

XVII Encontro Nacional de Pesquisa em Ciência da Informação

Declaração de Direito Autoral

Autores que submetem a esta conferência concordam com os seguintes termos:

- a) Autores mantêm os direitos autorais sobre o trabalho, permitindo à conferência colocá-lo sob uma licença Licença Creative Commons Attribution, que permite livremente a outros acessar, usar e compartilhar o trabalho com o crédito de autoria e apresentação inicial nesta conferência.
- b) Autores podem abrir mão dos termos da licença CC e definir contratos adicionais para a distribuição não-exclusiva e subsequente publicação deste trabalho (ex.: publicar uma versão atualizada em um periódico, disponibilizar em repositório institucional, ou publicá-lo em livro), com o crédito de autoria e apresentação inicial nesta conferência.
- c) Além disso, autores são incentivados a publicar e compartilhar seus trabalhos online (ex.: em repositório institucional ou em sua página pessoal) a qualquer momento antes e depois da conferência.

Fonte:

<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/viewFile/4163/2534>

REFERÊNCIA:

SAMPAIO, Ricardo Barros; VILAN FILHO, Jayme Leiro; FERNANDES, Jorge Henrique Cabral. Produção brasileira em Ciência da Informação: uma análise baseada em redes complexas. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador. **Anais...** Salvador: ANCIB, 2016. Disponível em: <<http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/viewFile/4163/2534>>. Acesso em 22 dez. 2016.



XVII Encontro Nacional de Pesquisa em Ciência da Informação (XVII ENANCIB)

GT 07 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

**PRODUÇÃO BRASILEIRA EM CIENCIA DA INFORMAÇÃO: UMA ANÁLISE
BASEADA EM REDES COMPLEXAS**

***BRAZILIAN PRODUCTION ON INFORMATION SCIENCE: AN ANALYSIS BASED ON
COMPLEX NETWORKS***

Ricardo Barros Sampaio¹, Jayme Leiro Vilan Filho², Jorge Henrique Cabral Fernandes³

Modalidade da apresentação: Comunicação Oral

Resumo: A pesquisa realizada aplica e aprimora métodos relacionados à cientometria com foco na análise de redes complexas e análise lexical de resumos. Utiliza como principal base empírica a base de dados ABCDM (Arquivologia, Biblioteconomia, Ciência da Informação, Documentação e Museologia), que contém aproximadamente dez mil referências completas de publicações em revistas científicas brasileiras e portuguesas, além das publicações presentes no principal evento brasileiro de ciência da informação, o ENANCIB. Essa abordagem permite uma avaliação diferenciada da produção científica brasileira na área de ciência da informação e utilizou como referência as seis revistas brasileiras com maior número de citações no recorte temporal dos últimos 16 anos (2000-2015). Os métodos são baseados na análise de redes sociais, análise de redes complexas e na análise semântica de conteúdo. Dentre os elementos de pesquisa desenvolvidos a coautoria científica é o elemento central do projeto, com foco nos grupos de colaboração científica. Leva em consideração não apenas a relação entre os pesquisadores, mas também as relações entre as instituições de ensino e pesquisa e as áreas do conhecimento às quais se vinculam. Ademais, foi utilizado um método para classificação textual, de modo a validar a formação dos grupos de pesquisa ou colégios invisíveis e os textos relacionados às suas áreas do conhecimento. Embora a avaliação da produção científica seja importante para o entendimento do desenvolvimento científico, tecnológico e de inovação de um país, o Brasil ainda não dispõe de muitas bases científicas e metodologias próprias para avaliar a sua própria produção. Como estamos lidando com um número elevado de comunicações de viés científico, e considerando a presença parcial da produção científica nacional nas bases de dados estrangeiras, estimada em cerca de 20% de representatividade, se

¹ Universidade de Brasília - UnB

² Professor, UnB - Faculdade de Ciência da Informação.

³ Professor, UnB - Faculdade de Ciência da Informação.

faz necessário o desenvolvimento de tecnologias nacionais que possam recuperar, tratar e disponibilizar os dados da produção brasileira de forma eficiente.

Palavras-chave: Coautoria. Análise de Redes. Análise de Conteúdo. Cientometria.

***Abstract:** The research applies and refines methods related to scientometrics and focusing on the analysis of complex networks and lexical analysis of abstracts. It used as the main empirical base the ABCDM database (Archival Science, Library Science, Information Science, Documentation and Museology) which contains approximately ten thousand full references of publications in Brazilian and Portuguese scientific journals, in addition to publications presented in the main Brazilian event in information science, the ENANCIB. This approach allowed a differentiated evaluation of the scientific production in information science and adopted as a reference the six Brazilian journals with the highest number of citations in the time frame of the last 16 years (2000-2015). The methods are based on social network analysis, analysis of complex networks and semantic analysis of content. Among the elements of research developed the scientific co-authorship is a central element of the project. It focused on scientific collaboration of groups and takes into account not only the relationship between researchers, but also the relationship between their affiliated research institutions and areas of knowledge. In addition, employed a method for text classification in order to validate the formation of research groups or invisible colleges and texts related to their areas of knowledge. Although the evaluation of scientific production is important to the understanding of the science, technology and innovation of a country, Brazil still does not have many specific scientific bases and methodologies to assess their own production. As we are dealing with a large number of scientific communications, and considering the partial presence of national scientific production in foreign databases - estimated at about 20% of representation - the development of national technologies is necessary to efficiently recover, treat and provide data about the Brazilian production.*

Keywords: Co-authorship. Network Analysis. Content analysis. Scientometrics.

1 INTRODUÇÃO

A Análise da produção científica é realizada de forma extensiva pela grande maioria dos países desenvolvidos, como forma de avaliar o seu aprimoramento científico e inovação. Essa é uma etapa estratégica na avaliação e aprimoramento de políticas públicas de um país. A forma de se avaliar essa produção difere da estratégia ou método utilizado, mas na sua grande maioria, esses métodos se valem de dados disponíveis em bases de dados de publicações científicas, patentes e outros. Como essas bases são, na sua maioria, localizadas e financiadas em países como Estados Unidos ou do Oeste Europeu, muitas das publicações brasileiras acabam não sendo avaliadas, por não serem indexadas nessas bases.

O fato de que a maior parte da produção científica brasileira não está catalogada em bases internacionais, limita a capacidade de avaliação do real potencial científico, tecnológico e

de inovação do país. Dessa maneira, a pesquisa demonstra o desenvolvimento de um método cientométrico com base na utilização de fontes de informação científica inteiramente brasileiras, em uma base que cobre toda a produção brasileira em determinado recorte, que é o caso da base de dados ABCDM, que indexa os artigos das áreas de Arquivologia, Biblioteconomia, Ciência da Informação, Documentação e Museologia.

A ABCDM é uma base de dados que possui aproximadamente sete mil referências completas de artigos, de mais de 30 títulos de periódicos científicos das áreas de Informação, publicados no Brasil e Portugal, no período compreendido entre 1963 e 2015. Além disso, a ABCDM contém registros sobre cerca de três mil trabalhos (comunicações e pôsteres) do principal evento científico de âmbito nacional: o Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), cobrindo atualmente todas as edições, até 2014.

