
Redes de produção científica *em Knowledge Organization: visualizações e análises possíveis relativas à autoria e citações*

*Redes de producción científica sobre organización del conocimiento:
visualizaciones y análisis alternativos de la autoría y las citas*

Scientific production networks in Knowledge Organization: possible author and citation visualization and analysis

Elaine de Oliveira LUCAS (1), Marilda Lopes Ginez de LARA (2)

1) Universidade do Estado de Santa Catarina (UDESC), Florianópolis-SC (Brasil), lani@udesc.br

(2) Universidade de São Paulo (USP), São Paulo-SP (Brasil), larama@usp.br

Resumen

Se analiza la producción científica sobre sistemas de organización del conocimiento entre 1997 y 2011 en español, francés y portugués, recogida en las bases de datos Library Information Science Abstract (LISA), Library, Information Science & Technology Abstracts (LISTA), Scopus y Web of Science. Para la organización de los datos y la identificación de las redes de agentes se utilizaron métodos bibliométricos, en particular, la Metodología de Análisis de Redes Sociales (ARS). Se presentan los resultados obtenidos sobre la distribución de artículos por ámbitos lingüísticos, publicaciones periódicas, co-palabras, grupos temáticos; y sobre la identificación de redes de autores, citas y grupos temáticos. También fue posible medir el grado de centralidad, modularidad, cubo y vector propio de los autores de la red; e identificar la élite de investigadores y su producción correspondiente.

Palabras clave: Organización del conocimiento. Representación del conocimiento. Lenguajes documentales. Sistemas de organización del conocimiento. Redes científicas. Análisis bibliométrico. Análisis de redes sociales.

1. Introdução

Conhecer a produção científica sobre *Knowledge Organization* (KO), que corresponde ao termo brasileiro Organização e Representação do Conhecimento (ORC) (Fujita, 2008), no período 1997-2011, identificar as relações existentes entre os pesquisadores (autores e coautores) e suas citações (autores citados), equivale a reunir dados para auxiliar o mapeamento do Campo Científico relativo ao domínio investigado. Nesta pesquisa, pretendemos contribuir com a discussão do tema investigando a literatura de KO sobre Sistemas de Organização do Conhecimento, incluindo Linguagens Documentárias e temas relacionados em língua francesa, portuguesa e espanhola publicada no período e reportada em bases de dados internacionais.

Abstract

The scientific production on knowledge organization systems in Spanish, French and Portuguese between 1997 and 2011, obtained from the reference databases Library Information Science Abstract (LISA), Library, Information Science & Technology Abstracts (LISTA), Scopus and Web of Science is analyzed. For the organization of data and identification of the scientific networks, we used bibliometric methods, notably, Social Network Analysis (SNA). The results allowed us to gather information on the distribution of articles by language, journals, co-occurrence of keywords, thematic groups, and the identification of networks among authors, citations, and thematic groups. For the authorship analysis, we used the law of elitism, which allowed us to identify a research elite and a group of corresponding search.

Keywords: Knowledge organization. Knowledge representation. Indexing languages. Knowledge organization systems. Scientific networks. Bibliometric analysis. Social network analysis.

A identificação das relações entre pesquisadores auxilia a construção de um mapa de tendências formalmente identificadas no escopo da KO (ou que poderiam integrá-la) focalizando, além da rede de relações entre os pesquisadores, as tendências de pesquisa.

Considerando-se a formação francesa, espanhola e brasileira de KO como continuidade à Documentação, é de se esperar que muitas das pesquisas contemplem, em alguma medida, formas de ver o campo que integrem referenciais clássicos da Documentação, ou que se fundamentem em alguns dos aspectos teórico-metodológicos de natureza linguístico-comunicacional.

A proposta explorou o tema abordado dentro da KO utilizando a metodologia da Análise de Re-

des Sociais (ARS) e da Bibliometria, de forma paralela de modo a identificar os agrupamentos e relações entre pesquisadores, incluindo sua bibliografia comum, como subsídio para posteriores pesquisas sobre o compartilhamento de teorias e métodos.

A importância de pesquisar sobre o desenvolvimento da KO nas línguas selecionadas reside, entre outros, no fato de que a França, a Espanha e o Brasil são países que compartilham, em grande medida, referenciais semelhantes oriundos da Documentação. Os contatos entre os três países, bem como entre eles e alguns países latino-americanos (Argentina, Colômbia, Uruguai, em especial) e o México, foram favorecidos pela língua. França, Espanha e Brasil contam, hoje, com um capítulo da *Internacional Society for Knowledge Organization* (ISKO), cujos eventos têm, frequentemente, atraído pesquisadores das línguas citadas. Assim, contribuem significativamente para o desenvolvimento dos estudos da KO, porém o alcance de contribuições fora da língua inglesa, é pouco visível e pouco estudada. Acreditamos, também, que a produção nestas três línguas influencia fortemente pesquisadores de outras comunidades francófonas, hispanófonas ou lusófonas, fora da delimitação geográfica destes três países.

O objetivo geral foi o de identificar os agentes presentes na produção científica sobre Sistemas de Organização do Conhecimento, incluindo Linguagens Documentárias e termos relacionados no campo da KO com base nas metodologias de Análise de Redes Sociais e Bibliometria. Na busca deste objetivo, alguns objetivos específicos foram formalmente identificados como necessários:

- Identificar os artigos sobre os temas citados na literatura de KO no período de 1997 a 2011, considerando-o como o *corpus* da pesquisa.
- Identificar os autores mais produtivos e os considerados 'autores de elite': autores com *alta produtividade* e identificados como a raiz quadrada da população dos autores produtores de documentos publicados.
- Identificar os autores mais citados e aqueles que fazem parte da 'frente de pesquisa': autores com *alta citação* e identificados como a raiz quadrada da população citada
- Identificar as centralidades de Entrada – medida do número de ligações que um ator *recebe* de outros atores que denota popularidade ou receptividade –, de Saída – medida do número de ligações que um ator *estabe-*

lece com outros atores, desta rede que denota expansividade –, e de Bonacich dos atores de 'elite' e 'frente' de pesquisa, formalmente identificados – medida que leva em consideração o número de ligações diretas que um ator possui, mas também a centralidade dos atores vizinhos, que denota um aspecto de *prestígio* mais global.

A pesquisa aqui apresentada trata de uma análise parcial levando-se em consideração todas as delimitações introdutórias, bem como, as opções metodológicas. Neste caso, ressaltamos que não abrange todo o domínio da KO e sim uma parte dele, restrita aos Sistemas de Organização do Conhecimento (SOC), e ainda com restrições em relação ao idioma, período abordado e até mesmo opção das bases de dados sobre as quais foram efetuados os levantamentos que originaram o *corpus* da pesquisa. Neste caso, os resultados obtidos irão refletir uma análise parcial do domínio e não uma análise completa do mesmo.

2. Knowledge Organization (KO)

Diferentes autores se reportam ao bibliotecário Henry Evelyn Bliss como primeira referência ao uso do termo 'Organização do Conhecimento', utilizado em dois livros: o primeiro, publicado em 1929, *The Organization of Knowledge and the System of the Sciences*; e o segundo, em 1933 *The Organization of Knowledge in Libraries* (Almeida, 2013; Ferreira e Guimarães, 2013; Dahlberg, 2006).

