



UMA ANÁLISE DO ACERVO DA REVISTA PRODUÇÃO ONLINE

AN ANALYSIS OF THE PRODUCTION ONLINE MAGAZINE COLLECTION

Gustavo Silva Semaan* E-mail: gustavosemaan@id.uff.br

Germano Lopes Ferraz* E-mail: germanoferraz@id.uff.br

Rodrigo Erthal Wilson* E-mail: rodrigoerthal@id.uff.br

Debora Alvernaz Corrêa** E-mail: deboradac@gmail.com

José André de Moura Brito*** E-mail: jambrito@gmail.com

* Universidade Federal Fluminense (UFF), Niterói, RJ

** Instituto Militar de Engenharia (IME), Urca, RJ

*** Escola Nacional de Ciências Estatísticas (ENCE-IBGE), Rio de Janeiro, RJ

Resumo: Para a realização de meta-análises do acervo da Revista Produção Online foram considerados o processo de Descoberta de Conhecimento em Bases de Dados e conceitos de Teoria dos Grafos. Os resultados obtidos possuem estatísticas gerais, análises das redes de colaboração, termos em destaque por meio do uso de mineração de texto e a identificação de autores considerados influentes, com base em frequência de publicação, e nas medidas de centralidade em grafos. O objetivo do trabalho é apresentar um panorama geral e de contribuições da revista, de maneira impessoal, apolítica e restrita a dados públicos. Os resultados obtidos apoiam interpretações, perspectivas e saberes deste importante acervo.

Palavras-chave: KDD. Teoria dos Grafos. Meta-análise. Medidas de Centralidade.

Abstract: In order to carry out meta-analyses of papers published in Production Online Magazine, the Knowledge Discovery in Databases process and Graph Theory were considered. The obtained results show general statistics, analysis of collaboration networks, text mining, and the identification of influential authors, based on centrality measures and the frequency of their publications. This paper aims to present an overview of the magazine collection, and the work was done impersonally, apolitical, and restricted to public data. The results obtained support interpretations, perspectives, and knowledge of this important collection.

Keywords: KDD. Graph Theory. Meta-analysis. Centrality Measures.

1 INTRODUÇÃO

A Revista Produção Online teve sua primeira edição no ano de 2001 e, até o presente momento, em seus 20 anos de colaborações, conta com 863 artigos publicados por 1.929 autores. Com foco na comunidade de Engenharia de Produção e áreas correlatas, a revista tem como objetivo promover a divulgação do conhecimento e o debate acerca de assuntos de interesse científico e da sociedade brasileira (PRODUÇÃO ONLINE, 2020).

O presente trabalho tem como objetivo principal fornecer uma visão geral da Revista por meio do uso de meta-análises, realizadas com base nos trabalhos publicados em seu acervo, de maneira impessoal, apolítica e restrita aos dados públicos. Para isso, além do cálculo de estatísticas gerais, foi considerado um processo de descoberta de conhecimento em bases de dados (KDD, do inglês *Knowledge Discovery in Databases*) (AGGARWAL, 2015) já utilizado em literatura recente, em diversos acervos de grande importância e de diversas áreas. Em especial, para a área de Engenharia de Produção, pode-se citar análises sobre o acervo do Simpósio de Engenharia de Produção (SIMPEP), que contempla mais de 11 mil artigos (SEMAAN, 2019b).

O conhecimento é construído a partir da troca de experiências e colaborações, que nem sempre são formalizadas em produtos como artigos científicos ou livros. De fato, em termos práticos, a identificação de tais experiências e colaborações ocorre mediante a aplicação de técnicas de meta-análises, que possibilitam determinar as relações de coautorias nos artigos ou mesmo em suas referências. Tendo em vista a demanda cada vez maior deste tipo de informação, estudos sobre comunidades em áreas específicas, colaborações e acervos têm sido alvo de intensa pesquisa em diversas áreas (NÓBREGA et al. 2019) (SEMAAN et al., 2019a) (OLIVEIRA, 2012).

Nesse sentido, de forma a possibilitar uma ampla e detalhada análise das redes de colaborações (do inglês *Collaborations Network*), incluindo a identificação de autores considerados mais influentes, foram utilizadas medidas de centralidade e outros conceitos de Teoria dos Grafos (CORTELETTE JUNIOR; CARNEIRO, 2012) (BOAVENTURA, 2001) (BORGATTI; EVERETT, 2006) (LIMA ET AL., 2017) (SEMAAN, 2019c).

Diversos trabalhos da literatura fornecem panoramas gerais de acervos importantes ou sobre uma temática específica. Para isso, metadados são utilizados como dados de entrada do processo, e comumente estatísticas gerais são construídas e formatadas com o intuito de fornecer diferentes perspectivas, que podem ser utilizadas tanto como indicadores importantes quanto no apoio à tomada de decisão.

Em relação a trabalhos relacionados, Vanz (2009) analisou a colaboração entre autores de cerca de 50 mil artigos brasileiros indexados na base *Web of*

Science. Rodrigues (2015) e Oliveira (2012) abordaram a comunidade de Sistemas de Informação (SBSI), enquanto Cavalcanti (2011) apresentou uma análise do Simpósio Brasileiro de Engenharia de Software (SBES). Lima *et al.* (2017) teve como foco as redes de colaborações do Simpósio Brasileiro de Banco de Dados (SBBD).

Fernandes (1999) organizou e mapeou os grupos atuantes na área de Informática na Educação, enquanto Magalhães *et al.* (2013), Procaci (2015) e Semaan *et al.* (2019a) analisaram o acervo (em diferentes momentos) e características do SBIE (Simpósio Brasileiro de Informática na Educação). Com mais aderência ao contexto Engenharia de Produção, Semaan *et al.* (2019b) e Semaan *et al.* (2019c) consideraram o acervo com 21 edições do SIMPEP (Simpósio de Engenharia de Produção), em que foi estabelecido um panorama geral sobre as publicações do evento e análises das redes de colaborações entre os autores.

Além do objetivo principal apresentado, seguem os objetivos específicos: (i) Processo de KDD: obter os dados no portal da revista, organizá-los e consolidá-los para formar uma base de dados íntegra e consistente; (ii) Estatísticas Gerais: construção e apresentação de dados estatísticos através da análise dos trabalhos com base nos atributos disponíveis; (iii) Palavras em destaque: os títulos dos trabalhos foram submetidos a tarefa de Mineração de Texto (do inglês *Text Mining*) do processo de KDD; (iv) Autoria e colaborações: após a consolidação do banco de dados é possível obter a quantidade de artigos publicados por autor e a relação de colaboração entre eles por meio de modelagens em grafos. (v) Construir uma análise que contemple os objetivos específicos, por meio de medidas de centralidade em grafos, da frequência de publicação e das colaborações dos autores; e, por fim, (vi) com base nas análises realizadas, identificar e apresentar uma relação de autores considerados influentes no contexto, de acordo com os critérios adotados.

