

Interpretação e validação de agrupamentos em análise de cocitação de autores: estudo exploratório e metodológico

Rodrigo Aquino de Carvalho

Mestre; Universidade Federal do Rio Grande, Rio Grande, RS, Brasil;
racfurg@gmail.com

Catarina Prestes de Carvalho

Especialista; Universidade Federal de Pelotas, Pelotas, RS, Brasil;
catarinaprestes@yahoo.com.br

Sônia Elisa Caregnato

Doutora; Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil;
sonia.caregnato@ufrgs.br

Resumo: No contexto da fase de interpretação e validação de uma Análise de cocitação de Autores, são formulados os seguintes objetivos: (i) comparar dados dos artigos citantes do *corpus* com dados que caracterizam os citantes e os autores citados de um agrupamento de uma ACA; (ii) propor indicadores adicionais relacionados aos autores que formam um agrupamento; e (iii) visualizar a rede de documentos cocitados dos autores que formam o agrupamento. O trabalho é exploratório e metodológico e os dados foram retirados de 151 artigos da área da Organização do conhecimento, do período de 2011 a 2015, recuperados na *Web Of Science*. Uma matriz simétrica com 50 posições (autores com nove ou mais artigos citantes) gerou cinco agrupamentos. O principal agrupamento (17 autores) é o objeto dessa análise. Os autores são cocitados em 69 artigos. Das 131 relações dos autores nas referências, 84 ocorrem no parágrafo. Há dados que comparam citações, citantes e documentos citados para cada um dos 17 autores. São citados 230 artigos distintos no agrupamento. Conclui-se que a proximidade da cocitação no parágrafo valida o agrupamento e que os demais dados ampliam a visão da formação e da caracterização das relações de cocitação. Os resultados demonstram que os agrupamentos em uma ACA podem ser analisados como unidades autônomas e que a descrição dos dados que sustentam a formação do agrupamento (artigos e autores citantes) e que caracterizam os autores participantes e a rede de documentos cocitados, colabora na validação e interpretação das relações, pois apresenta dados objetivos. Espera-se que essa abordagem exploratória possa ser aplicada em outras ACA, independente da área.

Palavras-chave: Cocitação. Análise de cocitação de autores. Análise de cocitação de documentos. Estudo metodológico.

1 Introdução

A Cientometria estuda e utiliza diversos indicadores para entender a ciência e o *habitus* de seus agentes, principalmente nos aspectos concernentes à comunicação dos resultados de pesquisa. O campo científico é complexo e depende de outras áreas do conhecimento para sua avaliação e evolução.

Os indicadores bibliométricos são instrumentos que ajudam nessa função de estudar e avaliar a ciência e são, basicamente, divididos em indicadores de produtividade, de citação e de colaboração científica (GLÄNZEL, 2003). A cocitação, um indicador analítico derivado da citação, é o tema desse trabalho, mais especificamente a análise de cocitação de autores (*Authors cocitation analysis* – ACA).

Estudos de cocitação são utilizados principalmente como técnica de análise de domínio, visualização e mapeamento de uma área específica do conhecimento (EOM, 2009; LIMA, 2015). Como um método de análise de citações (HJØRLAND, 2013), a cocitação herda todo o arcabouço teórico-metodológico que estuda o fenômeno de citar/referenciar, pois a citação (incluindo a referência) é a prova empírica das relações que formam os campos do saber (ALVARENGA, 1998), mapeando os domínios científicos, pois apresenta elementos objetivos (registros e seus rastros) e subjetivos (as intenções de quem cita, a partir dos objetivos do seu texto e do contexto que o cerca) dessas relações (SILVEIRA; CAREGNATO, 2017), sendo um ato que permeia implicações das mais variadas, sejam “[...] psicológicas, sociológicas, políticas e históricas, assim como influências de outras naturezas, tais como o narcisismo (autocitações), influências entre autores e instituições, adesão a paradigmas vigentes.” (ALVARENGA, 1998, p. 6).

As principais unidades de contagem para estudos de cocitação são os documentos (EOM, 2009; TRUJILLO; LONG, 2018), os periódicos (MACHADO, 2015) e, principalmente, os autores (GRÁCIO; OLIVEIRA, 2013), que respondem por grande parte dos estudos publicados (CHEN; IBEKWE-SANJUAN; HOU, 2010).

Assim, no contexto da fase de interpretação e validação dos agrupamentos formados em uma ACA, são apresentados os objetivos dessa pesquisa:

- a) comparar dados dos artigos citantes do *corpus* com dados que caracterizam os artigos citantes e os autores citados do principal agrupamento da ACA realizada na pesquisa;
- b) propor indicadores adicionais relacionados aos autores, como por ex: identificação e número de documentos citados dos autores no *corpus* e no agrupamento, proporção de cocitação entre os autores na proximidade por referência e parágrafo etc.;
- c) e comparar a rede de documentos cocitados dos autores que formam o principal agrupamento da análise com as relações de cocitação dos próprios autores (o trabalho não se caracteriza como uma análise de cocitação de documentos, pois o processo será feito como parte da validação de um agrupamento específico de uma ACA).

Assim, esse trabalho pretende contribuir com os estudos de cocitação de autores no sentido de observar e mensurar dados que balizam e caracterizam a formação dos agrupamentos, pois a ACA tradicional não parece ser suficientemente informativa (BU *et al.*, 2017). Além disso, lidar com apenas um agrupamento mostra que esses resultados de estudos de cocitação emergem como unidades que podem ser analisadas separadamente e que ampliar a descrição dos dados torna a interpretação e a validação mais objetiva e rica, ainda que a análise subjetiva dos pesquisadores envolvidos nesse tipo de estudo seja necessária e única.

O estudo é aplicado na subárea da Ciência da Informação “Organização do conhecimento”. Essa área foi escolhida pela familiaridade dos pesquisadores com o tema e seus principais autores, e é uma extensão de um projeto em andamento sobre questões metodológicas em ACA e sua relação com a análise de cocitação de documentos.

2 Aspectos conceituais e trabalhos relacionados

A ACA é uma técnica de relacionamento de referências e, por vezes, é confundida com o acoplamento bibliográfico. O acoplamento relaciona os documentos citantes (há estudos que acoplam a partir dos autores¹) a partir de um documento citado (GLÄNZEL, 2003), e, portanto, tem um olhar para frente, enquanto a cocitação relaciona dois documentos citados (a unidade pode ser autor ou periódico) a partir de um documento citante (SMALL, 1973), tendo, assim, um olhar para o passado.

A matriz da Figura 1, que é assimétrica e de ocorrência (LEYDES-DORFF; VAUGHAN, 2006), representa o relacionamento de publicações e suas referências, e mostra que tanto a cocitação como o acoplamento partem de uma mesma fonte. A linha “C1” mostra que os documentos “R1”, “R3” e “R6” estão cocitados, enquanto que a coluna “R1”, mostra que os documentos “C1”, “C3” e “C5” estão acoplados.

Figura 1 - Matriz representando o relacionamento entre documentos citantes e citados

		Documentos citados (Referências)					
		<i>R1</i>	R2	R3	R4	R5	R6
Documentos Ci- tantes	<i>C1</i>	<i>01</i>	<i>00</i>	<i>01</i>	<i>00</i>	<i>00</i>	<i>01</i>
	C2	<i>00</i>	01	01	01	01	01
	C3	<i>01</i>	01	00	00	00	00
	C4	<i>00</i>	00	00	00	00	00
	C5	<i>01</i>	01	01	01	01	01

Fonte: Adaptado de Glänzel (2003, p. 85).

A matriz de relacionamento demonstrada na Figura 1 pode derivar duas matrizes simétricas (Figura 2) que indicam com mais clareza as relações, pois a contagem considera apenas ou a cocitação ou o acoplamento, respectivamente.

Figura 2 - Matrizes quadradas de cocitação e acoplamento

Cocitação (documentos citados)							Acoplamento (documentos citantes)					
	R1	R2	R3	R4	R5	R6		C1	C2	C3	C4	C5
R1	-	03	02	01	01	02	C1	-	02	01	00	03
R2		-	02	02	02	02	C2		-	01	00	05
R3			-	02	02	03	C3			-	00	02
R4				-	02	02	C4				-	00
R5					-	02	C5					-
R6						-						

Fonte: Dados da pesquisa.