No entanto, o termo só obteve "destaque a partir de 1971, quando o alemão Dagobert Soergel o divulgou em uma conferência da área" (Ferreira e Guimarães, 2013, p.190).

Após esse período, aconteceram alguns eventos fundamentais: a tese de Dahlberg, em 1973, sob o título *Foundations of Universal Organization of Knowledge*; a publicação da revista *International Classification*, em 1974; e a criação, em 1989, da *International Society for Knowledge Organization* (Almeida, 2013). Para Dahlberg (2006), além do pioneirismo no uso do termo 'Organização do Conhecimento' feito por Bliss, o mesmo foi considerado o mais adequado para ser utilizado na criação da *International Society for Knowledge Organization* (ISKO), em 22 de julho de 1989.

Após a fundação da ISKO, o periódico científico da área o *International Classification* (1974-1992) altera seu título para *Knowledge Organization*, no ano de 1993. Gomes (2009, p.61) comenta que "Dahlberg, fundadora daquele título, apresentou longo artigo justificando a

alteração: focalizando seu argumento na teoria do conceito e na classificação”.

De fato, a própria Dahlberg (2006, p. 12, tradução livre), em artigo posterior, cita que a revista sob o novo título introduziu a ‘Organização do Conhecimento’ como compreendendo

[...] os objetos e as atividades da teoria do conceito, classificação e indexação e representação do conhecimento, onde por ‘representação do conhecimento’ Entende-se não somente a estrutura lógica da representação conceitual, mas também todas as questões relativas a nomear conceitos pelos termos mais adequados segundo o qual as questões de terminologia têm que ser consideradas também.

Dahlberg vê “a origem da Organização do Conhecimento (OC) nas preocupações de Otlet e La Fontaine, no início do século XX, às quais foram se juntando documentalistas e cientistas da informação e, desde a obra de Wüster, os terminólogos se envolveram também com o assunto” (Dahlberg, 1993, citada por Gomes, 2009, p. 61). Estabelece, assim, uma profunda relação entre a KO e a Documentação.

Hjørland (2003) afirma que é difícil esboçar o progresso teórico e científico na área de Organização do Conhecimento, por ela ser demasiadamente pragmática, muito fragmentada e com a coexistência de diferentes linhas de pensamento. O autor considera prioritário para a área explorações sobre os métodos e as bases metodológicas de Organização do Conhecimento. A falta de um mapeamento neste sentido acaba por limitar a área aos avanços relacionados ao progresso das tecnologias da informação e da comunicação (TICs), mais do que pelo desenvolvimento de sua própria pesquisa teórica (Hjørland, 2003, citado por Fujita, 2008; Ferreira e Guimarães, 2013).

Para Fujita (2008, p. 5), “por um lado, falta à área de Organização do Conhecimento uma consolidação científica, de outro, é inegável sua diversidade conceitual e o impacto de seus resultados para a organização do conhecimento de outras áreas científicas”. Nesse sentido, é fundamental ao Campo Científico conhecer seu domínio em Organização do Conhecimento.

Entre os diversos estudos anteriores que mapearam o campo da KO sob a ótica dos Estudos Métricos da Informação, podemos citar os de López-Huertas e Contreras (2004); Café e Bräschner (2008); Moneda-Corrochano, López-Huertas e Contreras (2011); Alves, Oliveira e Grácio (2012); e Souza (2013).

Dentre os inúmeros subtemas do domínio da KO temos os Sistemas de Organização do Co-

nhecimento que é o termo resultante da tradução literal de *Knowledge Organization System (KOS)*, termo adotado pela *International Society of Knowledge Organization (ISKO)* que engloba diferentes instrumentos que tratam do ‘conteúdo’ de documentos, entre eles, as Linguagens Documentárias (Lara, 2013).

Segundo Lara (2013), o termo *Knowledge Organization System* foi inicialmente utilizado por Hodge, em 2000, para abranger todos os tipos de esquemas de organização da informação e representação do conhecimento.

3. Metodologia

Em virtude do cenário apresentado anteriormente, definimos como objeto desta pesquisa os agentes presentes na produção científica que aborda questões relativas aos Sistemas de Organização do Conhecimento, além de conceitos relacionados, em estudos no campo da KO.

A abordagem adotada é de natureza exploratório-descritiva e delineada por um estudo de campo. A pesquisa é de natureza qualitativa no que concerne ao tipo de análise e usará diferentes métodos de acordo com as fases da pesquisa.

De acordo com as premissas, apresentadas na introdução, e que embasam a pesquisa, apresentaremos as opções e procedimentos metodológicos em três momentos distintos: levantamento nas bases de dados, gerenciamento de dados, e análise dos dados. Cada um desses momentos é composto de diferentes etapas e serão abordados e detalhados a seguir.

3.1. Levantamento nas bases de dados

O levantamento da produção científica a ser analisada foi feito entre novembro e dezembro de 2012 e compreendeu a literatura repertoriada pelas bases de dados multidisciplinares *Web of Science (WoS)* - disponibilizada pela Thomson Scientific - e *Scopus* - publicada pela Elsevier - ; e as especializadas em Ciência da Informação *Library, Information Science & Technology Abstracts (LISTA)* - disponibilizada pela plataforma EBSCOhost - e *Library Information Science Abstract (LISA)* - editada e disponibilizada pela ProQuest -.

A mesma equação de busca foi utilizada nas quatro bases de dados, mantendo-se as palavras-chave em inglês utilizadas por cada uma das bases de dados, mesmo sendo outros os idiomas originais dos artigos. O *corpus* da pesquisa foi definido a partir da equação de busca abaixo:

("Bibliographic classification" OR Categories OR Categorization OR Classification OR "Classification of Knowledge" OR "Classification scheme" OR "Classification system" OR "Controlled language" OR "Controlled vocabulary" OR "Document analysis" OR "Document organization" OR "Documentary language" OR "Facet analysis" OR "Facet classification" OR "Facet classification schemes" OR "Index languages" OR Indexing OR "Indexing language" OR "Information organization" OR "Knowledge Organization" OR "Knowledge Representation" OR "Knowledge Organization Systems" OR KOS OR "Markup language" OR "Natural languages" OR Ontologies OR Ontology OR "Retrieval Languages" OR "Semantic vocabularies" OR "Simple Knowledge Organization System" OR SKOS OR "Structured vocabulary" OR "Subject analysis" OR "Subject approach to information" OR "Subject cataloguing" OR "Subject heading" OR "Subject Headings Schemes" OR "Subject Indexes" OR "Subject Indexing" OR "Synonym rings" OR Taxonomies OR Terminologies OR Thesauri OR Thesaurus OR "Topic Maps" OR Vocabularies OR "Vocabulary control" OR "web semantics") AND (Language OR "Language theories" OR Linguistics OR "Linguistics theories" OR Semantics OR Terminology OR Lexicology)

Os campos de busca utilizados para a recuperação pelos termos acima descritos foram os de 'palavras-chave', 'título' e de 'resumo' como limites comuns nas bases de dados. Para que a formação do *corpus* desta pesquisa privilegiasse a análise de algumas premissas pré-estabelecidas, foram aplicadas algumas condições de filtragem, como segue: condição Idiomática – o levantamento considerou a produção científica nos idiomas espanhol, francês e português –; condição cronológica – o período abarcado para o levantamento considerou as obras publicadas entre 1997 e 2011 (15 anos) –; e a condição do tipo de documento – foram analisadas somente as produções do tipo 'artigo de periódico' e dentre estes, somente os que forneciam acesso ao texto completo –.

