



ANÁLISE DE DOMÍNIO DA PRODUÇÃO CIENTÍFICA SOBRE COLETA DE DADOS NO CONTEXTO DO INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS

Fábio Mosso Moreira

Graduado em Administração de Empresas pela Faculdade de Ciências e Engenharia (UNESP/Tupã). Mestrado concluído em Ciência da Informação - Faculdade de Filosofia e Ciências (UNESP/Marília). Doutorado em andamento Programa de Pós-Graduação em Ciência da Informação - Faculdade de Filosofia e Ciências (UNESP/Marília). Atua como membro do Grupo de Pesquisa Novas Tecnologias em Informação - GPNTI (UNESP/Marília) e Grupo de Pesquisa Tecnologia de Acesso a Dados - GPTAD (UNESP / Tupã). Editor de Conteúdo da Revista Eletrônica Competências Digitais para Agricultura Familiar (RECoDAF). E-mail: fabio.moreira@unesp.br

Fernando Assis Rodrigues

Doutor e Mestre em Ciência da Informação pela UNESP - Universidade Estadual Paulista. Especialista em Sistemas para Internet pela UNIVEM - Centro Universitário Eurípides de Marília. Bacharel em Sistemas de Informação pela USC - Universidade do Sagrado Coração. Membro dos grupos de pesquisa GPNTI - Novas Tecnologias em Informação e GPTAD - Tecnologias de Acesso a Dados (UNESP), GPIDT - Informação, Dados e Tecnologia (USP) e GPDM - Dados e Metadados (UFSCar). Editor do periódico RECoDAF - Revista Eletrônica Competências Digitais para a Agricultura Familiar. E-mail: fernando.assis@unesp.br

Ricardo César Gonçalves Sant'Ana

Professor Associado da Universidade Estadual Paulista - UNESP, Faculdade de Ciências e Engenharias - FCE, Campus de Tupã, em regime de dedicação exclusiva, onde é Presidente da Comissão de Acompanhamento e Avaliação dos cursos de Graduação - CAACG, Coordenador Local do Centro de Estudos e Práticas Pedagógicas - CENEPP e Ouvidor Local. Professor do Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista, Campus de Marília. E-mail: ricardo.santana@unesp.com

RESUMO

Este estudo teve como objetivo identificar pesquisas que versam sobre a temática de coleta de dados. Para tanto foi utilizada análise de domínio em publicações científicas, com a aplicação de uma análise de citações e cocitações de autores. A

identificação dos representantes da temática de coleta de dados e a interlocução existente entre eles foi obtida a partir do processamento dos conjuntos de metadados sobre publicações em periódicos disponíveis da base de conhecimento IEEE Xplore(r) Digital Library. Como estratégia de busca, utilizou-se na interface de busca avançada os termos 'Data Collecting', 'Data Collect' e 'Data Gathering', concatenados pelo operador booleano 'OU'. Foram recuperados dados sobre 2.278 publicações e a amostra foi delimitada aos artigos publicados em periódicos científicos, totalizando 281 publicações. Para cada artigo, foi coletado o conjunto de referências na forma de documento HTML. Sobre o conteúdo obtido foi aplicado um algoritmo para a conversão das referências de HTML para CSV e a serialização das referências para o padrão de estilo IEEE Editorial Style. Foram processadas 5.867 referências e identificados 8.267 autores, sendo descartadas 270 referências por estarem fora do padrão. Aplicou-se a Lei do Elitismo de Price para a delimitação do corpo de autores nas análises de citação e cocitação de autores, totalizando 91 autores, número este que foi arredondado para 94 para contabilizar mais 3 autores que possuíam o mesmo número de citação do 91º autor. Posteriormente, aplicou-se um algoritmo para a geração das matrizes i) citados e citantes e ii) de frequência absoluta de cocitação de autores. A partir destes dados, a identificação de informações sobre nacionalidade e associação institucional foram obtidas por processamento manual. Foram calculados os indicadores de rede social: i) Densidade da Rede, representando a intensidade de relacionamentos entre autores da rede e ii) Centralidade de Grau, representando o número de ligações incidentes em um autor. A análise dos dados resultou em uma Densidade de Rede no valor de 3,20% e desvio padrão amostral de 3,34%, ou seja, cada pesquisador têm aproximadamente 3% de fazer relacionamentos com pares da rede; e Centralidade de Grau no valor de 20,93%, demonstrando dispersão, com cada vértice possuindo 20,93% de probabilidade de receber alguma interação da rede. A dispersão está associada com a amplitude do domínio analisado, pois coleta de dados é um tema recorrente em diferentes áreas do conhecimento, mas aderente ao contexto de publicação dos periódicos da IEEE. Ao analisar a Centralidade de Grau dos autores individualmente, é possível observar uma relação com a quantidade de citações recebidas, uma vez que os 13 principais autores do indicador de Centralidade de Grau são também os mais citados. Conclui-se que a temática apesar de amplamente citada apresenta um núcleo estadunidense, ligado às instituições UC, USC e MIT.