Assim, a técnica de ACA, tradicionalmente, se realiza a partir de seis passos, conforme McCain (1990, p.434):

- seleção dos autores: os mais citados ou os com mais documentos citantes em uma área do conhecimento ou periódico, através de consulta a pesquisadores, listas etc. (McCAIN, 1990);
- recuperação das frequências de cocitação: onde deve ser definido se serão considerados todos os autores das referências ou apenas o primeiro (PERSSON, 2001; ROUSSEAU; ZUCCALA, 2004; CARVALHO; CAREGNATO, 2016), sendo que a coautoria pode ser considerada cocitação, conforme os objetivos e operacionalização do estudo (SCHNEIDER; LARSEN; INGWERSEN, 2009), e a proximidade da relação (LIU; CHEN, 2012), sendo que a proximidade por artigo, onde as relações são estabelecidas a partir das referências, é a opção mais utilizada;
- compilação da matriz de cocitação com valores absolutos: nessa etapa é necessário definir os valores da diagonal (GRÁCIO; OLIVEIRA, 2013), já que a matriz é quadrada e esse dado influencia na próxima etapa;
- conversão da matriz principal em uma matriz de correlação: utilizar índices de similaridade para normalizar os dados, sendo que os mais utilizados são o r de Pearson, o Cosseno de Salton e o Índice de Jaccard (GRÁCIO; OLIVEIRA, 2015). Leydesdorff e Vaughan (2006) sugerem que esta conversão deve ser feita em uma matriz assimétrica;
- análise multivariada da matriz de correlação/normalizada: as principais técnicas utilizadas são “análise de componentes principais”, “análise de

cluster” e “escalonamento multidimensional” (McCAIN, 1990; EOM, 2009);

- f) e interpretação e validação: comparação com outros dados, consulta aos especialistas da área e/ou validação estatística (McCAIN, 1990).

Além de técnicas estatísticas multivariadas, há estudos que optam pela análise de redes sociais (ARS; do inglês: *Social Network Analysis*) como ferramenta para evidenciar os agrupamentos e analisar a força das relações de cocitação de autores e documentos. Os trabalhos de Jeong, Song e Ding (2014), Lima (2015), Piovezan e Fujita (2015) e Muñoz-Muñoz e Mirón-Valdivieso (2017) são alguns exemplos de pesquisas que seguem essa opção.

Cada etapa apresentada envolve discussões e decisões metodológicas significativas sobre a ACA, mas também mostra a complexidade em operacionalizar os estudos. A seleção dos autores, por exemplo, implica em decidir sobre considerar todas as autorias dos documentos (inclusive as omitidas por normas de citação e de referência bibliográfica, regras das revistas e descrição dos registros em bases de dados) ou apenas as primeiras, como já salientado. A recuperação das frequências considera a proximidade da ocorrência da cocitação (artigo ou referência, seção do documento, parágrafo ou sentenças) (LIU; CHEN, 2012), e isso implica em observar como as referências são mencionadas ou recitadas no corpo texto (DING *et al.*, 2013; BU *et al.*, 2017; CARVALHO *et al.*, 2018), uma vez que a lista de referências bibliográficas iguala o uso dos documentos. Porém, há dados que demonstram que alguns são mais utilizados do que outros, além da tendência de os autores mais citados terem mais de um documento utilizado em um mesmo artigo citante, como demonstram Carvalho *et al* (2018).

No contexto dos objetivos dessa pesquisa, três trabalhos se destacam quanto às propostas e resultados que apresentam.

O primeiro trabalho é dos autores Chen, Ibekwe-Sanjuan e Hou (2010), que pontuam a grande dificuldade de interpretar os resultados oriundos de ACA e de análise de cocitação de documentos (DCA, do inglês *Document Cocitation Analysis*), que se concentram nas entidades citadas (autores/documentos) como

fonte primeira de evidência, e sugerem que se concentrar nos citantes pode melhorar a compreensão. O artigo objetivou apresentar uma perspectiva múltipla aos estudos de cocitação, no sentido de analisar padrões estruturais, temporais e semânticos, além de elementos dos citados e citantes para ampliar a interpretação e a confiabilidade dos agrupamentos formados pelo estudo. Os dados foram retirados de documentos publicados em 12 periódicos da área da Ciência da Informação, de 1996 a 2008, totalizando 10853 registros, que citaram 129060 referências únicas através de 206180 citações. Essas referências apresentaram 58711 autores únicos. Os agrupamentos da ACA foram definidos a partir de 663 autores e os agrupamentos da DCA a partir de 655 documentos. Um sistema automático rotulou os agrupamentos com dados dos documentos citantes (*citers*) e os autores concluíram que é recomendado realizar as duas técnicas juntas (ACA e DCA) sempre que possível. Um recorte menor da análise (2001-2005) validou tanto os agrupamentos em relação a um estudo² anterior, com uma sobreposição de 82% dos membros; como os rótulos gerados de forma automática com rótulos desenvolvidos manualmente pelo mesmo estudo anterior, que utilizava o mesmo *corpus* e recorte temporal.

Bu, Liu e Huang (2016) propõem um método de ACA modificado, chamado de MACA (*Modified Author Co-citation Analysis*). A pesquisa explora o uso de quatro metadados descritivos para qualificar o relacionamento que os dados de cocitação apresentam: i) autores citados/referenciados (contagem de cocitação); ii) o tempo (data) em que a citação/referência foi publicada; iii) suporte ou fonte onde a citação/referência foi publicada; e iv) palavras-chaves de uma citação/referência. O conjunto de dados para a pesquisa foi recuperado no periódico *Journal of the American Society for Information Science and Technology* (JASIST) entre os anos de 2003 e 2012. Vários metadados descritivos foram retirados de 2038 artigos e 68606 referências e, no final, os 100 autores mais citados foram utilizados para finalizar a pesquisa. Análise fatorial e escalonamento multidimensional (*Multidimensional Scaling* – MDS) foram as técnicas de estatística multivariada utilizadas, e as matrizes foram construídas com diagonal igual a zero e considerando apenas os primeiros autores das referências. Os autores concluíram que MACA oferece mais detalhes utilizando MDS como técni-

ca estatística, adiciona mais informações à ACA tradicional e que, dependendo da informação utilizada para ampliar a análise, o impacto é distinto.

Por último, Bu *et al* (2017³) apresentam um método de ACA combinado com metadados do texto completo, chamado de MFTACA (*Metadata-in-full-text-based ACA*). Os autores partem do pressuposto de que as menções a documentos no corpo do texto modificam a suposta “igualdade” apresentada na lista de referências e isso tem consequência nas análises de cocitação. O Quadro 3 desse trabalho ilustra essa preocupação dos autores apenas relacionando citações e documentos citantes e mostra que, assim como os documentos (CARVALHO *et al*, 2018), os principais autores são mencionados várias vezes por documentos citantes, até pelo número de textos utilizados de alguns autores específicos, como Birger Hjørland, que possui, nos dados dessa pesquisa, 59 diferentes documentos citados, que totalizam 228 citações em 70 artigos citantes.

Os dados da pesquisa de Bu *et al* (2017) são extraídos de um conjunto de 1420 artigos do periódico *Journal of the American Society for Information Science and Technology* (JASIST). Os 500 autores mais citados são utilizados para formar a matriz de cocitação, além do número de menções de cada referência, número de palavras do contexto de uso das referências e ano de publicação das referências. A matriz de correlação é construída pelo r de Pearson e a análise fatorial, a análise de redes sociais e o MDS são as técnicas utilizadas para gerar os agrupamentos. Os autores concluem que, apesar de não utilizar informação no nível do conteúdo, como fazem Jeong, Song e Ding (2014), o método proposto apresenta melhor desempenho na formação dos agrupamentos do que a ACA tradicional, além de fornecer mais detalhes. Por fim, consideram que a técnica tanto pode ser ampliada com o uso de outros metadados presentes nos textos, como pode ser utilizada também em estudos de acoplamento bibliográfico. Além disso, consideram que usar apenas o primeiro autor das referências é uma limitação considerável da pesquisa.

Os estudos, de maneira geral, têm buscado melhorar a qualidade de análise dos agrupamentos, para ajudar na interpretação do que realmente ocorre ou existe de relação entre os autores/documentos. Esse trabalho busca isso também,

mas emprega ações em um agrupamento já formado por uma ACA tradicional, identificando e descrevendo padrões estruturais dos dados que sustentam (documentos citantes) e caracterizam (informações dos autores e documentos citados) a cocitação de autores em um determinado contexto.