As características da ABCDM possibilitam que ela seja não só uma fonte de dados para o monitoramento de boa parte da produção científica das áreas de informação no Brasil, como também para a produção de indicadores de colaboração científica em vários níveis (autores, programas de PG, instituições, estados, regiões, países, temas e áreas do conhecimento). Nesse projeto, esses indicadores foram produzidos por meio da análise exploratória de redes sociais com diferentes níveis de complexidade. A base ABCDM encontra-se atualmente em processo de atualização, visando incluir artigos e trabalhos publicados em 2014 e 2015.

Para investigar a ciência, considerada como um processo social, onde as ações e procedimentos de seus atores dependem do contexto em que estão inseridos, a cientometria utiliza indicadores para medir as ações sistemáticas que têm relação com a produção, difusão, transferência e aplicação dos conhecimentos científicos e tecnológicos, que podem ser chamados de indicadores cientométricos (VANTI, 2012).

Na década de 1950, Eugene Garfield desenvolveu instrumentos de informação que permitiram aos pesquisadores agilizarem o processo de pesquisa, avaliarem o impacto do seu trabalho, identificarem as tendências científicas e traçarem a história do pensamento científico moderno (YANCEY, 2005).

Para Spinak (SPINAK, 1995), a cientometria investiga o crescimento quantitativo da ciência, o desenvolvimento de disciplinas e subdisciplinas, a relação entre ciência e tecnologia, a antiguidade dos padrões científicos, a estrutura de comunicação entre cientistas, a produtividade e criatividade dos pesquisadores e as relações entre desenvolvimento científico e crescimento econômico.

Segundo Vanti (2012), a cientometria utiliza técnicas matemáticas e estatísticas para descobrir as características da ciência, gerando dados e informações úteis para os estudos

conduzidos na sociologia da ciência.

De forma geral, os indicadores cientométricos representam uma medida ou um índice que permite avaliar ou acompanhar o desempenho de um fenômeno, da sua natureza, do seu estado e evolução. Isso incentiva a constante revisão e aperfeiçoamento dos indicadores que são utilizados para investigar e avaliar o desempenho e impacto da atividade científica, e também o fortalecimento dos sistemas de informação que processam a produção científica, tecnológica e inovadora de qualquer país.

Dentre os estudos cientométricos, a colaboração científica é um fenômeno marcante da pesquisa contemporânea. Ao colaborarem, os pesquisadores podem estabelecer redes de comunicação acadêmica, compartilhar ideias, recursos e informações, gerar e disponibilizar novos conhecimentos e, em última instância, criar inovações, reduzindo o custo e aumentando a produtividade da pesquisa (ROYAL SOCIETY, 2011).

O avanço da tecnologia tem permitido uma análise mais ampla dessas colaborações, especialmente em grandes bases de dados. A análise de redes sociais (HANNEMAN; RIDDLE, 2005; WASSERMAN; FAUST, 1994) com dezenas ou centenas de milhares de atores e relacionamentos era incomum e restrita a pequenos grupos de pesquisadores. Atualmente, com as ferramentas disponíveis, as atividades de pesquisa e os métodos propostos para esse estudo, várias oportunidades e mecanismos de gestão podem ser desenvolvidos para subsidiar o processo decisório, inclusive a governança de políticas públicas.

O estudo das redes, conhecido sob os vieses de análise de redes sociais e da análise de redes complexas, vem adquirindo dimensões importantes para a compreensão da dinâmica da informação mediada no dia a dia, em grupos de indivíduos. Muitos autores ressaltam que as relações sociais podem ser consideradas canais de mediação da informação, serviço e bens entre pessoas e organizações. A análise de redes permite colher os dados acerca dessas relações em que o fenômeno ocorre em uma determinada rede.

Análise de redes possibilita avaliar tanto os aspectos descritivos dos relacionamentos quanto as análises estatísticas causais de tais fenômenos (HANNEMAN; RIDDLE, 2005; WASSERMAN; FAUST, 1994). Escolheu-se, portanto, utilizar os métodos de análise de redes complexas, por incluir informações sobre o relacionamento entre os colaboradores de determinadas áreas de conhecimento.

Com o objetivo de trazer uma nova abordagem que valide as análises cientométricas e de redes é desenvolvido também um estudo semântico e lexical das palavras chaves recuperadas dos artigos. A análise léxica tem como foco avaliar a frequência e proximidade de palavras e assuntos que trazem alguma similaridade entre eles. A classificação apoia e por vezes valida a

criação dos grupos ou colégios invisíveis que, de certa maneira, tendem a discutir assuntos similares entre os membros do seu grupo. Entender esses discursos e buscar inovações que possam vir a surgir em determinados campos são possibilidades de uso dessa técnica no entendimento das publicações. Além disso, a análise de assuntos similares mais discutidos por diferentes grupos pode auxiliar na avaliação das propostas de pesquisa e potencial tema para colaboração.

2 METODOLOGIA

A metodologia a seguir apresentada define o recorte de pesquisa utilizado para a análise dentro da base de dados ABCDM. Em seguida é apresentado um resumo sobre a metodologia de análise de redes no contexto da colaboração científica e por fim é apresentado um resumo sobre o método e ferramenta empregados para a análise léxica.

2.1 RECUPERAÇÃO DOS DADOS

Há uma necessidade de desenvolvimento de novos métodos de avaliação da produção brasileira tendo em vista que grande parte da produção nacional não é indexada em bases de dados internacionais. Apesar das revistas escolhidas no trabalho serem indexadas na base Scielo e conseqüentemente na plataforma Web of Science em decorrência de um recente acordo de parceria, o método desenvolvido utilizou uma base nacional, a ABCDM. Esse método pode, portanto, ser replicado a demais revistas da ciência da informação, quando se fizer necessário, uma vez que a base tem como foco os periódicos brasileiros das áreas de informação.

O recorte da pesquisa levou em consideração os seis periódicos das áreas de informação com maior número de citações, além dos registros do congresso ENANCIB. O levantamento das citações em periódico levou em consideração o trabalho apresentado por Filho e Mueller (2014) que indicou os seguintes periódicos: Ciência da Informação com 34,58% das citações; Datagramazero com 14,54%, Perspectivas em Ciência da Informação com 13,85% ; Informação & Sociedade com 6,55%, Transinformação com 6,11% e Encontros Bibli com 4,10%. Os referidos periódicos possuem classificação Qualis extrato A1 (3 periódicos) e B1 (3 periódicos), na área de conhecimento da Ciências Sociais Aplicadas I.

O período de tempo avaliado foi de 16 anos (2000-2015). Nesse recorte foram encontrados um total de 2.307 publicações com 2.443 autores e 623 organizações. No caso dos autores foi realizado um trabalho de limpeza dos dados com a busca de sinonímias (mesmo autor com diferentes nomes) com o objetivo de adequar a contagem da produção. No caso das

organizações, fez-se a agregação das informações em nível organizacional mais elevado, uma vez que não foi possível identificar todos os departamentos e laboratórios como referência.

Foram extraídos para análise: os autores, as organizações indicadas pelos autores, as palavras chaves, os resumos incluindo as datas e os periódicos de publicação. Os resultados estatísticos dessa busca são apresentados na seção de resultados.

2.2 ANÁLISE DE REDES COMPLEXAS

Para a análise de redes complexas foi utilizado como referência o trabalho de Sampaio et al (2015), que utiliza o modelo exploratório de Nooy, Mrvar e Batagelj (2011), e na qual são definidas quatro etapas cíclicas, sendo elas: a definição da rede; o tratamento de dados da rede; a determinação de características estruturais; e a inspeção visual.