PALAVRAS-CHAVE: Dados. Coleta de Dados. Análise de Domínio. IEEE.

ANALYSIS OF DOMAIN OF SCIENTIFIC PRODUCTION ON DATA COLLECTION IN THE CONTEXT OF THE INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS

ABSTRACT

The goal of this study is to identify scientific studies about the thematic of data collecting. For this purpose, it was adopted the domain analysis method on the scientific papers, by an application of Citation and Co-citation Analysis. The identification of representatives from the thematic of data collecting and the existent dialog among them were obtained by authors and papers metadata sets processing, available on IEEE Xplore(r) Digital Library. As search strategy, it was used on advanced search the terms 'Data Collecting', 'Data Collect', and 'Data Gathering', concatenated by the boolean operator 'OR'. This process recovered 2,278 scientific papers and the sample was set only by scientific papers published in scientific journals, with a total of 281 papers. For each paper, the reference section was collected in HTML document format. It was applied an algorithm to convert formats from HTML documents to CSV files and also to serialize the IEEE Editorial Style found on collected reference data. The algorithm processed 5,867 references and discarded 270 because they not fit into the IEEE Editorial Style standards adopted on serialization. From this references, was identified a total of 8,267 authors. In Citation and Co-citation Analysis, it was applied the Price's square root law to delimit the authors' group to 91 participants, rounded to 94 participants because of the 91st participant had the same total of citation of his 3 successors. After that, the "Cited and Who cited" and the "Absolute Frequency of Co-citation" matrices was generated from an application of an algorithm. By those data, the identification of nationality and the institutional affiliation were obtained by a manual process. Was calculated the social networks indexes i) Network Density, representing the relationship intensity between authors on the network and ii) Centrality Degree, representing the number of relationships received by an author. The analyzed data resulted in a Network Density value of 3.20% with a standard deviation of 3.34%, that is, each researcher has approximately 3% to make interactions with other network nodes. Also, the resulted value of Centrality Degree was 20.93%, demonstrating dispersion on the network, once that each node has 20.93% of probability to receive some interaction from the network. This dispersion is associated with the analyzed domain amplitude, once that Data Collecting is a recurrent theme on distinct knowledge areas, but still adherent to IEEE scientific journals context. When results of the Centrality Degree of each author are analyzed, it is possible to observe a relationship between the results of received citations, indicating that the 13 best-ranked authors by Centrality Degree are also the most cited ones. It was concluded that this thematic, although widely cited, shows an American core, related to the institutions UC, USC, and MIT.

KEYWORDS: Data. Data Collecting. Domain Analysis. IEEE.

