

Grupo de Trabalho 7: PRODUÇÃO E COMUNICAÇÃO DA INFORMAÇÃO EM CT&I

BIBLIOMETRIA, CIENTOMETRIA, INFOMETRIA: CONCEITOS E APLICAÇÕES

BIBLIOMETRICS, SCIENTOMETRICS, INFORMETRICS: CONCEPTS AND APPLICATIONS

Raimundo Nonato Macedo dos Santos
Universidade Federal de Santa Catarina
e-mail: raimundo@cin.ufsc.br

Nair Yumiko Kobashi
Universidade de São Paulo
e-mail: nykobash@usp.br

Resumo

Análise da constituição e institucionalização da bibliometria, da cientometria e da infometria. O estudo teve como objetivos principais identificar os conceitos-chave dessas disciplinas, suas semelhanças e diferenças, suas potencialidades e limites e as tendências contemporâneas da pesquisa. De natureza exploratória, o estudo se apoiou em literatura clássica e recente da área. Observou-se que, num primeiro momento, com o nome de bibliometria, os estudos procuravam quantificar os produtos da atividade científica (livros, artigos e revistas) para fins de gestão de bibliotecas e bases de dados; a cientometria, por sua vez, se constituiu como modelo que se preocupa com a interpretação dos dados quantitativos, à luz das teorias construídas no âmbito das ciências humanas e sociais (CHS). Seu objetivo principal é fornecer insumos para o planejamento e a avaliação de políticas científicas. A infometria, modelo mais recente, se apropria dos métodos bibliométricos e cientométricos para apreender os aspectos cognitivos da atividade científica. Nesse sentido, tem como preocupação central conhecer o estado-da-arte dos diferentes domínios do conhecimento. Observou-se, também, interesse crescente por técnicas de visualização da informação, com a finalidade de elaborar mapas que possam representar adequadamente os aspectos quantitativos e cognitivos da ciência. Conclui-se que os estudos métricos da informação se aproximam, desde Price, das CHS em busca de teorias e modelos que sustentem a interpretação dos dados quantitativos. Nesse sentido, os estudos métricos da informação, em sua configuração contemporânea, se configuram como campo interdisciplinar que se fertiliza entrecruzando as teorias e os métodos de quantificação com as teorias sociais.

Palavras-chave

Estudos métricos da informação; Bibliometria; Cientometria; Infometria.

Abstract

Analysis of the constitution and institutionalization of bibliometrics, scientometrics and informetrics. The research's main objective was to identify the key concepts of these

scientific domains, their similarities and differences, their potentialities and limits, and their contemporary research trends. The study was exploratory in nature and based on classic and recent literature. It was observed that, at first, the research on Bibliometrics tried to quantify the products of the scientific activities (books, papers and journals), for the purpose of libraries and database management. Scientometrics, on the other hand, was established as a model concerned with the interpretation of quantitative data, in face of Humanities and Social Sciences theories. Its main goal was to supply information for the planning and evaluation of scientific policies. The most recent model, informetrics, borrowed methodologies from bibliométricos and cientométricos to study the cognitive aspects of scientific activities. To this end, informetrics central concern is the state of the art of the knowledge domains. In addition, a growing interest for information visualization techniques was observed, with the aim at preparing maps that may adequately represent the quantitative and cognitive aspects of science. It is concluded that the information measurement research is leaning, since Price, towards the Humanities and Social Sciences, in search of theories and models that may support the interpretation of quantitative data. Therefore, the information measurements research, in its contemporary form, is set as a multidisciplinary field that is fermented by the crossover of quantitative methods with social theories.

Keywords

Quantitative studies of information; Bibliometrics; Scientometrics; Informetrics.

Introdução

Aqueles que necessitam reunir informações sobre o desenvolvimento da ciência enfrentam, por vezes, enormes desafios para localizar os itens mais pertinentes para subsidiar determinada tarefa. Os desafios se tornaram ainda mais agudos, na sociedade contemporânea, provocados pela progressiva informatização dos métodos de trabalho e a crescente ampliação das formas de armazenamento e de circulação do texto escrito, seja ele impresso, digital ou eletrônico. (LE COADIC, 2004a, 2004b). Esse lado desafiador vem sendo compensado, porém, pela criação de métodos e técnicas de tratamento, análise e visualização de informação, de naturezas diversas, baseados em princípios estatísticos e / ou lingüísticos.

Segundo Boustany (1997), a análise estatística de informações bibliográficas e a formulação de modelos ou leis vêm sendo feitas desde o século XIX. Sua expressão mais sistemática, porém, teve início no século XX, com os trabalhos de Lotka. A partir daí, as informações bibliográficas ou factuais, reunidas em bancos de dados públicos, de acesso gratuito ou mantidos por serviços comerciais, foram objeto de inúmeros estudos que resgataram ou deram origem a novas designações, de acordo com o objeto em foco: cientometria, infometria, tecnometria, museometria, arquiometria, iconometria, biblioteconometria, webmetria, entre outras. (ROSTAIN, 1996). Essas designações, certamente sintomas de crescimento do campo, geram controvérsias. Mesmo entre os especialistas da área não há consenso sobre a terminologia empregada no campo e sobre os limites que demarcam cada denominação utilizada.

Esse cenário de crescimento expressivo dos estudos métricos da informação cria a oportunidade e, também, a necessidade de refletir criticamente sobre as

questões conceituais, terminológicas e metodológicas da área. Não se tem aqui a pretensão de apresentar quadro exaustivo dos estudos já realizados na área ou visão histórica abrangente. Procuramos, apenas, resgatar aspectos do processo de constituição da área, seus conceitos e termos centrais, ao mesmo tempo em que lançamos um olhar crítico sobre estudos e aplicações contemporâneos. Assim, o artigo apresenta, inicialmente, breve histórico da constituição do campo, os principais conceitos da bibliometria, da cientometria e da infometria para, em seguida, abordar algumas tendências atuais dos estudos métricos da informação.

Bibliometria, cientometria e infometria: breve histórico

O uso de métodos estatísticos e matemáticos para mapear informações, a partir de registros bibliográficos de documentos (livros, periódicos, artigos), não constitui fato novo. É o que Boustany (1997) aponta, ao se referir a um autor desconhecido, citado no *Manuel du Bibliophile* ou, *traité du choix des livres*, editado em 1823, de autoria de Gabriel Peignol, que pesquisou a produção universal de livros no período compreendido entre a metade do século XV e início do século XIX.

No entanto, é no século XX que esses métodos ganham densidade e legitimidade. É bastante conhecida, por pesquisadores da Ciência da Informação, a proposta de Paul Otlet, considerado por historiadores franceses como o criador da bibliometria. Na obra seminal *Traité de documentation: le livre sur le livre*, o autor belga definiu a bibliometria como a área que se ocupa da medida ou da quantidade aplicada a livros. (OTLET, 1934, *apud* BOUSTANY, 1997).