2 METODOLOGIA

O presente trabalho considerou um processo de KDD baseado ao relatado recentemente na literatura (SEMAAN *et al.*, 2019a) e (SEMAAN *et al.*, 2019b), adaptado conforme as características específicas do acervo da Revista Produção Online e de seu portal (PRODUÇÃO ONLINE, 2020). A Figura 1 apresenta o

fluxograma utilizado, desde a coleta dos dados aos resultados, e as conclusões da pesquisa. Uma vez que o processo original foi citado, será apresentado brevemente nessa seção, e informações adicionais podem ser obtidas em suas referências.

Figura 1 – Ilustração dos passos do processo de KDD da literatura.



Fonte: Fluxograma do processo de KDD, adaptado de Semaan et. al (2019a).

Um processo de KDD é composto, basicamente, por três etapas principais: (i) *Pré-processamento*: responsável pela coleta, o enriquecimento, a consolidação, a formatação e a organização dos dados; (ii) *Mineração de Dados*: para obtenção de informações não triviais, novas e potencialmente úteis. Nesse caso em específico, consiste na análise dos termos de interesse com base nos títulos dos artigos publicados e em relação aos autores e suas colaborações; (iii) *Pós-processamento*: consiste na produção de artefatos sobre os resultados obtidos (textos, planilhas, gráficos) capazes de fornecer as informações aos especialistas do domínio da aplicação (AGGARWAL, 2015).

Com base na Figura 1, cada atividade discriminada é apresentada para facilitar o entendimento e possibilitar a reprodução do processo por outros pesquisadores. Destaca-se, novamente, que a metodologia associada ao processo base foi publicada em trabalhos recentes, e mais informações sobre tal podem ser obtidas nas referências contidas no texto.

As letras utilizadas na Figura 1 indicam atividades entre as etapas do Processo. Assim, a atividade a (Figura 1 (a)) consiste em obter os dados dos trabalhos da revista em seu portal (PRODUCAO ONLINE, 2020). Nesse sentido, destaca-se a atuação da equipe editorial deste relevante periódico, que contribui com democratização do acesso à informação. O acesso a todo o repositório é simples, rápido e transparente.

Para a coleta automática dos metadados do portal, comumente são utilizados aplicativos *webcrawlers*, como realizado no projeto de pesquisa responsável por esse artigo. Contudo, devido às características específicas do portal da revista, foi

decidido que ela ocorresse manualmente (Figura 1 (b)). Assim sendo, os dados foram copiados do *site*, colados e formatados em uma planilha eletrônica (Figura 1 (c)).

Em seguida, ainda na etapa de pré-processamento, iniciou-se o processo responsável por organizar, enriquecer, consolidar e formatar os dados coletados (Figura 1 (d)). Os títulos dos trabalhos foram segmentados em palavras, sendo necessário eliminar as *stopwords*, que são as palavras sem relevância semântica como conjunções e preposições. Os termos de interesse foram unificados, e para isso, foi utilizado o processo de redução ao radical ou equivalência à palavra semelhante mais frequente (ARANHA, 2007). Em relação ao nome dos autores, além do uso do cálculo da distância de edição (NAVARRO, 2001) (ou distância *Levenshtein*), foram necessárias intervenções humanas para decidir se dois nomes correspondiam a um mesmo autor e o sentido oposto, ou seja, se dois nomes semelhantes são realmente de uma mesma pessoa.

Na referida intervenção foi utilizada uma planilha eletrônica, e os 2.585 autores, após a unificação, resultaram em um total de 1.929. A redução de mais de 600 unidades deve-se ao fato de alguns autores que possuem mais de um artigo na revista, além dos casos em que seus nomes possuem erros de digitação, de grafia, abreviações e(ou) omissões de parte(s) do sobrenome.

Por fim, a consolidação dos dados resultou em listas de termos de interesse das publicações, extraídos dos títulos e de autores unificados (Figura 1 (e)). Uma vez que os autores possuem um código único que possibilita a sua identificação, inicia-se a construção dos grafos de colaboração com base nas coautorias existentes (Figura 1 (f) e (g)), e a realização dos cálculos das medidas de centralidades. Por fim, a atividade da Figura 1 (h) consiste na produção dos artefatos que relatam os resultados obtidos, sejam eles: planilhas, gráficos, grafos e, inclusive, o presente artigo Figura 1 (i).

3 PANORAMA GERAL DO ACERVO

Para apresentar um panorama geral da revista, foi coletado todo seu acervo até o momento, que corresponde a 20 anos de publicações, em um total de 1.929 autores e 863 artigos. A Tabela 1 apresenta os quantitativos de trabalhos publicados

(coluna Artigos), de autores, médias de autores por artigo (MA) e autores que publicaram trabalhos de maneira isolada (coluna SA, do inglês *single-authors*).

Tabela 1 – Estatísticas Gerais dos metadados da revista Produção Online (PRODUÇÃO ONLINE, 2020)

Ano	Artigos	Autores	MA	SA	SA (%)
2001	11	33	3,00	0	0
2002	10	29	2,90	2	20
2003	35	91	2,60	4	11
2004	28	75	2,68	3	11
2005	41	105	2,56	4	10
2006	30	79	2,63	6	20
2007	27	80	2,96	2	7
2008	55	151	2,75	3	5
2009	40	104	2,60	3	8
2010	40	124	3,10	4	10
2011	48	127	2,65	3	6
2012	48	132	2,75	2	4
2013	60	169	2,82	1	2
2014	60	191	3,18	3	5
2015	60	199	3,32	2	3
2016	60	222	3,70	0	0
2017	60	191	3,18	1	2
2018	60	184	3,07	3	5
2019	60	201	3,35	5	8
2020	30	98	3,27	2	7

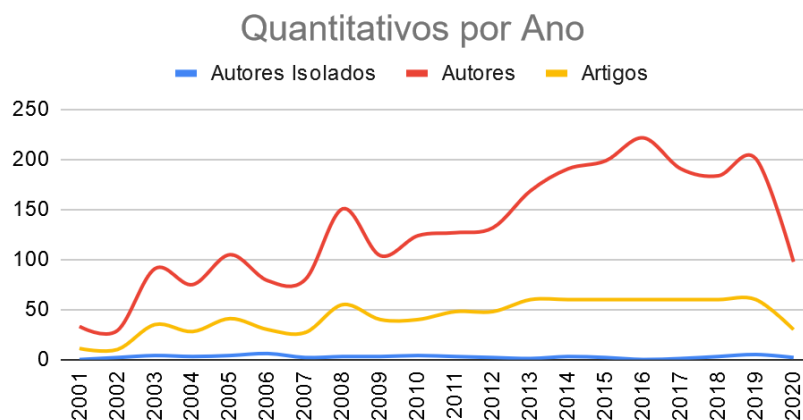
Fonte: Autoria própria

Com o intuito de melhorar a visualização e facilitar a interpretação das análises, a Figura 2 apresenta um gráfico com os quantitativos de autores, de artigos e de SA por ano. Em relação à quantidade de trabalhos publicados e de autores por ano, embora ocorram variações, é possível observar seus crescimentos ao longo dos anos. Apesar de terem sido consideradas, em um primeiro momento, análises das linhas de tendências por técnicas de regressão, optou-se por removê-las dos gráficos com o intuito de melhorar suas visualizações. Destaca-se que para o ano 2020 estão disponíveis apenas dois números do volume 20. Conseqüentemente, o gráfico reflete a queda nos quantitativos pela ausência dos volumes que ainda serão publicados nesse ano.