1 INTRODUÇÃO

Desde a invenção das ferramentas de pedra pelo homem primitivo até a massificação da disponibilidade de conexão com a Internet, nunca houve um volume tão expressivo de dados disponíveis (CASTELLS, 2008; IBM, 2017). Esse efeito é, em parte, reflexo de uma sociedade

em rede (CASTELLS, 2008), conectada e que adota em ritmo acelerado tecnologias ligadas às Tecnologias da Informação e Comunicação (TIC) (RODRIGUES, 2017), contribuindo para um aumento significativo no volume de dados gerados, em que a obtenção destes dados, ou seja, os processos de coleta de dados podem ser realizados por meio de uma combinação de tecnologias, desde o uso de satélites até micro sensores nos dispositivos pessoais ou instalados no ambiente (SALES; CAVALCANTI, 2015).

Além disso, o uso de TIC reduziu o tempo necessário entre a fase de obtenção do dado e a fase de armazenamento, demandando, para sua compreensão, ordens de magnitude na casa de milionésimos de segundo. Essa velocidade dos processos de coleta, associada a fatores como barateamento do custo de armazenamento, contribui para um crescimento exponencial da disponibilidade de novos dados, o que reflete resultados como o do contexto estudado pela IBM do início do ano de 2018, em que 90% dos dados armazenados foram obtidos nos anos de 2016 e 2017 (IBM, 2017).

Em perspectiva complementar, os sensores embarcados nos dispositivos móveis e instalados em ambientes públicos e privados, estão propiciando (i) a disseminação de uma grande variedade de tipos de dados e (ii) o surgimento de um processo retroalimentado (MANTOVANI; DANTAS, 2011) em que receptor e emissor se conjugam em um fluxo e refluxo, fonte de novas variedades e quebras de estruturas que limitavam a gênese de novas emissões.

As características de volume, velocidade e variedade de dados são componentes do fenômeno ora denominado *big data* (DAVENPORT, 2014), em que processos de coleta e de armazenamento de dados são realizados em larga escala, com uso de recursos tecnológicos variados para a modelagem das estruturas de dados e administração de suas instâncias.

O processo de acesso e uso de dados tem início na fase de coleta, ou seja, a partir da obtenção dos dados. Esta fase é identificada como aquela em que se inicia com a manifestação de necessidades informacionais que, por sua vez irão nortear escolhas e definições sobre quais dados serão necessários nas fases posteriores. É nessa fase que são estabelecidas e operacionalizadas estratégias sobre como localizar e capturar dados a partir de leitura de variáveis oriundas de um determinado contexto ou fato, a escolha de mecanismos utilizados para a obtenção, e, a adoção ou de metodologias e de ferramentas para consecução destes dados (SANT'ANA, 2016).

Compreender os desafios e as possibilidades proporcionadas pela fase de coleta é

requisito para tornar o processo de obtenção de dados mais eficiente e, para tanto, faz-se necessário identificar, sistematicamente, estudos que abordem a fase de coleta de dados, considerando a importância de estudos aderentes ao uso de TIC e, conseqüentemente, identificando nucleação dos pesquisadores e vertentes de temáticas. Neste sentido, esta pesquisa teve como objetivo identificar a estrutura em que se organizam pesquisadores dos estudos selecionados por meio da análise de domínio, analisando a interlocução existente entre eles a partir de indicadores provenientes de estudos bibliométricos.

Adota-se a visão de domínio como um reflexo de uma comunidade discursiva e de seu papel na ciência. Sua importância na caracterização e avaliação da ciência permite identificar as condições pelas quais o conhecimento científico se constrói e se socializa. Segundo Hjørland (2002), um domínio pode ser analisado por meio de 11 abordagens, dentre estas, utiliza-se nesta pesquisa os estudos bibliométricos. Dentro dessa abordagem, adotou-se, os procedimentos: análise de citações e análise de cocitações.

A análise de citações está relacionada ao processo de visualização de elementos de um campo científico enquanto um domínio, por meio da forma principal com a qual a comunidade desse domínio se legitima – as citações (PIOVEZAN; FUJITA, 2015). Essa análise permite identificar *clusters* (agrupamentos) de pesquisadores e de publicações, bem como as relações, evidenciando o impacto de pesquisadores de uma área de conhecimento (GLÄNZEL, 2003).