Autores de origem anglo-saxônica, no entanto, atribuem a criação do termo bibliometria a Pritchard, que propôs sua utilização em substituição à bibliografia estatística, cunhado em 1923 por Hulme, como Rostaing (1996) retoma. A bibliometria foi caracterizada por Pritchard (1969) como conjunto de métodos e técnicas quantitativos para a gestão de bibliotecas e instituições envolvidas com o tratamento de informação. Os resultados das análises bibliométricas foram, nessa medida, considerados importantes coadjuvantes da definição de estratégias de gestão de unidades de informação e de bases de dados.

No âmbito dos esforços para quantificar os produtos da atividade científica, citar Lotka, Bradford, Zipf e Price se torna obrigatório. A lei do quadrado inverso, de Lotka (1926), refere-se ao cálculo da produtividade de autores de artigos científicos. Segundo essa hipótese, numa especialidade científica, coexiste pequeno número de pesquisadores extremamente produtivos com uma grande quantidade de cientistas menos produtivos. A lei de Bradford (1934), por sua vez, trata da dispersão dos autores em diferentes publicações periódicas. O que lhe interessava, à época, era determinar o núcleo dos periódicos que melhor se concentrassem em determinado tema. Essa lei resultou de estudos que visavam propor critérios de seleção de periódicos para uma dada coleção, de modo a equilibrar custo x benefício. É, portanto, uma lei voltada para fins gerenciais. Em se tratando da lei de Zipf, apresentada em 1935, refere-se à frequência da ocorrência de palavras num texto longo. Lei quantitativa fundamental da atividade humana, Zipf extraiu sua lei de um princípio geral do “esforço mínimo”: palavra cujo custo de utilização seja pequeno ou cuja transmissão demande esforço mínimo são frequentemente usadas em texto grande. Por fim, no caso de Price, ele será objeto de discussão mais ampla em contraposição aos estudos dos três primeiros,

por demais conhecidos, os quais serão citados apenas quando necessários para sustentar pontos de vista aqui defendidos.

A partir de Price, que se valeu das propostas de Lotka, Bradford e Zipf para formular suas leis cienciométricas, os estudos quantitativos adquiriram novos contornos, centrando-se fundamentalmente, na análise da dinâmica da atividade científica, incluindo tanto os produtos quanto os produtores de ciência. Uma nota: a abordagem do modelo de Price, aqui apresentada, beneficia-se largamente do trabalho de Polanco (1995). Este conhecido pesquisador da infometria, radicado na França, apresenta excelente balanço crítico sobre os antecedentes e o estado atual dos estudos métricos, em particular, sobre as contribuições de Price.

Para Price (1969), a cientometria é o estudo quantitativo da atividade científica. Semelhante caracterização teve como ponto de partida a percepção de que certas leis econométricas, em especial, as relativas ao cálculo da mão-de-obra, no mundo do trabalho, poderiam explicar, igualmente, o comportamento da literatura científica. Com base nesse raciocínio analógico, Price afirmou que os dados quantitativos sobre revistas e artigos científicos obedecem a certas regras estáveis, configurando-se como indicadores do estado da ciência.

A análise estatística da literatura científica, na abordagem de Price, encontra na história e na sociologia da ciência os contextos que dão sentido aos dados quantitativos analisados. Segundo Meadows (1990, *apud* Polanco, 1995).

The key figure in this new quantitative studies was Price, whose writings, especially *Little Science, Big Science* had a major impact on thinking about the growth and evolution of scientific journals. In part, he drew together ideas already under discussion. For example, the rapid growth in the amount of scientific literature had been debated by librarians and others since the First World War. Similarly, but separately, there had been work on relevant statistical distributions, such as Lotka's work on scientific productivity and Zipf's on word distributions. Price extended this earlier work to provide an integrated, quantitative picture of the scientific literature [...] One important area of Price's work covered the applications of citation analysis. In this, he relied on the contemporaneous activities of Garfield in developing the concept of a citation index.

O pioneirismo de Price manifesta-se, portanto, na elaboração de novo modelo que integrou a bibliometria ao estudo da atividade científica. Deve-se a ele a formulação teórica da ciência da ciência, campo que entrecruza a cientometria com a Ciência da Informação, como Garfield já percebera. Da mesma forma, Small e Garfield (1986) afirmaram que Price antecipara, nos anos 60, a idéia de evidenciar, por meio de mapas da ciência, o grau de cobertura e a importância relativa das revistas científicas. Os referidos mapas têm por base, segundo Price (1965), as "relações estruturais de uma rede de referências e citações", projeto que será desenvolvido no *Institute for Scientific Information* (ISI), na década de 70, por meio de métodos de co-citação. (POLANCO, 1995).

Os modelos bibliométricos e cientométricos têm os artigos científicos como objetos empíricos privilegiados. Pode-se, certamente, discutir à exaustão a pertinência destes últimos para examinar a atividade científica e realizar inferências sobre o

estado da ciência. Embora não discutamos as razões que levaram à eleição desse tipo de produto como objeto principal dos estudos métricos, fazemos uma ressalva: não se pode reduzir a atividade científica à produção, à circulação e ao consumo de artigos de periódicos e, muito menos, confundir o crescimento quantitativo de artigos com o desenvolvimento cognitivo da ciência. De fato, muitos autores já assinalaram que os modelos tradicionais tinham como foco a quantificação de artigos, de autores mais produtivos, de periódicos mais citados, etc., porém, não penetraram nas questões cognitivas, ou seja, no conhecimento contido nesses objetos. Como assinala Polanco (1995),

[...] le modèle bibliométrique de la science de Price ne saisissent pas directement le contenu cognitif des articles scientifiques, autrement dit les connaissances qu'ils véhiculent. Les informations présentes dans le titre de l'article, dans son résumé, ou dans le texte lui-même sont en réalité ignorées dans l'approche scientométrique instauré par Price [...] De ce fait, un tel dispositif ne réalise qu'une scientométrie "externaliste". Par conséquent, le défi est maintenant de développer une scientométrie "internaliste" si l'on peut ainsi s'exprimer. Ce qui s'est fait en France avec la mise au point de la méthode des mots associés au début des années quatre-vingts, utilisant justement les mots-clés comme indicateurs de connaissance [...]

Depreende-se do que foi dito, que a análise multidimensional, aplicada às palavras-chave de registros bibliográficos, configura-se como uma das contribuições teóricas mais recentes aos métodos quantitativos. Baseado no cálculo matricial e na álgebra linear, esse método supõe a classificação automática dos dados e sua representação por meio de cartografias temáticas.

