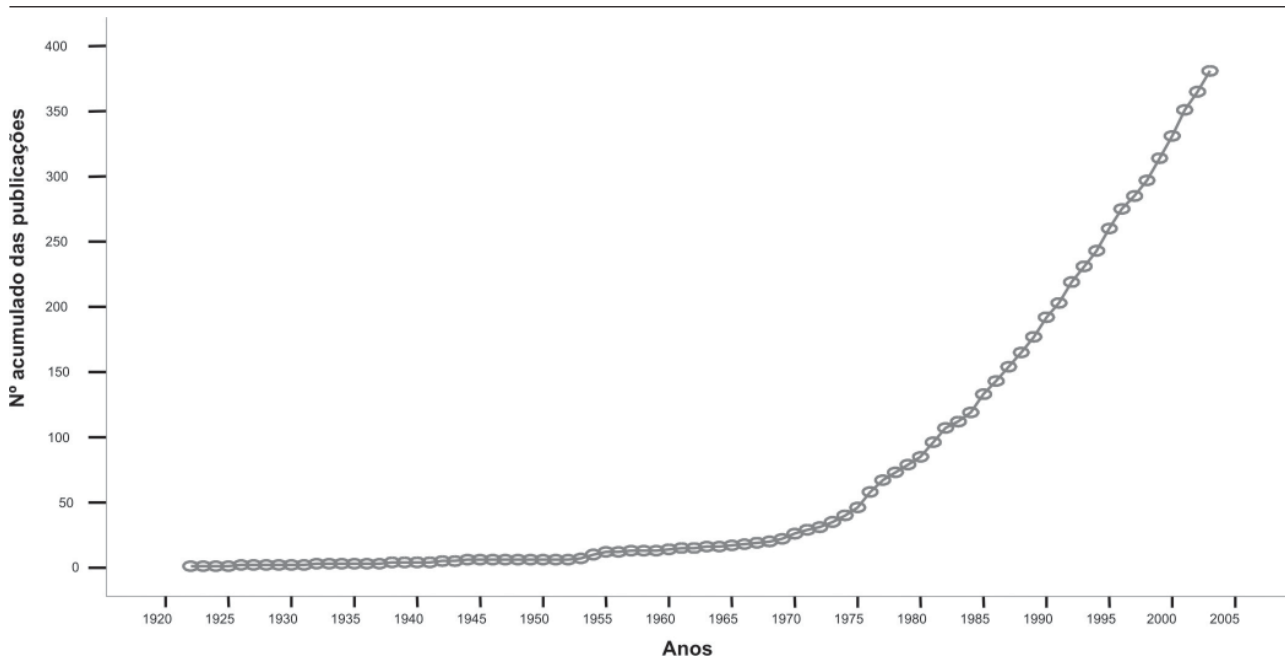


FIGURA 2

Número acumulado das publicações sobre a Lei de Lotka segundo os anos



no campo da lei de Lotka, tomando a forma da equação seguinte:

$$C(t) = 1,738 \times 1,070^t$$

Esta equação indica que a literatura cresce a uma proporção média de 7% ao ano. Uma vez conhecida essa taxa de crescimento, pode-se estimar o período de duplicação da literatura, da seguinte maneira:

$$(1,070)^n = 2,0$$

Tomando os logaritmos de ambos os lados dessa equação, temos que

$$n (\log 1,070) = \log 2,0$$

$$n = \frac{\log 2,0}{\log 1,070}$$

$$n = \frac{0,69315}{0,06766}$$

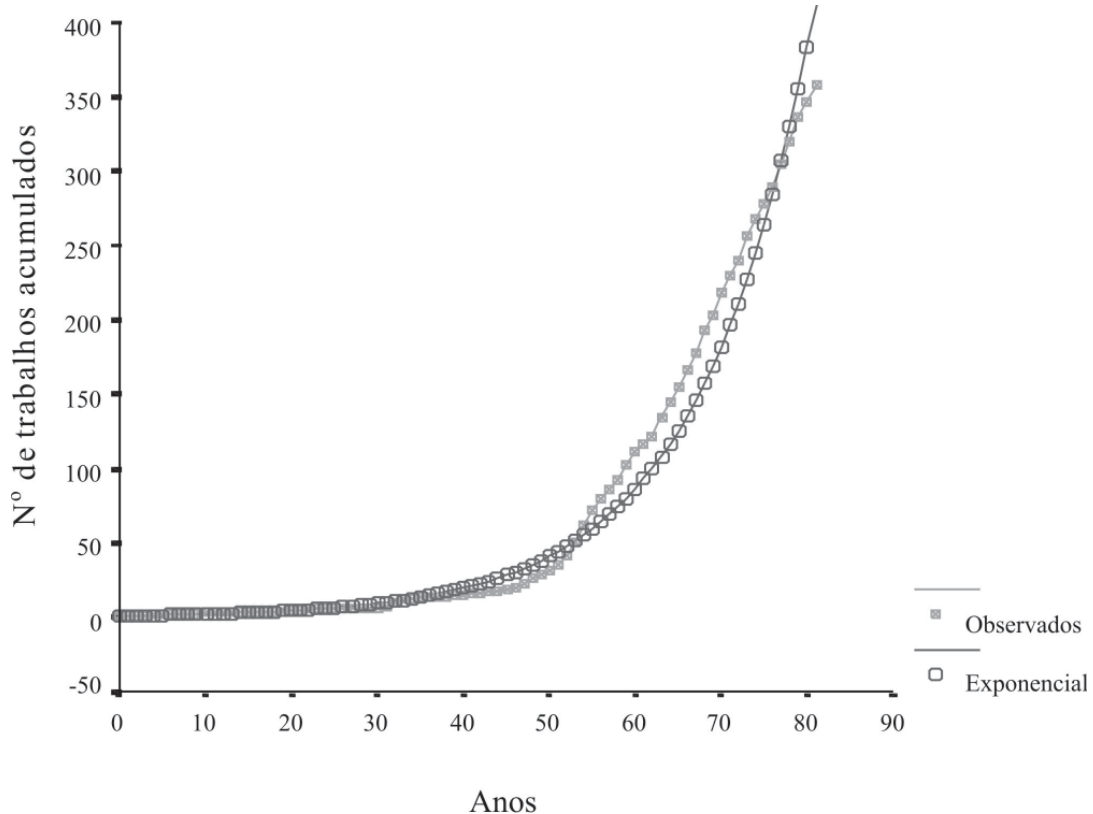
$$n = 10,2446$$

Portanto, o crescimento da literatura sobre a Lei de Lotka tem um período de duplicação a cada 10,2 anos. Também a estatística indica bom ajuste ao modelo exponencial, sendo a representação gráfica dos dados observados e esperados obtidos pela aplicação do modelo exponencial não linear mostrados na figura 3. Ravichandra Rao (1998) tem sugerido que o crescimento da literatura científica confirma, quase sempre, o crescimento exponencial, e que isso se reflete no processo de interação social através do qual o contato entre cientistas contribui para o crescimento cumulativo do conhecimento. Uma típica curva de crescimento é caracterizada por três estágios de crescimento:

Estágio 1: lento crescimento inicial, com muito pouco número de artigos, instituições e cientistas, caracterizado pelo fato de a informação chegar muito demoradamente aos cientistas. Há uma luta constante e permanente por comunicar os resultados das pesquisas.

Estágio 2: etapa de rápido crescimento da literatura produzida.

FIGURA 3

Dados observados e esperados do modelo exponencial

Estágio 3: etapa de estabilização, saturação e declínio do crescimento da produção da literatura científica.

Seguindo esses estágios, a figura 3 mostra que, até 1970, o crescimento da literatura produzida era lento, revelando todas as características de estar no estágio 1, com lento crescimento inicial de poucos trabalhos produzidos. Para esse estágio estimou-se taxa de crescimento de 6,1% ao ano e com período de duplicação a cada 11,7 anos. Porém, de 1970 a 2003, o crescimento da literatura parece entrar no estágio 2. Nesse período a literatura cresceu à proporção média de 6,6% ao ano e uma taxa de duplicação mais rápida a cada 10,8 anos.

Crescimento dos autores

Os autores produtores de trabalhos, ordenados cronologicamente segundo os anos, bem como seu número acumulado, podem ser observados na tabela 2, a seguir. Para a contagem da entrada dos autores no campo, considerou-se a data do primeiro artigo publicado, portanto; para autores com múltiplos trabalhos publicados, foi definida a sua entrada no campo na data da publicação do primeiro trabalho.

TABELA 2

Autores produtores de publicações sobre a Lei de Lotka, 1922-2003

Anos	Nº de autores	Autores acumulados	Anos	Nº de autores	Autores acumulados	Anos	Nº de autores	Autores acumulados
1922	1	1	1950	0	6	1978	5	68
1923	0	1	1951	0	6	1979	6	74
1924	0	1	1952	0	6	1980	3	77
1925	0	1	1953	1	7	1981	7	84
1926	1	2	1954	2	9	1982	13	97
1927	0	2	1955	0	9	1983	9	106
1928	0	2	1956	0	9	1984	6	112
1929	0	2	1957	1	10	1985	14	126
1930	0	2	1958	0	10	1986	8	134
1931	0	2	1959	0	10	1987	12	146
1932	1	3	1960	2	12	1988	6	152
1933	0	3	1961	1	13	1989	6	158
1934	0	3	1962	0	13	1990	9	167
1935	0	3	1963	1	14	1991	13	180
1936	0	3	1964	0	14	1992	9	189
1937	0	3	1965	0	14	1993	12	201
1938	1	4	1966	1	15	1994	13	214
1939	0	4	1967	1	16	1995	22	236
1940	0	4	1968	1	17	1996	25	261
1941	0	4	1969	4	21	1997	8	269
1942	1	5	1970	2	23	1998	6	275
1943	0	5	1971	3	26	1999	19	294
1944	1	6	1972	1	27	2000	22	316
1945	0	6	1973	3	30	2001	18	334
1946	0	6	1974	5	35	2002	15	349
1947	0	6	1975	5	40	2003	27	376
1948	0	6	1976	13	53			
1949	0	6	1977	10	63			

O traçado dos autores produtores de documentos sobre a Lei de Lotka de 1922 a 2003 está mostrado na figura 4, a seguir. Esse traçado ascende lentamente de 1922 até aproximadamente 1970 e, mais rapidamente, conforme o tempo expressado em anos aumenta a partir dessa data. O crescimento dos autores se assemelha a um crescimento exponencial. Nesse período observou-se o rango mínimo de um autor e o rango máximo de 27 autores por ano, com média de 4,6 autores e um desvio padrão de 6,6 autores por ano. Pode-se observar que a quantidade de autores atuando neste campo começa a crescer também a partir dos anos 1970.

Com a ideia de mostrar essas flutuações visualmente, elaborou-se a figura 5, a seguir, mas considerando

somente os autores produtores de literatura sobre a Lei de Lotka entre 1922 e 1970. Pode-se notar que, nesse período (1922-1970), há um crescimento muito pequeno e oscilante: observou-se o rango mínimo de um autor e o rango máximo de quatro autores por ano, com média de 0,47 autores e com erro padrão de 0,793 autores. Até 1970 foram acumulados somente 5,8% do total dos 376 autores identificados até 2003 como produtores neste campo.