A análise de cocitação de autores trata sobre a identificação da frequência com que dois elementos de um domínio possuem em comum. A frequência é calculada a partir da identificação de co-ocorrências em que estes elementos são citados na literatura científica, onde estes elementos podem ser constituídos por documentos, autores, periódicos ou países, entre outros. O enfoque é permitir identificar as citações que dois estudos recebem na literatura posterior a sua publicação, onde o fato de serem citados em um mesmo estudo retrata uma proximidade entre os citados, na perspectiva dos citantes (GLÄNZEL, 2003).

A análise de cocitações tem base no conceito que a frequência das citações recebidas por um artigo é um indicador de sua importância científica. A análise de co-citações pode ser um elemento chave para representar na estrutura da ciência em termos geográficos (SMALL, 1977; LE COADIC, 1996).

A identificação de autores relacionados com a temática de coleta de dados e a interlocução entre eles foram obtidas a partir do processamento de conjuntos de metadados sobre publicações em periódicos do IEEE Xplore(r) Digital Library. O critério de escolha desta

base de conhecimento levou em consideração o enfoque dado a periódicos e publicações de áreas e de disciplinas relacionadas ao desenvolvimento de recursos tecnológicos e a coleta de dados em ambientes digitais.

2 PROCEDIMENTOS METODOLÓGICOS

O levantamento foi realizado durante os meses de maio e junho de 2018, por meio das interfaces de pesquisa e busca avançada. A estratégia de busca utilizou os termos ‘*Data Collecting*’, ‘*Data Collect*’ e ‘*Data Gathering*’, concatenados pelo operador booleano ‘OU’. Foram recuperados conjuntos de metadados sobre 2.278 publicações e a amostra foi delimitada aos artigos publicados em periódicos científicos, totalizando 281 publicações. Para cada artigo, foi coletado o conjunto de referências na forma de documento HTML - HyperText Markup Language. Os conjuntos de dados analisados estão disponíveis no endereço <http://dadosabertos.info/data/collection_ieee_analysis_2018>.

Foi aplicado aos metadados coletados um algoritmo para a conversão das referências de HTML para CSV - Comma-Separated Values e a serialização das referências para o padrão de estilo IEEE Editorial Style. Foram processadas 5.867 referências e identificados 8.267 autores, sendo descartadas 270 referências por estarem inconsistentes ao padrão de estilo.

Aplicou-se a Lei do Elitismo de Price para a delimitação do corpo dos principais autores nas análises de citação e de cocitação de autores, totalizando 91 autores, número este que foi arredondado para 94 para contabilizar mais 3 autores que possuíam o mesmo número de citação do 91º autor.

Posteriormente, foi desenvolvido e aplicado um algoritmo na linguagem de programação Python, versão 3.5.4, para a geração das matrizes de citados por citantes e de ocorrência absoluta de cocitação de autores entre os citados. A partir destes dados, a identificação de informações sobre nacionalidade e associação institucional foram obtidas por processamento manual também de forma online.

Por meio do software UCINET, versão 6.636, que inclui partes da ferramenta NetDraw, foi elaborada a rede de cocitação de autores estabelecida entre os pesquisadores mais citados, gerando uma visualização que aproxima os autores mais similares em relação aos valores absolutos de cocitação de autores. Foram calculados dois indicadores de rede social: i) Densidade da Rede, representando a intensidade de relacionamentos entre autores da rede e ii) Centralidade de Grau, representando o número de ligações incidentes em um autor.