3 Procedimentos metodológicos

O trabalho é metodológico, exploratório e uma extensão dos resultados publicados por Carvalho e Caregnato (2017). Os dados empíricos foram extraídos de um conjunto de 151 artigos de pesquisa, recuperados na base de dados *Web of Science* em maio de 2017. A busca foi realizada pelo termo “*Knowledge organization*” no campo “TÓPICO” e na categoria “*INFORMATION SCIENCE LIBRARY SCIENCE*”. A Tabela 1 apresenta a distribuição dos artigos recuperados por ano de publicação e periódico.

Tabela 1 - Distribuição dos artigos recuperados por ano de publicação e periódicos

PERIÓDICOS	2011	2012	2013	2014	2015	Total	%
<i>Knowledge Organization</i>	07	12	18	20	26	83	54,97
<i>Journal of the Association for Information Science and Technology – JASIST</i>	02	02	04	01	02	11	07,28
<i>Journal of Documentation</i>	02	02	03	01	02	10	06,62
<i>Journal of Information Science</i>	-	-	04	02	01	07	04,64
Perspectivas em Ciência da Informação	04	03	-	-	-	07	04,64
<i>Library Trends</i>	-	02	02	-	02	06	03,97
<i>Scientometrics</i>	-	-	02	01	-	03	01,99
TransInformação	01	-	-	01	01	03	01,99
<i>El profesional de la información</i>	-	02	-	-	-	02	01,32
Informação & Sociedade	-	02	-	-	-	02	01,32
<i>Information Research</i>	-	01	-	01	-	02	01,32
<i>Information Processing and Management</i>	-	-	02	-	-	02	01,32
<i>Qualitative and Quantitative Methods in Libraries</i>	-	-	-	-	02	02	01,32
<i>The Electronic Library</i>	-	01	-	01	-	02	01,32
Outros ⁴	01	02	02	01	03	09	05,96
Total por ano	17	29	37	29	39	151	100,00

Fonte: Dados da pesquisa.

Os dados extraídos desse *corpus* de análise foram estruturados em três diferentes conjuntos, sendo que o primeiro caracteriza as referências a partir de sete indicadores: 1) nº de referências; 2) nº de referências com entrada por autoria pessoal; 3) nº de referências que possuem apenas um autor; 4) nº de referências com dois ou mais autores; 5) nº de referências com outros tipos de entrada

(jurisdição, título, entidade coletiva etc.); 6) nº de autorias; e 7) referências desconsideradas. A Tabela 2 apresenta os valores relativos a esses indicadores com as principais medidas estatísticas de centralidade e de dispersão.

Tabela 2 - Estatísticas descritivas dos indicadores das referências dos artigos (N=151)

INDICADORES	Total	Medidas de centralidade			Medidas de dispersão		
		Média	Moda	Mediana	Mínimo	Máximo	Desvio padrão
Referências	5771	38,22	28	34	05	120	22,50
Referências autoria pessoal	5449	36,06	21	32	04	111	21,81
Referências com um autor	3469	23,08	07	20	02	88	17,09
Referências com dois ou mais autores	1980	13,11	02	09	-	59	11,72
Referências autoria não pessoa	318	2,11	-	01	-	55	5,29
Autorias	9311	61,66	33	49	06	213	43,00
Referências desconsideradas	04	0,03	-	-	-	03	0,26

Fonte: Dados da pesquisa.

O segundo conjunto de dados é formado por um *ranking* de citação com 5333 autores, feito manualmente a partir das 5449 referências com entrada por autoria pessoal. A limpeza dos dados foi realizada manualmente, identificando erros de digitação e entradas por sobrenomes compostos. Além da identificação do número de citações de cada autor, foi identificado o número de artigos citantes no *corpus*, bem como a posição do autor na referência de documentos publicados em coautoria (primeira ou secundária) ou autoria única. A Tabela 3 apresenta as dez primeiras posições do *ranking*, organizado a partir do número de citações.

Tabela 3 - Distribuição das dez primeiras posições do *ranking* de citação

Autores	Citações	Citantes	Autoria única*	Autoria principal*	Autoria secundária*
1 Hjørland, B	228	70	180	38	10
2 Smiraglia, RP	68	29	51	08	09
3 Olson, HÁ	62	31	52	07	03
4 Szostak, R	59	13	54	01	04
5 Mai, J-E	47	25	46	-	01
6 Dahlberg, I	44	28	44	-	-
7 Gnoli, C	35	18	22	12	01
8 Tennis, JT	34	23	27	02	05
9 Berners-Lee, T	32	18	20	10	02
10 Guimarães, JAC	32	13	08	16	08

* esses dados não se referem necessariamente ao nº de documentos que esses autores possuem no corpus, ou seja, podem ser derivados de documentos repetidos dependendo dos artigos citantes.

Fonte: Dados da pesquisa.

O terceiro conjunto é uma matriz simétrica de cocitação de autores com 50 posições (autores com nove ou mais artigos citantes). Diferente de Bu, Liu e Huang (2016) e Bu *et al* (2017), considerou-se o número de artigos citantes mais interessante que o número de citações para a formação da matriz de cocitação, pois esse indicador apresenta um autor mais “espalhado” no *corpus* e alguns dados mostram que há autores bastante citados em poucos documentos.

A matriz foi analisada e gerou cinco agrupamentos a partir dos seguintes parâmetros:

- contagem dos pares considerando todos os autores das referências (ROUSSEAU; ZUCCALA, 2004) e coautoria como cocitação (SCHNEIDER, LARSEN E INGWERSEN, 2009);
- normalização dos dados utilizando o *Cosseno de Salton* (GRÁCIO; OLIVEIRA, 2015);
- valor da diagonal da matriz (autor *vs* ele mesmo) igual a zero; d) análise fatorial utilizando a técnica de componentes principais, sem definir número fixo de fatores, com rotação *varimax* e suprimindo coeficientes menores que “0,5” (BECKER, 2015; FIELD, 2009).

O Quadro 1 apresenta a distribuição dos quatro principais fatores, que explicam 89,23% da variância, assim como a carga fatorial das relações.

Quadro 1 - Distribuição dos fatores e cargas fatoriais da ACA

1º Fator (46,40%*)		2º Fator (22,75%*)		3º Fator (13,12%*)		4º Fator (6,96%*)	
Autores	CF	Autores	CF	Autores	CF	Autores	CF
López-Huertas	,953	Ranganathan	,976	Berners-Lee, T	,960	Bowker, GC	,874
Albrechtsen, H	,919	La Barre, K	,955	Greenberg, J	,935	Star, SL	,862
Thellefsen, TL	,907	Gnoli, C	,954	Lassila, O	,925	Berman, S	,762
Thellefsen, MM	,891	Spiteri, LF	,947	Taylor, A	,924	Beghtol, C	,696
Frohmann, B	,870	Vickery, BC	,941	Hendler, J	,923	Olson, HA	,691
Guimarães, JAC	,867	Mazzocchi, F	,899	McGuinness	,914	Tennis, JT	,657
Buckland, MK	,858	Broughton, V	,823	Noy, NF	,910	Foucault, M	,628
Hansson, J	,846	Soergel, D	,822	Gruber, TR	,906	Furner, J	,581
Andersen, J	,843	Szostak, R	,800	Smith, B	,714		
Friedman, A	,778	Jacob, EK	,771	Zeng, ML	,637		
Mai, J-E	,754	Kuhn, TS	,760				
Dahlberg, I	,715	Lancaster, FW	,709				
Bates, MJ	,712	Green, R	,590				
Smiraglia, RP	,678						
Hodge, G	,675						
Hjørland, B	,664						
Svenonius, E	,624						

CF = Carga fatorial | 5º fator formado pelos autores ‘Bawden, D’ (0,876) e ‘Chan, LM’ (0,702).

Fonte: Dados da pesquisa.

Os objetivos desse trabalho são desenvolvidos a partir dos resultados do primeiro agrupamento e se dividem em três grupos distintos de dados:

- a) dados genéricos e de caracterização do agrupamento, para comparação com dados do *corpus*, por exemplo, número de artigos dos autores citados no *corpus* e no agrupamento, informações dos artigos citantes etc.;
- b) dados relativos aos autores para detalhamento das relações entre eles, além, por exemplo, do número de artigos citados no *corpus* e no agrupamento etc.;
- c) caracterização e rede dos artigos mais citados que formam o agrupamento.