Para demonstrar as flutuações visualmente, elaborou-se a figura 6, a seguir, mas considerando somente os autores produtores de literatura entre 1970 e 2003. Pode-se observar que nesse período a entrada de novos autores no campo começa a

FIGURA 4
Número de autores segundo os anos, 1922-2003

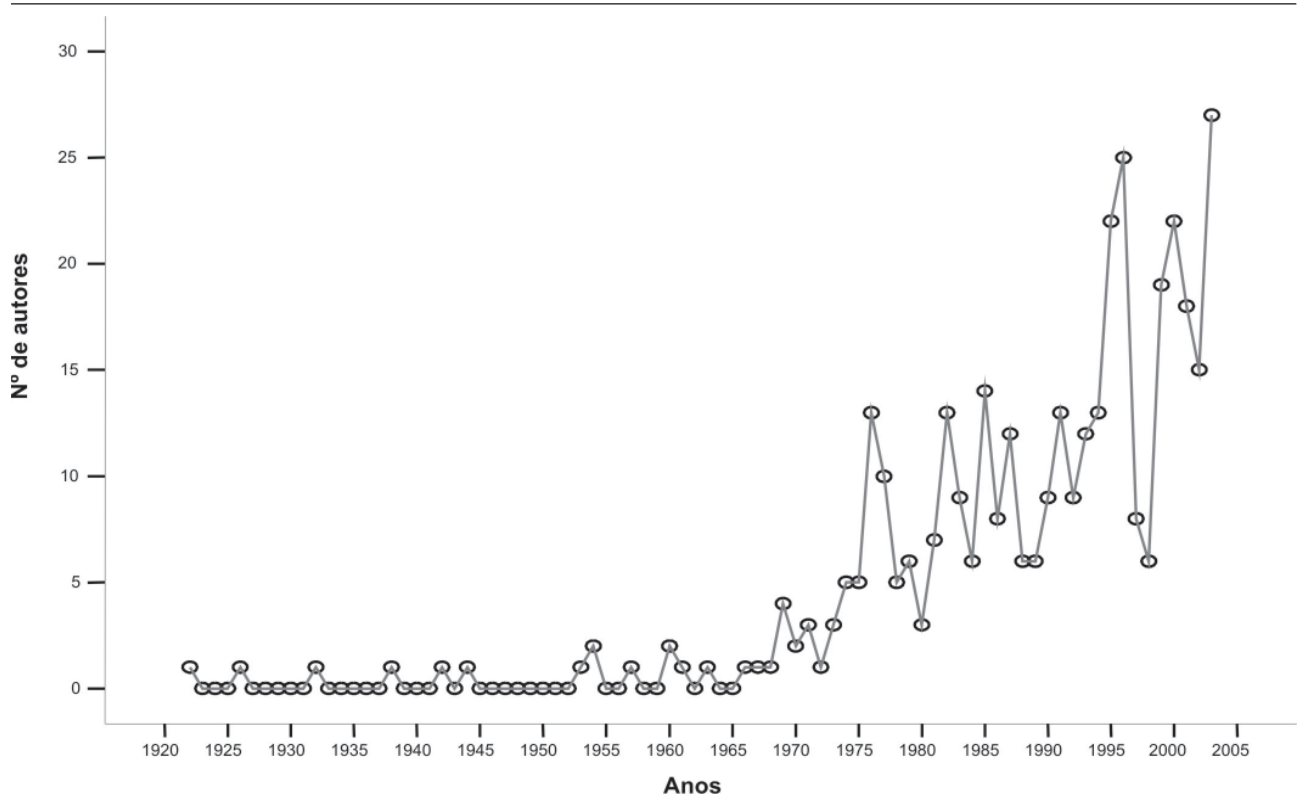


FIGURA 5
Crescimento do número de autores, 1922-1970

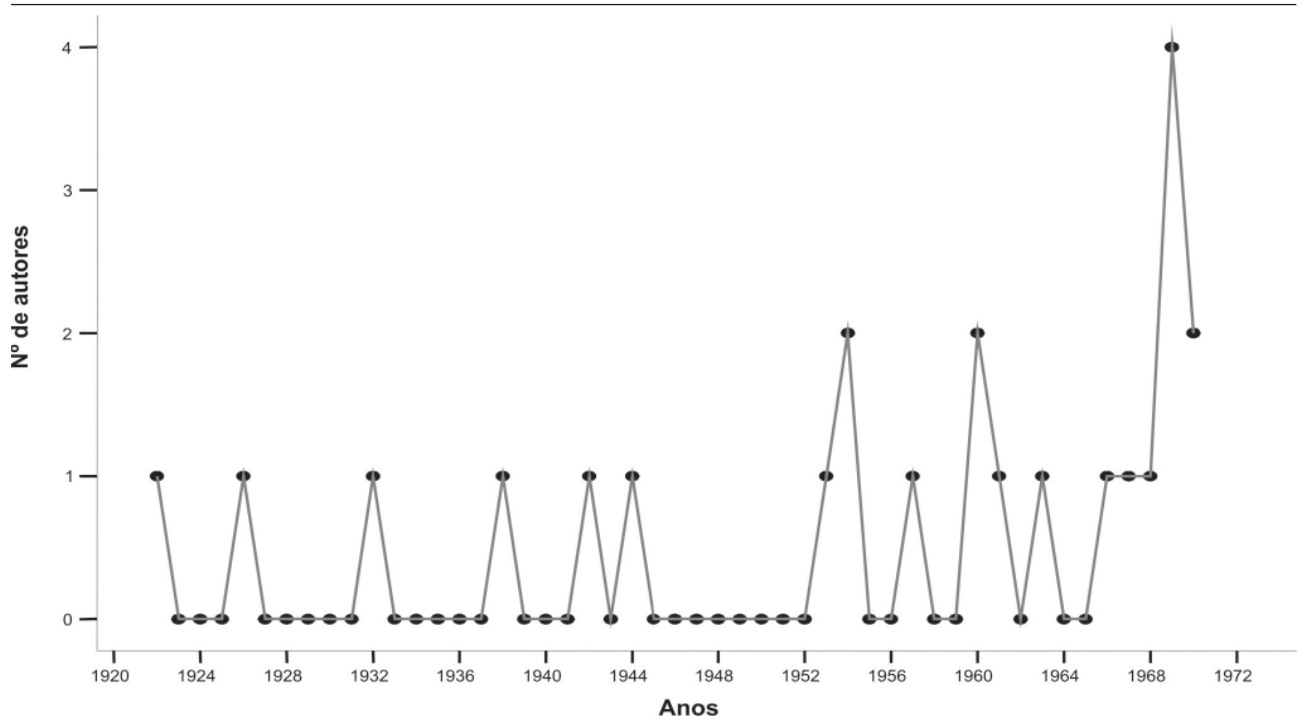
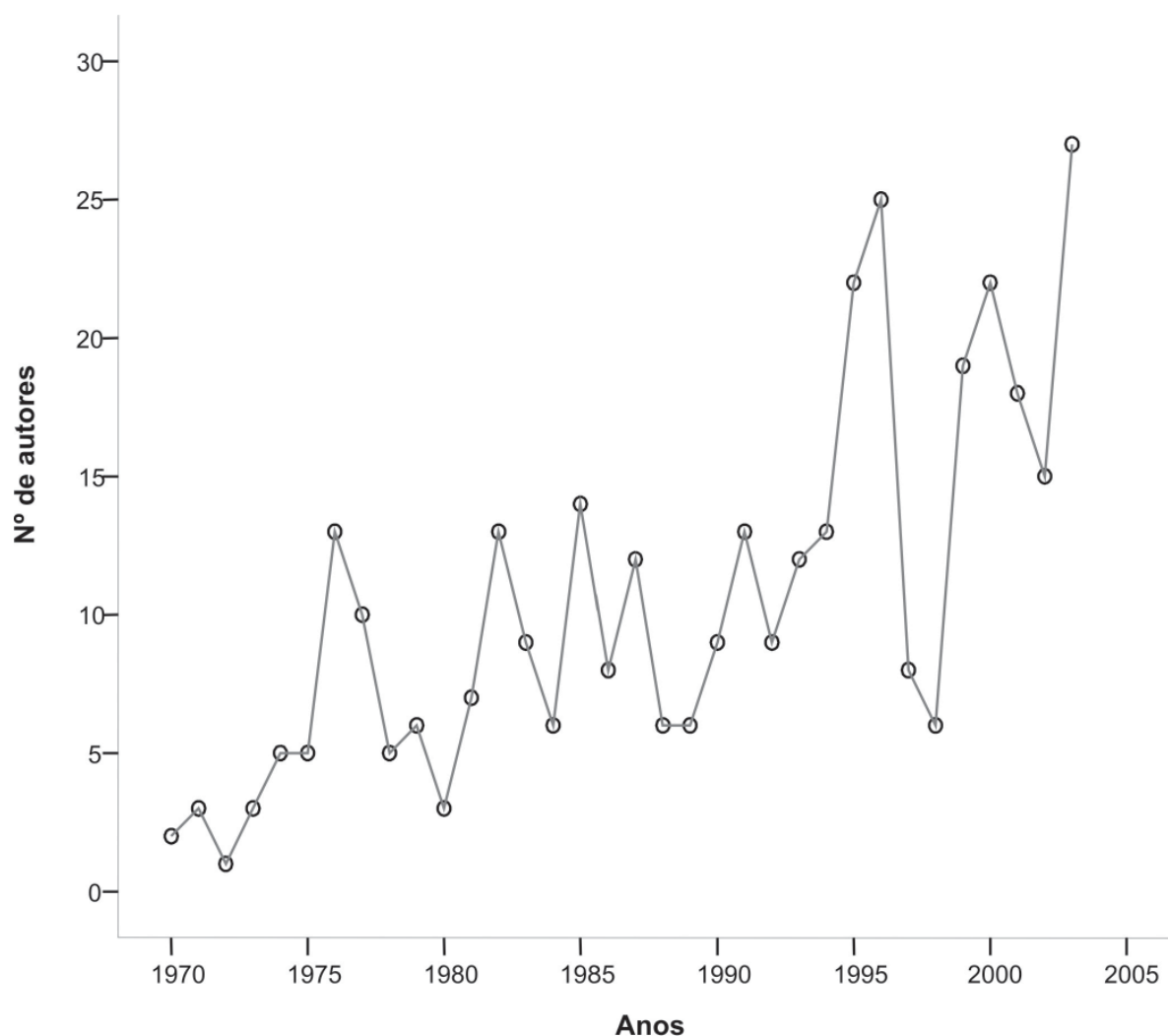