3.2. Gerenciamento de dados

Feita a busca dos dados em cada uma das quatro bases de dados elencadas anteriormente, e respeitando as peculiaridades de busca de cada interface, obtivemos como resultado 536 referências bibliográficas, conforme descrição na tabela abaixo.

Devido às diversas origens dos dados (SCOPUS, WoS, LISA e LISTA) estes não estavam preparados para que a mineração fosse aplicada diretamente. Foram realizadas algumas ações necessárias no processo de limpeza dos dados que envolveram filtragem, combinações

e/ou agrupamentos e preenchimento de itens vazios.

Base de dados	Referências
LISA	241
LISTA	184
WoS	40
SCOPUS	71
Total	536

Tabela 1. Resultado do levantamento bibliográfico que comporia o *corpus* de análise

Com a recuperação dos artigos, apresentada na Tabela 1, foram feitos ajustes conforme segue: verificação de duplicatas nas base de dados – foram retiradas as duplicatas da mesma obra em uma mesma base de dados; verificação de tipo de publicação – mesmo com a recuperação delimitada nas interfaces das bases de dados para 'artigos de periódicos', foram identificadas e retiradas as publicações no formato de resenhas, resumos de teses e/ou dissertações, editoriais e artigos de anais de eventos; e verificação de pertinência – foram retirados os artigos destoantes da pertinência requerida pela pesquisa. Para este aspecto de verificação foram analisados os títulos e *abstracts* e retirados os artigos que tratem da 'linguagem' em uma abordagem que não a da *Knowledge Organization*.

Com os ajustes acima descritos, foram retiradas 73 referências bibliográficas do *corpus inicial*, resultando em 463 referências bibliográficas que passaram a compor o *corpus* definitivo.

BD	Referências
LISA	224
LISTA	179
WoS	33
SCOPUS	27
Total	463

Tabela 2. *Corpus* definitivo

É importante salientar que as duplicatas em diferentes bases de dados foram mantidas para posterior análise de visibilidade e relação entre elas. As 463 referências bibliográficas foram agrupadas em duplicatas recuperadas em diferentes bases de dados, resultando no *corpus* definitivo de 354 *artigos originais* a serem analisados.

3.2.1. Análise dos dados

As metodologias de Análise de Redes Sociais (ARS) e a Bibliometria auxiliaram na organização dos dados e identificação das relações en-

tre os artigos que compunham o *corpus* analisado, com intuito de responder aos questionamentos levantados nos objetivos específicos. A utilização de *softwares* é considerada essencial para esta fase da pesquisa. Nas análises bibliométricas foram utilizados os indicadores de autoria (coautoria, autores mais produtivos, elite de pesquisa) e da análise de citações da elite de pesquisa (autores mais citados, frente de pesquisa).

Em relação a descrição da elite de pesquisa no campo do presente estudo, nos baseamos em Price (1976, p. 30), para quem “o número de produtores prolíficos parece equivaler à raiz quadrada do número total de autores”. Neste caso, a elite da pesquisa apresentada foi definida com base na raiz quadrada ($\sqrt{\quad}$) da quantidade de autores responsáveis pelos artigos que compunham o *corpus* desta pesquisa. Da mesma forma, a frente de pesquisa também é uma variável dependente e, para sua definição, procedemos da mesma forma que para a elite de pesquisa. Porém, tivemos como referência todos os autores identificados mediante a contagem de citações nos artigos da elite de pesquisa anteriormente definida.

Na aplicação da metodologia da ARS utilizamos os grafos desenvolvidos e seus atributos para análise dos indicadores de autoria – colaboração científica (relação nas autorias múltiplas) – e da análise de citação da elite de pesquisa.

<i>Indicadores métricos</i>	<i>Significado</i>
Centralidade de grau	Conta o número de arestas incidentes em um Nó do grafo.
Hubs	Relação de intermediação (Broker). Nó que articula comunidades.
Autoridade	Autoridade do Nó naquele grafo (soma de todos os Hubs).
Modularidade	Possibilidade de comunidades (colégios invisíveis).
Eigenvector ou Bonacich	Aumenta de acordo com o alto grau de conectividade dos outros nós com que está conectado.
Grau de entrada	Quantas arestas entram nos nós.
Grau de saída	Quantas arestas saem dos nós.

Quadro 1. Indicadores métricos pertinentes na visualização dos grafos

Dentro da metodologia de ARS os sistemas de visualização dos atores e suas relações, e os fenômenos decorrentes destas relações a partir de atributos estabelecidos pela própria metodologia de ARS são os pontos chave desta perspectiva. O *Gephi*® foi utilizado como sistema de visualização para todos os grafos apresentados, bem como para a determinação dos atributos extraídos das relações estudadas neste item,

como exposto no quadro 1. Com base nos atributos apresentados no quadro acima, definimos quais seriam utilizados para análises construídas a partir do olhar bourdieusiano adotado. Foram elaboradas planilhas de análise para os indicadores métricos na visualização dos grafos de acordo com sua categoria relacional. Isso se faz necessário pois (Martins, 2012, p. 144):

Entendemos que as escolhas de como realizar as etapas envolvidas na análise não são apenas escolhas técnicas, mas modos de tratamento da informação que influenciam em nossa capacidade de olhar e identificar padrões que fornecem pistas das tendências e fenômenos sociais daquilo que estudamos.

A seguir, estão apresentadas as planilhas utilizadas de acordo com os indicadores de cada grafo relacional.

A *colaboração científica*, representada neste estudo utilizando a metodologia de ARS por meio de um grafo que apresenta as relações de coautoria identificadas no *corpus*, foi explorada quanto aos seus indicadores métricos do grafo, conforme demonstra o quadro, abaixo.

<i>Indicadores métricos</i>	<i>Significado</i>
Centralidade de grau	Conta o número de arestas incidentes em um Nó do grafo.
Hubs	Relação de intermediação (Broker). Nó que articula comunidades.
Autoridade	Autoridade do Nó naquele grafo (soma de todos os Hubs).
Modularidade	Possibilidade de comunidades (colégios invisíveis).
Eigenvector ou Bonacich	Aumenta de acordo com o alto grau de conectividade dos outros nós com que está conectado.

Quadro 2. Indicadores métricos pertinentes na visualização do grafo – Colaboração científica

<i>Indicadores métricos</i>	<i>Significado</i>
Modularidade	Possibilidade de comunidades (colégios invisíveis)
Grau de saída	Quantas arestas saem dos nós (Elite de pesquisa)
Grau de entrada	Quantas arestas entram nos nós (Frente de Pesquisa)

Quadro 3. Indicadores métricos pertinentes na visualização do grafo - Análise de citação da elite de pesquisa

Os indicadores métricos utilizados na análise do grafo desenvolvido para *análise de citação da elite de pesquisa* do *corpus* desta pesquisa são apresentados no Quadro 3. Para este grafo, onde as arestas possuem direção e sentido (direcionadas) foi possível analisar os indicadores de entrada e de saída.