Esse procedimento é genérico e pode ser utilizado em ACA aplicada a qualquer área do conhecimento, desde que a extração dos dados permita uma relação clara dos autores citados com os documentos que os citam. A matriz simétrica de cocitação de autores para esse estudo foi elaborada a partir de uma matriz assimétrica gerada através de um *ranking* de citação que rotulava documentos citantes.

Além disso, outros dois aspectos merecem destaque para a utilização dessa abordagem em ACA: a técnica de estatística multivariada utilizada para gerar os agrupamentos e o acesso aos textos completos dos artigos citantes. A utilização de outra técnica de estatística ou a utilização da análise de redes sociais pode implicar em resultados distintos à aplicação de uma análise fatorial e somente o acesso ao texto completo vai permitir determinar a proximidade da cocitação no parágrafo.

Por fim, os softwares *Microsoft Excel*, *SPSS* e *Gephi* foram utilizados na estruturação e análise dos dados.

4 Resultados e discussão

Para entender a dimensão do agrupamento em relação ao *corpus* de análise (N=151 artigos), são apresentados alguns dados gerais e a relação dos principais

documentos citantes. Esses dados atendem ao primeiro objetivo da pesquisa e evidenciam que um estudo de cocitação deve considerar a relação com o contexto geral de análise.

Os autores que formam o agrupamento analisado são citados em 106 artigos (70,20%), mas somente em 69 (45,70%) há cocitação, ou seja, há pelo menos dois autores citados, e a Tabela 4 demonstra a distribuição desses artigos relacionando o número de autores citados com os periódicos publicados.

Tabela 4 - Distribuição do nº de autores citados por artigo no agrupamento

Revistas	Distribuição por artigos (n=69)									Total	%	% Acumulado
	02 Autores	03 Autores	04 Autores	05 Autores	06 Autores	07 Autores	08 Autores	09 Autores	12 Autores			
Knowledge Organization	08	15	06	04	03	01	02	02	01	42	60,87	60,87
Journal of Documentation		03		01		01	02		01	08	11,59	72,46
JASIST*	02	04					01			07	10,14	82,61
Library Trends		01	01							02	2,90	85,51
PCI*	01	01								02	2,90	88,41
EBLIS*		01								01	1,45	89,86
El profesional de la información	01									01	1,45	91,30
Information Research					01					01	1,45	92,75
IPM*		01								01	1,45	94,20
JoLIS*	01									01	1,45	95,65
Online Information Review				01						01	1,45	97,10
REDC*					01					01	1,45	98,55
TransInformação	01									01	1,45	100,00
Total	14	26	07	06	05	02	05	02	02	69		
%	20,29	37,68	10,14	8,70	7,25	2,90	7,25	2,90	2,90			100,00
% Acumulado	100,00	79,71	42,03	31,88	23,19	15,94	13,04	5,80	2,90			

* JASIST: *Journal of the Association for Information Science and Technology*; PCI: *Perspectivas em Ciência da Informação*; EBLIS: *Evidence Based Library and Information Practice*; IPM: *Information Processing and Management*; JoLIS: *Journal of Librarianship and Information Science*; REDC: *Revista Española de Documentación Científica*.

Fonte: Dados da pesquisa.

Os dados indicam que quase 80% dos artigos citam pelo menos três autores e que 16 artigos (23,19%) citam seis ou mais. O periódico *Knowledge Organization* é a fonte de 60,87% dos artigos, e trata-se de uma revista científica especializada. As revistas *Journal of Documentation* e *Journal of the Association for Information Science and Technology* são mais genéricas em relação à Ciên-

cia da Informação, mas podem atender o escopo do periódico *Knowledge Organization* e são as fontes de publicação de outros oito e sete artigos, respectivamente. Utilizar os dados desses documentos e de suas fontes pode ajudar no desafio de rotular os agrupamentos, como afirmam Chen, Ibekwe-Sanjuan e Hou (2010). O Quadro 2 apresenta alguns metadados dos 16 artigos que mais citam autores do agrupamento analisado.

Observa-se que seis autores (Friedman, A; Guimarães, JAC; Hjørland, B; Smiraglia, RP; Thellefsen, TL; e Thellefsen, MM) dos principais documentos citantes também fazem parte do agrupamento como citados. Os dados de autocitação não foram retirados da análise, pois esse não é um procedimento usual nos trabalhos aplicados de cocitação de autores e nem é recomendado por trabalhos metodológicos, mas evidencia-se como elemento para estudos futuros.

Quadro 2 - Lista de artigos que citaram seis ou mais autores

Autores - Título - Data - Nº autores citados – Periódico	
1.	Thellefsen, M.; Thellefsen, T.; Sørensen, B. <i>A Pragmatic Semeiotic Perspective on the Concept of Information Need and Its Relevance for Knowledge Organization</i> – 2013 (12) – <i>Knowledge Organization</i> .
2.	Friedman, A.; Thellefsen, M. <i>Concept theory and semiotics in knowledge organization</i> – 2011 (12) – <i>Journal of Documentation</i> .
3.	Guimarães, JAC.; Tognoli, N. <i>Provenance as a Domain Analysis Approach in Archival Knowledge Organization</i> – 2015 (09) – <i>Knowledge Organization</i> .
4.	Hjørland, B. <i>User-based and Cognitive Approaches to Knowledge Organization: A Theoretical Analysis of the Research Literature</i> – 2013 (09) – <i>Knowledge Organization</i> .
5.	Smiraglia, R.; Heuvel, Cvd. <i>Classifications and concepts: towards an elementary theory of knowledge interaction</i> – 2013 (08) – <i>Journal of Documentation</i> .
6.	Thellefsen, T.; Thellefsen, M.; Sørensen, B. <i>Emotion, Information, and Cognition, and Some Possible Consequences for Library and Information Science</i> – 2013 (08) – <i>Journal of the Association for Information Science and Technology</i> .
7.	Guimarães, JAC.; Pinho, FA. <i>Male Homosexuality in Brazilian Indexing Languages: Some Ethical Questions</i> – 2012 (08) – <i>Knowledge Organization</i> .
8.	Friedman, A.; Smiraglia, R. <i>Nodes and arcs: concept map, semiotics, and knowledge organization</i> – 2013 (08) – <i>Journal of Documentation</i> .
9.	Almeida, CC.; Fujita, MSL.; Reis, DM. <i>Peircean Semiotics and Subject Indexing: Contributions of Speculative Grammar and Pure Logic</i> – 2013 (08) – <i>Knowledge Organization</i> .
10.	Castanha, RCG.; Grácio, MCC. <i>Bibliometrics Contribution to the Metatheoretical and Domain Analysis Studies</i> – 2014 (07) – <i>Knowledge Organization</i> .
11.	Hjørland, B. <i>Is classification necessary after Google?</i> – 2012 (07) – <i>Journal of Documentation</i> .
12.	Martínez-Ávila, M.; San Segundo, R.; Zurian, FA. <i>Challenges and opportunities for knowledge organization at the intersection with information technologies</i> – 2014 (06) – <i>Revista Española de Documentación Científica</i> .
13.	Tennis, JT. <i>Foundational, First-Order, and Second-Order Classification Theory</i> – 2015 (06) – <i>Knowledge Organization</i> .
14.	Martínez-Ávila, M. <i>Knowledge Organization in the Intersection with Information Technologies</i> – 2015 (06) – <i>Knowledge Organization</i> .
15.	Almeida, CC. <i>The Methodological Influence of Peirce's Pragmatism on Knowledge Organization</i> – 2012 (06) – <i>Knowledge Organization</i> .
16.	Friedman, A. <i>The relationship between research method and visual display: a study of conference proceedings in the field of knowledge organization</i> – 2014 (06) – <i>Information Research</i> .

Fonte: Dados da pesquisa.

Os trabalhos focam em aspectos conceituais da Organização do conhecimento e a relação com outras áreas, além de algumas aplicações. Vale destacar os estudos sobre classificação e teoria do conceito. Um estudo de acoplamento bibliográfico poderia evidenciar com mais clareza a relação dos documentos.

