

Um Panorama Geral das Contribuições e das Redes de Colaboração do SBIE

Gustavo Silva Semaan^{1,2}, Débora Alvernaz Corrêa³, Sávio Gurgel Ribas¹,
Bruno José Dembogurski⁴, Edelberto Franco Silva⁵,
José André de Moura Brito⁶, Luiz Satoru Ochi⁷

¹ Instituto do Noroeste Fluminense de Educação Superior (INFES)
Universidade Federal Fluminense (UFF) - Santo Antônio de Pádua - RJ

² Escola de Engenharia Industrial e Metalúrgica de Volta Redonda (EEIMVR)
UFF - Volta Redonda - RJ

³ Faculdade Metodista Granbery - Juiz de Fora - MG

⁴ Universidade Federal Rural do Rio de Janeiro (UFFRJ) - Nova Iguaçu - RJ

⁵ Departamento de Ciência da Computação (DCC)
Universidade Federal de Juiz de Fora (UFJF) - Juiz de Fora - MG

⁶ Escola Nacional de Ciências (ENCE) - Rio de Janeiro - RJ

⁷ Instituto de Computação - UFF - Niterói - RJ

gustavosemaan@gmail.com, deboradac@gmail.com, savio_ribas@id.uff.br,
brunodembogurski@gmail.com, edelberto.franco@gmail.com,
jambrito@gmail.com, luiz.satoru@gmail.com

Abstract. *The Brazilian Symposium on Informatics in Education celebrates its 30th edition in 2019, and the present paper presents an overview of the contributions and collaboration networks created by this important event in the area of education and information technology. To do so, the processes of knowledge discovery in databases, the task of text mining, methods of data analysis and concepts of graph theory were considered. The database used has information related to the bibliographic data of the Annals of the SBIE of the last 18 years and consists of 2,147 articles, 4,286 authors and the collaboration among their peers. It is a neutral presentation, restricted to the data obtained from the main scientific event of the area in Brazil, which aims to foster the importance of the theme in the academic scenario.*

Resumo. *O Simpósio Brasileiro de Informática na Educação comemora sua 30ª edição em 2019, e o presente trabalho apresenta um panorama geral das contribuições e das redes de colaboração criadas por este importante evento na área de educação e informática. Para tanto, foram considerados os processos de descoberta de conhecimento em bases de dados, a tarefa de mineração de textos, métodos de análises de dados e conceitos de teoria de grafos. A base de dados utilizada possui informações referentes aos dados bibliográficos dos Anais do SBIE dos últimos 18 anos, e é formada por 2.147 artigos, 4.286 autores*

e a colaboração entre seus pares. Trata-se de uma apresentação neutra, restrita aos dados obtidos do principal evento científico da área no Brasil, que visa fomentar a importância do tema no cenário acadêmico.

1. Introdução

O Simpósio Brasileiro de Informática na Educação (SBIE) teve sua primeira edição no Rio de Janeiro em 1990, e se consolidou como o principal evento científico da área no Brasil, comemorando sua 30^a edição em 2019. Atualmente o evento ocorre como parte do itinerante Congresso Brasileiro de Informática na Educação (CBIE)[CBIE 2019], proporcionando um ambiente propício à troca de experiências e ideias. Em virtude de sua importância, este trabalho se apoia no crescimento de outra comunidade científica, aquela que estuda áreas relacionadas à análise de dados para informações relevantes à área de educação e informática. Baseado, portanto, na chamada análise de redes sociais (do inglês *Social Networks Analysis* [BORGATTI 2006]) e novas formas de apresentação (como grafos) [LIMA 2017], para celebrar essas conquistas, o presente trabalho coletou, organizou, formatou e apresenta dados de todos os artigos disponíveis no repositório do SBIE [SBIE 2019], referentes aos dados bibliográficos de seus Anais dos últimos 18 anos, no período entre 2001 e 2018.

A produção de trabalhos no SBIE é predominantemente colaborativa [PROCACI 2015] [MAGALHÃES 2013]. Mesmo antes das tecnologias de comunicação modernas e redes sociais digitais, o uso de modelagens baseadas em teoria dos grafos para tratar redes de colaboração já eram largamente utilizadas, em que ocorre um mapeamento entre autores e suas colaborações em trabalhos. Em especial, análises que consideram conceitos de centralidade de vértices com o objetivo de mensurar e identificar influências foram introduzidas por [FREITAS 2010] *apud* [BAVELAS 1950]. No contexto atual, a atuação de um pesquisador de maneira isolada tende a ser reduzida, e a produção científica comumente requer efetiva colaboração entre especialistas [SILVA 2002] [CRONIN 2005].