Por fim, salientamos que para a análise do *corpus* foram utilizados alguns *softwares*, de acordo com as diferentes fases da pesquisa: o *Microsoft office Excel 2013*® foi utilizado para separar e minerar os dados; o *Software R*®, para desenvolver os pares relacionais (matrizes) necessários para o desenvolvimento dos grafos; e, em seguida, para o desenvolvimento dos grafos com atributos e métricas de acordo com as relações estabelecidas, o *software* livre *Gephi*® versão 0.8.2 – Beta.

4. Apresentação dos resultados

Os artigos recuperados de acordo com a metodologia descrita anteriormente resultaram em um *corpus* de 354 artigos, sendo 167 no idioma espanhol, 110 em francês e 77 em português. Para o ano de 1997, recuperamos 16 artigos e, em 2011, o resultado foi de 39 artigos, número que demonstra um grande aumento na literatura da área nos quinze anos correspondentes ao período coberto pelo estudo. O ano com maior quantidade de artigos dentro dos parâmetros estabelecidos, foi o de 2009, com um total de 44 artigos recuperados. A distribuição anual e idiomática dos artigos é apresentada na tabela 3.

Ano	Idiomas			Total
	Esp.	Fra.	Por.	
1997	4	12	-	16
1998	7	15	1	23
1999	14	4	1	19
2000	10	7	2	19
2001	5	2	1	8
2002	7	5	3	15
2003	5	7	1	13
2004	9	2	5	16
2005	7	4	4	15
2006	22	10	3	35
2007	16	15	11	42
2008	15	3	9	27
2009	27	7	10	44
2010	9	1	13	23
2011	10	16	13	39
Total Geral	167	110	77	354

Tabela 3. Distribuição anual dos artigos, no *corpus* analisado, por idioma, 1997-2011

4.1. Coautoria

Dos 354 artigos do *corpus* desta pesquisa, 204 artigos têm autoria única e 150 tem autoria em colaboração. O total de autores presentes no *corpus* são 462, dos quais 141 publicaram sozinho e 321 deles publicaram, em pelo menos

um dos artigos do conjunto analisado, em colaboração. É sobre estes artigos em colaboração que nos debruçaremos no grafo a seguir.

O Gráfico 2, na página seguinte, apresenta todas as relações de coautoria existentes no *corpus* analisado. Alguns agrupamentos em colaboração se destacam por terem uma grande quantidade de relações – representadas por arestas - e/ou por terem alta densidade relacional – grau da aresta. Cada um dos nós neste grafo representa um dos 321 autores que trabalharam em coautoria, sendo que os que apresentam uma quantidade maior de relações tem o nó em tamanho proporcional a este número de relações. É possível perceber 10 agrupamentos bastante significativos no grafo acima, com destaque para cada um dos autores que possui maior quantidade de relações. Para analisarmos este grafo recorremos, seguindo a metodologia de Análise de Redes Sociais, aos atributos disponibilizados pelo *software* Gephi®, a partir do qual foi possível identificar algumas métricas relacionais importantes sobre a autoria do *corpus* em análise.

Para análise das métricas da rede de colaboração científica no *corpus* foram verificados os atributos descritos na metodologia: centralidade de Grau; modularidade; autoridade; hub e Eigenvector (Bonacich). Estes atributos estão apresentados no Gráfico 1, disposto em ordem decrescente de centralidade de grau, onde podemos perceber que o fato de os autores terem uma elevada centralidade de grau na rede, não está relacionado numericamente (proporcionalmente) à modularidade e/ou ainda ao Eigenvector.