Além desses dados, vale salientar que os 17 autores possuem 650 citações no *corpus*, sendo que 586 (90,15%) são nos 69 artigos do agrupamento. Assim, o agrupamento se mostra muito coeso na estrutura dos dados que o sustentam, como as fontes onde os citantes são publicados, que, apesar de a análise não ser no nível do conteúdo, demonstram, como sugerem Bu *et al.* (2017), uma ligação temática.

Um conjunto de dados relacionados aos autores é apresentado para atender ao segundo objetivo. O Quadro 3 descreve diversas informações que caracterizam a participação dos autores no agrupamento e no *corpus* de análise da pesquisa, são eles: valores absolutos de cocitação na proximidade por artigo (referências), que considera coautoria como cocitação, técnica de contagem utilizada por Schneider, Larsen e Ingwersen (2009); valores absolutos de cocitação na proximidade por parágrafo, onde são contados os artigos onde há ocorrência e não o número de parágrafos, e desconsideradas as coautorias; número de citações que os autores receberam nos artigos que formam o agrupamento e no total; número de artigos citantes no agrupamento e no total; e número de documentos distintos dos autores citados no agrupamento e no total, destacando as datas de publicação dos mais antigos e dos mais atuais. Esses dados, ainda que descritivos, evidenciam a importância e a participação de cada autor no agrupamento. Essa abordagem pode colaborar com estudos onde os pesquisadores tem pouca familiaridade com o domínio, além de ajudar a tornar as comunicações dos resultados mais autônomas para a comunidade interessada.

As coautorias mais significativas da análise são entre os autores “Hjørland, B” e “Albrechtsen, H” e entre os autores “Thellefsen, MM” e “Thellefsen, TL”. Realizar estudos considerando todos os autores das referências enriquece a análise, pois evidencia autores que colaboram muito (CARVALHO; CAREGNATO, 2016), mas acarreta um pouco de ruído informacional dependendo do

objetivo proposto, mesmo que essas relações sejam separadas, pois a unidade documental é a mesma.

A ocorrência da cocitação no parágrafo mostra com mais clareza a relação entre os autores e se revela efetivamente temática, assim como na proximidade por sentença (LIU; CHEN, 2012). Os dados mostram que das 131 relações entre os autores, dentre 136 possíveis (pois há autores que não são cocitados), 64,12% também ocorrem no parágrafo e isso valida o agrupamento. Verificar a ocorrência dos pares de cocitação em outros tipos de proximidade se mostra uma alternativa interessante à ACA tradicional, principalmente na fase de validação e interpretação dos agrupamentos ou de um agrupamento de destaque, uma vez que operacionalizar estudos no texto completo se mostra complexo, principalmente para desambiguação dos nomes dos autores (DING *et al.*, 2013) e pela forma como a citação ocorre no corpo do texto.

Quadro 3. Distribuição de dados relativos aos autores do agrupamento analisado.

Autores*	Matrizes de cocitação (referência vs parágrafo)																	Indicadores adicionais dos autores											
	Hjørland, B	Smiraglia, RP	Dahlberg, I	Mai, J-E	Hodge, G	Albrechtsen, H	Svenonius, E	López-Huertas, MJ	Guimarães, JAC	Frohmann, B	Buckland, MK	Thellefsen, MM	Bates, MJ	Andersen, J	Hansson, J	Thellefsen, TL	Friedman, A	Citações no corpus	Citações no agrupamento	Diferença	Citantes no corpus	Citantes no agrupamento	Diferença	Documentos citados no corpus (datas de publicação do mais antigo e do mais atual)	Documentos citados no agrupamento (datas de publicação do mais antigo e do mais atual)	Diferença			
Hjørland, B		19	23	20	14	19	14	13	10	10	09	10	07	09	07	08	08	228	202	26	70	59	11	59	1988-2015	56	1992-2015	03	
Smiraglia, RP	14		09	08	06	06	07	06	03	03	03	03	02	02	02	02	05	68	62	06	29	24	05	48	1987-2015	44	=	04	
Dahlberg, I	10	05		08	07	09	06	05	06	07	05	06	04	03	03	05	02	44	41	03	28	26	02	18	1974-2009	17	=	01	
Mai, J-E	11	05	02		06	08	06	04	02	05	04	07	02	06	04	05	06	47	44	03	25	22	03	20	1997-2015	19	=	01	
Hodge, G	05	01	02	01		04	04	03	-	03	04	04	04	04	04	03	03	25	16	09	23	16	07	04	1999-2013	02	2000-2013	02	
Albrechtsen, H	12	03	01	04	-		02	04	05	06	05	05	04	02	03	05	02	23	23	-	19	19	-	05	1993-1998	05	=	-	
Svenonius, E	06	03	01	02	-	-		02	-	03	03	02	-	05	04	01	04	25	25	-	18	18	-	08	1981-2004	08	=	-	
López-Huertas, MJ	05	-	02	01	02	01	-		01	02	04	04	01	02	04	04	03	21	19	02	16	14	02	10	1996-2012	09	=	01	
Guimarães, JAC	05	-	02	-	-	01	-	01		03	02	04	01	-	01	04	-	32	25	07	13	11	02	26	2000-2015	22	=	04	
Frohmann, B	03	02	-	01	-	-	-	-	01		01	03	02	04	03	03	03	21	19	02	12	11	01	10	1983-2009	08	1983-2008	02	
Buckland, MK	04	02	02	02	-	01	-	01	-	-		07	02	02	05	05	04	14	11	03	12	11	01	09	1969-2012	05	=	04	
Thellefsen, MM	05	01	04	02	02	02	-	02	01	01	01		02	02	04	09	05	24	24	-	11	11	-	11	2003-2013	11	=	-	
Bates, MJ	03	01	01	-	-	-	-	-	-	-	-	-		02	03	02	01	11	09	02	11	09	02	08	1989-2007	07	=	01	
Andersen, J	02	-	-	04	01	01	-	-	-	01	-	01	-		04	02	01	12	12	-	10	10	-	07	2002-2008	07	=	-	
Hansson, J	01	-	-	-	02	-	01	-	-	-	02	01	01	01		04	02	12	12	-	10	10	-	06	1997-2006	06	=	-	
Thellefsen, TL	04	01	03	01	01	02	-	01	01	01	01	04	-	01	01		04	27	27	-	09	09	-	14	2002-2013	14	=	-	
Friedman, A	01	02	02	01	02	-	-	01	-	-	-	-	-	-	01	01		15	15	-	10	10	-	05	2006-2013	05	=	-	

OBSERVAÇÕES: * ordem definida a partir do número de artigos citantes | ** n° de artigos e não de parágrafos onde há ocorrência da cocitação nesse tipo de proximidade, além de coautorias desconsideradas |

Fonte: Dados da pesquisa, 2017.

O autor “Hjørland, B” é o único do agrupamento cocitado, na proximidade por parágrafo, com todos os outros autores, e isso pode ser explicado pelos dados que complementam o Quadro 3, como o número de citações (202), o número de artigos citantes (59) e, principalmente, o número de documentos citados (56), em uma produção bibliográfica, que se estende de 1992 a 2015, falando especificamente dos dados do agrupamento. Já os autores “Svenonius, E”, “Guimarães, JAC” e “Bates, MJ” são os que menos se relacionam no nível do parágrafo, e isso pode ser explicado pela distribuição das citações nos 69 artigos citantes que formam o agrupamento. Ou seja, ampliar o nível de descrição ajuda a perceber esses detalhes que caracterizam o agrupamento.

Um destaque interessante é o autor “Hodge, G”, que possui apenas dois artigos citados no corpus, mas um deles está entre os mais citados. O artigo foi publicado em 2000 e seu título é “*Systems of knowledge organization for digital libraries*” (ver Figura 3). Essa informação evidencia que a unidade bibliográfica é muito significativa para a área, mesmo para um autor pouco produtivo diante do contexto analisado.

Os dados mostram, também, que sete autores são citados apenas em artigos que formam o agrupamento e que exceto pelos autores “Hjørland, B”, “Smiraglia, RP” e “Hodge, G” a diferença se mostra pequena entre o número de citações, documentos citantes e documentos citados dos outros sete autores. Vale salientar que os artigos citantes do agrupamento também podem influenciar as relações dos outros agrupamentos gerados pela análise fatorial, principalmente os que citam apenas dois e três autores.

O autor “Hjørland, B”, por exemplo, apresentou uma carga fatorial alta em relação ao segundo agrupamento, apresentado no Quadro 1. O número de documentos que o citam e de documentos de sua autoria citados na análise evidenciam sua influência no contexto geral desse *corpus*. Aplicar essa abordagem no segundo agrupamento poderia mostrar outros detalhes e oferecer uma base comparativa que enriqueceria a análise.