Ainda segundo Sampaio, a análise de redes complexas com foco nas colaborações e parcerias científicas e tecnológicas proporciona uma visão dos padrões de cooperação entre indivíduos e organizações (MELIN; PERSSON, 1996; GLÄNZEL, 2002; NEWMAN, 2004). A coautoria de um documento é um registro oficial do relacionamento entre dois ou mais autores ou organizações (GLÄNZEL; SCHUBERT, 2004), e apesar do debate antigo a respeito do seu significado e interpretação (KATZ; MARTIN, 1997; LAUDEL, 2002), a análise de coautoria tem sido amplamente utilizada para entender e avaliar os padrões de colaboração científica. A colaboração científica pode, então, ser definida como a interação que ocorre em um contexto social entre dois ou mais pesquisadores, que facilita o compartilhamento de significado e a realização de tarefas com relação a um objetivo mutuamente compartilhado (SONNENWALD, 2007). Em redes de coautoria, os vértices representam os autores ou as organizações às quais esses estão vinculados, e dois ou mais autores estão conectados se eles compartilham a autoria de uma publicação (NEWMAN, 2004). Em uma perspectiva interorganizacional, a cooperação pode ser definida em termos de publicações que tenham autoria compartilhada por indivíduos afiliados a duas ou mais organizações (NAGPAUL, 2002).

2.3 ANÁLISE DE CONTEÚDO

Para a realização da análise do conteúdo foi utilizado o software IRAMUTEQ (Interface R para Análise Multidimensional de Textos e Questionários), com base em uma metodologia desenvolvida por Max Reinert na década de 1990 no laboratório de psicologia social da *l'Ecole des hautes études en sciences sociales*. O IRAMUTEQ é um software desenvolvido na França e possui um manual⁴ em português. Este software vem sendo amplamente utilizado em

⁴ <http://www.iramuteq.org/documentation>

sociologia, psicologia e ciências políticas e da saúde para a análise quantitativa de dados textuais ou textuais estatísticos (GUARNACCIA et al., 2015; RODRIGUES et al., 2015).

O software IRAMUTEQ tem a capacidade de analisar textos em diferentes línguas com base nos dicionários que possui instalados. Uma vez feita a entrada de dados no software este realiza uma contagem das palavras e a sua classificação morfológica. Como o software não possui a capacidade de interpretar palavras compostas, no caso das palavras chaves se fez necessário a junção destas trocando os espaços vazios entre as palavras por *underscore*. Com o texto analisado estatisticamente o usuário pode realizar em seguida análises com base nas classificações do próprio texto, como ano ou revistas por exemplo, ou utilizar o algoritmo de classificação desenvolvido por Reinert, também conhecido como Classificação Hierárquica Descendente. Os resultados das análises são apresentados em formato visual utilizando a Análise Fatorial do Componente (AFC) para organizar sua distribuição espacial em um plano cartesiano. As palavras então são agrupadas utilizando-se da classificação desejada e apresentadas em cores diferentes para cada grupo e dispostas no plano cartesiano utilizando a AFC.

Outros tipos de análise utilizando o software podem ser feitos como coocorrência de termos e nuvem de palavras. No entanto essas estratégias de análise não foram utilizadas nesse trabalho.

Foram analisadas portanto as palavras chaves dos artigos encontrados e apresentados em formato gráfico para análise. Foi possível demonstrar algumas contribuições da análise de dados lexicais na extração de informações relevantes em documentos de textos, para as pesquisas em ciência da informação.

3 RESULTADOS ENCONTRADOS

Esta seção é composta por três partes, sendo a primeira focada na apresentação dos dados estatísticos e bibliométricos obtidos pela busca realizada dentro do recorte temporal e científico; a segunda parte apresenta resultados e explicações pontuais sobre as redes de colaboração encontradas; e por último são apresentados os temas centrais e redes de palavras quanto à análise de conteúdo.

3.1 DADOS BIBLIOMÉTRICOS

Apresenta-se a seguir o número de publicações por revista, as publicações por ano, as organizações com maior número de publicações, os autores com maior número de publicações e as palavras chaves mais comuns. No total foram encontradas 2.307 publicações divididas em 16

anos e em seis revistas. Um total de 2.442 autores estão filiados a 623 organizações. A média de autores por artigo foi de 1,96.

O número de artigos por revista é apresentado na Tabela 1.

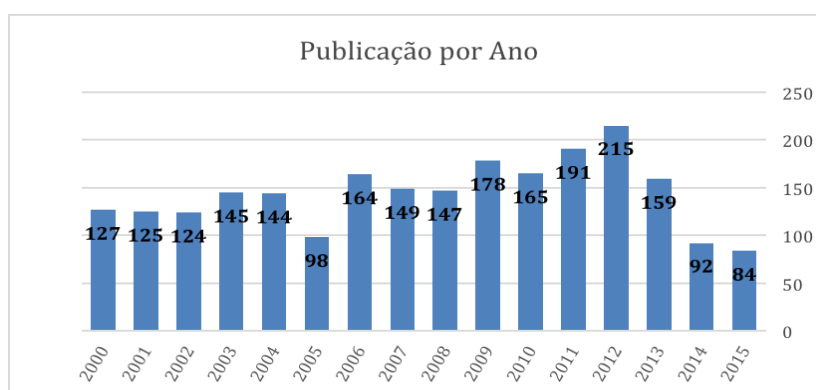
Tabela 1 – Número de publicações em seis revistas na ABCDM (2000-2015).

Revista	Publicações	Qualis
Perspectivas em CI	530	A1
DataGramZero	431	B1
Ciência da Informação	395	B1
Informação e Sociedade	379	A1
Encontros Bibli	305	B1
Transinformação	267	A1

Fonte: autores

O número total de publicações por ano é apresentado no Gráfico 1.