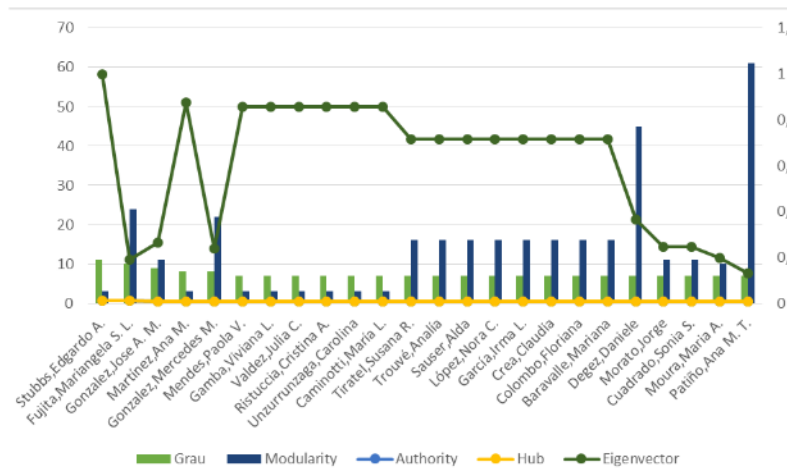


Gráfico 1. Atributos da rede de colaboração científica

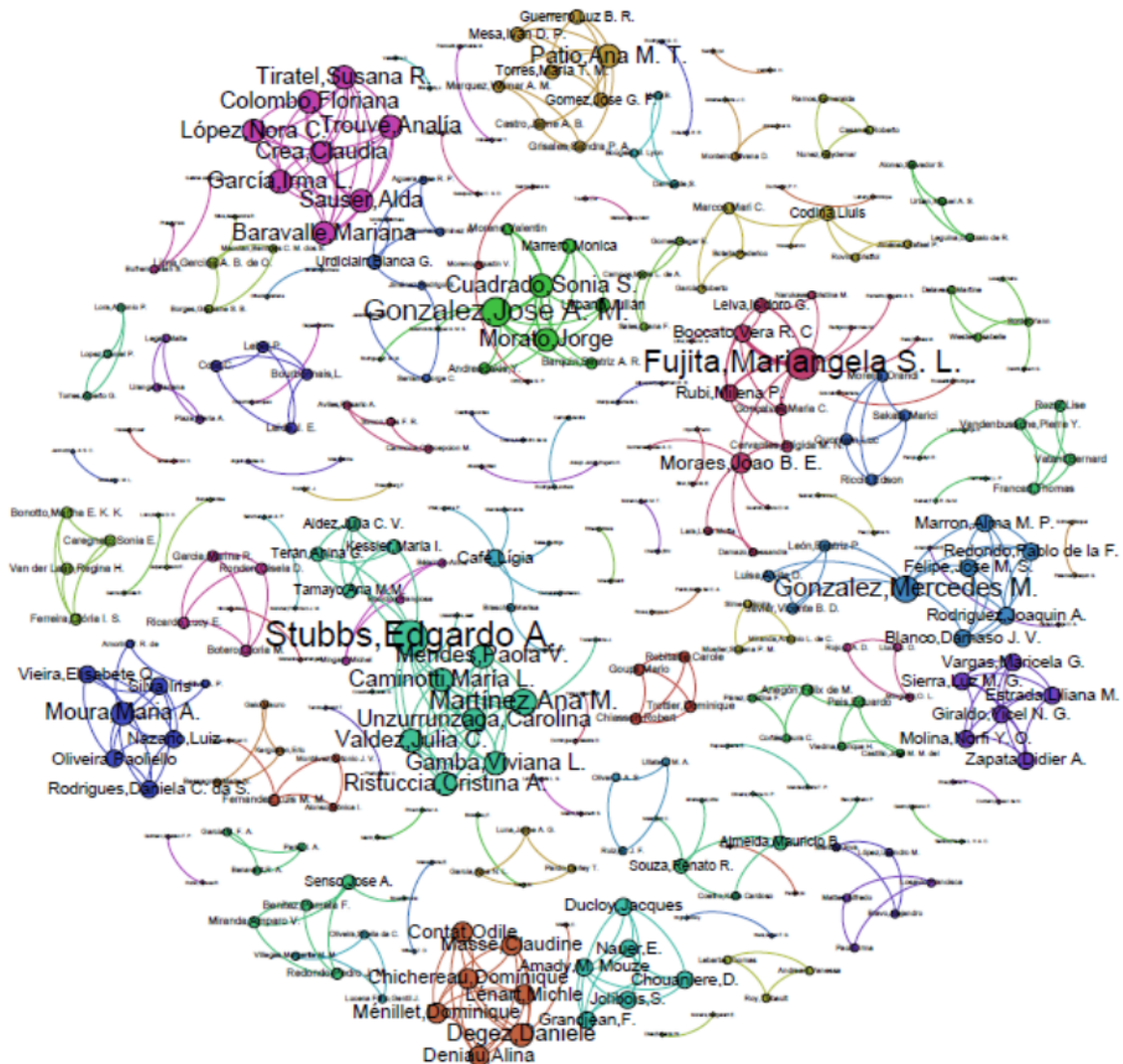


Gráfico 2. Grafo da rede de colaboração científica no corpus

Isso ocorre porque o indicador de Grau na rede demonstra a quantidade de arestas, ou seja, de conexões com coautores diferentes que cada autor da rede possui, enquanto o indicador de modularidade (modularity) demonstra a possibilidade de pertencer a comunidades.

Neste gráfico, o autor com maior centralidade de grau na rede, conforme apresentado, é 'Stubbs, Edgardo A.' seguido de 'Fujita, Mariangela S. L.'. Já quanto à modularidade, esta posição se inverte. É possível perceber ainda que há outros autores com centralidade de grau baixa e que possuem altos níveis de modularidade, casos de autores 'Degez, Daniele' e 'Patiño, Ana M. T.'

Com base nisso, podemos dizer que, em redes de coautoria como esta, a centralidade de grau representa o número de relações de um autor

com os outros autores pertencentes à rede, em relação à quantidade de relações. Neste caso, trata-se de coautoria de autores que publicaram juntos, enquanto no caso da modularidade, ainda que para autores com baixas quantidade de relações de coautoria, apresenta-se a capacidade que eles têm em formar comunidades – quer dizer, que têm uma posição importante na rede do ponto de vista de articulação entre pares.

Quanto ao Eigenvector, podemos afirmar que este tampouco indica ligação proporcional à modularidade e/ou à centralidade de grau, sendo completamente autônomo em relação aos outros atributos, uma vez que ele é proporcional ao grau de conectividade dos outros nós com que está conectado. A quantidade de relações perde foco se comparada à observação de 'com quem' o autor está ligado.

Para este gráfico, os valores de hub de todos os autores são muito baixos e não apresentam necessidade de análise. Isso porque o hub é um atributo mensurado pelas relações de intermediação que um nó é capaz de estabelecer em uma rede, caracterizando a capacidade deste nó em articular comunidades na rede. Entre os autores da nossa rede de colaboração não há nenhum que apresente capacidade de articulação em níveis de destaque. Por consequência, também não há valores significativos para o atributo de autoridade dos autores presentes na rede, já que a autoridade de um nó refere-se a soma de todos os hubs.

4.2. Elite e frente de pesquisa

Também consideramos a possibilidade de identificar os autores mais produtivos ou a 'elite de pesquisa' no domínio do presente estudo. Baseado em Price (1976, p. 30), para quem "o número de produtores prolíficos parece equivale à raiz quadrada do número total de autores", temos então $\sqrt{462} = 21,49$ – onde 462 é o número total de autores presentes no *corpus* do estudo e 21,49 o equivalente à quantidade de autores onde deve se inserir o corte que identifica a 'elite de pesquisa'. Por motivos de adequação, foi feito arredondamento no corte para 19 (corte em autores com 4 ou mais artigos).

Autores	Artigos
Fujita, Mariangela S. L.	13
Boccatto, Vera R. C	7
Café, Lígia	6
Gonzalez, José A. M.	6
Codina, Lluis	5
Martínez, Ana M.	5
Peña, Catalina N.	5
Almeida, Mauricio B.	4
Brun, Ricardo E.	4
Cuadrado, Sonia S.	4
Gonzalez, Mercedes M.	4
Lima, Gercina A. B. de O.	4
Maniez, Jacques	4
Moraes, Joao B. E.	4
Morato, Jorge	4
Moura, Maria A.	4
Quintana, Ania R. H.	4
Rubi, Milena P.	4
Urdiciain, Blanca G.	4
Total	95

Tabela 4. Elite de pesquisa e produtividade nos artigos recuperados (*corpus*), 1997-2011

Dos 462 autores envolvidos no estudo, 19 correspondem à elite de pesquisa, ou seja, 4,11% dos autores responde pela maior produção no domínio estudado com base nos parâmetros estabelecidos. Esses autores, distribuídos segundo sua produtividade, são apresentados na tabela 4, a seguir.

Temos assim que os autores de elite aparecem em 95 artigos e muitas vezes de forma colaborativa. Trata-se, portanto, de 95 vezes nas quais os nomes dos autores aparecem. Como a colaboração entre autores da elite da pesquisa foi identificada, em uma verificação posterior chegamos à quantidade de artigos produzidos pelos autores de elite: Eles produziram 75 artigos.

Com o objetivo de conhecer melhor a elite de pesquisa, procuramos identificar e analisar sua frente de pesquisa, indicador bibliométrico rico para análise. Tal indicador revela as redes que se estabelecem por meio de citações nos artigos que compõem o *corpus*, constituindo também uma forma confiável para estudar relações entre a produção de autores específicos que ficam obscurecidas por outros índices ou indicadores mais utilizados nos estudos bibliométricos.

A forma de mensuração usada foi a frequência de citações, seguindo as propostas de Price segundo a qual a raiz quadrada do número total de autores citados (\sqrt{N}) resultaria na quantidade dos autores que compõem a frente de pesquisa (Price, 1963).

Utilizando o mesmo método de Urbizagastegui-Alvarado (2007), foram atribuídos pontos a cada um dos autores citados, de forma que o número de pontos obtidos por um autor representará o número de vezes que este foi citado em todos os artigos analisados, independentemente da forma de autoria, como autor ou coautor. Na contagem, não foram consideradas as múltiplas citações de um mesmo autor e obra, com ocorrência no mesmo artigo.