De forma a embasar tecnicamente a motivação já apresentada neste trabalho, deve-se levar em consideração as definições a seguir. Embora diversas medidas de centralidade em estudos de correlação em redes de colaboração tenham sido propostas na literatura, de acordo com os critérios considerados em casos particulares, [FREEMAN 1978] propõe que sejam consideradas três definições clássicas: **(I)** a centralidade de grau, **(II)** de proximidade e **(III)** de intermediação. Em busca de identificar autores, além das medidas de centralidade foram consideradas as frequências de publicações. Assim, autores em destaque publicam com frequência e possuem influência conforme diferentes critérios de centralidade. Entretanto, antes de iniciar a análise das colaborações, é de grande importância apresentar um panorama geral em relação aos artigos publicados, o processo de descoberta de conhecimento em bases de dados que considerou a tarefa mineração de textos bem como a identificação dos termos de interesse em destaque.

O objetivo deste trabalho é brindar o evento com a apresentação de um panorama geral sobre as publicações no SBIE com base em todo acervo de edições anteriores disponível no site, de maneira impessoal, apolítica e restrita aos dados públicos. Para isso, o artigo segue organizado em 6 seções, a partir desta introdução. A seção 2 apresenta os Trabalhos Relacionados. Na seção 3 são apresentados os procedimentos metodológi-

cos de pesquisa adotados para realização do trabalho. A seção 4 apresenta Estatísticas Gerais e análises de dados. A seção 5 aborda as Redes de Colaboração, as análises, e os resultados do estudo são apresentados e discutidos. Por fim, na seção 6 são feitas as considerações finais e propostas para trabalhos futuros.

2. Trabalhos Relacionados

O conhecimento é construído a partir troca de experiências, e, sendo assim, trabalhos relacionados podem ser destacados ambientes diversos. Em cada contexto, as configurações específicas determinam suas formas particulares de colaborar. Entender como ocorrem as colaborações, buscar por influenciadores e relacioná-los a temas e produtos de interesse é alvo de intensa pesquisa. Desta forma é comum se considerar conceitos de teoria dos grafos e de redes de colaboração para embasar percepções e indicar correlações.

Dados interessantes confirmados por estas pesquisas mostram que: estudos relacionados à coautoria e colaboração indicam que os trabalhos teóricos tendem a possuir menos autores em relação a trabalhos de natureza experimental; trabalhos colaborativos internacionais têm maior impacto e visibilidade; e também que a colaboração aumenta a produtividade dos pesquisadores [GLANZEL 2001][LEE and BOZEMAN 2005][CRONIN 2005].

Diversas pesquisas, seguindo o objetivo principal do presente trabalho, fornecem panoramas gerais em relação a acervos de importantes eventos e/ou a revisões sistemáticas da literatura. Em [CAVALCANTI 2011] é apresentada uma análise abrangente e sistemática de 24 edições do Simpósio Brasileiro de Engenharia de Software (SBES), em que os dados dos artigos foram catalogados, analisados e sintetizado. Já em [OLIVEIRA 2012] foi analisada a colaboração entre autores em 6 edições do Simpósio Brasileiro de Sistemas de Informação (SBSI), referentes aos anos 2006 a 2011. De maneira semelhante, em [LIMA 2017] é apresentada uma análise detalhada da rede de colaborações em uma homenagem à comemoração pelas 30 edições do Simpósio Brasileiro de Banco de Dados (SBBD), em que foram analisados dados de 674 artigos de 1034 autores em 2299 colaborações entre eles. [VANZ 2009] analisou 49.046 artigos brasileiros indexados na base *Web of Science* no período entre 2004 e 2006 em relação também a colaboração entre autores. Em [FERNANDES 1999] foram mapeados perfis de grupos atuantes em Informática na Educação com o intuito de estimular o compartilhamento de informações e colaborações nos projetos desenvolvidos.

Em especial, o trabalho de [MAGALHÃES 2013] apresenta um mapeamento sistemático com uma visão ampla de trabalhos primários realizados em 12 edições referentes aos anos 2001 a 2012 do SBIE. A análise contou com 835 artigos com autores de cerca de 250 instituições. Já [PROCACI 2015] analisou o período entre 2001 e 2013 (13 edições). Uma vez que tais trabalhos foram considerados como parte do conjunto total de dados abordado no presente trabalho, eles serão citados e irão corroborar para a análise realizada ao longo do texto.