3 RESULTADOS E DISCUSSÃO

Quanto à nacionalidade dos principais autores, observou-se um destaque dos EUA - Estados Unidos da América (70,2%), seguido da China - República Popular da China (17%), da Noruega (3,2%), da Itália, Suíça e Nova Zelândia (2,1%), e da Romênia, Tailândia e Coreia do Sul (1,1%). A predominância dos estadunidenses pode estar associada ao fato do país ser a vanguarda do desenvolvimento de tecnologias aderentes a este estudo, tais como a Internet, além de ser país sede das maiores empresas relacionadas a este contexto.

Com relação às instituições na qual o grupo dos principais autores pertenciam, observou-se a relevância da Universidade do Sul da Califórnia, com 6 autores associados a esta instituição. Também foi possível observar a importância de outras instituições estadunidenses, como a Universidade da Califórnia, Universidade de Stony Brook e do Instituto de Tecnologia da Geórgia.

No topo do *ranking* de autores mais citados, destacam-se os treze primeiros, cujos quais representam 25,61% de todas as citações recebidas da rede dos 94 autores. O destaque ficou para os estudos de Deborah Estrin, da Cornell University (EUA), citada em 50 dos 281 artigos (17,80%). A pesquisadora atua na área da Ciência da Computação e trabalha com tópicos ligados a *small data*, que é a coleta de dados como pistas ou que buscam responder a perguntas específicas, e com *Sensing Networking*, prática de detecção de informações em redes.

Hamsa Balakrishnan, do Instituto de Tecnologia do Massachusetts (EUA), foi o segundo autor mais citado, totalizando citações em 43 artigos, o que representa 15,30% do total. O pesquisador atua na área de Aeronáutica e Astronáutica, desenvolvendo algoritmos de *Machine Learning* para coleta de dados com a finalidade de maximizar a capacidade de controle de congestionamentos de aeroportos, roteamento de tráfego, previsão do tempo e previsão de atraso.

Ramesh Govindan, da Universidade do Sul da Califórnia (EUA), foi o terceiro autor mais citado, identificado em 39 artigos, o que representa 13,87% do total. Este autor atua na área da Ciência da Computação, e estuda atividades de coleta de dados em redes de sensores sem fio e sua sincronização para transporte e armazenamento cêntrico.

Anantha P. Chandrakasan, do Instituto de Tecnologia de Massachusetts (EUA), obteve 37 citações, sendo assim o quarto autor mais citado, com representatividade de 13,16% do total. Seus projetos de pesquisa na área da Ciência da Computação abordam a coleta de dados

a partir da energização e do carregamento sem fio para a Internet das Coisas, e o desenvolvimento de circuitos e sistemas energeticamente eficientes para processamento de multimídia.

Wendi Heinzelman, da Universidade de Rochester (EUA), foi citado em 36 artigos, o que o coloca como o quinto autor mais citado, com representatividade de 12,81% do total. Este pesquisador atua na área de Engenharia Elétrica e da Computação, abordando a coleta de dados em redes de comunicação sem fio e computação móvel com armazenamento em nuvem. Martin Vetterli, da Escola Politécnica Federal de Lausanne (Suíça), foi citado em 31 artigos, ou seja, 11,03% do total, figurando como sexto colocado entre os mais citados. Engenheiro elétrico, o pesquisador aborda sua formação em pesquisas de coleta de dados para processamento de sinais e para codificação de canais de comunicação de vídeo.

John Heidemann, da Universidade do Sul da Califórnia (EUA), também obteve 31 citações e representatividade de 11,03% do total, contudo, figura como sétimo colocado pois possui um indicador de centralidade de grau menor comparado ao anterior com a mesma quantidade de citação. Este pesquisador atua na área da Ciência da Computação, desenvolvendo infraestruturas voltadas para coleta de dados, além de serviços de segurança na rede.

Yuanyuan Yang, da Universidade de Stony Brook (EUA), também obteve citações em 31 artigos e representatividade de 11,03% do total, contudo, possui indicador de centralidade de grau menor comparado aos dois anteriores, constando assim como oitavo colocado do *ranking*. Atua na área da Ciência da Computação, conduzindo pesquisas para coleta de dados em redes móveis e sem fio, visando armazenamento descentralizado na Computação em Nuvem.