O recorte do *corpus* utilizado para a análise da frente de pesquisa foi o total dos 75 artigos produzidos pelos 19 autores considerados como a elite de pesquisa. Esta elite utilizou para suas citações um total de 1.536 referências válidas, produzidas por 2.189 citados. Aplicando novamente a regra de Price (\sqrt{N}) temos $\sqrt{2.189} = 46,78$ – onde 2.189 é o número total de autores citados pela elite de pesquisa, e 46,7, o equivalente à quantidade de autores, onde deve se inserir o corte que identifica a 'frente de pesquisa'. Por motivos de adequação, foi feito arredondamento no corte para 47 (corte em autores com 4 ou mais citações). Essa frente de pesqui-

sa, distribuída segundo sua frequência, é apresentada na tabela 5, na página seguinte.

<i>Autores</i>	<i>No.</i>
Lancaster, F. W.	22
Fujita, Mariângela Spotti Lopes	14
Dahlberg, Ingetraut	11
Campos, Maria Luiza Almeida	10
Gruber, Thomas. R.	10
Hendler, J.	9
Vickery, Brian Campbell	9
Cabré, María Teresa	8
Foskett, Antony Charles	8
Gilchrist, Alan	8
Guimarães, J. A. C.	8
LARA, Marilda Lopes Ginez de	8
Lassila, Ora	8
Lee, Tim Berners	8
Slype, Georges Van	8
Urbician, Blanca Gil	8
Chaumier, Jaques	7
González, J. A. Moreiro	7
Guarino, Nicola	7
Hjorland, Birger	7
Leiva, Isidoro Gil	7
Marco, Francisco Javier García	7
Nardi, Maria Isabel Asperti	7
Aitchison, Jean	6
Fagundes, Silvana Aparecida	6
Guinchat, Claire	6
Noy, N. F.	6
Ranganathan, S. R.	6
Svenonius, Elaine	6
Boccatto, V. R. C.	5
Cintra, A.M.M.	5
Giaretta, P.	5
Mcguinness, Deborah L.	5
Menou, Michel	5
Peña, Catalina Naumis	5
Soergel, Dagobert	5
Vizcaya, Dolores	5
Sowa, J. F.	5
Café, Lígia	4
Dik, Simon	4
Floridi, L.	4
Gomes, Hagar Espanha	4
Herrera, Antonia Heredia	4
Ingwersen, Peter	4
Robredo, Jaime	4
Sager, Juan C.	4
Uschold, Mike	4

Tabela 5. Frente da pesquisa e frequência nos artigos produzidos pela elite de pesquisa (no. de citações recebidas)

Esta elite de pesquisa é autora – com colaborações entre si – de 75 artigos e utilizou, para suas citações, referências válidas produzidas por 2.189 citados, dos quais 47 correspondem aos mais citados, ou à frente de pesquisa. Um dos autores da elite de pesquisa, 'Brun, Ricardo E.', não utilizou em suas citações nenhuma referência válida, por isso, para análises relacionais, a elite foi reduzida para 18 autores. Para conhecer os laços entre a elite e a frente de pesquisa do *corpus* desenvolvemos grafos e gráficos que mostram indicadores desta rede social.

É importante perceber que podem existir casos em que um autor da elite de pesquisa também pertence à frente de pesquisa. Como forma de minimizar a confusão que essa coincidência poderia causar, os autores da elite estão indicados com seus sobrenomes expressos em letra minúscula, enquanto a frente de pesquisa tem os nomes dos autores com o sobrenome em maiúscula.

No Gráfico 3 (na página seguinte) a elite de pesquisa (mais produtivos) está representada pelos nós de cor vermelha, enquanto a frente de pesquisa (mais citados), pelos nós de cor azul.

Já o Gráfico 4 apresenta os laços relacionais entre a elite e a frente de pesquisa de forma a destacar as comunidades possíveis. Pode-se observar seis comunidades evidenciadas por meio das cores rosa, amarelo, azul escuro, vermelho, verde e azul claro. Estas comunidades representam os nós por proximidades de acoplamento. Os autores de elite que citam os mesmos autores da frente de pesquisa, estarão representados de forma mais próxima no Grafo e formam comunidades que podem representar colégios invisíveis identificados por estes acoplamentos.

Para ambos os grafos, buscamos proceder a análise com base em outros atributos métricos importantes na metodologia de ARS, como os graus de saída ou de entrada e a modularidade.

Os autores de elite com maior quantidade de laços com os autores da Frente de pesquisa apresentam um maior Grau de Saída. É com base neste atributo que o Gráfico 5, na página 43, está organizado, de forma decrescente.