O termo infometria, por sua vez, foi proposto em 1987, pela *International Federation of Documentation* / Federação Internacional de Documentação (FID), para designar o conjunto das atividades métricas relativas à informação, cobrindo tanto a bibliometria quanto a cientometria. (EGGE; ROUSSEAU, 1990). Semelhante caracterização é confirmada por Polanco (1995):

[...] a infometria comporta uma síntese da bibliometria e da cientometria, mas também como Brookes destacou tão bem, ela significa uma abertura ao estudo matemático da informação e sobre suas formas documentárias (Ciência social da informação) seja eletrônica ou física [...]

A partir do exposto acima, pode-se fazer uma primeira distinção entre bibliometria, cientometria e infometria. A bibliometria tem como objetos de estudo os livros ou as revistas científicas, cujas análises se vinculam à gestão de bibliotecas e bases de dados. A cientometria preocupa-se com a dinâmica da ciência, como atividade social, tendo como objetos de análise a produção, a circulação e o consumo da produção científica. A infometria, por sua vez, abarca as duas primeiras, tendo desenvolvido métodos e ferramentas para mensurar e analisar os aspectos cognitivos da ciência. Constata-se, nessa medida, a aproximação crescente dos estudos métricos da informação das ciências humanas e sociais (CHS). Trata-se, desse modo, não apenas de quantificar e constatar, mas de atribuir sentido aos dados, qualificando-os para que

possam ter melhor uso em políticas de ciência e tecnologia (C&T), por cada especialidade ou grupo de pesquisa, ou em contextos mais abrangentes, regionais, nacionais ou mundiais.

Em resumo, permanecem como conceitos operacionais centrais: a produtividade de autores de artigos científicos (com base nas leis de Lotka e Price); o núcleo e a dispersão de artigos em periódicos científicos (lei de Bradford); e a frequência de palavras em textos longos (lei de Zipf). O cálculo de co-ocorrências (de autores, de palavras, de instituições), fundada em métodos de análise multidimensional é uma das áreas que vem crescendo de forma acentuada nos estudos métricos contemporâneos.

Os estudos métricos na contemporaneidade

A partir do início deste milênio, as técnicas de visualização de informação, em apoio aos métodos e técnicas de tratamento e análise de informação, passaram a ser utilizadas de forma vigorosa e recorrente. Essas técnicas têm sido importantes para melhor perceber e compreender dados manipulados por meios estatísticos.

Com efeito, é possível, por meio de mapas da ciência e da técnica, fazer múltiplas inferências sobre os fatos que ocorrem na pesquisa científica. Essa tendência é evidente no programa da *12th International Conference on Scientometrics and Informetrics*, Rio de Janeiro, julho de 2009. O evento, reunindo pesquisadores nacionais e internacionais do domínio da cientometria e infometria, incorporou em sua programação a realização de três *workshops* sobre a visualização da informação:

1. WK1 (*This workshop aims to contrast perspectives on how to measure and map interdisciplinary research*) – este *workshop* objetiva identificar e avaliar pesquisas interdisciplinares: métricas e mapas;
2. WK2 (*This workshop aims to provide an interactive forum for researchers and educators to share their experiences and needs for understanding and analyzing scientific literature using CiteSpace*) – este *workshop* objetiva visualizar e analisar a literatura científica com *CitySpace*. Ressalta-se que *CiteSpace* é uma aplicação Java para analisar e visualizar redes de citação, como Chen (2004) esclarece. A meta central do *CiteSpace* é facilitar a análise de tendências emergentes num determinado campo de conhecimento. Permite ao usuário se informar com rapidez das novidades na área e, então, tirar proveito.
3. WK3 (*This workshop aims to provide an interactive forum to discuss how maps of science can assist in the teaching of graduate-level courses*) – este *workshop* objetiva utilizar mapas de ciência para ensinar ciência.

Segundo Nascimento e Ferreira (2005), os estudos das técnicas de visualização de informações distribuem-se em dois eixos principais. O primeiro leva em consideração o modo como essas técnicas exploram o substrato visual, as marcas e as propriedades visuais do desenho. No segundo caso, de particular interesse para os principais usuários de informação científica (comunidade científica e gestores de C&T),

as técnicas são classificadas de acordo com as características dos dados a serem visualizados.

As Figuras 1 e 2 são formas tradicionais de visualização de informação. A Figura 1, estudo exploratório de Momm (2009), mostra a distribuição de orientadores de dissertações e teses de programas de pós-graduação *stricto sensu* de uma área de avaliação da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Dos 334 trabalhos defendidos, num período de seis anos, 20 docentes foram responsáveis pela orientação de 66,2% enquanto os demais, aproximadamente 40, orientaram 33,8%.

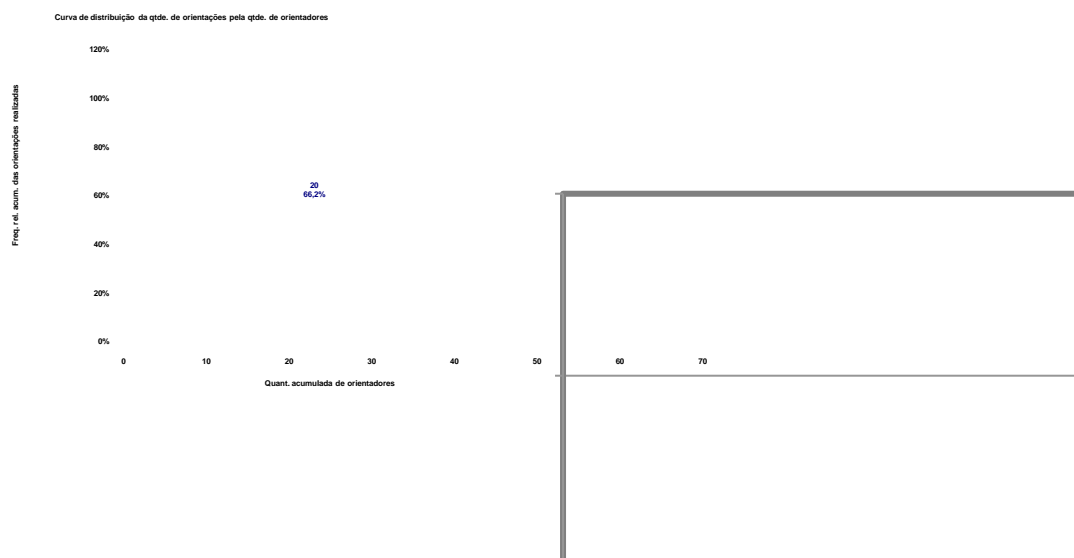


Figura 1 – Curva de distribuição de orientadores por dissertações produzidas

A Figura 2, por sua vez, indica a distribuição mundial dos artigos em periódicos científicos por região, em 1993, obtido da fonte *Science & Engineering Indicators*, 1995.