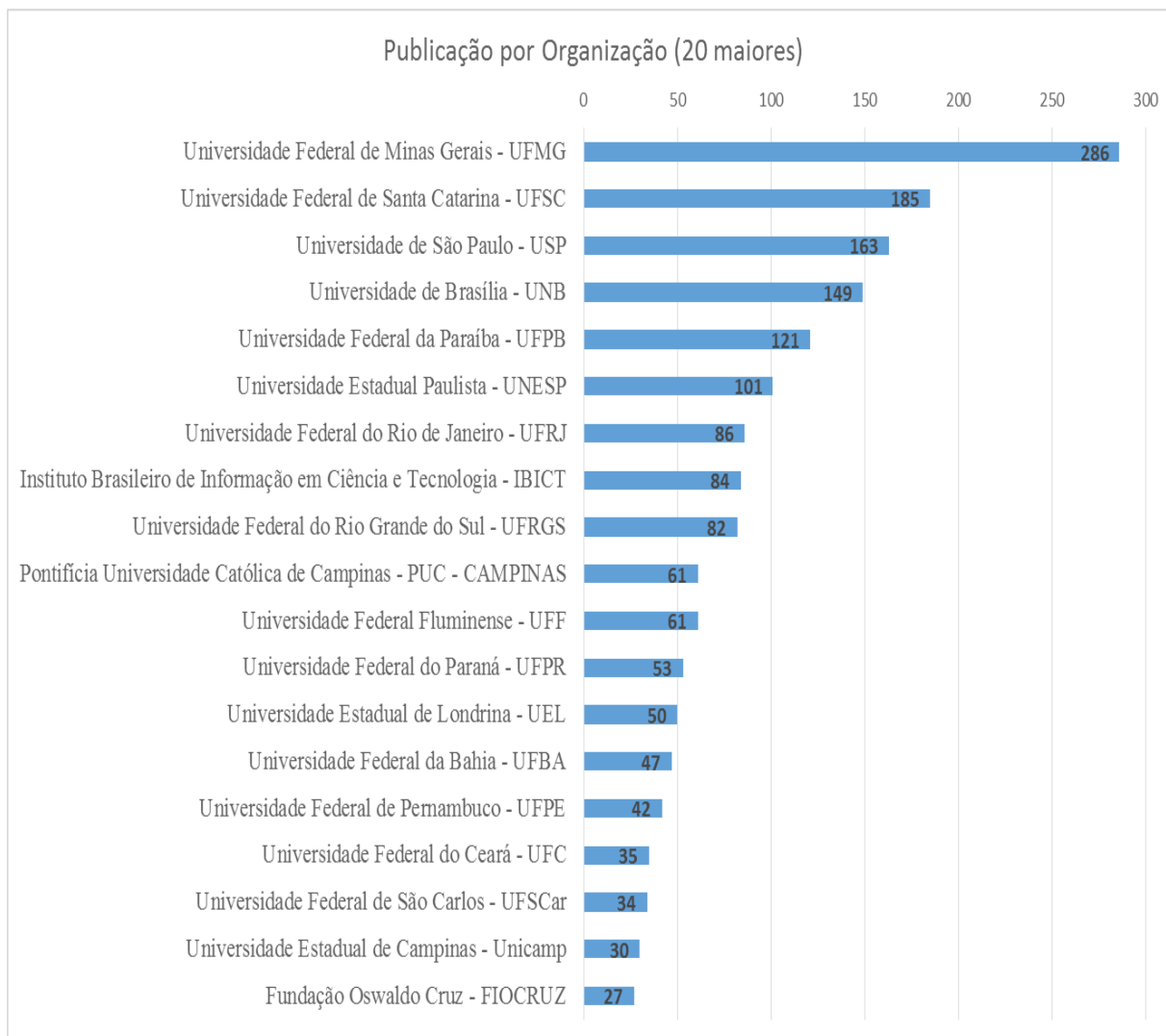
Gráfico 1 – Número de publicações anuais em revistas na ABCDM (2000-2015).



Fonte: autores – Obs: O período 2014-2015 inclui apenas os periódicos “Perspectivas em Ciência da Informação” e “Informação e Sociedade”. Os demais periódicos ainda não tinham sido indexados pela base ABCDM quando da realização deste trabalho.

As publicações por ano não aparentam ter tido grandes mudanças, uma vez que as revistas avaliadas possuem um histórico de publicação até o ano de início dos estudos. Ressaltasse que nos anos 2014-2015 a produção foi menor em decorrência da fonte de dados não estar completa, ou seja, nem todos os artigos destes anos tinham sido incluídos na base ABCDM.

Gráfico 2 – Quantidade de publicações por organização de filiação dos autores registradas na ABCDM (2000-2015).

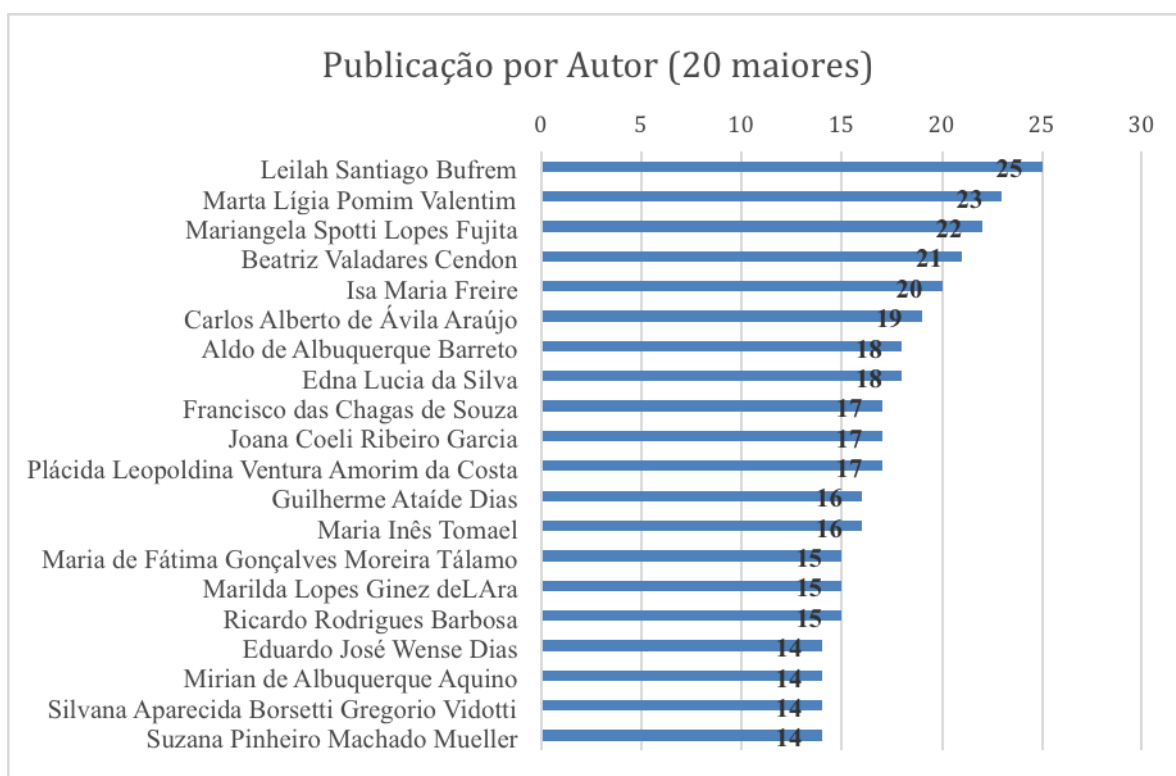


Fonte: autores

Dentre as organizações que mais publicam nas revistas avaliadas a UFMG aparece em primeiro, seguida da UFSC, USP e UnB. Além disso, cerca de metade das publicações está relacionada com apenas 8 instituições, em sua maioria universidades federais. É importante ressaltar que muitas publicações possivelmente não foram adequadamente classificadas tendo em vista que em alguns casos somente o nome do departamento ou laboratório estavam registrados.

O gráfico 3 apresenta as publicações por autor, para os 20 autores com maior número de publicações nessas revistas. Mais da metade deles (12) são vinculados a apenas três instituições: UNESP, UFMG e UFPB.

Gráfico 3 – Vinte autores com maior número de publicações em revistas na ABCDM (2000-2015)



Fonte: autores

A Tabela 2 apresenta as 40 palavras chaves mais comuns encontradas nos registros da ABCDM. Destaque para o peso científico da Ciência da Informação (240) em relação às demais áreas (Biblioteconomia e Arquivologia), da Gestão do Conhecimento (91) e Comunicação Científica (86) que se destacam dos demais. De forma agrupada, os termos relacionados com gestão (conhecimento, informação e inteligência competitiva) e comunicação científica (bibliometria, produção científica e redes sociais) também se destacam dos demais.