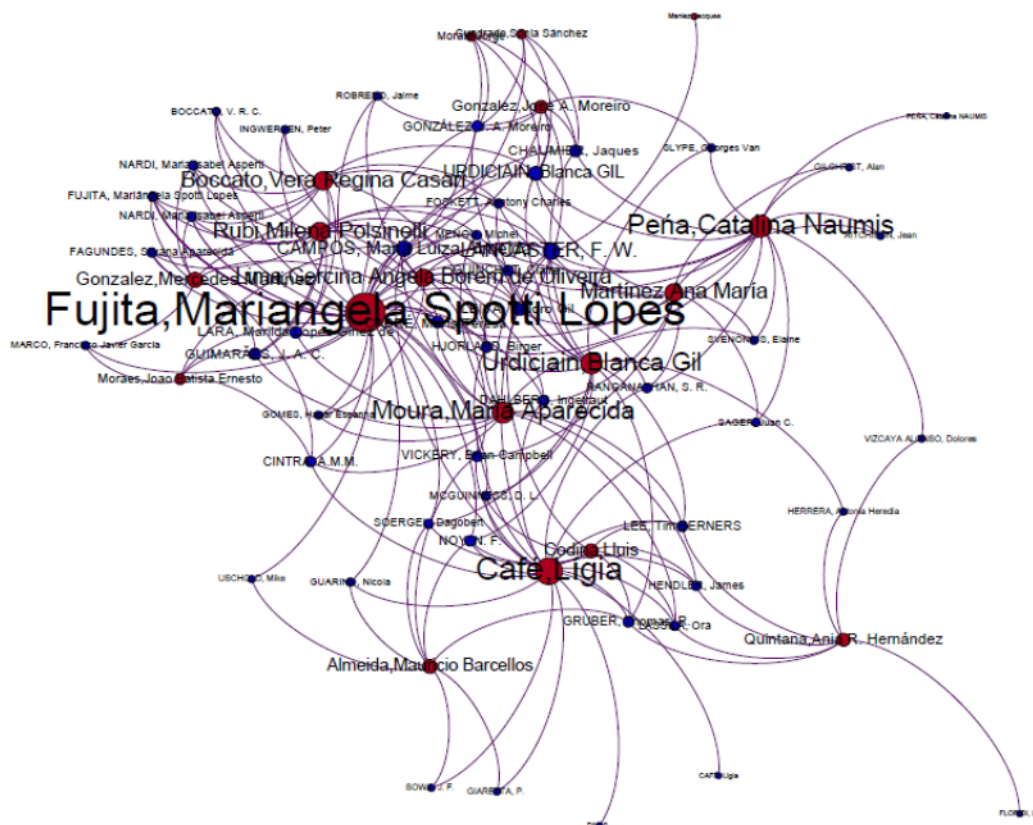


Gráfico 3. Rede de relações da elite e frente de pesquisa – por tipo do nó

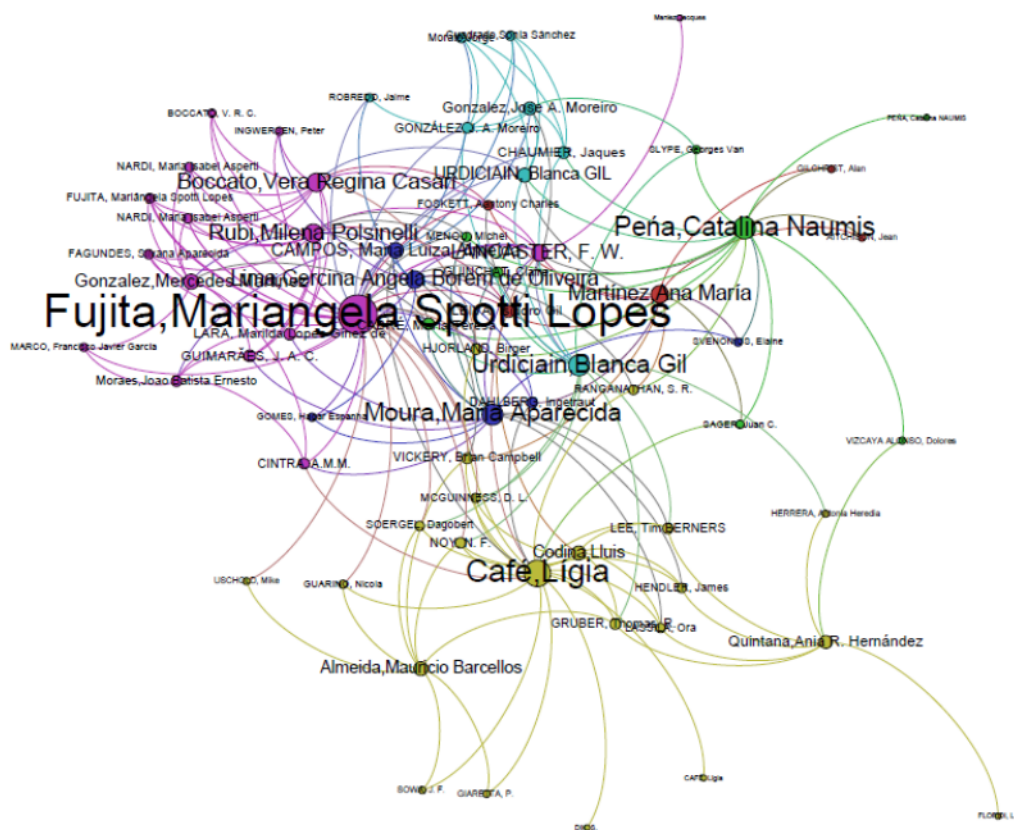


Gráfico 4. Rede de relações entre a elite e a frente de pesquisa – por comunidades

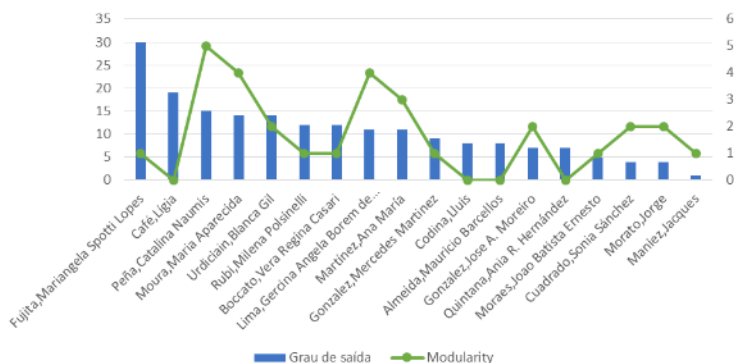


Gráfico 5. Atributos da rede - Elite de pesquisa

Podemos perceber que a modularidade dos autores da elite, também apresentada no Gráfico 5, não acompanha os valores do grau de saída de forma proporcional. Isso porque o grau de saída demonstra o número de citações efetuadas por cada um dos autores de elite, enquanto a modularidade apresenta a capacidade de gerar comunidades de cada um destes autores.

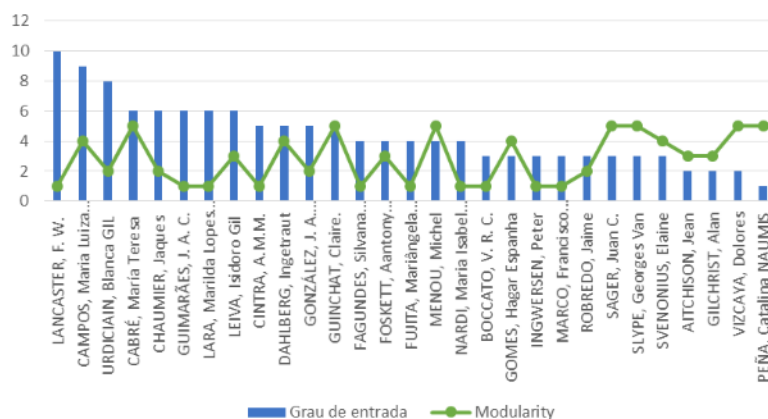


Gráfico 6. Atributos da rede - Frente de pesquisa

Para análise da frente de pesquisa, utilizamos o grau de entrada como atributo para organizar o Gráfico 6, abaixo. Dos 47 autores pertencentes à frente de pesquisa, apenas 29 foram apresentados no gráfico abaixo, pois 18 dos citados possuíam modularidade igual a zero. São eles: Café, Lígia; Dik, S.; Floridi, L.; Giaretta, P.; Gruber, Thomas. R.; Guarino, Nicola; Hendler, James; Herrera, Antonia Heredia; Hjørland, Birger; Lassila, Ora; Lee, Tim Berners; McGuinness, D. L.; Noy, N. F.; Ranganathan, S. R.; Soergel, Dagobert; Sowa, J. F.; Uschold, Mike; Vickery, Brian Campbell.

Assim como no gráfico referente a elite de pesquisa (Gráfico 5), no gráfico Gráfico 6, podemos perceber igualmente que a

Modularidade dos autores da Frente de pesquisa não acompanha os valores do Grau de saída de forma proporcional.

5. Considerações finais

Partimos da premissa de que uma das origens da Ciência da Informação é a Documentação francesa, primeiramente desenvolvida por Paul Otlet. Isso provavelmente, influenciou a produção das literaturas francófona, hispanófona e lusófona e desenvolveu um legado que teve, entre suas referências, a linguagem, em seus aspectos semânticos, sintáticos e pragmáticos. Muito provavelmente, essa produção não ficou imune às teorias anglo-saxônicas, considerando que a ciência é viva e as relações diretas e indiretas entre pesquisadores existe. A partir de tal premissa buscamos, na literatura científica do campo em análise – a Ciência da Informação, artigos que abordassem questões sobre os Sistemas de Organização da Informação, incluindo as linguagens documentárias, formalmente identificadas à KO ou que lhe fossem próximas, considerando a produção científica nos idiomas citados e repertoriada por bases de dados internacionais no período 1997-2011.