Ainda sobre a Figura 2, observa-se um baixo quantitativo de Autores Isolados em relação aos valores dos outros atributos. Portanto, torna-se necessário, para uma melhor visualização, um gráfico com o percentual de SA por ano (Figura 3).

Novamente linhas de tendências por técnicas de regressão foram utilizadas e confirmam queda.

Figura 2 – Gráfico Quantitativo de Artigos por Ano

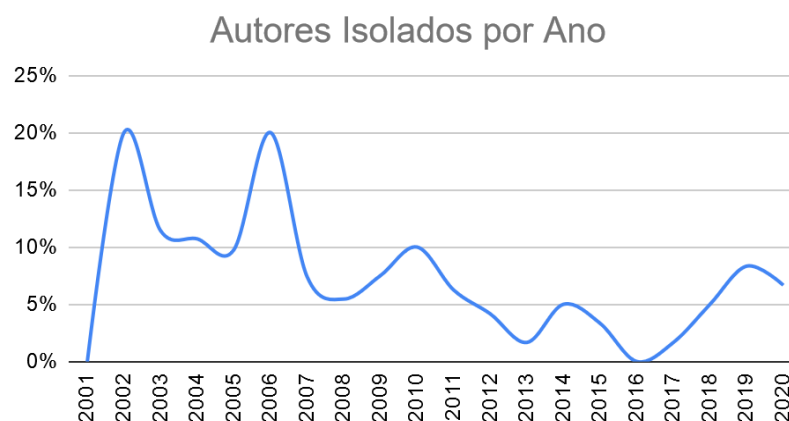


Fonte: Autoria própria.

Conforme Figura 3, existem picos nos percentuais de SA nos anos de 2002 e 2006 (de 20%), enquanto o percentual médio do acervo é de 6,1%. Apenas como exemplo, conforme Semaan *et al.* (2019c), o percentual de SA do acervo do SIMPEP teve pico de 30% em 2004, mas o percentual geral é de apenas 3%, em um acervo com mais de 11 mil artigos. Destaca-se, entretanto, que existem características específicas nos processos avaliativos da revista e do evento considerado, inclusive, o próprio modelo editorial (entre uma revista/periódico e anais de um evento científico).

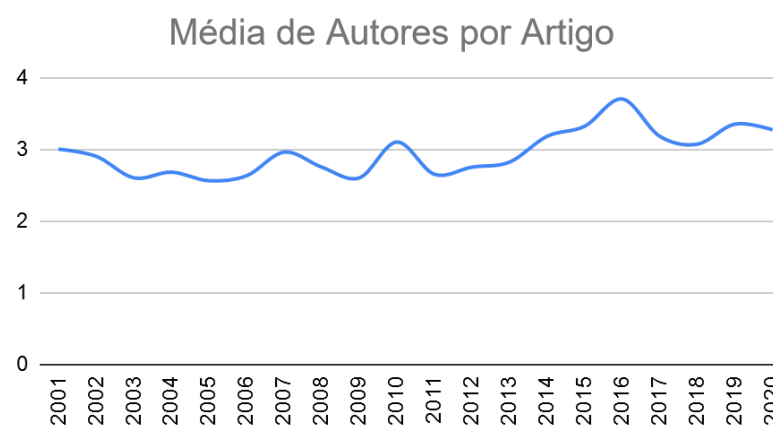
A Figura 4 ilustra o gráfico Média de Autores por Artigo, em que é possível observar um pequeno crescimento na colaboração entre coautores. De fato, novamente, linhas de tendências por técnicas de regressão indicam crescimento. O crescimento da média de autores por artigo em conjunto à tendência de declínio de SA pode ser considerado um indício de aumento de colaborações entre autores. Destaca-se que para o ano 2020 foram considerados apenas 2 números publicados.

Figura 3 – Gráfico Autores Isolados (SA) por Ano



Fonte: Autoria própria.

Figura 4 – Média de Autores por artigo

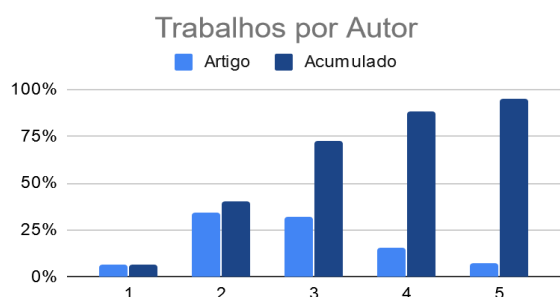


Fonte: Autoria própria.

Ainda em relação ao ponto de vista colaboração, conforme o gráfico Trabalhos por Autor(es) (Figura 5), cerca de 90% dos trabalhos possuem entre dois e cinco coautores, e apenas 6% são SA. De modo atípico, embora mais de 95% dos trabalhos tenham até cinco autores, existem 2 trabalhos com 9 (co)autores.

A Figura 6 apresenta o gráfico Autores por Trabalho, em que é possível observar que cerca de 85% dos autores participaram de apenas um trabalho, e cerca de 10% dos autores publicaram dois artigos. Novamente existem exceções, com sete autores que publicaram dez artigos ou mais, e um autor que possui 26 trabalhos no acervo.

Figura 5 – Gráfico Trabalhos por Autor(es)



Fonte: Autoria própria.

Figura 6 – Gráfico Autor(es) por Trabalho



Fonte: Autoria própria.

Destaca-se que os gráficos das Figuras 5 e 6 os valores no eixo das abscissas foram limitados a 5 unidades devido à visualização, sem prejuízo às análises. Além disso, para ambos os gráficos, os valores acumulados foram apresentados com o intuito de facilitar a visualização e fornecer de maneira mais clara informações adicionais.

As Figuras 7 e 8 relatam as 20 palavras mais frequentes nos títulos dos artigos disponíveis no acervo da revista, em que a primeira detalha o percentual por ano, enquanto a Figura 8 apresenta os percentuais em relação à todo o acervo. Em destaque, a palavra *análise* consta em cerca de um a cada cinco trabalhos do acervo, com pico de cerca de 33% de ocorrência nos anos de 2017 e 2018.