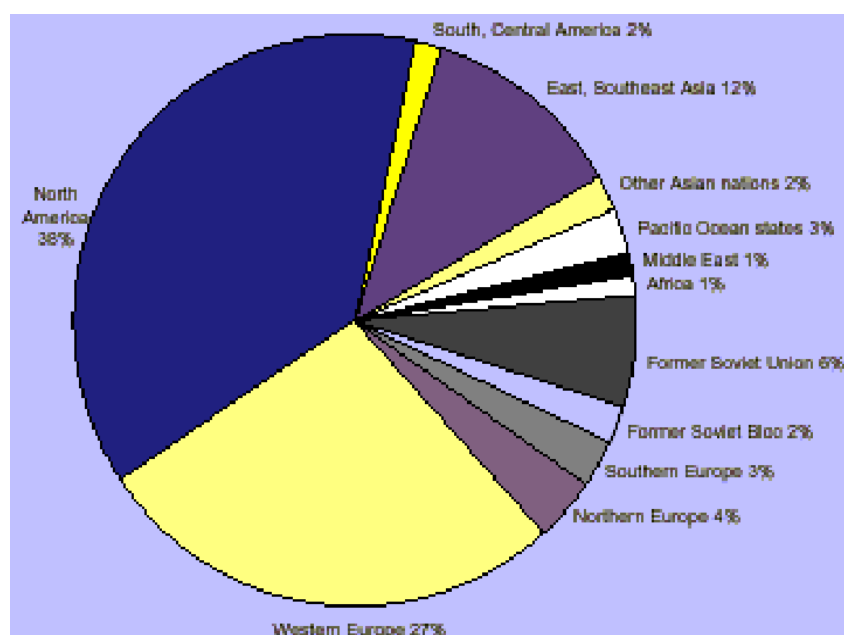


Figura 2 – Percentual de artigos científicos publicados em 1993, por regiões do mundo

Técnicas mais elaboradas de visualização são aplicadas a dados que possuem múltiplas dimensões (n dimensões, sendo n maior que três). Uma técnica consiste em mapear apenas dois atributos dos dados para as coordenadas X – Y de um plano cartesiano e associar os demais atributos à forma, cor e tamanho de objetos visuais. Pode-se, ainda, projetar o espaço n -dimensional no plano cartesiano, através da técnica conhecida por *Multidimensional scaling*. (BORG; GROEMEN 199,7 *apud* NASCIMENTO; FERREIRA, 2005).

Outras técnicas difundidas na literatura são: (a) a das coordenadas paralelas, que consiste na associação das dimensões dos dados a eixos paralelos verticais eqüidistantes, denominados coordenadas; (b) a dos *Glyphs* (ou ícones), entendidos como representações simbólicas que evidenciam as características essenciais de um dado, segundo colocações de Nascimento e Ferreira (2005).

A Figura 3, por seu turno, consiste em exemplo de técnica de visualização. Representa o *ranking* de produção de artigos científicos de países da América do Sul, América Central e Caribe de Herrero-Solana (2005).

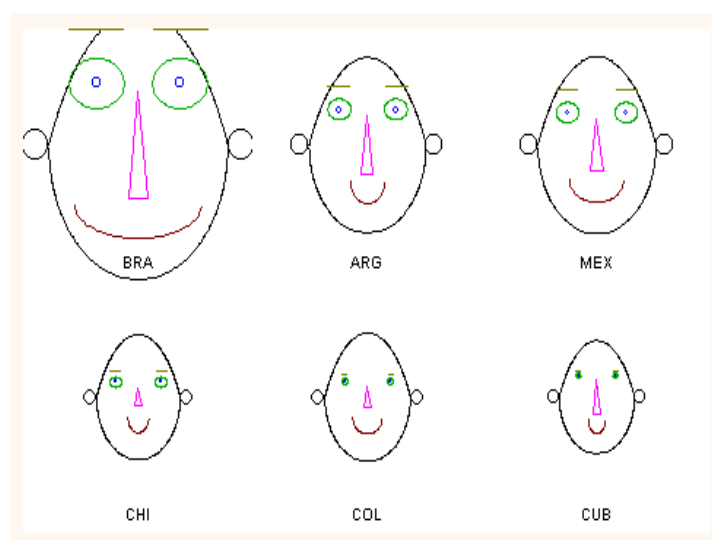


Figura 3 – *Ranking* da produção de artigos científicos de países da América Latina, América Central e Caribe

Há diversas formas de apresentação gráfica de dados relacionados hierarquicamente. O chamado *Tree-Map* (Figura 4), criado por Johnson e Shneiderman (1991), do Laboratório de Interação Humano-Computador (*Maryland University*), é um exemplo de técnica de representação hierárquica de dados.

Os dendogramas constituem outra técnica de visualização de dados. Também conhecidos como diagrama de árvore, permitem visualizar, hierarquicamente, resultados de análise de agrupamentos.

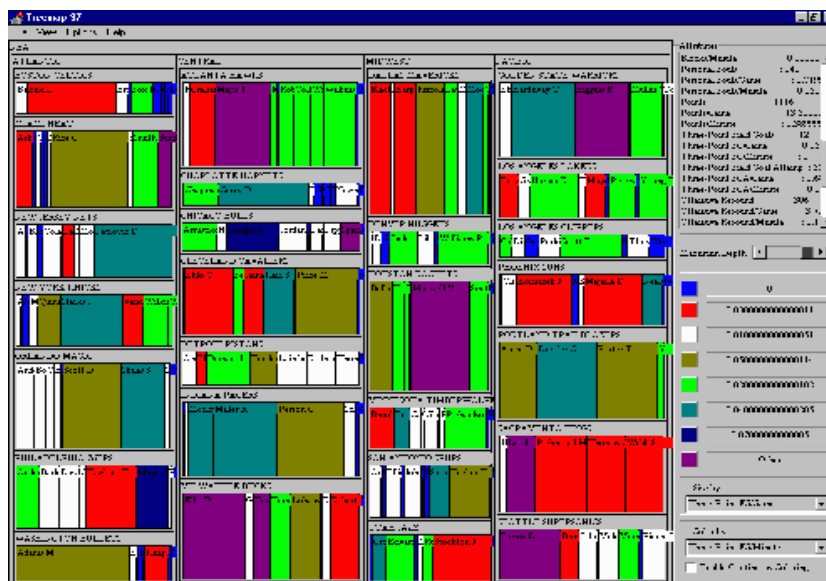


Figura 4 – Exemplo de imagem gerada com a técnica *Tree-Map*

As visualizações apresentadas nas Figuras 5 e 6 são dendogramas que exemplificam experiências de análise de dados, realizadas por Fachin e Santos (2009), que analisaram, com fins exploratórios, os registros da área de organização da informação e do conhecimento, armazenados em *Library and Information Science Abstracts* (LISA).