A diferença entre as citações e o número de documentos citantes mostra que os autores são citados mais de uma vez no mesmo artigo, como já salientado

(BU *et al.*, 2017; CARVALHO *et al.*, 2018). Essa informação é omitida por ACA tradicionais, pois a contagem dos pares é feita pela ocorrência, mesmo que um determinado autor possua 10 documentos citados em um mesmo artigo. No presente estudo, apenas os autores “Hodge, G” (16), “Buckland, MK” (11) e “Bates, MJ” (09) possuem citações iguais ao número de documentos citantes no agrupamento e apenas a última tem uma relação idêntica nos dados relativos ao *corpus*. Os Autores “Hjørland, B”, “Smiraglia, RP”, “Dahlberg, I” e “Mai, J-E” são os que apresentam maior diferença nessa relação e estão também entre os cinco autores que mais possuem documentos distintos citados no agrupamento, respectivamente 56, 44, 17 e 19.

Outro dado que merece destaque se refere às datas de publicação dos documentos citados de cada autor, apresentadas na última parte do Quadro 3. O trabalho de Bu, Liu e Huang (2016) utilizou essa informação para qualificar os dados apresentados em uma ACA. No presente estudo, apenas as datas dos documentos mais antigos e mais atuais são apresentadas. Assim, há documentos que foram publicados a partir de 1969 e isso é, certamente, um elemento de destaque. Apenas o autor “Albrechtsen, H” não possui documento citado que tenha sido publicado no século 21. Há quatro autores (Hjørland, B, Smiraglia, RP, Mai, J-E e Guimarães, JAC) que possuem documentos publicados em 2015, quatro (Hodge, G, Thellefsen, MM, Thellefsen, TL e Friedman, A) que possuem documentos publicados em 2013 e dois (López-Huertas, MJ e Buckland, MK) em 2012. Essas datas foram destacadas por fazerem parte do recorte temporal do *corpus* de análise, e mostram a atualidade da pesquisa na área de Organização do Conhecimento, no contexto dos artigos utilizados por essa pesquisa.

Por fim, para atender o terceiro objetivo da pesquisa, foram identificados os documentos citados dos 17 autores que formam o agrupamento analisado. Eles possuem 252 documentos distintos citados no *corpus*, sendo que 230 (91,27%) são citados no agrupamento, totalizando 532 citações.

A Tabela 5 apresenta a distribuição por frequência de citação e a Figura 3 apresenta uma rede dos 32 artigos mais citados, que possuem pelo menos quatro citações.

Tabela 5 - Distribuição das frequências de citações dos documentos citados no agrupamento

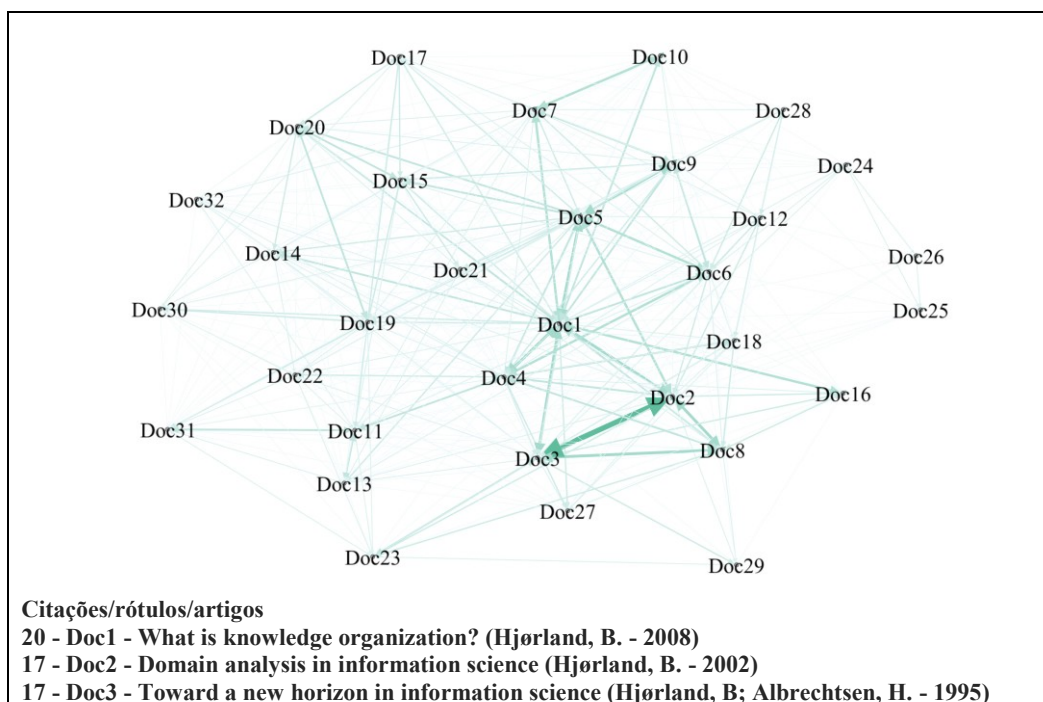
Frequência de citações	Distribuição por documento			Distribuição por citação		
	Número de artigos	%	% acumulado	Citações	%	% acumulado
20 citações	01	00,43	00,43	20	03,76	03,76
17 citações	02	00,87	01,30	34	06,39	10,15
15 citações	02	00,87	02,17	30	05,64	15,79
11 citações	01	00,43	02,61	11	02,07	17,86
10 citações	01	00,43	03,04	10	01,88	19,74
09 citações	03	01,30	04,35	27	05,08	24,81
07 citações	02	00,87	05,22	14	02,63	27,44
06 citações	07	03,04	08,26	42	07,89	35,34
05 citações	05	02,17	10,43	25	04,70	40,04
04 citações	08	03,48	13,91	32	06,02	46,05
03 citações	20	08,70	22,61	60	11,28	57,33
02 citações	49	21,30	43,91	98	18,42	75,75
01 citações	129	56,09	100,00	129	24,25	100,00
Total	230	100,00		532	100,00	

Fonte: Dados da pesquisa.

Os dados da tabela mostram que há uma grande dispersão de documentos citados apenas uma vez nos 69 artigos que formam o agrupamento e que 32 artigos respondem por 46,05% das 532 citações.

Essa informação evidencia a complexidade de lidar com dados de cocitação de autores, pois mesmo os mais citados, por produzirem muito, possuem documentos pouco citados e que ampliam consideravelmente o ruído informacional presente nos dados.

Figura 3 - Rede e distribuição dos principais artigos citados (04 ou mais citações)



- 15 - Doc4 - knowledge organization: a new science? (Dahlberg, I. - 2006)
- 15 - Doc5 - Systems of knowledge organization for digital libraries (Hodge, G. - 2000)
- 11 - Doc6 - Concept theory (Hjørland, B. - 2009)
- 10 - Doc7 - The intellectual foundation of information organization (Svenonius, E. - 2000)
- 09 - Doc8 - Fundamentals of knowledge organization (Hjørland, B. - 2003)
- 09 - Doc9 - Semantic and knowledge organization (Hjørland, B. - 2007)
- 09 - Doc10 - The epistemological foundations of knowledge representation (Svenonius, E. - 2004)
- 07 - Doc11 - Semiotics and indexing: an analysis of the subject indexing process (Mai, J-E - 2001)
- 07 - Doc12 - User-based and cognitive approaches to knowledge organization (Hjørland, B. - 2013)
- 06 - Doc13 - Concept Theory and Semiotics in Knowledge Organization (Friedman, A.; Thellefsen, M. - 2011)
- 06 - Doc14 - Epistemology and the sociocognitive perspective in information science (Hjørland, B. - 2002)
- 06 - Doc15 - Information seeking and subject representation (Hjørland, B. - 1997)
- 06 - Doc16 - Some current research questions in the field of knowledge organization (López-Huertas, M. - 2008)
- 06 - Doc17 - The foundation of the concept of relevance (Hjørland, B. - 2010)
- 06 - Doc18 - The importance of theories of knowledge (Hjørland, B. - 2011)
- 06 - Doc19 - The subject indexing process (Mai, J-E, 2000)
- 05 - Doc20 - Analyzing the role of knowledge organization in scholarly communication (Andersen, J. - 2004)
- 05 - Doc21 - Knowledge organization (Broughton, V.; Hansson, J.; Hjørland, B.; Lopez-Huertas, M. - 2005)
- 05 - Doc22 - Knowledge organization: its scope and possibilities (Dahlberg, I. - 1993)
- 05 - Doc23 - Pragmatic semiotics and knowledge organization (Thellefsen, M.; Thellefsen, T. - 2004)
- 05 - Doc24 - The modernity of classification (Mai, J-E, 2011)
- 04 - Doc25 - A substantive theory of classification for information retrieval (Hjørland, B. - 2005)
- 04 - Doc26 - Classification in a social world: bias and trust (Mai, J-E, 2010)
- 04 - Doc27 - Discourse Analysis as a Research Method in Library and Information Science (Frohmann, B. - 1994)
- 04 - Doc28 - Facet analysis: The logical approach to knowledge organization (Hjørland, B. - 2013)
- 04 - Doc29 - Information as thing (Buckland, M. - 1991)
- 04 - Doc30 - Rules of indexing: a critique of mentalism in information retrieval theory (Frohmann, B. - 1990)
- 04 - Doc31 - Semiotic knowledge organization: theory and method development (Thellefsen, T. - 2002)
- 04 - Doc32 - The progress of theory in knowledge organization (Smiraglia, R., - 2002)