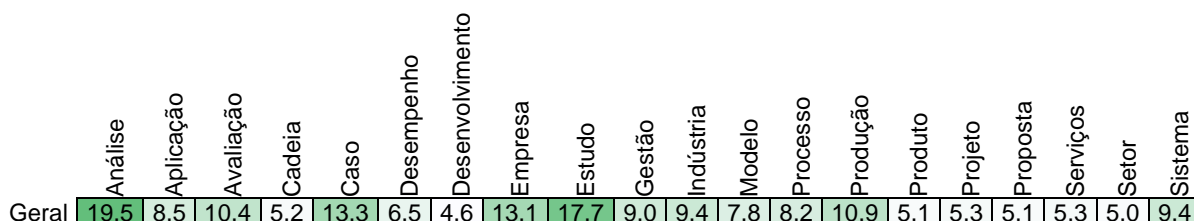
Figura 7 – Percentual das 20 palavras mais frequentes por ano

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Análise	27	10	17	14	10	7	11	13	15	18	8	21	17	28	28	12	32	33	23	23
Aplicação	9	0	0	14	2	17	4	0	10	8	15	10	2	8	3	8	10	8	20	20
Avaliação	9	0	3	0	7	10	7	5	13	13	10	21	15	17	8	10	15	8	8	10
Cadeia	9	0	6	4	7	7	4	7	3	10	4	6	5	3	7	3	5	5	5	3
Caso	18	0	9	11	17	17	26	15	13	23	19	10	13	7	8	12	8	10	18	20
Desempenho	0	0	6	0	7	0	15	11	3	10	6	13	10	8	7	7	3	7	3	0
Desenvolvimento	9	0	11	7	2	3	0	7	5	3	0	8	12	3	7	0	2	7	3	0
Empresa	18	20	14	7	5	13	7	11	10	28	13	8	20	22	15	13	8	8	12	13
Estudo	27	10	6	14	20	7	11	20	8	25	25	23	18	12	23	15	17	17	23	27
Gestão	9	10	6	4	7	13	7	18	3	5	10	6	12	15	8	5	8	10	12	3
Indústria	9	0	6	7	10	7	11	11	8	5	19	8	7	7	10	13	3	12	13	13
Modelo	9	10	3	7	7	3	0	5	18	10	8	10	12	5	7	10	7	8	8	3
Processo	0	0	3	4	7	3	7	5	0	15	13	8	12	17	13	5	5	12	7	7
Produção	0	0	17	4	12	3	15	7	3	20	8	6	13	12	10	13	10	17	8	23
Produto	0	0	14	11	5	7	4	7	0	8	4	6	8	7	2	5	3	2	2	7
Projeto	0	10	14	7	7	0	4	5	3	8	10	2	8	3	5	7	3	2	5	3
Proposta	9	0	6	0	0	10	7	4	8	3	4	8	5	2	8	8	3	2	8	7
Serviços	0	10	0	0	10	3	0	5	5	5	4	6	7	7	12	5	8	2	5	3
Setor	9	0	0	0	2	7	7	7	8	8	6	6	5	3	5	7	3	3	8	0
Sistema	9	0	11	7	17	13	22	5	13	10	8	6	3	13	5	7	10	12	8	10

Fonte: Autoria própria.

Destacam-se, também, com ocorrências superiores a 10% na Figura 8, as palavras *estudo*, *caso*, *empresa*, *produção* e *avaliação*, relatadas em ordem decrescente de frequência. De fato, essas palavras e a palavra *análise* têm grande importância e aderência ao foco da revista e à área de Engenharia de Produção.

Figura 8 – Percentual das 20 palavras mais frequentes no acervo



Fonte: Autoria própria.

Com o intuito de apresentar os termos em destaque, foram geradas duas nuvens de palavras (do inglês *word cloud*). Trata-se de uma maneira de exibir a frequência de um conjunto de palavras em um texto ou em um conjunto de dados. No presente contexto, autores mais frequentes ficam em destaque em relação à disposição e ao tamanho da fonte.

A nuvem da Figura 9 contém todas as palavras de todos os artigos do acervo da revista, em um total de 2.036 termos após o processo de unificação e consolidação da base de dados, e pode-se observar o destaque das palavras que possuem alta frequência, conforme relatado anteriormente. De modo adicional, a nuvem da Figura 10 apresenta apenas os termos com frequência igual ou superior a cinco, em um total de 338 unidades.

Figura 9 – Nuvem de Palavras com todos as palavras dos títulos do acervo da revista



Fonte: Autoria própria.

Figura 10 – Nuvem de Palavras com todas as palavras com frequência 5 ou superior



Fonte: Autoria própria.

4 REDES DE COLABORAÇÃO E INFLUÊNCIA DE AUTORES

Embora as Estatísticas Gerais sejam importantes para análises de repositórios em geral e de comunidades, não são suficientes para fornecer uma visão holística. Deve-se buscar informações complementares, como análises das redes de colaborações entre autores e suas influências individuais. (BORGATTI; EVERETT, 2006) (SEMAAN *et al.*, 2019a) (BAVELAS, 1950) (BEAUCHAMP, 1965).

O conhecimento é construído a partir da troca de experiências e colaborações entre autores e podem ocorrer de diversas maneiras. Por exemplo, desde os *feedbacks* de revisores artigos submetidos a um breve encontro informal “*pelos corredores*” de uma instituição de ensino, uma simples dica pode ser suficiente para repensar a abordagem considerada e redirecionar uma trajetória de pesquisa (OLIVEIRA, 2012).

Nem todas as formas de colaboração são formalizadas, e trata-se de um ponto fraco ao considerar metodologias que atuam com metadados, uma vez que eventuais contribuições não são passíveis de detecção se não foram explicitadas no texto, seja por meio da adição da pessoa como coautora ou do uso de seu(s) trabalho(s) como referência(s) (OLIVEIRA, 2012) (NÓBREGA *et al.* 2019).

Assim, embora não seja possível a obtenção de informações que poderiam enriquecer as análises, o uso de metadados é de grande importância, e não diminui a qualidade dos resultados obtidos com técnicas de meta-análises. Mais especificamente, tais técnicas possibilitam identificar as colaborações entre pesquisadores, principalmente por meio das relações de coautoria e com base nas referências utilizadas, comumente através de modelagens e conceitos de Teoria dos

Grafos (SANTOS; SILVA, 2019) (PROCACI *et al.*, 2015) (MAGALHÃES *et al.*, 2013) (NEWMAN, 2001).