3. Metodologia

Esta seção introduz a metodologia utilizada desde a coleta dos dados brutos, as etapas de pré-processamento, a construção dos modelos de dados necessários e, por fim, a geração dos grafos considerados para a apresentação dos resultados, conforme ilustra a Figura 1.

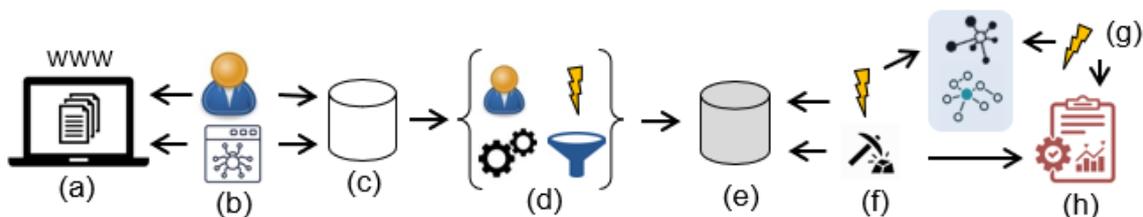


Figura 1. Fluxograma do processo de KDD

Em um primeiro momento foi necessário obter os anais de edições anteriores do SBIE, disponíveis de maneira organizada, simples e transparente no portal da Comissão Especial de Informática na Educação [CBIE 2019]. Os conteúdos dos anais (chamados de metadados) e os artigos podem ser acessados diretamente pelo portal do SBIE [SBIE 2019] (Figura 1 (a)). Desta forma, é relevante considerar a grande importância que anais de eventos abertos à comunidade e disponibilizados de forma simples, como faz o SBIE, auxiliam na democratização do acesso à informação.

Embora o acesso pelo portal seja simples, foi necessário automatizar a obtenção de dados devido ao seu volume e aos diferentes formatos dos arquivos do portal (páginas em HTML, do inglês *Hypertext Markup Language*). Assim, através de um aplicativo de automatização de recuperação de conteúdo, também conhecido como *web crawler*¹, foi possível tratar requisições HTTP (*Hypertext Transfer Protocol*) ao site [PROCACI 2015] e realizar a coleta dos mesmos (Figura 1 (b)). Foram obtidos os seguintes dados: ano do evento, tipo do artigo (Completo, Resumo ou Editorial), título do artigo e os nomes dos autores (Figura 1 (c)). Embora de grande interesse, na página e nos artigos não existe a relação de palavras-chave e instituições ou empresas em que os autores atuam.

Após a coleta foi necessário iniciar a etapa de pré-processamento, em que os dados devem ser organizados, enriquecidos e formatados (Figura 1 (d)). Nesse sentido foi considerado o Processo de Descoberta de Conhecimento em Bases de Dados e, mais especificamente, técnicas de mineração de texto [GOLDSCHMIDT 2005]. Os dados ano e tipo de artigo já estavam em formatos adequados e não foi necessário nenhum tratamento adicional. O título de um trabalho científico geralmente expressa seu propósito de maneira objetiva, e em uma etapa denominada *tokenização* ele foi fragmentado em palavras para que, após a exclusão de *stopwords* (palavras sem relevância semântica) como conjunções e preposições, fossem obtidos os termos de interesse. De maneira a consolidar esses termos, antes do cálculo de suas frequências, foi realizada unificação por meio de um processo de redução ao radical ou equivalência a palavra semelhante mais frequente [ARANHA 2007] [PROCACI 2015]. Com a ausência de especificação de palavras-chave, o uso do título foi de grande importância para direcionar e determinar os assuntos mais frequentes.

Para consolidar o conjunto de dados e manter sua consistência, foi necessário realizar uma desambiguação de nomes de autores, para evitar a divisão de publicações de uma pessoa que possui diferentes grafias nos anais, devido a erros de digitação ou mesmo as diferentes maneiras de escrever os nomes (com nomes/sobrenomes ausentes,

¹Desenvolvido em linguagem Java especificamente para este fim de se coletar os dados de forma automática.

abreviados ou escritos completamente). Nesse sentido foi necessário utilizar algoritmos para atuar na identificação e unificação dos autores, com o uso do cálculo da distância de edição [NAVARRO 2001] (ou distância *Levenshtein*) e a busca por equivalências entre autores com base no primeiro nome e no sobrenome. Mesmo com os softwares utilizados foi constatada a necessidade de uma verificação humana para decidir se dois nomes correspondem ao mesmo autor, e ela foi feita através de uma planilha eletrônica.