Tabela 2 - Palavras Chave mais comuns

Palavra chave	Frequência	Palavra Chave	Frequência
ciência da informação	240	ontologia	30
Informação	142	interdisciplinaridade	29
gestão do conhecimento	91	bibliotecas digitais	28
comunicação científica	86	arquivologia	27
Conhecimento	65	biblioteca digital	27
gestão da informação	64	leitura	27
sociedade da informação	64	metadados	27
Bibliometria	61	organização da informação	27
Internet	58	bibliotecas universitárias	26
produção científica	52	competência informacional	26
Biblioteconomia	38	sistemas de informação	26
inteligência competitiva	37	terminologia	26

organização do conhecimento	37	representação do conhecimento	25
tecnologia da informação	37	hipertexto	24
Epistemologia	33	inovação	24
profissional da informação	33	tecnologia	24
redes sociais	33	indexação	23
Bibliotecário	32	biblioteca escolar	22
Brasil	32	biblioteca pública	22
recuperação da informação	31	inclusão digital	22

3.2 REDES DE COLABORAÇÃO CIENTÍFICA

Para a criação das redes de colaboração científica foram analisadas 2.307 publicações com 2.442 autores, conforme já descrito. No entanto para a aplicação da maioria das métricas de análise de redes é necessário que sejam feitos alguns recortes no universo pesquisado, começando pelos 357 autores que não tiveram nenhuma publicação com outro autor. Excluídos esses 357 autores sem colaboração, foi identificado um total de 308 componentes ou grupos de autores no restante da rede, sendo que um desses componentes continha 1.039 autores ou 42,5% da rede total. Esse componente com 1.039 autores direta ou indiretamente relacionados por colaborações científicas é denominado, segundo o método de Análise de Redes, componente gigante, e possui ocorrência comum em redes de colaboração científica. Dos 307 componentes ou subgrupos restantes, aquele que possuía maior número de autores continha 18 elementos. Pode-se assumir com isso que os 307 subgrupos formados representam grupos de pesquisa mais específicos ou mesmo artigos pontuais com autores com poucas publicações nas revistas estudadas.

As métricas e figuras apresentadas a seguir sobre os autores levam em consideração o universo de pesquisadores do componente gigante. O total de linhas ou relações encontradas totalizam 2.053.

Dentre as métricas aplicadas ao componente gigante com 1.039 autores temos o coeficiente de clusterização com a média de 0,726. O coeficiente de clusterização é calculado como a média dos coeficientes de cada pesquisador que corresponde ao número de colaboradores de cada autor que também colaboram entre si. A centralidade de grau médio da rede é de 3,952 e a centralidade grau médio ponderado é de 4,76. Isso quer dizer que em média cada um dos 1.039 autores possui 3,9 colaboradores e 4,76 colaborações uma vez que dois autores podem possuir mais de uma colaboração entre eles.

Quanto às métricas individuais por autor, a Tabela 3 apresenta as centralidades de grau e grau ponderado, centralidades de intermediação e proximidade normalizadas, a centralidade de autovetor e o coeficiente de clusterização dos 20 autores com maior número de artigos

publicados nas revistas em análise, em ordem decrescente quanto ao número de publicações.

As definições das métricas podem ser apresentadas de acordo com Sampaio (2015)

Diferentes medidas e algoritmos foram resumidos, expandidos e desenvolvidos em um artigo escrito por Freeman (FREEMAN, 1979), no qual ele introduz a tipologia moderna de medidas de redes ao especificar três definições clássicas de centralidade: centralidade de grau (*degree centrality*), centralidade de proximidade (*closeness centrality*) e centralidade de intermediação (*betweenness centrality*). O grau é uma medida da influência direta que um vértice tem em relação a seus contatos, a proximidade está relacionada com o tempo que uma informação leva para ser compartilhada por todos os vértices na rede, e a intermediação de um vértice pode ser considerada o controle da comunicação entre todos os demais pares de vértices da rede. As medidas de centralidade de proximidade e centralidade de intermediação são baseadas na suposição de que a informação (ou qualquer conteúdo da ligação) é transmitida somente ao longo dos possíveis caminhos mais curtos. Bonacich (BONACICH, 1987) propôs uma medida denominada centralidade de autovetor (*eigenvector centrality*), cujo objetivo é medir a importância de um nó em função da importância de seus vizinhos. Isso quer dizer que, mesmo se um nó estiver conectado somente a alguns outros vértices da rede (tendo assim uma baixa centralidade de grau), esses vizinhos podem ser importantes e, conseqüentemente, o nó será também importante, obtendo uma centralidade de autovetor elevada.

Tabela 3 – Métricas de centralidade e clusterização para os 20 autores com maior número de publicações na ABCDM (2000-2016).

	Grau	Grau P.	Prox.	Inter.	AV	CoefClust
Leilah Santiago Bufrem	19	34	0,20	0,16	0,28	0,11
Marta Lígia Pomim Valentim	31	51	0,20	0,16	0,89	0,12
Mariangela Spotti L. Fujita	17	29	0,20	0,12	0,24	0,08
Beatriz Valadares Cendon	27	36	0,18	0,09	0,52	0,09
Isa Maria Freire	19	23	0,17	0,09	0,12	0,07
Carlos Alberto de Á. Araújo	32	37	0,19	0,06	1,00	0,14
Aldo de Albuquerque Barreto	9	10	0,11	0,00	0,24	1,00
Edna Lucia da Silva	13	22	0,16	0,04	0,08	0,06
Joana Coeli Ribeiro Garcia	13	19	0,18	0,13	0,14	0,12
Plácida L.V.A.Costa Santos	13	18	0,17	0,02	0,08	0,03
Guilherme Ataíde Dias	30	33	0,19	0,15	0,35	0,07
Maria Inês Tomael	26	58	0,19	0,10	0,96	0,23
Maria de Fátima G.M.Tálamo	7	18	0,19	0,07	0,12	0,10
Marilda Lopes Ginez deLara	7	9	0,17	0,03	0,05	0,05
Ricardo Rodrigues Barbosa	23	27	0,18	0,04	0,86	0,17
Eduardo José Wense Dias	13	16	0,19	0,10	0,11	0,08
Mirian de A. Aquino	11	12	0,17	0,02	0,07	0,07
Silvana A.B. Gregorio Vidotti	15	16	0,18	0,08	0,13	0,03
Suzana Pinheiro M. Mueller	10	12	0,14	0,06	0,06	0,07

Fonte: autores Onde: Grau P.= Grau Ponderado, Prox. = Proximidade, Inter. = Intermediação, AV = Auto Vetor e CoefCuster = Coeficiente de Clusterização

Uma série de análises pode ser feitas quanto às métricas de centralidade apresentadas.

Destacamos aqui algumas considerações gerais quanto aos autores com maiores valores para cada métrica apresentada. As centralidades de grau e grau ponderado demonstram a capacidade do autor de trabalhar com outros autores. Carlos Araújo, Marta Valetim e Guilherme Dias possuem o maior número de colaboradores, com destaque para o Guilherme Dias, que apesar de ter um número de publicações até 30% menor que os primeiros autores, ainda assim possui muitas colaborações. Essa fato pode decorrer do autor realmente trabalhar com número elevado de colaboradores ou um ou mais artigos publicado com grande número de co-autores.