Um grafo simples $G = (V, A)$ é definido como uma estrutura composta de um conjunto de vértices V e de arestas A , na qual cada aresta conecta dois vértices distintos. A modelagem utilizada nesse trabalho é amplamente utilizada na literatura sobre a temática redes de colaboração, em que cada autor é representado por um vértice e os autores que atuaram em um trabalho são conectados por arestas. Com frequência utiliza-se também o termo rede para em referência a grafo (BOAVENTURA NETTO, 2001).

A interação entre pessoas em um contexto específico, por exemplo em uma empresa, é considerada de grande importância. Muito antes do advento das tecnologias modernas de comunicação, a noção de centralidade em redes sociais foi introduzida por (BAVELAS, 1950), como uma ferramenta capaz de identificar e mensurar a “*importância*” de um dado indivíduo com base em sua posição estrutural na rede.

Segundo Bavelas (1950) “[...] *num grupo de pessoas, um particular indivíduo que se encontra estrategicamente localizado num caminho mais curto de comunicação entre pares de indivíduos, está numa posição mais central da rede [...]*”. De maneira adicional, devido a sua localização privilegiada, “[...] *esse indivíduo deve ser responsável por transmitir, modificar, ou reter a informação entre membros do grupo [...]*”. As medidas de centralidades estão relacionadas à *influência* de indivíduos em uma rede, o que no presente contexto se traduz em mensurar autores em destaque na comunidade da Revista Produção Online com base em seu acervo.

Mais especificamente, em relação ao tópico centralidade em redes, embora existam diversas medidas que são aplicadas de forma a atender diferentes áreas e objetivos (BORGATTI; EVERETT, 2006), em Freeman (1978) é proposta a redução para três medidas clássicas: a Centralidade de Grau (do inglês *Degree Centrality*), de Proximidade (do inglês *Closeness Centrality*) e de intermediação (do inglês *Betweenness Centrality*).

Conforme Freitas (2010) e Moxley e Moxley (1974), redes sociais reais são comumente modeladas em grafos de grande porte e desconexos. Assim, Freeman (1977) pesquisou e propôs medidas baseadas em interligações entre pares de vértices que poderiam ser aplicadas também em grafos desconexos.

Mais especificamente, a medida de centralidade de proximidade não é adequada a grafos desconexos, e no contexto de modelagem de redes de colaborações em acervos, os grafos gerados frequentemente não são conexos. Em outras palavras, como um exemplo, cada artigo forma uma *clique*, que corresponde a um (sub)grafo em que todos os pares de vértices (autores) estão conectados por arestas (colaborações), mas não há garantia que algum desses autores colaborou em ao menos um segundo trabalho com qualquer outro autor (BOAVENTURA NETTO, 2001). Tal clique, portanto, poderia estar isolado, o que por consequência torna o desconexo grafo que modela o acervo.

A medida de centralidade de grau, ou simplesmente grau de um vértice v_k , $C_D(v_k)$, considera a quantidade de arestas conectadas a v_k , e pode ser interpretada como a probabilidade desse vértice receber alguma informação da rede. Nesse contexto, trata-se da *influência* direta que um vértice tem em relação a seus vizinhos (vértices adjacentes, VA). Assim, destacam-se autores que possuem mais relações com outros autores, que são considerados os mais influentes conforme esse critério (FREEMAN, 1977).

Conforme Freitas (2010), formalmente, dado um grafo G (conexo ou desconexo) com n vértices e v_k um vértice de G , a centralidade de grau desse vértice corresponde à quantidade de arestas incidentes a ele, em que a_{kj} são elementos de uma matriz de adjacência (Equação 1).

$$C_D(v_k) = \sum_{j=1}^n a_{kj} \quad (1)$$

Já a medida de centralidade de intermediação busca identificar vértices importantes em relação ao controle da comunicação, e formalmente é definida como dado um grafo G (conexo ou desconexo) com n vértices e v_k é um vértice de G . Considere um par de vértices v_i e v_j em G , tal que $i \neq j$, $i \neq k$ e $j \neq k$. A intermediação parcial de v_k com respeito a v_i e v_j é dada pela Equação 2, em que g_{ij} corresponde à quantidade de distâncias geodésicas (menor caminho entre dois vértices) entre v_i e v_j , e $g_{ij}(v_k)$ denota o número de geodésicas entre v_i e v_j que passam por v_k .

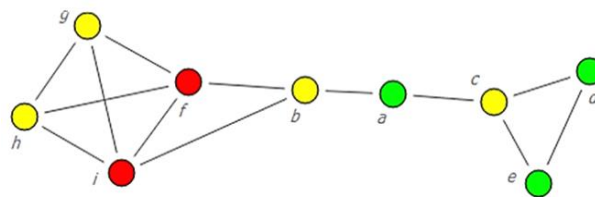
$$b_{ij}(v_k) = \begin{cases} 0, & \text{se não existir caminho entre } v_i \text{ e } v_j; \\ \frac{g_{ij}(v_k)}{g_{ij}}, & \text{caso contrário,} \end{cases} \quad (2)$$

$$c_B(v_k) = \sum_{\substack{1 \leq i < j \leq n \\ i, j \neq k}} b_{ij}(v_k) \quad (3)$$

De modo a introduzir conceitos necessários para a análise realizada nesse trabalho, um acervo exemplo com cinco artigos e nove autores é usado, conforme ilustra a Figura 11. Além da tabela com as relações de autoria, em que o artigo com *Id* 1 foi publicado pelos autores *a* e *b*, é apresentado uma modelagem em grafos.

Figura 11 – Acervo e grafo utilizados como exemplos

Id	Artigo
1	a, b
2	a, c
3	c, d, e
4	b, f, i
5	f, g, h, i



Fonte: Autoria própria.

Ainda com base na Figura 11, o grafo possui uma representação em que as cores dos vértices indicam seus graus. A Figura 12 apresenta um quadro com as frequências (coluna Freq.), os artigos que cada autor participou, as medidas de centralidade de grau (C_D), os vizinhos adjacentes (VA) e os valores de centralidade de intermediação (C_B) para cada autor.

Figura 12 – Resultados obtidos para o acervo exemplo e a Matriz *M* para cálculo de $C_B(a)$

Autor	Freq.	Artigos	C_D	VA	C_B
a	2	1 e 2	2	b, c	15
b	2	1 e 3	3	a, f, i	16
c	2	2 e 3	3	a, d, e	12
d	1	3	2	c, f	0
e	1	3	2	c, d	0
f	2	4 e 5	4	b, g, h, i	5
g	1	5	3	f, h, i	0
h	1	5	3	f, g, i	0
i	2	4 e 5	4	b, f, g, h	5

$C_B(a)$	a	b	c	d	e	f	g	h	i
a	0	0	0	0	0	0	0	0	0
b		0	1	1	1	0	0	0	0
c			0	0	0	1	1	1	1
d				0	0	1	1	1	1
e					0	1	1	1	1
f						0	0	0	0
g							0	0	0
h								0	0
i									0

Matriz *M* para cálculo de $C_B(a)$

Fonte: Autoria própria.