FIGURA 6
Crescimento do número de autores, 1970-2003

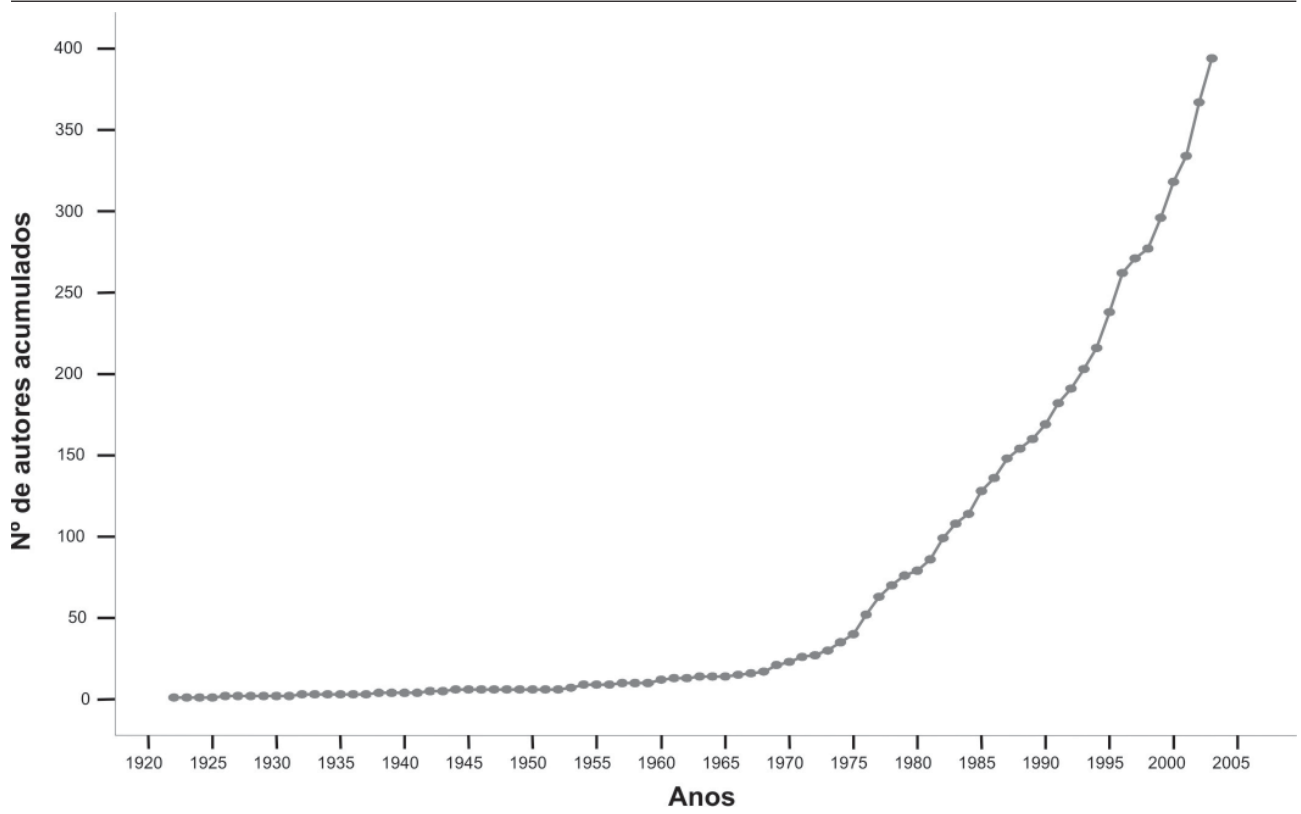


crescer mais claramente. Nesse espaço de tempo (1970-2003) observou-se rango mínimo de um autor e rango máximo de 27 autores, com média de 9,0 autores e desvio padrão de 6,7 autores por ano. O número de autores neste campo começa a crescer a partir de 1970, mas sempre com pequenas flutuações. Neste período apareceram 94,2% dos autores.

A representação gráfica dos autores acumulados segundo os anos pode ser observada na figura 7. Este traçado mostra forma côncava no início da distribuição, que ascende lentamente de 1922 até aproximadamente 1970, e um pouco mais rapidamente após essa data. No geral, a função é crescente e a curva inicial é côncava, ascendendo conforme o tempo, expressado em anos, que aumenta para transformar-se numa quase reta. A nuvem de pontos da distribuição do número de autores mostra crescimento de forma exponencial.

FIGURA 7

Número acumulado de autores segundo os anos, 1922-2003



O R^2 ajustado estimado foi igual a 0,992, indicando bom ajuste da distribuição ao modelo exponencial não linear. O valor estimado de c foi igual a 1,335, com erro padrão de 0,116. O valor estimado de g foi igual a 1,073, com erro padrão de 0,001. Com esses valores, pode-se construir a equação seguinte, como valores capazes de prever o crescimento exponencial dos autores atuando no campo da lei de Lotka:

$$C(t) = 1,335 \times 1,073^t$$

A equação indica que os autores crescem à proporção média de 7,3% ao ano. Uma vez conhecida essa taxa de crescimento, pode-se estimar o período de duplicação do crescimento dos autores da seguinte maneira:

$$(1,073)^n = 2,0$$

Tomando-se os logaritmos de ambos os lados da equação, obtém-se

$$n(\log 1,073) = \log 2,0$$

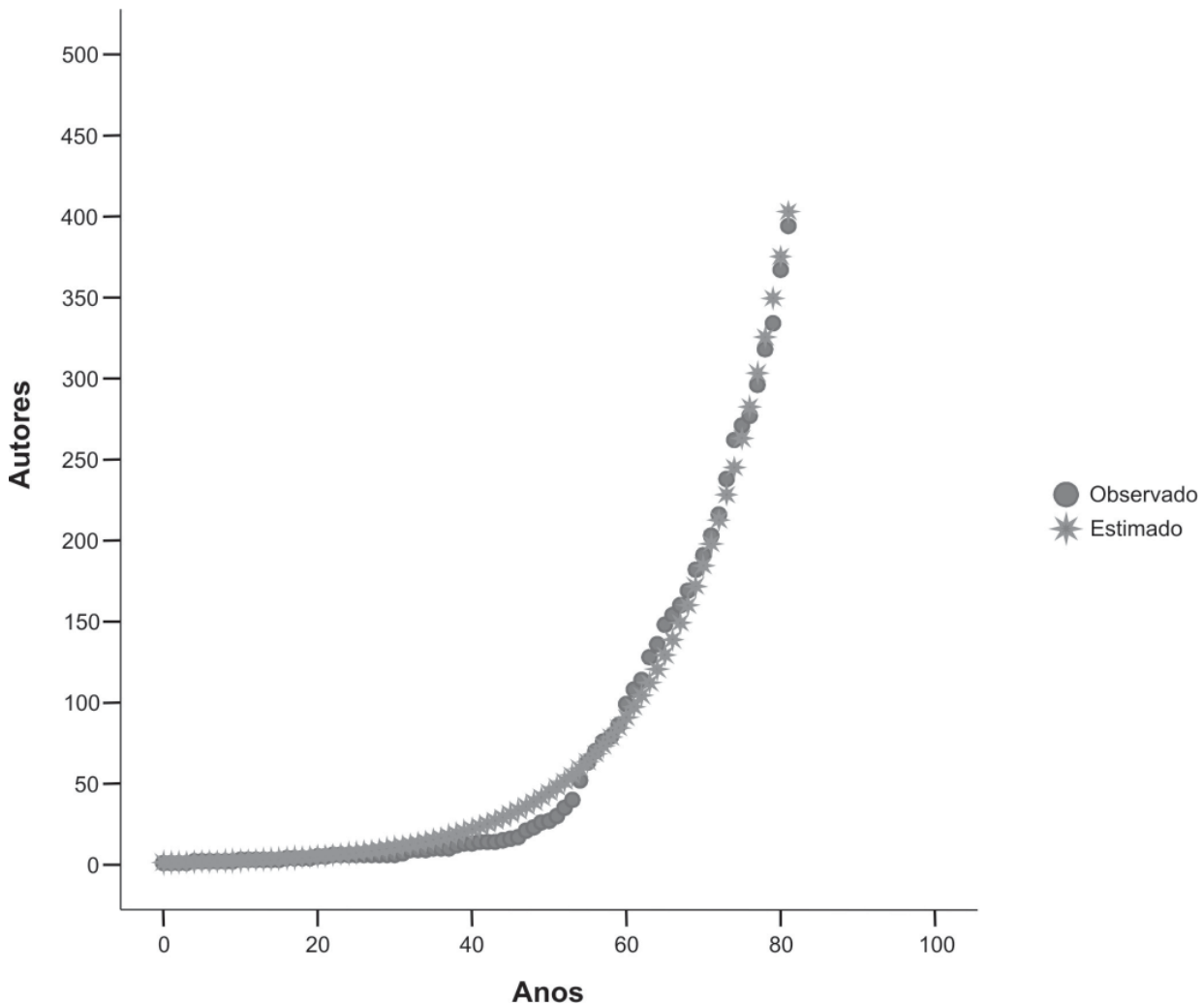
$$n = \frac{\log 2,0}{\log 1,073}$$

$$n = \frac{0,69315}{0,07046}$$

$$n = 9,83750$$

Portanto, os produtores de literatura sobre a Lei de Lotka duplicam-se a cada 9,8 anos (mais ou menos a cada 10 anos). Em outras palavras, os autores não mostram rápido crescimento e não muito longo período de duplicação, duplicando dez vezes em um século. Também a estatística indica bom ajuste ao modelo exponencial, sendo a representação gráfica dos valores observados e esperados obtidos pela aplicação do modelo exponencial não linear, mostrada na figura 8, a seguir.

FIGURA 8
Valores observados e esperados do modelo exponencial

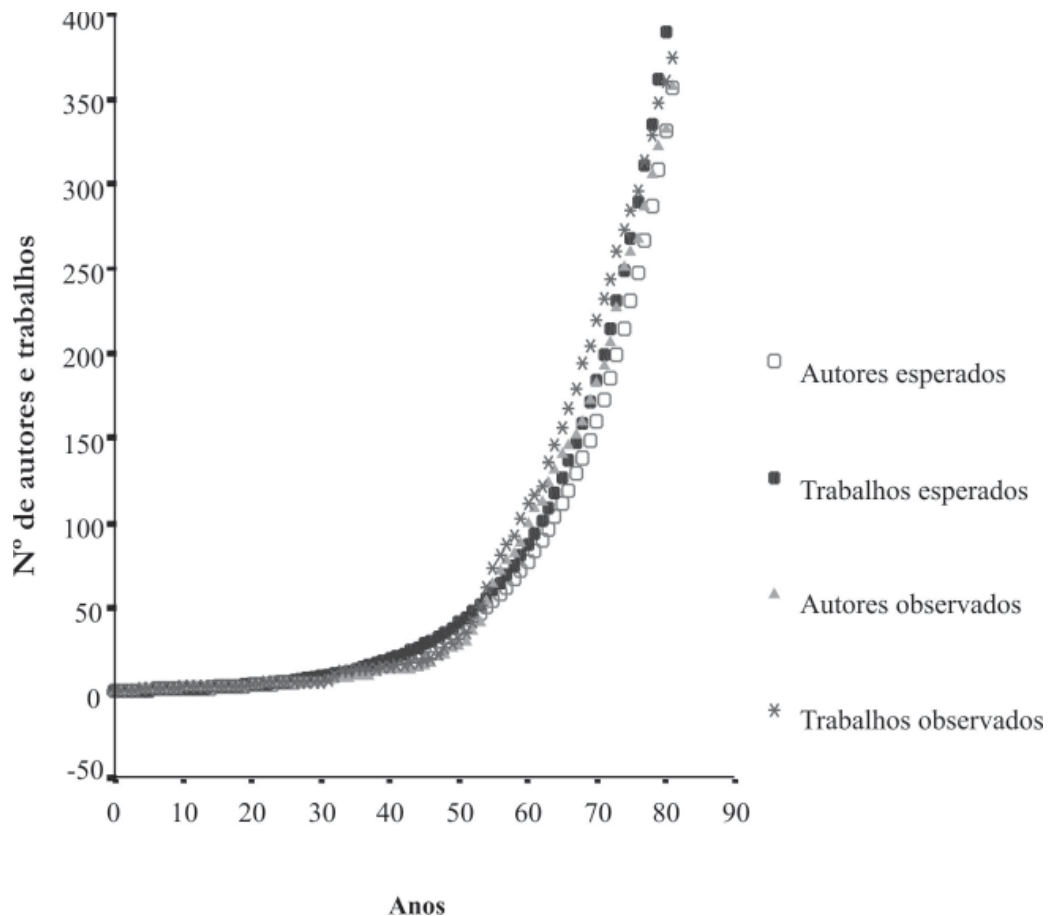


Entretanto, tudo indica que existe forte correlação entre o número de autores e o número de trabalhos produzidos. Tal correlação já tinha sido alertada por Brookes (1966), que chamava a atenção para o fato de que, se a literatura cresce, não é simplesmente porque os cientistas aumentam sua produtividade, senão também porque mais cientistas entram no campo e contribuem com taxa média de produtividade nesse campo.