Fonte: Dados da pesquisa.

A Figura 3 apresenta um núcleo de oito documentos que se relacionam com mais força, com destaque para os documentos “Doc2” e “Doc3”, como se pode ver na lista que complementa a figura. O autor “Hjørland, B” é responsável por seis desses documentos, sendo um deles em coautoria com “Albrechtsen, H”, os outros são de “Dahlberg, I” e “Hodge, G”, sendo que o último possui apenas dois documentos citados no agrupamento, como já salientado.

Desses oito documentos, três são sobre o que é a Organização do conhecimento e seus fundamentos: *What is knowledge organization?* (Hjørland, B. - 2008); *Knowledge organization: a new science?* (Dahlberg, I. - 2006); e *Fundamentals of knowledge organization* (Hjørland, B. - 2003). Outros documentos também abordam a área com esse enfoque sobre fundamentos e questões epistemológicas, além de relacionar a Organização do conhecimento com a Ciência da Informação e outras disciplinas, como a Semiótica.

Outros temas decorrentes dos documentos mais citados são a classificação, a indexação e a teoria do conceito. Essa avaliação inicial é feita pelos títulos e, mesmo parecendo claro, não é simples rotular um agrupamento (CHEN; IBEKWE-SANJUAN; HOU, 2010) que se mostra muito complexo, como evidenciam os dados. Um caminho possível é verificar e entender como se dá o uso desses documentos nos artigos citantes, enquanto outro é verificar a citação ou cocitação na seção do documento citante por evidenciar um rótulo ou um foco de uso.

Outra observação é que, dentre os documentos mais citados, há quatro que também são documentos citantes do corpus, apresentados a seguir por ordem de citação: (i) *User-based and cognitive approaches to knowledge organization* (Hjørland, B. – 2013); (ii) *Concept Theory and Semiotics in Knowledge Organization* (Friedman, A.; Thellefsen, M. - 2011); (iii) *The importance of theories of knowledge* (Hjørland, B. - 2011); e (iv) *Facet analysis: The logical approach to knowledge organization* (Hjørland, B. - 2013).

Vale salientar que os autores “Guimarães, JAC” e “Bates, MJ” não possuem documentos no recorte de análise da Figura 1 e que os autores “Albrechtsen, H” e “Hansson, J” possuem apenas um documento citado na análise e publicado em coautoria, onde ambos não estão na primeira posição das referências. Esses dados evidenciam a importância de definir sobre o uso de todos os autores das referências e os ganhos ou perdas informacionais que acarretam.

Por fim, a unidade de contagem merece uma discussão com base nos dados. Contar cocitação por autores simplifica e agrega mais informações (230 documentos representados por 17 autores), pois há um destaque na produção, mas também pode ser um problema para rotular os agrupamentos posteriormente. O autor “Smiraglia, RP”, por exemplo, possui 34 documentos citados no agrupamento, com apenas uma citação dos 44 citados na análise, e apenas um está entre os mais citados listados na Figura 3 (*The progress of theory in knowledge organization, 2002*). Em um estudo de cocitação de documentos, esses dados poderiam ser desconsiderados, pois a dispersão é maior utilizando essa unidade de contagem. O ideal, como sugerem Chen, Ibekwe-Sanjuan e Hou

(2010), é combinar as técnicas de cocitação de autores e cocitação de documentos. Além de “Smiraglia, RP”, os autores “Dahlberg, I” (11), “Guimarães, JAC” (21) e “Hjørland, B” (16) são outros destaques, com vários documentos citados uma única vez no agrupamento analisado.

Os resultados mostram que analisar agrupamentos pode ser uma abordagem válida em ACA, mesmo em um nível descritivo, e que parece fundamental dimensionar o agrupamento em relação aos dados gerais do *corpus* de análise. É necessário comparar dois ou mais agrupamentos utilizando a mesma abordagem para enriquecer esse entendimento.

5 Considerações finais

Validar e interpretar agrupamentos em estudos de cocitação não é uma tarefa simples. Qualquer escolha na execução de um estudo dessa natureza pode implicar em elementos que deixam de ser analisados ou em muita informação para medir e condensar, e isso já se evidencia na escolha da unidade de contagem, principalmente as mais comuns: autores e documentos.

Esse trabalho buscou, a partir de três objetivos, apresentar uma perspectiva nesse sentido, especificamente para agrupamentos de ACA.

Os dados dos artigos citantes mostram a extensão do agrupamento e destacam os artigos que mais cocitam os autores, além de indicar as fontes desses documentos, que, no caso, são os periódicos e os autores. No presente trabalho, é possível perceber que essas informações validam o agrupamento e mostram um pouco de homogeneidade e coerência do *corpus* do estudo, ainda que possam haver questionamentos sobre a sua formação e tamanho, aspectos relevantes e limitantes em qualquer pesquisa sobre citação e cocitação.

Os indicadores adicionais relacionados aos autores contextualizam a participação de cada um deles no agrupamento e no *corpus* de maneira geral, principalmente pela relação entre citação, citante e documentos citados. A identificação das relações de cocitação na proximidade por parágrafo pode validar as cocitações recuperadas na lista de referências, pois, ainda que a análise não seja

feita no nível do conteúdo, é bastante improvável que dois autores cocitados no mesmo parágrafo não tenham alguma relação temática.

Nos dados dessa pesquisa, os indicadores adicionais serviram à descrição dos padrões estruturais de caracterização do agrupamento, esclarecendo a importância e a participação de cada autor, e a cocitação, na proximidade por parágrafo, foi alta, validando os dados das referências. Sugere-se que essa abordagem seja utilizada na descrição mais detalhada dos resultados de ACA em qualquer área e para todos os agrupamentos, pois amplia a autonomia informacional dos estudos.

Apresentar os documentos dos autores amplia o nível informacional e evidencia a natureza temática do agrupamento. Nos dados dessa pesquisa, a rede de cocitação dos documentos mostrou uma relação omitida pela ACA, qual seja, a ligação de documentos do mesmo autor, além de um documento muito citado de outro autor que parecia sem muita relevância no agrupamento. Combinar as técnicas de cocitação, ainda que como uma forma de interpretar um agrupamento, mostrou-se válido.

Aplicar essa abordagem para comparar todos os agrupamentos de uma ACA é uma sugestão para futuros trabalhos, assim como incluir uma rede de acoplamento bibliográfico na fase de levantamento dos dados estruturais de formação do agrupamento (documentos citantes).

Espera-se que os resultados desse trabalho possam colaborar com o avanço da pesquisa sobre cocitação de autores e de documentos no Brasil, além da aplicação da técnica, independentemente da área, pois a citação é um fenômeno que caracteriza a ciência de forma geral, sendo um dos mecanismos que mais aproximam as distintas áreas na forma de ser utilizado.