Bhaskar Krishnamachari, da Universidade do Sul da Califórnia (EUA), foi citado em 28 artigos, com representatividade de 9,96%, sendo o nono autor mais citado. Realiza pesquisas na área da Engenharia Elétrica e da Computação, com interesses focados no desenvolvimento e na análise de algoritmos, protocolos e aplicativos para redes sem fio de próxima geração.

Rodica Cristescu, do Instituto Nacional para Lasers, Plasma e Radiação Física (Romênia), também obteve citações em 28 artigos e representatividade em 9,96% do total, contudo, figura como décimo colocado pois possui indicador de centralidade de grau menor comparado ao anterior. O pesquisador atua na área da Física, mas não aborda diretamente o processo de coleta de dados, e sim a manipulação dos dados que são gerados a partir de

experimentos com materiais via *lasers* e sistemas térmicos.

Baltasar Enrique Beferull Lozano, da Universidade de Agder (Noruega), também foi citado em 28 artigos, com representatividade de 9,96% do total, contudo, foi considerado o décimo primeiro colocado pois possui indicador de centralidade de grau menor comparado aos dois anteriores com a mesma quantidade de citações. Atua na área de Engenharia Elétrica, estudando redes de comunicação descentralizadas e desenvolvimento de algoritmos de inteligência artificial para coleta e processamento de dados multisensores.

O penúltimo autor mais citado no *ranking* dos treze primeiros autores foi David E. Culler, da Universidade da Califórnia (EUA), citado em 27 artigos, ou seja, 9,60% do total. Suas pesquisas na área da Ciência da Computação são direcionadas ao tema da coleta de dados em redes sem fio; e sistemas de grande escala para Internet, especialmente o tema de arquitetura de alta performance.

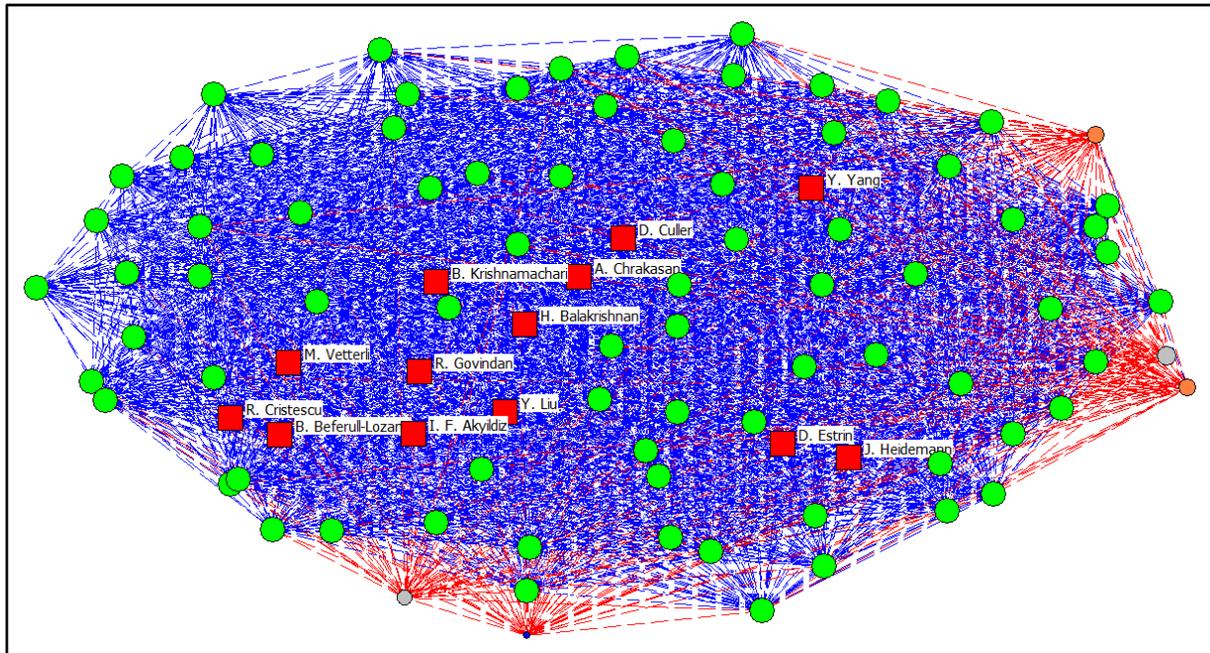
O último integrante do conjunto dos treze primeiros autores citados foi İlhan Fuat Akyıldız, do Instituto de Tecnologia da Geórgia (EUA), citado em 26 artigos, o que corresponde a 9,25% de representatividade. O pesquisador trabalha na área de Engenharia Elétrica e da Computação, e seus interesses de pesquisa atuais estão na coleta de dados em redes móveis 5G e redes de sensores sem fio com nano-sensores.