Neste caso, a correlação de Pearson é igual a 0,999 ao nível de significância de 0,01, indicando forte correlação entre ambas as variáveis. A figura 9, a seguir, sumariza a regressão entre os autores observados e esperados, bem como entre os trabalhos observados e esperados da distribuição estratificada na produção de literatura sobre a Lei de Lotka.

FIGURA 9

Valores observados e esperados de autores e trabalhos



CONCLUSÕES

No período estudado, o crescimento da literatura sobre a Lei de Lotka segue o modelo exponencial, com taxa média de crescimento de 7,5% ao ano e taxa de duplicação a cada 9,6 anos. Também o número de autores produtores da literatura sobre a Lei de Lotka cresce à proporção de 7,3% ao ano, com período de duplicação a cada 10 anos. A estatística indica bom ajuste ao modelo exponencial, com R^2 de 98,5% para a literatura e 99,2% para os autores, ambos ao nível de significância de 0,01.

A taxa de duplicação desta literatura é muito similar aos 10 anos observados por Brookes (1973) na literatura científica, e no campo da geologia por Menard (1971), bem como muito próximo aos 9 anos observados por Fleischer (1984), no campo

da química geológica, e aos 9,9 anos notados por Nijagunappa; Gunjal, & Parvathamma (1996) na geologia indiana. Estes resultados contradizem as afirmações de Egghe & Ravichandra Rao (1992), no sentido de que a taxa de duplicação nas ciências sociais é menor do que nas ciências puras e aplicadas. Se este é o caso, a bibliometria e, em especial, a subárea da Lei de Lotka, estaria mais próxima das ciências puras e aplicadas do que das ciências sociais ou humanas. Portanto, considera-se que o crescimento da literatura independe da área estudada e parece ser mais consequência da intensidade da pesquisa e do volume de pesquisadores trabalhando nessas áreas.

Artigo submetido em 19/02/2009 e aceito em 17/03/2010.

REFERÊNCIAS

- AKHIBGE, Funmi O. Kolanut: the characteristics and growth of its literature in Nigéria. *Quarterly Bulletin of the International Association of Agricultural Librarians and Documentalists*, v. 33, n. 2, p. 47-52, 1988.
- BAKER, Dale B. Recent trends in chemical literature growth. *Chemical and Engineering News*, v. 59, n. 22, p.29-34, Jun. 1981.
- BALDWIN, Ira Lawrence; MCCOY, Elizabeth. *Root nodule bacteria and leguminous plants*. Madison : University of Wisconsin, 1932.
- BOTTLE, R. T.; EFTHIMIADIS, E. N. Library and information science Literature: authorship and growth patterns. *Journal of Information Science*, v. 9, n. 3, p. 107-116, 1984.
- BRAUN, Tibor; LYON, W.S.; BUJDOSÓ, Ernő. Literature growth and decay: an activation analysis resume. *Analytical Chemistry*, v. 49, n.8, p.682A-688A, Jul. 1977.
- BROOKES, Bertram C. Numerical methods of bibliographical analysis. *Library Trends*, v. 22, n. 1, p. 18-43, Jul. 1973.
- CHAKRABORTY, Ratan. Citation analysis of petrological literature. *Lucknow Librarian*, v. 15, n. 3, p. 99-104, 1983.
- COLE, F. J.; EALES, Nellie B. The history of comparative anatomy: a statistical analysis of the literature. *Science Progress*, n. 11, p. 578-596, 1917.
- CONRAD, G. Miles. Growth of biological literature and the future of Biological Abstracts. *Federation Proceedings*, v. 16, n. 2, o. 711-715, Jul. 1957.
- DRESDEN, A. A report on the scientific work of the Chicago section, 1897-1922. *Bulletin of the American Mathematical Society*, n. 28, p. 303-307, Jul. 1922
- EFTHOMIADIS, Efthimis Nikolaos. The growth of the OPAC literature. *Journal of the American Society for Information Science*, v. 41, n. 5, p. 342- 347, Jul. 1990.
- EGGHE, Leo; RAVICHANDRA Rao, I. K. Classification of growth models base on growth rates and its applications. *Scientometrics*, v. 25, n. 1, p. 5-46, 1992.
- FLEISCHER, Michael. Rate of growth of the literature of chemical geology, 1946-1980, and a peep into the future. In: MEETING OF THE GEOSCIENCE INFORMATION SOCIETY, 17., 1982. *Proceedings*. Washinton D.C.: Geoscience Information Society, v. 13, p. 121-129, 1984.
- GUPTA, B. M. Growth and obsolescence of literature in theoretical population genetics. *Scientometrics*, v. 42, n. 3, p. 335-347, 1998.
- _____; KARISIDDAPPA, C. R. Modeling the growth of literature in the area theoretical population genetics. *Scientometrics*, v. 49, n. 2, p. 321-355, 2000.
- _____; KUMAR, Suresh; Sangam, S. L.; KARISIDDAPPA, C. R. Modeling the growth of world social science literature. *Scientometrics*, v. 53, n. 1, p. 161-164, 2002.
- _____; SHARMA, Lalita; KUMAR, Suresh. Literature growth and author productivity in Indian physics. *Information Processing and Management*, v. 34, n. 1, p. 121-131, 1998.
- HU, CHENGREN. Microcomputers in library and information services (1973-1986): a bibliometric analisys of literature growth. *Microcomputers for Information Management*, v. 4, n. 3, p. 183-208, Sep. 1987.
- HULME, E. Windham. *Statistical bibliography in relation to the growth of modern civilization*. London: Butler and Tunner, 1923.
- LOTKA, Alfred. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, v. 16, n. 12, p. 317-323, Jun. 1926.
- MAHAPATRA, M. On the validity of the theory of exponential growth of scientific literature In: IASLIC CONFERENCE, 15., 1985, Dec. 26-29, Bangalore, Índia.; Bibliometrics studies: application of quantitative method to study the published literature; Current information: needs and resources. *Papers ...* I. N. Sengupta and S. K. Kapoor (Ed). Calcutá: Indian Association of Special Libraries & Amp, Information Centres, c1985. p. 61-70.
- MAY, Kenneth O. Quantitative growth of the mathematical literature. *Science, New Series*, v. 154, n. 3757, p. 1672-1673, Dec. 1966.
- MENARD, H. W. *Science: growth and change*. Cambridge, Mass.: Harvard University, 1971.
- MITEV, Nathalie Nadia; EFTHIMIADIS, Efthimis Nikolaos. *A classified bibliography on online public access catalogues*. London: Dept of Information Science, City University; Wolfeboro, N. H.: Longwood, 1987.
- NIJAGUNAPPA, R.; GUNJAL, S. R.; PARVATHAMMA, N. Indian Earth Science Literature (1978-88): a bibliometric study. In: GUPTA, B. M. (Ed) et al. *Handbook of libraries, archives and information centres in Índia*. New Delhi, Índia: Information Industry Publications, 1996. *Bibliometrics, Scientometrics and Informetrics*, v. 13, p. 237-247.
- PRICE, John Derek de Solla. *Little science, big science*. New York, N. Y.: Columbia University, 1963.
- QUIXIANG, Ding. Bamboos: the characteristics and growth of its literature in China from 1975-1986. *Quarterly Bulletin of the International Association of Agricultural Librarians and Documentalists*, v. 35, n. 3, p. 135-137, 1990.
- RANGARAJAN, K. S.; BHATNAGAR, Poonam. Mossbauer effect studies: some deductions from a bibliometric analysis. *Annals of Library Science and Documentation*, v. 28, n. 1/4, p. 32-38, 1981.