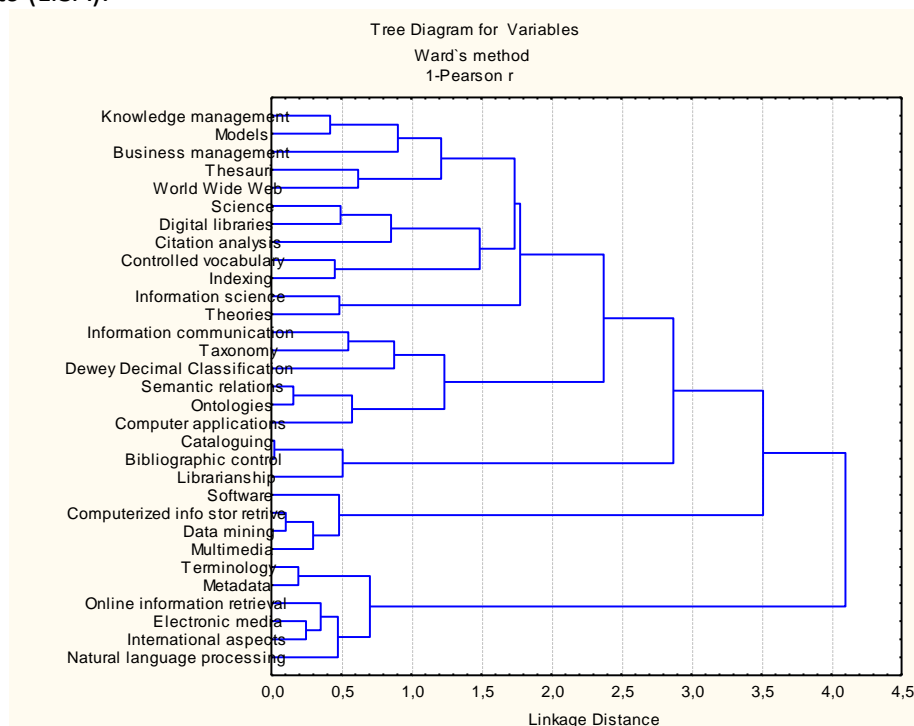


Figura 5 – Representação hierárquica de registros sobre *knowledge organization*, da base *Library and Information Science Abstracts* (LISA)

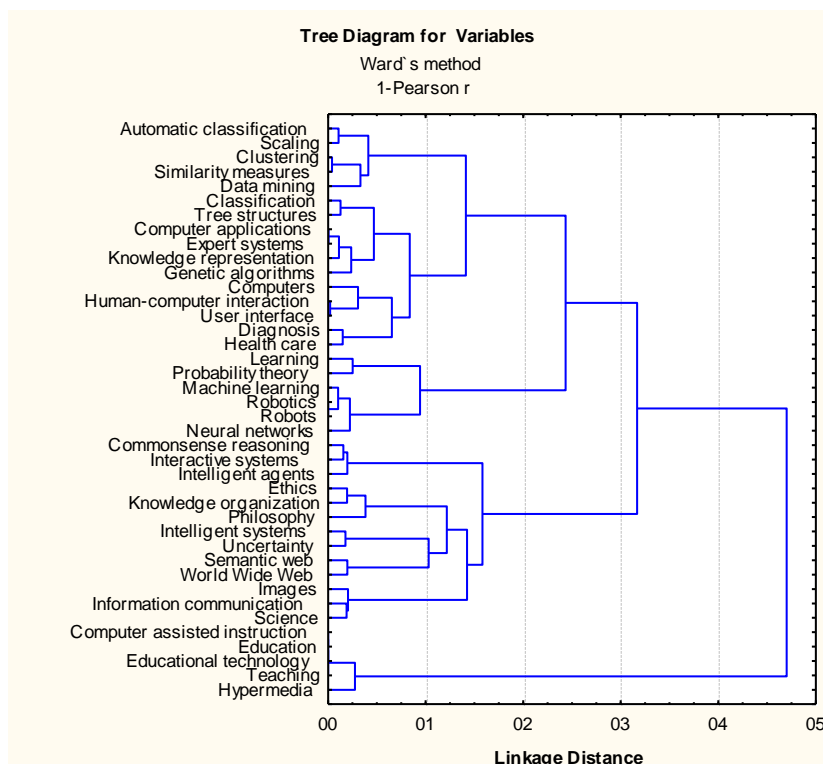


Figura 6 – Representação hierárquica de registros sobre *Knowledge representation* da base *Library and Information Science Abstracts (LISA)*

As Figuras 5 e 6 sistematizam dados coletados utilizando os descritores *Knowledge organization* e *knowledge representation*, considerados representativos do objeto em análise. Esses termos são coordenados, ou seja, são termos específicos de mesmo nível, subordinados ao termo genérico *Knowledge*, no tesauro LISA.

As referências bibliográficas coletadas foram armazenadas numa base de dados *ad hoc* e reformatadas para processamento bibliométrico (contagem de objetos) e análise cientométrica (análises multidimensionais dos descritores correlacionados). Nos dendogramas gerados, os descritores foram agrupados utilizando-se o método Ward, a partir da distância de 1-r de Pearson. Para comparar os agrupamentos, foi adotada a distância igual a 1. Observa-se, nas duas figuras, que os nós estão agrupados dois a dois (*thesauri*, *WWW*), (*controlled vocabulary*, *indexing*) ou três a três (*knowledge management*, *models*, *business management*), (*science*, *digital library*, *citation analysis*).

Da análise dos agrupamentos, infere-se que, na Base LISA, os artigos indexados com o descritor *knowledge organization* têm conteúdos relativos aos fundamentos da organização do conhecimento, enquanto os artigos indexados com o descritor *knowledge representation* são relativos às aplicações ditas tecnológicas da organização do conhecimento.

Os dendogramas têm sido utilizados também em pesquisas de institucionalização social e cognitiva da ciência, desenvolvidos pelo Grupo de Pesquisa *Scientia*, cadastrado no Diretório de Grupos de Pesquisas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Plataforma Lattes, sob a coordenação dos pesquisadores Raimundo Nonato Macedo dos Santos (Universidade Federal de Santa Catarina) e Nair Yumiko Kobashi (Universidade de São Paulo). Nessas pesquisas, busca-se identificar, por meio da análise de descritores de um *corpus* de registros bibliográficos, o conhecimento neles inscrito

e interpretá-lo à luz das teorias sociais da ciência. Esses resultados podem ser úteis para gestão da política científica ao nível dos programas de pesquisa, particularmente em relação às prioridades a serem estabelecidas nos grupos de pesquisa, como também para fazer análises do estado-da-arte de campos científicos.