Adicionalmente, na Figura 12 foi apresentada a estrutura de dados (matriz *M* com entradas $M[i,j]$) considerada para o cálculo da centralidade de intermediação do autor *a* ($C_B(a)$), em que $M[i,j] = M[j,i]$ e $M=[i,i] = 0$. Destaca-se que a ilustração tem

caráter didático, e que há outras estruturas de dados mais adequadas, que não desperdiçam espaço (memória) são utilizadas na prática.

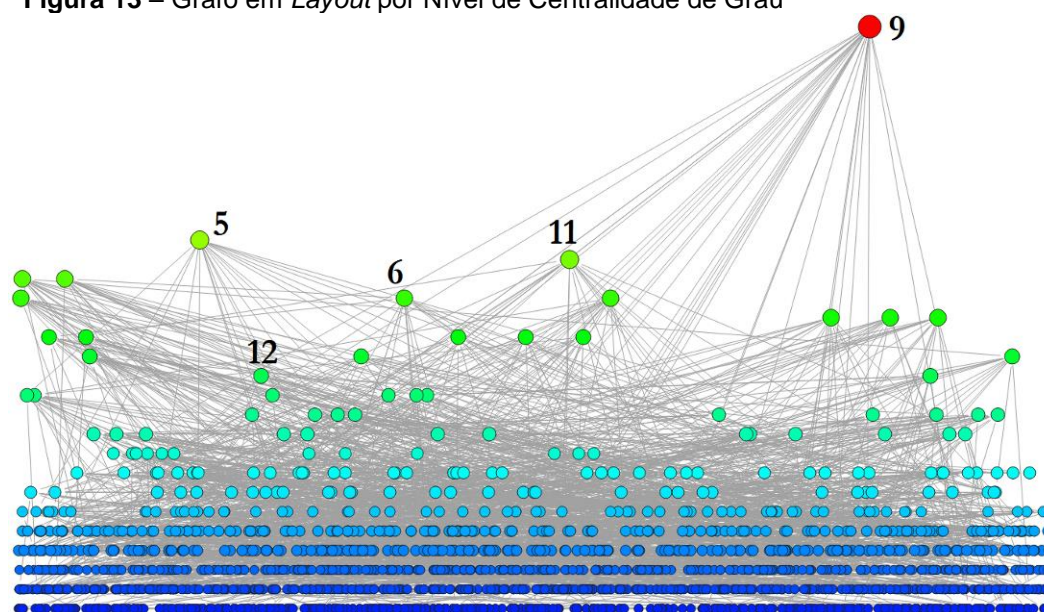
Com base no acervo exemplo e em seu grafo, os vértices *a* e *b* estão em destaque do ponto de vista de intermediação, uma vez que são utilizados como *pontes* no caminho geodésico entre muitos pares de autores. Tais vértices estão em destaque também conforme sua frequência, e possuem 2 artigos. De modo adicional, os autores *f* e *i* são considerados influentes com base na centralidade de grau, e possuem mais colaborações, com $C_D=4$.

Conforme o exemplo ilustrado pelas figuras 11 e 12, o presente trabalho irá destacar autores influentes com base em: (i) sua a frequência de publicação na revista, informação não contemplada nos grafos; (ii) na centralidade de grau, capaz de mensurar sua influência com base na quantidade de coautores que atuaram em parceria em seus artigos; e (iii) na centralidade de intermediação, que verifica se o autor está no caminho de comunicação entre muitos pares de autores na rede.

As Figuras 13 e 14 apresentam o Grafo 1 em diferentes *layouts*, construído com todos os 1.886 autores (vértices) que publicaram artigo em colaboração com outros autores, após o processo de unificação e consolidação de seus nomes. Na Figura 13 o grafo possui layout em nível por centralidade de grau, e quanto mais próximo ao limite superior da imagem um vértice está localizado, maior é a influência do autor nele representado.

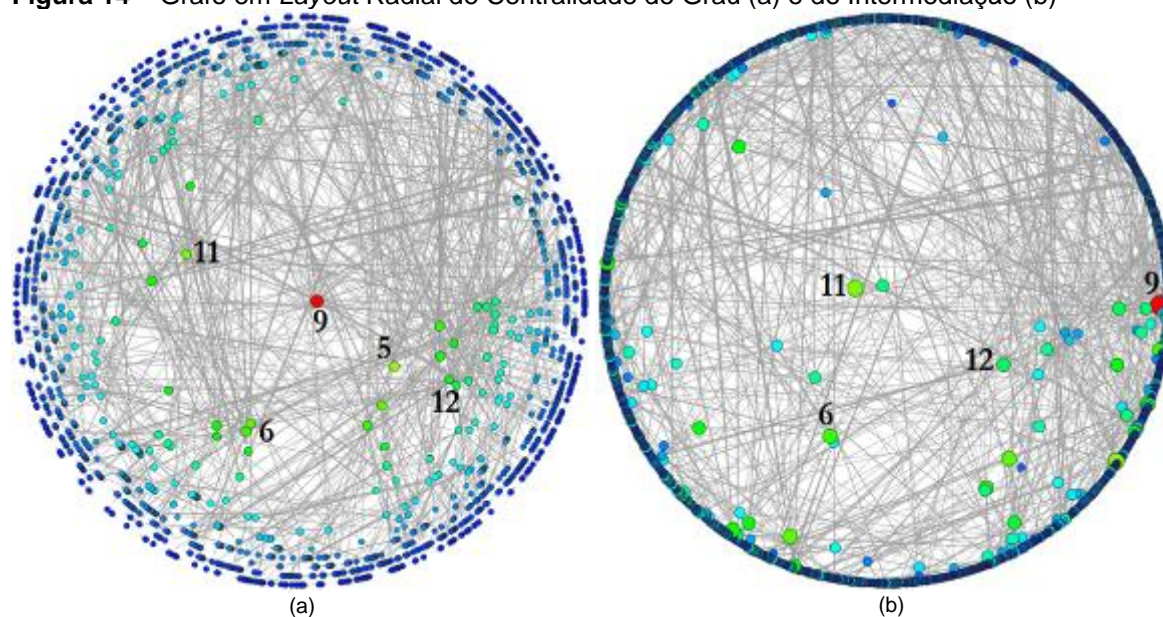
Os grafos ilustrados nas Figuras 14 (a) e 14 (b) estão em *layout* Radial por centralidade de grau e de intermediação, respectivamente. Nesse layout os vértices mais influentes, conforme a centralidade considerada, estão mais centrais na imagem, enquanto os menos influentes ficam nas margens do círculo formado. Destaca-se que nas Figuras 13 e 14 as cores e os tamanhos dos vértices estão relacionados aos seus graus, independente do layout e da medida de centralidade utilizados nas ilustrações.