Financiamento

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Referências

- ALVARENGA, L. Bibliometria e arqueologia do saber de Michel Foucault – traços de identidade teórico-metodológica. **Ciência da Informação**, Brasília, v. 27, n. 3, 1998.
- BECKER, J. L. **Estatística básica**: transformando dados em informação. Porto Alegre: Bookman, 2015.
- BU, Y. *et al.* MFTACA: an author co-citation analysis method combined with metadata in full text. In: INTERNATIONAL CONFERENCE ON SCIENTOMETRICS AND INFORMETRICS, 16., 2017, Wuhan, China. **Proceedings** [...] Wuhan: International Society of Scientometrics and Informetrics – ISSI, 2017.
- BU, Y.; LIU, T-Y; HUANG, W-B. MACA: a modified author co-citation analysis method combined with general descriptive metadata of citations. **Scientometrics**, Dordrecht, v. 108, n. 1, p. 143-166, 2016.
- CARVALHO, R. A. *et al.* A relação entre referências e menções: estudo exploratório em artigos na base de dados BRAPCI. **Prisma.Com**, Porto, n. 36, p. 99-115, 2018.
- CARVALHO, R. A.; CAREGNATO, S. E. Análise de cocitação de autores – ACA: estudo exploratório comparando proximidade nas referências, seção do artigo e parágrafo. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. **Anais** [...] Marília: ANCIB, 2017.
- CARVALHO, R. A.; CAREGNATO, S. E. Primeiro vs todos os demais autores citados: estudo empírico em artigos na base BRAPCI. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016, São Paulo. **Anais** [...] São Paulo: USP, 2016.
- CHEN, C.; IBEKWE-SANJUAN, F.; HOU, J. The structure and dynamics of co-citation clusters: a multiple-perspective co-citation analysis. **Journal of the American Society for Information Science and Technology**, New Jersey, v. 61, n. 7, p. 1386-1409, 2010.
- DING, Y. *et al.* The distribution of references across texts: some implications for citation analysis. **Journal of Informetrics**, Amsterdam, n. 7, p. 583-592, 2013.
- EOM, S. B. **Author cocitation analysis**: quantitative methods for mapping the intellectual structure of an academic discipline. Hershey: Information Science Reference, 2009.
- FIELD, A. **Descobrimos a estatística usando o SPSS**. 2. ed. Porto Alegre: Artmed, 2009.

GLÄNZEL, W. **Bibliometrics as a research field**: A course on theory and application of bibliometric indicators. [S. l.]: Course Handouts, 2003.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. de. Estudos de análise de cocitação de autores: uma abordagem teórico-metodológica para a compreensão de um domínio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais [...]** Florianópolis: ANCIB, 2013.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. de. Indicadores de proximidades em análise de cocitação de autores: um estudo comparativo entre coeficiente de Correlação de Pearson e Cosseno de Salton. **Informação & Sociedade: Estudos**, João Pessoa, v. 25, n. 2, p. 105-116, maio/ago. 2015.

HJØRLAND, B. Citation analysis: A social and dynamic approach to knowledge organization. **Information Processing and Management**, Kidlington, n. 49, p. 1313–1325, 2013.

JEONG, Y.; SONG, M.; DING, Y. Content-based author co-citation analysis. **Journal of Informetrics**, Amsterdam, n. 8, p. 197-211, 2014.

LEYDESDORFF, L.; VAUGHAN, L. Co-occurrence matrices and their applications in Information Science: extending ACA to the web environment. **Journal of the American Society for Information Science and Technology**, New Jersey, v. 57, n. 12, p. 1616–1628, 2006.

LIMA, L. S. **Produção científica em Organização do conhecimento**: uma análise de domínio via cocitações de autores. Marília. 2015. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista. 2015.

LIU, S.; CHEN, C. The proximity of co-citation. **Scientometrics**, Dordrecht, n. 91, n. 2, p.495-511, 2012.

MACHADO, R. N. **Estrutura intelectual da literatura científica do Brasil e outros países dos BRICS**: uma análise de cocitação de periódicos na área de célula-tronco. 2015. Tese (Doutorado em Ciência da Informação) - Escola de Comunicação, Universidade Federal do Rio de Janeiro. Programa de Pós-graduação em Ciência da Informação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 2015.

McCAIN, K. Mapping author intellectual space: a technical overview. **Journal of the American Society for Information Science**, New Jersey, v. 41, n. 6, p. 433-443, 1990.

MUÑOZ-MUÑOZ, A. M.; MIRÓN-VALDIVIESO, M. D. Analysis of collaboration and co-citation networks between researchers studying violence involving women. **Information Research**, Sheffield, v. 22, n. 2, 2017.

PERSSON, O. All author citations versus first author citations. **Scientometrics**, Dordrecht, v. 50, n. 2, p. 339-344. 2001.

PIOVEZAN, L. B.; FUJITA, M. S. L. Análise de cocitação de autores: uma aplicação em estudos de indexação. **Em Questão**, Porto Alegre, v. 21, n. 1, p. 110-129, 2015.

ROUSSEAU, R.; ZUCCALA, A. A classification of author co-citations: definitions and search strategies. **Journal of the American Society for Information Science and Technology**, New Jersey, v. 55, n. 6, p. 513-529, 2004.

SCHNEIDER, J. W.; LARSEN, B.; INGWERSEN, P. A comparative study of first and all author co-citation counting, and two different matrix generation approaches applied for author co-citation analyses. **Scientometrics**, Dordrecht, v. 80, n. 1, p. 105-132, 2009.

SILVEIRA, M. A. A.; CAREGNATO, S. E. Demarcações epistemológicas dos estudos de citação: o fenômeno da citação. **Informação & Sociedade**, João Pessoa, v. 27, n. 3, p. 145-154, 2017.

SMALL, H. Co-citation in the scientific literature: a new measure of the relationship between two documents. **Journal of the American Society for Information Science**, New Jersey, v. 24, n. 4, p. 265-269, 1973.

TRUJILLO, C. M.; LONG, T. M. Document co-citation analysis to enhance transdisciplinary research. **Science Advances**, Washington, D.C., v. 4, n. 1, 2018.

Interpretation and validation of groupings in author cocitation analysis: exploratory and methodological study

Abstract: In the context of the interpretation and validation of an authors cocitation analysis, the following objectives are formulated: (i) to compare data from the citing articles of the corpus with data characterizing the citing and the cited authors of an ACA cluster; (ii) to propose additional indicators related to the authors that form the grouping; and (iii) to visualize the network of cocitation documents of the authors that form the cluster. The work is exploratory and methodological and the data were extracted from 151 articles on Knowledge Organization, from 2011 to 2015, retrieved in the Web Of Science database. A symmetric matrix with 50 positions (authors with nine or more citing articles) generated five clusters. The main cluster (17 authors) is the object of this analysis. The authors are cocited in 69 articles. From the 131 authors' relations in references, 84 occur also in the paragraph. There are data comparing citations, citing and cited documents for each of the 17 authors. The cluster has 230 articles cited. It is concluded that the proximity of the cocitation in the paragraph validates

the cluster and that the other results broaden the view of the formation and the characterization of cocitation relations between authors. The results demonstrate that the clusters in an ACA can be analyzed as autonomous units and that the description of the data that support the formation of the cluster (citing articles and authors), which characterize the participating authors and the cocited document network, collaborates in the validation and interpretation relations, because it presents objective data. It is hoped that this exploratory approach can be applied in other ACA, regardless of the area.

Keywords: Cocitation. Author co-citation analysis – ACA. Document co-citation analysis (DCA). Methodological study.

Recebido: 13/04/2018

Aceito: 11/08/2018



¹ ZHAO, D.; STROTMANN, A. Evolution of Research Activities and Intellectual Influences in Information Science 1996–2005: Introducing Author Bibliographic-Coupling Analysis. **Journal of the American Society for Information Science and Technology**, v. 59, n. 13, p. 2070–2086, 2008. Disponível em: <<http://doi.org/10.1002/asi.20910>>. Acesso em: 18 fev. 2018.

² Referência da nota 1.

³ O texto “MFTACA: an author co-citation analysis method combined with metadata in full text” foi apresentado no “ISSI Conference” em 2017. Não há *link* nem para os metadados do texto nos mês de fevereiro de 2018.

⁴ Outros periódicos: *Aslib Proceedings*, *Evidence Based Library and Information Practice*, *Investigación Bibliotecológica*, *Information Technology and Libraries*, *Journal of Librarianship and Information Science*, *Library Quarterly*, *Online Information Review*, *Program: electronic library and information systems* e *Revista Española de Documentación Científica*.