O núcleo e a dispersão na distribuição de dados

Dois fenômenos podem ser destacados nas representações visuais dos estudos bibliométricos, cientométricos e infométricos: a conjunção e a disjunção de dados. A Figura 7 ilustra, de forma genérica, os fenômenos da conjunção (núcleo) e disjunção (dispersão)

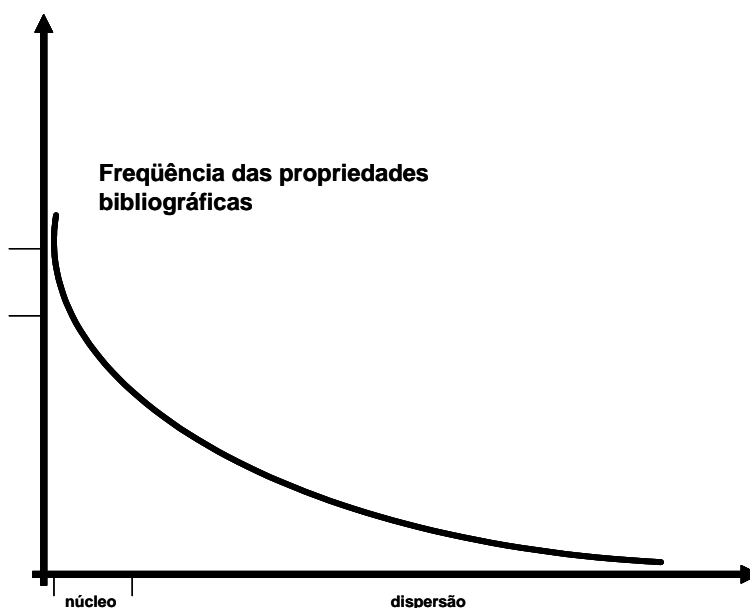


Figura 7 – O núcleo e a dispersão de uma distribuição bibliométrica

Na figura 7,

- O núcleo representa o grupo de dados que aparecem com as maiores freqüências, relativamente ao conjunto dos itens analisados. Levando-se em consideração a Lei de Lotka, o núcleo simboliza os autores mais férteis numa área de especialidade.
- A dispersão representa os dados de baixa freqüência relativamente ao conjunto de dados. Segundo a mesma lei de Lotka, a dispersão corresponde à grande diversidade de autores que publicam muito pouco dentro da área de especialidade estudada.

Nesta modelização, o núcleo é o espaço no qual se distribuem os elementos bibliográficos redundantes e / ou que identificam certa área ou certo domínio de especialização. Na zona da dispersão, é onde se distribuem os elementos que representam a individualidade ou a variação da área.

O mesmo princípio da distribuição em núcleo e dispersão é observado nos estudos de citação: é comum um autor reconhecido citar, principalmente, seus

próprios trabalhos, demonstrando que o núcleo cultiva a identidade. Por outro lado, os demais autores freqüentemente citados guardam vínculos intelectuais com o autor e, assim, alimentam uma identidade social.

O *Science Citation Index* (SCI) do ISI é, talvez, o exemplo mais ilustrativo da aplicação dessa regra. Criado nos anos 1960, em sua concepção inicial, o SCI pretendia se colocar como alternativa para recuperar informações científicas por meio da indexação dos elementos bibliográficos do campo *citation index*. Esse recurso permitia formular estratégias de busca pelo nome do autor e / ou do periódico mais citado, em determinada área de conhecimento: trata-se, novamente, da idéia de um núcleo que cultiva a identidade. (WORMELL, 1998).

A facilidade de compreensão e aplicação dessa regra pode gerar ao menos um questionamento: porque os diferentes atores da comunicação de conhecimento científico tenderiam a se submeter a tal regra? Dos estudos realizados para responder a esta indagação, decorreram duas regras:

1. A **vantagem do acúmulo** (*cumulative advantage*) – subentende-se que quanto maior a freqüência de uma referência bibliográfica, maior a tendência a reutilizá-la;
2. A **especialização** – quanto maior a freqüência de uma referência bibliográfica, menor a quantidade de informação a ela associada e, portanto, maior é a probabilidade de que seja rejeitada em favor de uma referência mais específica ou inovadora.

Em princípio, essas regras de comportamento antagônico parecem aplicáveis aos demais elementos bibliográficos de um registro: a vantagem do acúmulo tende para o efeito do núcleo e a especialização, para o efeito da dispersão.

Os colégios invisíveis também se comportam segundo as regras do núcleo e da dispersão. O comportamento dos pesquisadores, conforme as regras abaixo descritas, sinalizam, também, o que a área entende como uma “característica perversa da ciência”:

Efeito núcleo – quando se pesquisa, se interage dentro de um grupo de trabalho, o que pode gerar o monopólio de um periódico.

Efeito dispersão – é próprio, do processo de criação científica, o desafio permanente para a novidade. Promove-se, assim, a extensão contínua de redes de trabalho como forma de evitar a redundância.

Ligações ou relacionamentos entre dados

A análise das redes de colaboração é um campo em crescimento. Conhecida também como pesquisas de redes sociais (MARTELETO, 2001), a noção remonta aos estudos de citações e de co-autoria. Contemporaneamente, os estudos de redes sociais se beneficiam de modelos matemáticos e de *softwares* de tratamento e manipulação de dados, cujo cerne são os conceitos de entidade e relacionamentos.

Certamente, entidades de apreensão instantânea, como nomes de autores, de instituições e de revistas são mais fáceis de ser mapeados.

A Figura 8 é um exemplo de grafo ou rede de relacionamento entre empresas depositantes de patentes sobre catalisadores. O estudo de Santos (1998) apresenta rede de relacionamentos entre grandes empresas petrolíferas que, até 1996, compartilhavam a prioridade de depósito de patentes de tecnologias de produção de catalisadores de refino de petróleo para fabricação de óleo diesel. As Figuras 9 e 10, por seu turno, constituem exemplos de grafos de relacionamentos entre palavras-chave. (SANTOS, 2009).

Os estudos com entidades semânticas, como os descritores de uma base de dados, requerem intenso tratamento prévio de consolidação dos dados antes de submetidos ao processamento bibliométrico. Desse modo, é necessário olhar, com cautela, a classificação automática de dados porque podem fornecer resultados pouco confiáveis.