Os resultados permitiram ainda reunir informações sobre a distribuição dos artigos quanto às redes científicas existentes no assunto pesquisado. Para a análise da autoria e citação utilizamos a lei do elitismo, que permitiu identificar uma elite de pesquisa e a frente de pesquisa correspondente. A análise quantitativa dos resultados obtidos nos permite enunciar alguns resultados interessantes na pesquisa, que sintetizamos a seguir.

De um *corpus* de 354 artigos repertoriados e recuperados nas bases internacionais discriminadas, 167 são em idioma espanhol, 110 em francês e 77 em português. Os autores considerados de elite, no entanto, publicaram na sua maioria em língua portuguesa, indicando a necessidade de novas análises para explicar a diferença, bem como as variáveis que implicaram esse resultado.

Os 354 artigos recuperados são de responsabilidade de 462 autores, com uma média de 0,77 artigos por autor. Do conjunto total de artigos recuperados, 204 têm autoria única e 150 têm autoria coletiva. Dos 462 autores recuperados, 19 deles correspondem à elite de pesquisa, ou seja, 4,1% dos autores respondem pela maior produção no *corpus*.

Na elite de pesquisa, entre os autores com alta produtividade dentro do *corpus*, encontramos: Fujita, Mariângela S. L.; Boccatto, Vera R. C;

Café, Lígia; Gonzalez, Jose A. M.; Codina, Lluís; Martínez, Ana M.; Peña, Catalina N.; Almeida, Mauricio B.; Brun, Ricardo E.; Cuadrado, Sonia S.; Gonzalez, Mercedes M.; Lima, Gercina A. B. de O.; Maniez, Jacques; Moraes, Joao B. E.; Morato, Jorge; Moura, Maria A.; Quintana, Ania R. H.; Rubi, Milena P.; Urdiciain, Blanca G. Essa mesma elite foi responsável pela produção de 75 artigos do *corpus* (21,2%), e, em vários deles, assina em coautoria essa produção. As relações de coautoria da elite foram verificadas quanto à quantidade e quanto à intensidade das relações.

A elite citou, em seus 75 artigos, 1.536 referências produzidas por 2.189 citados. Foram identificados 47 autores citados como sendo a frente de pesquisa desta produção da elite. Dos 354 artigos do *corpus*, 150 foram feitos em coautoria e são de responsabilidade de 321 autores. As relações decorrentes destas coautorias foram investigadas utilizando ARS. Foram identificados, nos gráficos, dez agrupamentos bastante significativos entre os coautores (Gráfico 2) que sugere uma forma de colégio invisível.

As relações entre os autores da elite de pesquisa e os autores da frente de pesquisa também foram analisados utilizando ARS (Gráficos 3 e 4). Foi possível perceber, por meio dos gráficos, as autocitações, a formação de comunidades entre os agentes das duas categorias, além de verificar que a modularidade da rede não obedece a uma proporcionalidade aos graus de saída ou de entrada nela presentes.

Concluindo, podemos afirmar que a validade de uma pesquisa tem de ser considerada em relação ao universo delimitado e às suas metodologias. Isso significa dizer que não é possível, a partir dos resultados, fazer generalizações. Os recortes operacionalizados e as opções metodológicas para sua observação impõem limitações, mas são ferramentas importantes para a análise de problemas específicos. No entanto, o modelo metodológico aplicado se prestaria ao estudo de vários aspectos da produção científica. Sua adaptação a diferentes realidades, e/ou campos, dependeria da introdução de outras variáveis que atuam sobre as interações, conforme o interesse do estudo.

Nota

Integrando uma pesquisa mais ampla, este trabalho foi originado na tese de doutorado da autora, desenvolvido na Universidade de São Paulo (USP) que teve o apoio financeiro da CAPES (Brasil).

Referencias

- Almeida, Carlos Cândido (2013). Notas sobre os conceitos semióticos fundamentais à Organização do Conhecimento. // Congresso ISKO Espanha, Portugal, 1.; Congresso ISKO España, 11., Porto, Anais eletrônico... Porto: Faculdade de Letras da Universidade do Porto.
- Alves, B. H.; Oliveira, E. F. T.; Grácio, M. C. C. (2012). Análise bibliométrica da revista Scire: um estudo de colaboração científica institucional. // Scire. 18:2, 43-48.
- Café, Lígia; Bräscher, Marisa (2008). Organização da informação e bibliometria. // Encontros Bibli. 13:0, 54-75.
- Dahlberg, I. (2006). Knowledge organization: a new science? // Knowledge Organization. 33:1, 11-19.
- Ferreira, G. M.; Guimarães, J. A. C. (2011). Análise da presença e articulação da concepção teórica da catalogação de assunto na literatura científica da International Society for Knowledge Organization – ISKO: uma análise de domínio. // Revista EDICIC. 1:4, 188-207.
- Fujita, Mariângela S. L. (2008). Organização e Representação do Conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. // Tendências da Pesquisa Brasileira em Ciência da Informação. 1:1, 1-32.
- Gomes, Hagar Espanha (2009). Tendências da pesquisa em organização do conhecimento. // Tendências da Pesquisa Brasileira em Ciência da Informação. 2:1, 60-88.
- Laboratório de Estudos Métricos da Informação (LEMI) (2014). <http://lemi.uc3m.es/> (28 fev. 2014).
- Lara, Marilda Lopes Ginez. (2013). Documentary languages and knowledge organization systems in the context of the semantic web. // Transinformação. 25:2, 145-150.
- López-Huertas, M. J., Contreras, E. J. (2004) Spanish research in knowledge organization (1992-2001). // Knowledge Organization. 31:3, 136-150.
- Lucas, Elaine Rosângela de Oliveira (2014). Capital Social e Capital Científico na produção científica sobre Linguagens Documentárias e Sistemas de Organização do Conhecimento no campo da Knowledge Organization (KO) nos idiomas espanhol, francês e português. São Paulo, 2014. 165f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, SP.
- Martins, Dalton Lopes (2012). Análise de redes sociais de colaboração científica no ambiente de uma federação de bibliotecas digitais. 2012. 256 f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo.
- Moneda Corrochano, M.; López Huertas, M. J.; Contreras, E. J. (2011). La investigación sobre Organización del Conocimiento en España (2002-2010). // Congreso ISKO España, 10., Ferrol, Anais eletrônico... Ferrol: Universidade da Coruña.
- Price, Derek de Solla (1976). O desenvolvimento da ciência. Rio de Janeiro: Livros técnicos e Científicos, 1976.
- Souza, C. (2013). La organización del conocimiento: Estudio bibliométrico en la base de datos ISI Web of Knowledge. // Biblios. 0:51, 20-32.
- Software R® (2014). <http://www.r-project.org/> (28 fev. 2014).
- Software livre Gephi® versão 0.8.2 – Beta (2014). <https://gephi.org/> (28 fev. 2014).
- Urbizagastegui-Alvarado, Ruben. (2007). A Lei de Lotka e a Produtividade dos autores. 2007. 237f. Tese (Doutorado em Ciência da Informação) - Programa de Pós-graduação em Ciência da Informação. Universidade Federal de Minas Gerais, Belo Horizonte.

Enviado: 2014-04-01. Segunda versão: 2014-08-18.
 Aceptado: 2014-04-30.