No caso do Grau ponderado, o destaque ficaria com a autora Maria Inês Tomael, que possui 58 colaborações com média 3,6 de autores por artigo publicado.

Quanto à centralidade de proximidade, os autores possuem valores similares, com exceção de Aldo Barreto, que possui baixa centralidade de proximidade. Isso pode indicar a existência de um grupo de pesquisa com maior foco em uma área específica e menor relação com os demais pesquisadores. No caso da intermediação, os maiores valores denotam autores com maior capacidade de se relacionar com grupos diferentes, também conhecido na área da ciência da informação como *gate keepers*. Os autores com maior centralidade de intermediação são Leilah Bufren, Marta Valetim e Guilherme Dias .

Quanto ao autovetor, Carlos Araújo possui valor 1 e Maria Inês Tomael com valor 0,96. Isso quer dizer que esses autores possuem não só um grande número de colaborações, mas também colaborações com outros pesquisadores também relevantes. Por último o Coeficiente de Clusterização com valor 1 para Aldo de Albuquerque demonstra que todos os seus colaboradores também colaboram entre si, sugerindo que esse autor atua em um grupo bastante coeso.

A validação e aprofundamento dos resultados apresentados na tabela 3 carece de uma avaliação mais qualitativa das colaborações entre pesquisadores.

Na rede de colaboração apresentada na Figura 1 são colocados em destaque 1.039 autores do componente gigante.

Figura 1 – Colaboração Científica entre 1.039 autores do componente gigante da rede de coautorias registradas na ABCDM (2000-2016).



Fonte: autores

Por conta da impossibilidade de visualização de 1.039 nomes em uma mesma figura, se optou por apresentar apenas os nomes dos autores com 10 ou mais publicações registradas. Assim, destacam-se pelo tamanho da produção os autores da UFMG, UFSC, USP e UnB, conforme também foi relatado no Gráfico 2. Cada nó da rede representa um autor e cada linha uma relação entre dois coautores. O tamanho do nó (e também o tamanho da fonte de texto na grafia do nome) está relacionado ao número de publicações. Quanto maior o tamanho do nó, maior o número de publicações. A espessura das linhas representa o número de relações entre quaisquer dois autores. As cores da figura representam os subgrupos formados pelo algoritmo de clusterização que busca, por meio do número de linhas entre os autores, definir subgrupos coesos, ou seja, que possuam maior número de relações entre eles do que com autores fora de determinado grupo.

Foram encontrados 20 subgrupos coesos dentro do componente gigante, sendo que o maior deles apresenta 99 autores, e o menor 23. Um estudo utilizando modelos estatísticos pode ser aplicado para avaliar os elementos internos e externos a rede que permite e reforça a

Fonte: autores

Seguindo uma lógica similar à da Figura 1, cada nó da rede representa uma instituição, e as linhas as relações entre duas organizações. Quanto maior a espessura da linha, maior o número de relações de colaboração entre duas instituições. O tamanho de cada nó (e do nome) é proporcional o número de publicações de cada organização. Para possibilitar a visualização, optou-se por mostrar apenas as instituições com 10 ou mais publicações registradas.

No que diz respeito às colaborações, seguem algumas considerações quanto à Figura 2. O número de colaborações entre organizações é pequeno, se comparado com as colaborações dentro de uma mesma organização. As relações da UnB com a UFMG, e também com a UFSC, estão acima da média, com 11 e 9 colaborações respectivamente e podem ser vistas ao centro da figura. Outras relações importantes que denotam a internacionalização são: a da UFPE com uma forte colaboração com a Université de Mairsele II e a UFRGS com a Universidade Carlos III de Madrid, e dessa com a Universidade de Granada. Além disso, pode-se perceber um grande número de colaborações entre as organizações de grande porte (com alto número de publicações) com outras menores quanto ao número de publicações na área. Esse último tipo de relação deve ser mais estudado, mas pode sugerir uma relação de orientação, onde professores de universidades menores tendem a realizar seus estudos de doutoramento em universidades maiores, gerando este tipo de colaboração.

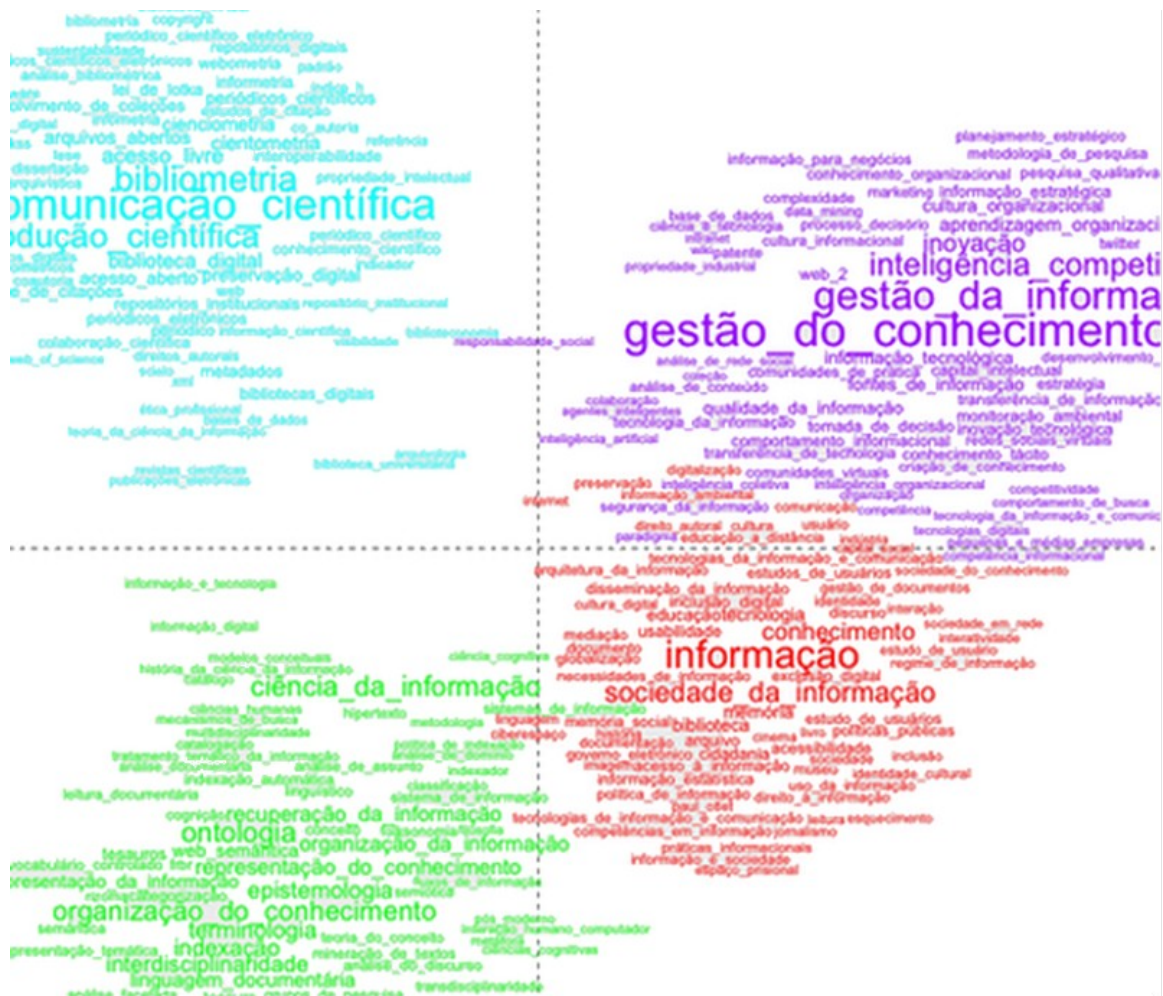
3.3 TEMAS CENTRAIS

Para o estudo dos temas centrais foram avaliados os resumos dos textos apresentados e as palavras chaves de cada texto. Devido a necessidade de aprofundamento quanto a análise dos textos se optou por apresentar os resultados da análise das palavras chaves com o uso da ferramenta IRAMUTEQ.