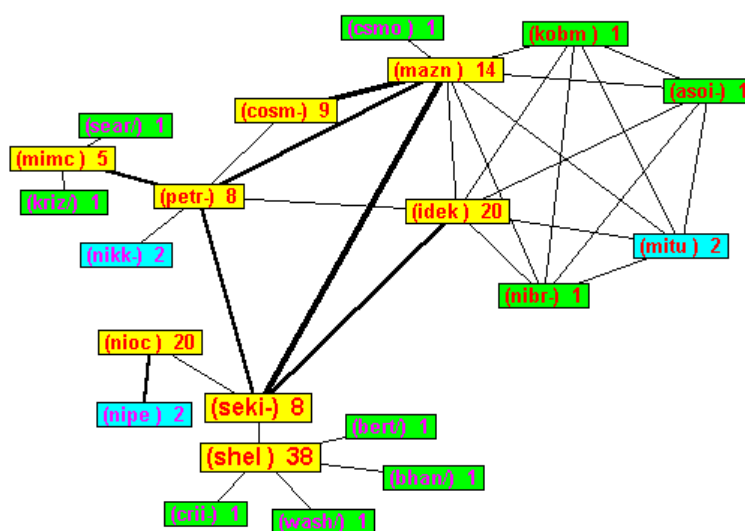


Figura 8 – Grafo ou rede de relacionamentos entre empresas depositantes de patentes sobre catalisadores

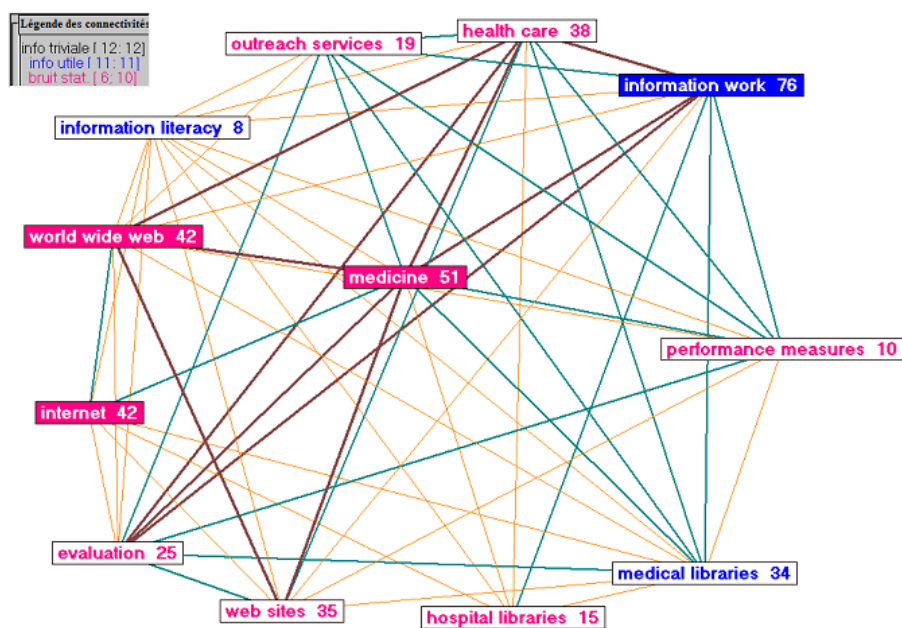


Figura 9 – Rede de relacionamentos de descritores dos registros de uma base de dados

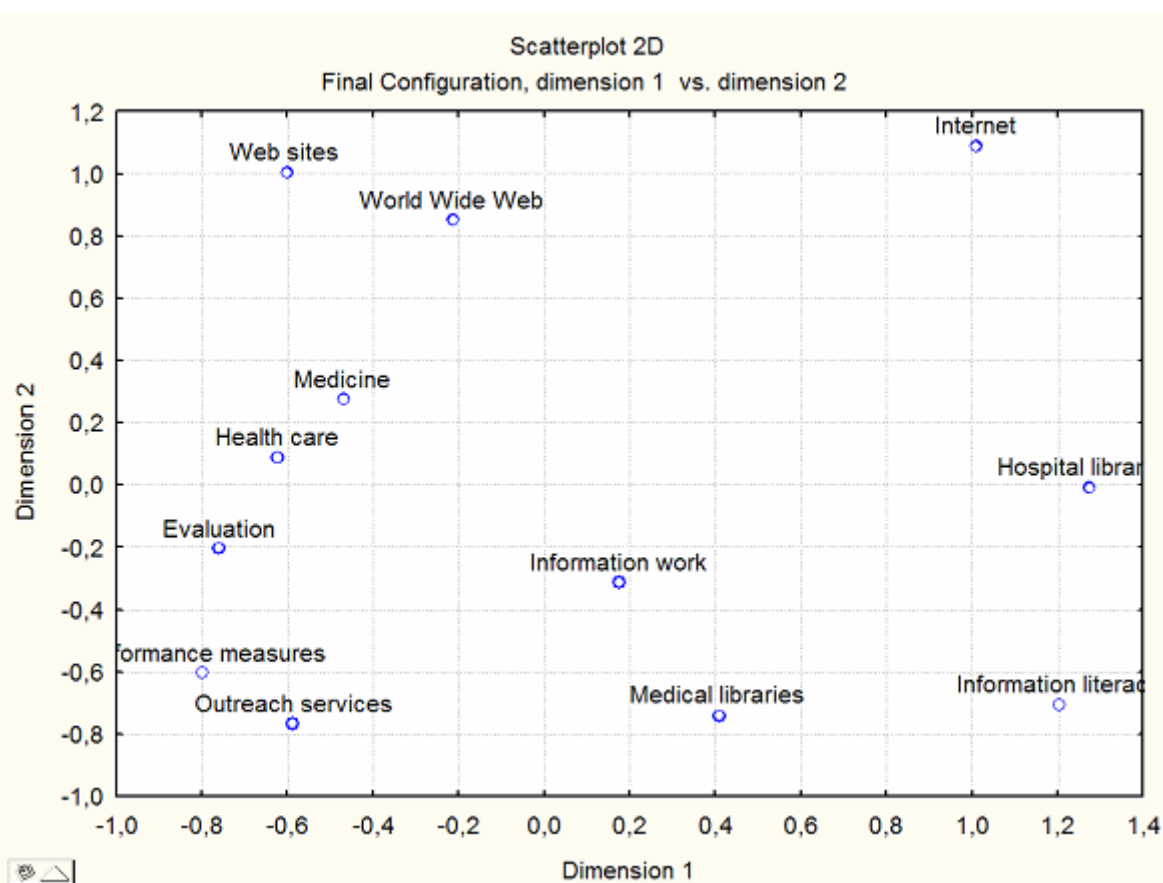


Figura 10 – Rede de relacionamentos dos descritores de registros de uma base de dados

Considerações finais

Apresentamos, neste texto, breve histórico dos estudos quantitativos da informação para, em seguida, abordar as tendências dos estudos contemporâneos sobre o tema. Esse percurso permite tirar algumas conclusões:

1. O campo dos estudos métricos, embora tenha origens remotas, se consolidou a partir da década de 50.
2. Os objetos de estudo passaram, gradativamente, da contagem de livros e revistas, para fins de gestão, para a interpretação da atividade científica e para orientar políticas de ciência.
3. Novos métodos de abordagem e tecnologias foram sendo criados para correlacionar dados, como estudos de citação, co-autorias e *co-words*.
4. A visualização de informações ganha, desde os anos 90, lugar de destaque.
5. Os estudos métricos se aproximam, cada vez mais, das ciências ditas “moles” (CHS) porque estas últimas oferecem teorias e modelos que permitem interpretar os dados em contextos culturais, políticos, ideológicos e econômicos distintos.
6. Ao lado dos estudos de natureza externalista, vêm crescendo os estudos internalistas. Significa, portanto, que aos estudos de quantificação de autores, de artigos e de citações, agregam-se estudos sobre os conteúdos dos trabalhos científicos.