Figura 13 – Grafo em *Layout* por Nível de Centralidade de Grau



Fonte: Autoria própria.

Figura 14 – Grafo em *Layout* Radial de Centralidade de Grau (a) e de Intermediação (b)



Fonte: Autoria própria.

Os números existentes nessa figura, assim como nas figuras 13 e 14 correspondem aos identificadores únicos (coluna Id) de alguns autores reportados na Tabela 2, e têm como objetivo ilustrar de maneira efetiva a participação e influência deles em cada layout considerado.

Tabela 2 – Autores em destaque, considerados influentes no acervo da Revista

ID	Nome	Freq.	C _D	C _B
1	Cláudia F. Gohr	8	18	32 ^a
2	Edson P. Lima	7	15	97 ^a
3	João L. Kovaleski	9	14	34 ^a
4	José L. D. Ribeiro	11	17	19 ^a
5	Leonardo Ensslin	13	20	104 ^a
6	Luciano C. Santos	7	17	7 ^a
7	Luiz Alberto Pilatti	10	18	18 ^a
8	Marly M. Carvalho	10	15	26 ^a
9	Miguel A. Sellitto	26	31	63 ^a
10	Oswaldo L. G. Quelhas	9	16	115 ^a
11	Paulo A C. Miguel	8	19	2 ^a
12	Tarcisio A. Saurin	12	13	6 ^a

Fonte: Autoria própria.

A partir da consolidação da base de dados e da criação dos grafos, foi realizado o cálculo das centralidades dos vértices, sendo criadas duas listas com autores em destaque. A primeira lista possui os 20 autores com mais artigos publicados, enquanto a segunda lista relaciona os 20 autores com maior grau de colaboração. A Tabela 2 apresenta a interseção dessas listas, ou seja, 12 autores que pertencem as duas relações, em destaque tanto pela quantidade de artigos publicados quanto pela quantidade de colaboradores (NEWMAN, 2004a) (NEWMAN, 2004b).

Adicionalmente, além da quantidade de artigos (coluna Freq.) e da centralidade de grau (coluna C_D), a Tabela 2 possui a coluna C_B, que indica a posição (colocação) do autor em relação à medida centralidade de intermediação, dentre todos os 1.886 autores representados no Grafo 1. Embora não seja o objetivo da pesquisa, tal tabela presta uma singela homenagem ao destacar autores influentes, que contribuíram com o crescimento da Revista Produção Online.

Novamente nuvens de palavras são utilizadas, porém para apresentar os autores em destaque. Assim, autores mais frequentes ficam em destaque em relação à disposição e ao tamanho da fonte. A nuvem apresentada na Figura 15 possui todos os 1.928 autores do acervo, enquanto a Figura 16 apresenta os 315 autores que possuem ao menos 2 artigos publicados na revista.

Figura 15 – Nuvem de Palavras com todos os autores do acervo da Revista Produção Online



Fonte: Autoria própria.

Figura 16 – Nuvem de Palavras com autores com 2 artigos publicados na revista



Fonte: Autoria própria.

5 CONCLUSÕES E TRABALHOS FUTUROS

O presente trabalho teve como o objetivo principal apresentar uma visão geral do acervo da revista Produção online e, para isso, foram considerados a apresentação de estatísticas gerais, o processo de descoberta de conhecimento em bases de dados, conceitos de teoria de grafos.

Os metadados foram coletados diretamente do portal da revista, com simplicidade de acesso, organização e transparência. De fato, trata-se de um acervo composto de trabalhos de alta qualidade e disponível à comunidade, contribuindo com a democratização do conhecimento.

O acervo da revista conta, no momento, com 71 edições em 20 anos de publicações, com 863 artigos publicados por 1.929 autores. A seção Metodologia apresentou o processo de KDD da literatura, adaptado conforme as características específicas da revista, além de apresentar conceitos de Teoria dos Grafos necessários para o entendimento dos resultados obtidos, em especial, sobre medidas de centralidade e de Redes de Colaboração.

Na Introdução, além de uma breve apresentação da revista, foram citados trabalhos relevantes que abordam repositórios de importantes periódicos e eventos, além de ressaltar a importância da troca de experiências e das colaborações no processo de construção do conhecimento.

Na seção Panorama Geral foi possível observar o crescimento dos quantitativos de autores e de artigos ao longo dos anos. Com base nos gráficos apresentados foram detectados indícios do aumento da colaboração entre autores, em que por linhas de tendências e por técnicas de regressão linear (HAIR, 2018) confirmam, concomitantemente, o crescimento da média de autores por artigo e a redução de trabalhos publicados por SA.

Nas análises de redes de colaborações foram apresentados grafos com layout Radial por Centralidade de Grau e de Intermediação. A Tabela 2 relaciona os 12 autores que pertencem as duas listas formadas, sejam elas: (i) autores que publicaram mais artigos; e (ii) autores que possuem mais colaborações. Portanto, tal relação corresponde aos 12 autores considerados mais influentes de acordo com os critérios adotados nessa pesquisa.

De maneira adicional, uma nuvem de palavras contém todos os autores que publicaram na revista, e fornece uma significativa imagem da comunidade da Revista Produção Online.

O panorama apresentado possui estatísticas gerais, análises em relação aos termos frequentes em títulos dos trabalhos, as colaborações através de modelagem em grafos e medidas de centralidade, e destaque aos autores selecionados como influentes conforme os critérios adotados. A abordagem considerou análises neutras, impessoais, apolíticas e restritas aos dados coletados em seu portal, e o trabalho apresenta-se como uma importante referência para pesquisas futuras.

Como trabalhos futuros pretende-se: (i) obter mais dados para análises das colaborações interinstitucionais, (ii) o uso das referências como fonte de dados para analisar as colaborações; (iii) avaliar outras modelagens em grafos ponderados ou de dígrafos; (iv) e continuar as pesquisas abordando outros acervos importantes no cenário nacional e internacional; (v) usar a técnica de análise de agrupamentos para segmentar a base em função das variáveis associadas à publicação e à cooperação.