Foram encontradas nos textos analisados um total de 4.574 palavras chaves, das quais 90% apareceram uma única vez e 45 palavras apareceram mais de 20 vezes.

A Figura 3 apresenta as palavras agrupadas utilizando o método Reinert de classificação de texto que leva em consideração o número de vezes que duas palavras quaisquer aparecem juntas em um mesmo texto além da semântica da palavra. Foram encontrados 4 grandes grupos de palavras apresentados em um plano cartesiano.

Figura 4 – Distribuição das palavras por classificação Reinert



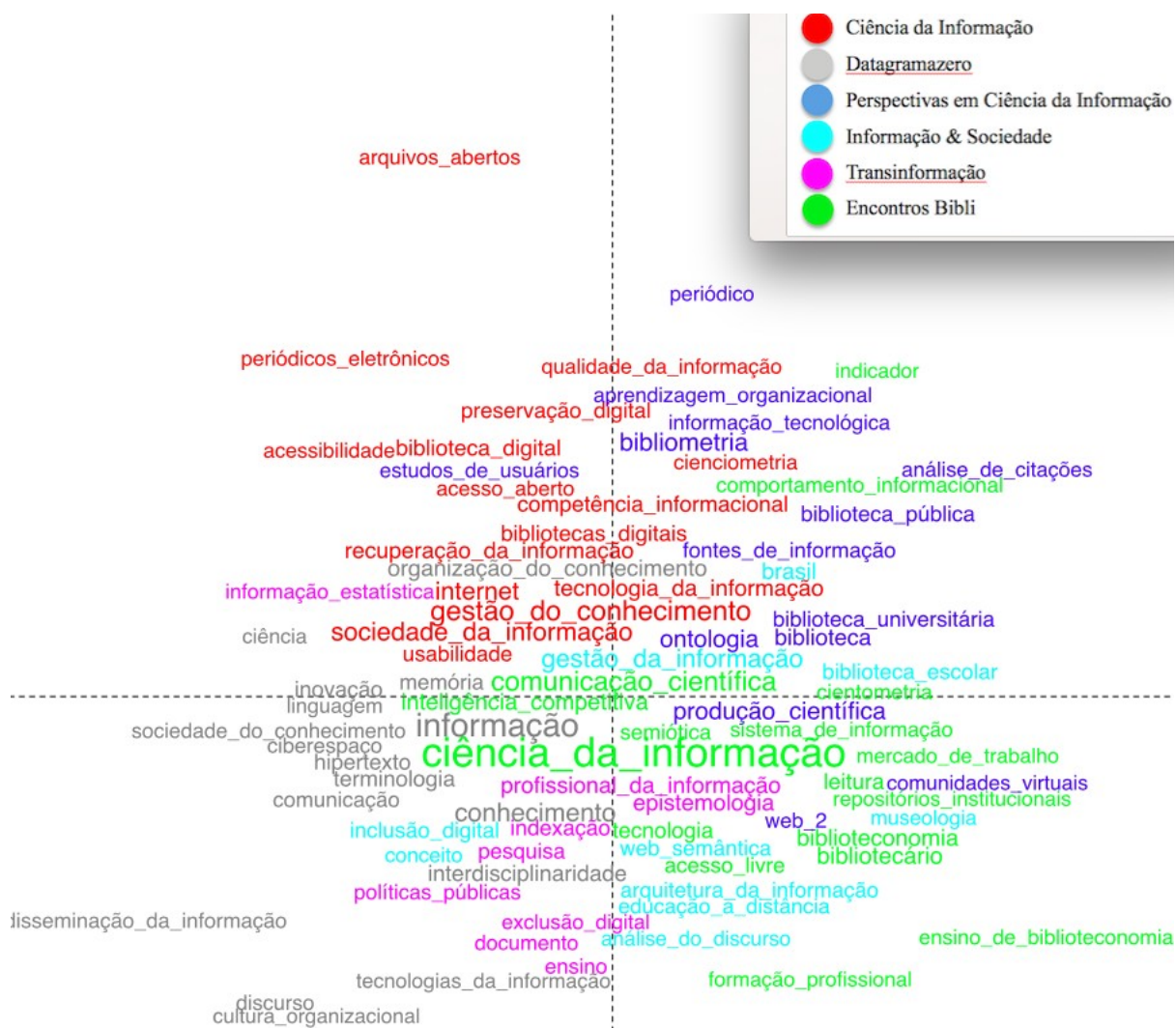
Fonte: autores

O tamanho de cada palavra está relacionado ao número de vezes que a palavra apareceu nos textos. Quanto maior o número de vezes maior a fonte. O posicionamento das palavras é feito de acordo com a frequência com que a palavra se posiciona dentro de determinado grupo ou se é também a outros grupos. No caso, por exemplo, do grupo verde a palavra “ciência_da_informação” está mais ao centro que a palavra “indexação” ou “linguagem_documentada” sendo esta primeira mais comum aos demais grupos.

A segunda análise textual focou nas instituições para analisar as diferenças entre elas quanto a área do conhecimento, no entanto não foi notada nenhuma diferença relevante entre os diferentes temas por instituição exceto no caso da Fiocruz que teve as palavras “acesso_à_informação” e “educação_em_saúde” em maior evidência que as demais instituições.

Por fim, foram analisadas as palavras com base nas revistas para se entender se há algum direcionamento para este ou aquele assunto para cada uma das revistas e o resultado é apresentado na figura 5.

Figura 5 – Distribuição das palavras chaves por revista



Fonte: autores

A figura 5 apresenta dividido por cores as palavras mais representativas de cada revista com a legenda de cores por revista na parte superior direita da figura. A disposição das palavras segue o padrão apresentado na figura 4 onde as palavras mais próximas ao eixo central fazem parte não só do grupo com a qual a cor a identifica mas também a outros grupos como é o caso da palavra “ciência_da_informação”. No caso da palavra “ensino_de_biblioteconomia” como está mais longe do eixo central é uma palavra comum quase que exclusivamente a revista Encontros Bibli.

4 DISCUSSÃO DOS RESULTADOS

Os resultados apresentados no estudo são exploratórios e possibilitam um aprofundamento qualitativo em estudos posteriores em relação a vários aspectos, como por exemplo, o tipo de relação entre os autores das instituições. Levando-se em consideração o

recorte realizado, pode-se notar a representatividade de algumas instituições de ensino e pesquisa, que além de serem os maiores produtores de conhecimento científico por meio de publicações, ainda são reconhecidas pelos pares por meio de avaliações e outros como sendo destaque nessa área do conhecimento. Isso pode ser interpretado quanto a relevância dos periódicos Brasileiros para a produção nacional.