É crescente, portanto, a consciência de que os estudos da ciência não podem se orientar por critérios meramente quantitativos. A incorporação da sociologia e da história da ciência inaugura a aproximação entre estudos quantitativos e estudos qualitativos. Atualmente, os trabalhos de mapeamento cognitivo da ciência se colocam como tendência importante para aprofundar a compreensão sobre a dinâmica da ciência. Como o texto científico se expressa pela linguagem, parece ser imprescindível incorporar, ao campo dos estudos métricos, as teorias e métodos da área da organização da informação e do conhecimento para construir os *corpora* de análise. Nesses estudos, os mapas de *co-words*, para serem confiáveis, devem resultar de compreensão rigorosa dos processos semânticos e pragmáticos da representação de informação, abandonando os procedimentos ingênuos de contagem da frequência de palavras do texto.

Para finalizar, destaca-se um fato auspicioso. O Brasil tem seguido a tendência internacional de crescimento dos estudos quantitativos da ciência. No período 1990-2006, segundo Meneghini e Packer (2008), as publicações em cientometria e bibliometria, de autores brasileiros, cresceram exponencialmente. Da mesma forma, a quantidade de estudos em nível de pós-graduação (doutorado, mestrado), as comunicações em congresso e demais formas de manifestação e expressão da produção científica demonstram o vivo interesse da comunidade científica, dos mais

diferentes campos de pesquisa, em utilizar os procedimentos metodológicos e as técnicas próprias dos estudos métricos da informação. É importante, porém, realizar estudos empíricos para identificar o grau de participação e a natureza dos trabalhos dos pesquisadores do campo da Ciência da Informação, nesse conjunto.

REFERÊNCIAS

- BOUSTANY, Joumana. **La production des imprimés non-périodiques au Liban de 1733 à 1920: étude bibliométrique**. 1997. Tese (Doutorado em *Sciences de l'Information et de la Communication*) – Université Michel de Montaigne – Bordeaux III, Bordeaux. 1997.
- CHEN, Chaomei. Searching for intellectual turning points: progressive knowledge domain visualization. **Proceedings of the National Academy of Sciences**, [S. l.], n. 101 Suppl., p. 5303-5310, 2004.
- EGGE; Leo; ROUSSEAU, Ronald. **Introduction to informetrics**: quantitative methods in Library, Documentation and Information Science. Amsterdam: Elsevier Science, 1990.
- FACHIN, Gleisy R. B; SANTOS, Raimundo Nonato Macedo. **Scientific communication and ontology**: bibliometric analysis of the Library and Information Science Abstracts. In: INTERNATIONAL CONFERENCE ON SCIENTOMETRICS AND INFORMETRICS 12., 2009, Rio de Janeiro. 2009. (No prelo).
- GARFIELD, E. In tribute to Derek John de Solla Price: a citation analysis of Little Science, Big Science. **Scientometrics**, [S. l.], v. 7, n. 3-6, p. 487-503, 1985.
- HERRERO-SOLANA, Victor. **A representação gráfica da informação como meio de acesso**. Marília: UNESP / *campus* de Marília, 2005. 56 eslaides: color. Eslaides gerados a partir do *software* PowerPoint.
- HULME, E. Wyndham. **Statistical bibliography in relation to the growth of modern civilization**. London: Grafton, 1923. p. 44.
- JOHNSON, Brian; SHNEIDERMAN, B. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. CONFERENCE ON VISUALIZATION, 2., 1991, San Diego. **Proceedings...** San Diego: [s. n.], 1991.
- LE COADIC, Yves-François. **A Ciência da Informação**. 2. ed. Brasília – DF: Briquet de Lemos, 2004a.
- _____. Princípios científicos que direcionam a ciência e a tecnologia da informação digital. **Transinformação**, Campinas, v.16, n. 3, p. 205-213, set. / dez. 2004b.

MARTELETO, Regina Maria. Análise de redes sociais: aplicação nos estudos de transferência da informação. **Ciência da Informação**, Brasília – DF, v. 30, n. 1, p. 71-81, jan. / abr. 2001.

MEADOWS, A. J. Theory in information science. **Journal of Information Science**, [S. l.], v. 16, n. 1, p. 59-63, 1990.

MENEZHINI, Rogério; PACKER, Abel. **A amplitude da autoria multidisciplinar de artigos sobre cienciometria no Brasil**. São Paulo: BIREME OPAS–OMS, 2008.

MOMM, Christiane Fabíola. **O conhecimento científico em turismo no Brasil**: cursos de pós-graduação (*stricto sensu*); período de 2000 a 2006. 2009. 128 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Santa Catarina, Florianópolis. 2009.

NASCIMENTO, Hugo A. D.; FERREIRA, Cristiane, B. R. Visualização de informações: uma abordagem prática. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 25., 2005, São Leopoldo. **Anais...** São Leopoldo: UNISINOS, 2005.

POLANCO, Xavier. **Aux sources de la scientométrie: bibliometrie, scientometrie, infometrie**, [S. l.], n. 2, 1995. Disponível em: <<http://www.info.unicaen.fr/bnum/jelec/Solaris/d02/2polanco1.html>>. Acesso em: 16 jan. 2003.

PRICE, D. J. de Solla. The structures of publication in science and technology. In: GRUBER, H.; MARQUIS, D. G. (Org.). **Factors in the transfer of technology**. Cambridge, Mass.: MIT Press, 1969. p. 91-104.

PRITCHARD, Alan. Statistical bibliography or bibliometrics. **Journal of Documentation**, [S. l.], v. 25, n. 4, p.348-349, 1969.

ROSTAIN, Hervé. **La bibliométrie et ses techniques**. Toulouse: Sciences de la Société, 1996. 131 p.

SANTOS, Raimundo N. Macedo. **Grafo de relacionamento entre descritores**: referências bibliográficas da LISA (busca "DE=CONSUMER HEALTH INFORMATION"). Florianópolis: UFSC, 2009. 6 eslaides: color. Esclaides gerados a partir de *softwares* gráficos.

_____. **Propriedade industrial**. Natal. UFRN, 1998. 13 eslaides: color. Esclaides gerados a partir do *software* PowerPoint.

SMALL H., GARFIELD E. The geography of science: disciplinary and national mappings. **Journal of Information Science**, [S. l.], v. 11, n. 4, p. 147-159, 1986.

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

BIBLIOGRAFIA

CALLON, M.; COURTIAL, J.P.; PENAN, H. **La scientométrie**. Paris: Presses Universitaires de France, 1993.

COURTIAL, J. P. **Introduction à la scientométrie**: de la bibliométrie à la veille technologique. Paris: Anthropos-Economica, 1990.