Na Figura 1, apresenta-se uma visualização da rede de cocitação dos 94 autores mais citados, na qual o tamanho dos elementos refletem a intensidade do índice de Centralidade de Grau na rede. Na rede é possível observar a formação de um *cluster* maior (representado pelas esferas na cor verde) contendo pesquisadores que se destacam quanto seu indicador de Centralidade de Grau (representado pelos quadriláteros na cor vermelha).

Verificam-se também a ocorrência de alguns *clusters* periféricos, que possuem proporção menor quando comparado ao *cluster* principal, representados pelos elementos de cores laranjas, cinza e azul. As linhas de cor azul representam as relações derivadas do *cluster* principal e as linhas de cor vermelha as relações derivadas dos *clusters* periféricos.

A análise da Centralidade de Grau resultou na média de 20,93%, demonstrando certa dispersão na rede, o que significa que cada vértice possui um quinto de probabilidade de receber alguma interação de algum dos outros nós desta rede.

Figura 1 - Rede de cocitação de autores



Fonte: Elaborado pelos autores.

O autor mais citado também foi o que obteve maior indicador de Centralidade de Grau de saída – Deborah Estrin (25,52%), demonstrando a importância da pesquisadora na vanguarda deste domínio. Observou-se uma relação direta deste indicador com a quantidade de citações recebidas, uma vez que os 13 principais autores com maior Centralidade de Grau também são os mais citados.

Quanto o indicador Densidade de Rede, o resultado de 3,20% com desvio padrão amostral de 3,34%, indica que cada pesquisador possui aproximadamente 3% de probabilidade de receber interação com algum dos pares da rede.

Esta dispersão da rede, verificada pelos resultados da Densidade de Rede, está associada com a amplitude do domínio analisado, pois coleta de dados é um tema recorrente em diferentes áreas do conhecimento, ou seja, com caráter interdisciplinar.

4 CONSIDERAÇÕES FINAIS

A partir dos resultados obtidos na pesquisa, conclui-se que a temática de coleta de dados, apesar de ser amplamente citada por pesquisas publicadas na IEEE, apresenta um pequeno núcleo de autores, com onipresença de instituições estadunidenses: Universidade da Califórnia, Universidade do Sul da Califórnia e Instituto de Tecnologia de Massachusetts. Secundariamente, destaca-se a presença de pesquisadores da China, demonstrando um interesse

de 14 instituições chinesas no domínio observado, de forma pulverizada.

Com base na investigação dos metadados dos estudos dos principais autores, destaca-se que tópicos abordados sobre coleta de dados no *corpus* analisado têm aderência aos aspectos ligados ao uso de recursos tecnológicos para obtenção de dados, principalmente pesquisas sobre (i) a comunicação de dados coletados pelos recursos via redes de computadores, (ii) aspectos ligados à segurança, (iii) definições de estruturas e propostas de processos de coleta de dados em dispositivos embarcados em ambientes e por dispositivos móveis, e (iv) aprendizagem de máquina.