No que diz respeito as palavras chaves encontradas, se faz necessário uma avaliação mais aprofundada junto a pesquisadores com larga experiência na área e de preferência que tenha tido inserção em diferentes subáreas, para que possa ser feita uma interpretação mais adequada dos resultados. Perguntas como surgimento e consolidação de novas áreas de pesquisa ou a incorporação de novos conhecimentos em áreas específicas podem ser direcionamentos importantes para políticas públicas em ciências. Os subgrupos encontrados na Figura 4 são relevantes para uma primeira análise, mas insuficientes para se entender a dinâmica e fluxo de informações.

Quanto aos dados dos autores e suas colaborações as possibilidades de aprofundamento desse estudo são diversas. A política institucional pode promover novos direcionamentos para parcerias dos mais diversos tipos com objetivos também diversos. As avaliações, que hoje tem sido objeto de discussão podem usar dessas métricas elementos que apoiem novas metodologias de análise.

Por ultimo, a análise das revistas é essencial para se entender tanto no nível de publicação quanto no nível de estudo onde e como se deve buscar informações que atendam as suas necessidades de disponibilização quanto de recuperação da informação.

5 CONCLUSÃO

O que se pretendeu com o estudo foi principalmente demonstrar que aliado aos dados bibliométricos é possível trazer novas soluções que podem enriquecer o entendimento sobre alguns aspectos da literatura científica da área de Ciência da Informação e consequentemente da comunidade científica brasileira nesta área.

Os resultados mostraram que a produção anual de artigos dos seis periódicos brasileiros mais citados cresceu paulatinamente desde 2000 de pouco mais de 120 artigos para pouco mais de 200 artigos em 2012. Apontou como instituições com maior produção (Quartil 1) a UFMG, a UFSC e a USP, estando metade da produção (Quartil 2) vinculada a apenas 8 das 623 instituições. Foram identificados os 20 autores com maior produção, destacando-se que mais da metade destes (12) se concentram na UNESP, UFMG e UFPB. Em relação aos temas, foram apresentadas tabelas e figuras que destacam a gestão do conhecimento e a comunicação

científica e seus subtemas.

A discussão mais detalhada de cada uma das análises de redes e textual deve ser levada a prova e validado quanto ao que se presume sobre cada uma delas de forma a melhor se entender os seus resultados e implicações.

Como elementos de proposta para novos estudos existe a possibilidade de entendimento específico de cada sub área do conhecimento na CI como a Biblioteconomia, Museologia, Arquivologia e outros e como estas discutem e usam conhecimentos diversos por meio da análise textual. Cabe também estudar o nível das colaborações “intra” e “inter” institucionais a luz de políticas públicas ao ponto de se analisar ao atendimento de suas propostas.

REFERÊNCIAS

BONACICH, P. Power and Centrality: A Family of Measures. **American Journal of Sociology**, v. 92, n. 5, p. 1170–1182, 1 mar. 1987.

FREEMAN, L. C. Centrality in social networks: conceptual clarification. **Social networks**, v. 1, n. 3, p. 215–239, 1979.

GLÄNZEL, W. Coauthorship Patterns and Trends in the Sciences (1980-1998): A Bibliometric Study with Implications for Database Indexing and Search Strategies. **Library Trends**, v. 50, n. 3, p. 461, Winter 2002.

GLÄNZEL, W.; SCHUBERT, A. Analysing Scientific Networks Through Co-Authorship. In: MOED, H. F.; GLÄNZEL, W.; SCHMOCH, U. (Eds.). **Handbook of Quantitative Science and Technology Research: : The Use of Publication and Patent Statistics in Studies of S&T Systems**. Netherlands: Springer, 2004. p. 257–276.

GUARNACCIA, C. et al. Differences in social representation of blood donation between donors and non-donors: an empirical study. **Blood Transfusion = Trasfusione Del Sangue**, p. 1–7, 4 nov. 2015.

HANNEMAN, R. A.; RIDDLE, M. **Introduction to Social Network Methods**. Riverside, CA: University of California, 2005. Disponível em: <<http://faculty.ucr.edu/~hanneman/nettext/>>.

KATZ, J. S.; MARTIN, B. R. What is research collaboration? **Research Policy**, v. 26, n. 1, p. 1–18, mar. 1997.

LAUDEL, G. What do we measure by co-authorships? **Research Evaluation**, v. 11, n. 1, p. 3–15, 2002.

MELIN, G.; PERSSON, O. Studying research collaboration using co-authorships. **Scientometrics**, v. 36, n. 3, p. 363–377, 1 jul. 1996.

NAGPAUL, P. S. Visualizing cooperation networks of elite institutions in India. **Scientometrics**, v. 54, n. 2, p. 213–228, 1 jun. 2002.

NEWMAN, M. **Networks: An Introduction**. 1. ed. USA: Oxford University Press, 2010.

NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration. **Proceedings of the National Academy of Sciences**, v. 101, n. suppl_1, p. 5200–5205, 23 jan. 2004.

NOOY, W. DE; MRVAR, A.; BATAGELJ, V. **Exploratory Social Network Analysis with Pajek**. USA: Routledge, 2005.

RODRIGUES, M. M. D. et al. Activities Offered to the Elderly in Primary Health Care: Social Representations of Elderly and Professionals. **International Archives of Medicine**, v. 8, n. 0, 17 nov. 2015.

ROYAL SOCIETY. **Knowledge, networks and nations: global scientific collaboration in the 21st century**. UK: Royal Society, 2011.

SAMPAIO, R. B. **As estruturas globais e regionais do campo de pesquisa, desenvolvimento e inovação das doenças negligenciadas leishmaniose e tuberculose sob a ótica das redes complexas**. Doutorado (Ciência da Informação)—Brasília: Universidade de Brasília, 2015.

SAMPAIO, R. B. et al. A colaboração científica na pesquisa sobre coautoria: um método baseado na análise de redes. **Perspectivas em Ciência da Informação**, v. 20, n. 4, p. 79–92, 17 dez. 2015.

SONNENWALD, D. H. Scientific collaboration. **Annual Review of Information Science and Technology**, v. 41, n. 1, p. 643–681, 1 jan. 2007.

SPINAK, E. Quantitative analyses of scientific literature and their validity for judging Latin American Production. **Bulletin of the Pan American Health Organization (PAHO)**, v. 29, n. 4, p. 352–60, 1995.

VANTI, N. A Cientometria revisitada à luz da expansão da ciência, da tecnologia e da inovação. **PontodeAcesso**, v. 5, n. 3, p. 5–31, 1 fev. 2012.

VILAN FILHO, J. L.; MUELLER, S. P. M. Periódicos brasileiros das áreas de informação: prestígio e interdisciplinaridade. **Encontro Brasileiro de Bibliometria e Cientometria**, v. 4, p. POSTER-03, 2014 .

WASSERMAN, S.; FAUST, K. **Social Network Analysis: Methods and Applications**. 1. ed. USA: Cambridge University Press, 1994.

YANCEY, R. Fifty years of citation indexing and analysis. **Science**, set. 2005.