REFERÊNCIAS

- AGGARWAL, C.C. **Data Mining**: the Textbook, Heidelberg:Springer, 2015.
Disponível em: <https://link.springer.com/book/10.1007%2F978-3-319-14142-8>
- ARANHA, C. N. **Uma abordagem de pré-processamento automático para mineração de textos em português**: sob o enfoque da inteligência computacional. Tese (Doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brasil, 2007.
- BAVELAS, A. Communication patterns in task oriented groups. **Journal of the Acoustical Society of America**, 22, 1950. <https://doi.org/10.1121/1.1906679>
- BEAUCHAMP, M. A. An improved index of centrality. **Behavioral Science**, v. 10, 1965. <https://doi.org/10.1002/bs.3830100205>
- BOAVENTURA NETTO, P. O. **Grafos**: teoria, modelos, algoritmos. Editora Blucher, 2001.
- BORGATTI, S.P. EVERETT, M.G. A Graph-theoretic perspective on centrality. **Social Networks**, v. 28, n. 4, 2006. <https://doi.org/10.1016/j.socnet.2005.11.005>
- CAVALCANTI, T. R; DA SILVA, F. Q. B. Historical, conceptual, and methodological aspects of the publications of the brazilian symposium on software engineering: a systematic mapping study. *In*: BRAZILIAN SYMPOSIUM ON SOFTWARE ENGINEERING (SBES), 25., 2011. **Anais [...]**. São Paulo, 2011. <https://doi.org/10.1109/SBES.2011.36>
- CORTELETTE JUNIOR, M., CARNEIRO, T.C.J. Análise de redes sociais na indústria farmacêutica: um estudo com farmácias e drogarias. **Revista Produção Online**, v.12, n. 4, 2012. <https://doi.org/10.14488/1676-1901.v12i4.1006>
- FERNANDES, C. T.; SANTOS, N. Pesquisa e desenvolvimento em informática na educação no Brasil – Parte I. **Revista Brasileira de Informática na Educação**, 4, 1999. <http://dx.doi.org/10.5753/rbie.1999.4.1.9-32>.
- FREEMAN, L. C. A set of measures of centrality based on betweenness. **Sociometry**, v. 40, n. 1, pp. 35-41, 1977. <https://doi.org/10.2307/3033543>
- FREEMAN, L. C. Centrality in social networks: conceptual clarification. **Social Networks**, 1978.[https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- FREITAS, L. Q. **Medidas de centralidade em grafos**. Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia de Produção, COPPE, da Universidade Federal do Rio de Janeiro, 2010.
- HAIR, J.F; BABIN, W.C.; ANDERSON, J.B. and BLACK, W.C. **Multivariate data analysis**, 8th Edition, 2018.

LIMA, L.H.C., G. PENHA, L.M.A. ROCHA, M.M. MORO, A.P.C. SILVA, A. H. F. LAENDER, J.P. M.. The collaboration network of the brazilian symposium on databases - 30 editions of history. **Journal of the Brazilian Computer Society** v. 23, n. 10, 2017. <https://doi.org/10.1186/s13173-017-0059-6>

MAGALHÃES, C.V.C, SANTOS, R.E.S., SILVA, F.Q.B., Gomes, A.S. Caracterizando a Pesquisa em Informática na Educação no Brasil: Um Mapeamento Sistemático das Publicações do SBIE. *In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO - SBIE*, 24., 2013. **Anais [...]**. <http://dx.doi.org/10.5753/cbie.sbie.2013.22>

NAVARRO, G. (2001). A guided tour to approximate string matching. *ACM Surveys* 33. <https://doi.org/10.1145/375360.375365>

NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration. *In: Proceedings of the National Academy of Sciences*, 2004a.

NEWMAN, M. E. J. The structure of scientific collaboration networks. **Proc. Natl. Acad. Sci.USA**, v. 98, 2001. <https://doi.org/10.1073/pnas.98.2.404>

NEWMAN, M. E. J. *Who is the best connected scientist? a study of scientific coauthorship networks*. *In: BEN-NAIM, E.; FRAUENFELDER, H.; TOROCZKAI, Z. (Ed.). Complex Networks*. Springer Berlin / HEIDELBERG, 2004, (Lecture Notes in Physics, v. 650), 2004b. https://doi.org/10.1007/978-3-540-44485-5_16

NÓBREGA, B.M.; MOURA, N.R.; GOHR, C.F.; GONÇALVES, J.M.S. Capacidades colaborativas e criação de valor: revisão sistemática e agenda de pesquisa. **Revista Produção Online**, v.19, n.4, 2019. <https://doi.org/10.14488/1676-1901.v19i4.3214>

OLIVEIRA, E. A. **Sobre a colaboração na comunidade de sistemas de informação através dos simpósios SBSI**. Dissertação (Mestrado) - UNIRIO, Rio de Janeiro, Brasil, 2012.

PROCACI, T.B., SIQUEIRA, S.W.M., PINHATI, F., NUNES, B.P. Estudo exploratório das produções e colaborações entre pesquisadores em informática na educação: uma análise de publicações do simpósio brasileiro de informática na educação de 2001 a 2013. SBIE, 2015. <http://dx.doi.org/10.5753/cbie.sbie.2015.1323>

PRODUÇÃO ONLINE. **Portal da revista produção online**, Disponível em <https://www.producaoonline.org.br/>. Acesso em: 01 ago. 2020.

RODRIGUES, N.S.; RALHA, C.G. Conhecendo a comunidade de sistemas de informação no brasil: um estudo comparativo utilizando diferentes abordagens de banco de dados. *In: BRAZILIAN SYMPOSIUM ON INFORMATION SYSTEM*, 11., 2015. **Anais [...]**. Goiânia, GO, 2015. <https://doi.org/10.5753/sbsi.2015.5861>

SANTOS, T.F.; SILVA, M.A.V. Transporte público e sustentabilidade: uma análise das publicações à luz da bibliometria. **Revista Produção Online**, v.19, n.1, 2019. <https://doi.org/10.14488/1676-1901.v19i1.3068>

SEMAAN, G.S.; CORRÊA, D.A.; RIBAS, S.G.; DEMBOGURSKI, B.J.; SILVA, E.F.; BRITO, J.A.M.; OCHI, L.S. Um panorama geral das contribuições e das redes de colaboração do SBIE. *In*: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO - SBIE, 2019(a). <http://dx.doi.org/10.5753/cbie.sbie.2019.1361>

SEMAAN, G.S.; TAVARES, M.V; CORRÊA, D.A.; WILSON, R.E.; BRITO, J.A.M. **Um panorama dos trabalhos do SIMPEP**: 25 anos de contribuições. Simpósio de Engenharia de Produção - SIMPEP, 2019(b).

SEMAAN, G.S.; TAVARES, M.V; CORRÊA, D.A.; WILSON, R.E.; BRITO, J.A.M.; **Uma análise das redes de colaboração do SIMPEP**. Simpósio de Engenharia de Produção - SIMPEP, 2019(c).

MOXLEY, R. L., MOXLEY, N. F. Determining point-centrality in uncontrived social networks. *Sociometry*, v. 37, p. 120-133. <https://www.doi.org/10.2307/2786472>

VANZ S. A. S. **As redes de colaboração científica no Brasil (2004-2006)**. Tese (Doutorado) - Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 2009.



Artigo recebido em: 07/08/2020 e aceito para publicação em: 05/12/2020
DOI: <http://dx.doi.org/10.14488/1676-1901.v20i4.4102>