- _____. Mossbauer effect studies: some deductions from a bibliometric analysis. In: GUPTA, B. M. et al (Ed.). *Handbook of libraries, archives and information centres in Índia*. New Delhi, Índia: Information Industry Publications, 1996. *Bibliometrics, Scientometrics and Informetrics*, v. 13, p. 285-293.
- RAVI, S. Growth and collaboration trends in nuclear science research literature: a case study of Índia, 1980-1994. In: INTERNATIONAL SOCIETY FOR SCIENTOMETRIC AND INFORMATICS CONFERENCE, 8., 2001. Sidney, Austrália. *Proceedings ...* Sidney, Austrália: Mari Davis, Concepción Wilson Ed. *Bibliometric & Informetric Research Group*, v. 2, p. 573-585, 2001.
- RAVICHANDRA RAO, I. K. Informetrics: scope, definition, methodology and conceptual questions. In: DRTC WORKSHOP ON INFORMETRICS AND SCIENTOMETRICS. *Paper AA*. p. 1-14, 1998.
- _____. Classical laws of informetrics: an overview. In: DRTC WORKSHOP ON INFORMETRICS AND SCIENTOMETRICS. *Paper AB*. P. 1-37 1998
- SAXENA, Anurag; GUPTA, B. M.; JAUHARI, Monika. Forecasting growth of literature: All models are wrong, some are useful! In: INTERNATIONAL SOCIETY FOR SCIENTOMETRICS AND INFORMETRICS CONFERENCE, 8., 2001. *Proceedings ...*Sidney, Austrália, Mari Davis, Concepción Wilson. Sidney, Austrália: Bibliometric & Informetric Research Group, v. 2, p. 647-653, 2001.
- SHARMA, Praveen; GUPTA, B. M.; KUMAR, Suresh. Application of growth models to science and technology literature in research specialties. *DESIDOC Bulletin of Information Technology*, v. 22, n. 2, p. 17-25, Mar. 2002.
- STEPHENSON, Mary Sue. The research method used in subfields and the growth of published literature in those subfields: vertebrate, palentology and geochemistry. *Journal of the American Society for Information Science*, v. 36, n. 2, p. 130-133, Mar. 1985.
- SULLIVAN, Daniel et al. The state of the science: indicators in the speciality of weak Interactions. *Social Studies of Science*, v. 7, n. 2, p. 167- 200, May, 1977.
- TAGUE, Jean; BEHESHTI, J.; REES-POTER, L. The laws of exponential growth: evidence, implications and forecasts. *Library Trends*, v. 31, p. 125-149, Summer 1981.
- TAMIYA, Hiroshi. Eine mathematische Betrachtung über die Zahlenverhältnisse der in der Bibliographie von Aspergillus zusammengestellten Publikationen. *The Botanical Magazine*, v. 45, n. 530, p. 62-71, 1931.
- WILSON, P. W.; FRED, E. B. The growth curve of a scientific literature: nitrogen fixation by Plants. *The Scientific Monthly*, v. 41, n. 3, p. 240-250, Sep. 1935.
- WOLFRAM, D.; CHU, C. M., & LU, X. Growth of knowledge: bibliometric analysis using online database data. In: *Informetrics 89/90*: selection of papers submitted for the second International Conference on Bibliometrics, Scientometrics, and Informetrics, London, Ontario, Canada, 5-7 July 1989. New York : Elsevier, 1990. p. 355-372.
- WYLLYS, Ronald E. On the analysis of growth rates of library collections and expenditures. *Collection Management*, v. 2, n. 2, p.115-128, Summer 1978.