Conclui-se que, a partir da sistematização desses estudos e da identificação dos principais pesquisadores, estes resultados podem servir como subsídio para a construção de um referencial teórico voltado para pesquisas que tratam ou tratarão o tema de coleta de dados ou sobre tópicos relacionados a emergência deste novo domínio.

As pesquisas destes autores também são recursos informacionais importantes para investigações interdisciplinares, fato justificado pelo fenômeno da coleta de dados ser objeto de estudo de diferentes áreas do conhecimento, tais como as ciências documentárias e a Ciência da Informação.

É importante destacar também a necessidade de estender estudos desta natureza utilizando também artigos recuperados por outras bases de dados em contextos de áreas distintas, a fim de comparar ou corroborar os resultados obtidos por esta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

CASTELLS, M. *A sociedade em rede*. São Paulo: Paz e Terra, 2008.

DAVENPORT, T. H. *Big data at work: dispelling the myths, uncovering the opportunities*. Harvard: Harvard Business School Publishing, 2014.

GLÄNZEL, W. *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. Bélgica: Course Handouts, 2003.

HJØRLAND, B. Domain analysis in information science: eleven approaches-traditional as well as innovative. *Journal of Documentation*, v. 58, n. 4, p. 422-462, 2002.
<https://doi.org/10.1108/0022041021043>

1136

IBM. International Business Machines Corporation. *10 Key Marketing Trends For 2017*. Estados Unidos da América, 2017. 18 p. Disponível em: <<https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN>>. Acesso em: 27 ago. 2018.

LE COADIC, Y-F. *A ciência da informação*. Brasília: Briquet de Lemos, 1996.

MANTONAVI, C.; DANTAS, G. G. C. Os Fluxos Informacionais nos Dispositivos Móveis. In: MOURA, M. A. (Org.). *Cultura informacional e*

liderança comunitária: concepções e práticas. 1 ed. Belo Horizonte: PROEX/UFMG, 2011.

PIOVEZAN, L. B.; FUJITA, M. S. L. Análise de cocitação de autores: uma aplicação em estudos de indexação. *Em Questão*, v. 21, n. 1, p. 110-129, 2015. <https://doi.org/10.19132/1808-5245211.110-129>

RODRIGUES, F. A. *Coleta de dados em redes sociais: privacidade de dados pessoais no acesso via Application Programming Interface*. Tese — Marília, Brasil: Universidade Estadual Paulista, 3 mar. 2017.

SANT'ANA, R. C. G. Ciclo de Vida dos Dados: Uma perspectiva a partir da

ciência da informação. *Informação e Informação*, v. 21, n. 2, p. 116-142, 2016. <http://dx.doi.org/10.5433/1981-8920.2016v21n2p116>

SALES, L. F.; CAVALCANTI, M. T. Seleção e avaliação de coleções de dados digitais de pesquisa: uma possível abordagem metodológica. *Informação & Tecnologia (ITEC)*, v. 2, n. 2, p. 88-105, 2015. Disponível em: <<http://www.periodicos.ufpb.br/ojs/index.php/itec/article/view/34134>>. Acesso em: 28 ago. 2018.

SMALL, H. A cocitation model of a scientific specialty: a longitudinal study of colleges research. *Social Studies of Science*, n.7, 1977

MOREIRA, F.M.; RODRIGUES, F.A.; SANT'ANA, R.C.G. *Análise de Domínio da Produção Científica Sobre Coleta de Dados no Contexto do Institute of Electrical and Electronic Engineers*. *Complexitas - Rev. Fil. Tem.* Belém, v. 3, n. 1, p. 28-39, jan./jun. 2018. Disponível em: <<http://www.periodicos.ufpa.br/index.php/complexitas/article/view/6634>>. Acesso em: 20 de fevereiro de 2019.