Por fim, na última parte do pré-processamento, os dados estão consolidados e foram organizados em tabelas que possuem os atributos ano, tipo de artigo, listas de termos de interesse do título e de autores unificados (Figura 1 (e) e (f)). É necessário, então, gerar grafos para que seja possível identificar as perspectivas sociais de coautoria (parcerias). Embora o trabalho não aprofunde em temas de teoria dos grafos e redes de colaboração (Figura 1 (g)), a seção 4 apresenta de maneira objetiva os conceitos considerados na modelagem e nas análises dos resultados obtidos (Figura 1 (h)).

4. Estatísticas Gerais

A rede de colaboração SBIE, construída a partir dos dados coletados todos os trabalhos publicados nas últimas 18 edições, possui um total de 4.286 autores distintos (vértices) e 9.596 colaborações (arestas), compreendendo um total de 2.147 trabalhos organizados por tipo (artigo completo, resumo ou editorial).

A Tabela 1 apresenta os quantitativos de trabalhos e de autores, médias de autores por artigo, autores que publicaram trabalhos de maneira isolada (sozinhos, do inglês *single-authors* (SA)) e percentuais autores estreatantes. Com base nos quantitativos de autores e de trabalhos é possível observar o crescimento do evento, com destaque para a edição de 2018, com 229 trabalhos e 872 autores. Em relação a média de autores por artigo, observa-se que não houve grande variação ao longo dos anos.

Tabela 1. Estatísticas Gerais

Ano	Trabalhos				Autores	Média	SA*	Estreatantes(%)
	Completo	Resumo	Editorial	Total				
2001	58	0	0	58	182	3,14	3	100*
2002	54	17	0	71	205	2,89	6	70
2003	71	0	0	71	216	3,04	3	63
2004	56	30	0	86	278	3,23	4	58
2005	63	10	0	73	231	3,16	4	64
2006	60	30	0	90	304	3,38	4	58
2007	60	75	0	135	439	3,25	8	56
2008	84	0	0	84	300	3,57	2	54
2009	84	0	0	84	271	3,23	7	59
2010	88	44	0	132	431	3,27	5	70
2011	91	47	0	138	480	3,48	6	57
2012	100	42	8	150	491	3,27	8	55
2013	102	7	0	109	395	3,62	0	60
2014	152	0	0	152	560	3,68	1	60
2015	139	0	0	139	481	3,46	0	57
2016	129	17	0	146	508	3,48	4	58
2017	172	28	0	200	729	3,65	5	48
2018	170	59	0	229	872	3,81	3	52

Com base nos dados obtidos, cerca de 4% dos artigos possui apenas um autor (SA), 30% possuem até 2 autores e cerca de 70% possui ao menos 3 autores. Assim

como em [PROCACI 2015] [MAGALHÃES 2013], pode-se verificar que os trabalhos publicados no SBIE são colaborativos em sua grande maioria.

Entretanto, cerca de 74% dos autores participaram de apenas um artigo. Assim, surge a necessidade de avaliar os quantitativos de autores estreantes por ano, conforme apresenta a Tabela 1. É considerado um autor estreante aquele que não possui nenhum artigo publicado em anos anteriores. Uma vez que os dados correspondem ao período entre 2001 e 2018, autores que publicaram trabalhos no SBIE antes desse período são considerados estreantes. Ainda conforme a Tabela 1, com exceção da edição de 2017 (com 48%), mais da metade dos autores dos eventos dos anos relacionados são estreantes, com pico de 70% em 2010.

A Figura 2 apresenta as 20 palavras mais frequentes entre os títulos dos 2.147 artigos. A palavra *ambiente*, por exemplo, foi utilizada no título de mais de 15% dos artigos publicados no SBIE de 2002. Ainda em 2002, ao menos uma das palavras em destaque foi utilizada em mais de 71% dos artigos.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
alunos	1,1%	1,0%	2,3%	1,1%	0,0%	0,7%	0,9%	1,3%	1,5%	0,9%	1,3%	1,4%	0,8%	2,3%	2,1%	0,6%	2,1%	1,7%
ambiente	11,0%	15,6%	7,9%	5,8%	9,1%	5,3%	6,4%	5,3%	8,5%	3,7%	5,8%	4,9%	3,8%	4,6%	4,8%	3,0%	4,1%	2,1%
análise	0,5%	0,5%	1,9%	0,4%	0,4%	0,3%	1,4%	1,0%	0,7%	1,2%	1,9%	1,4%	0,8%	2,5%	2,7%	1,6%	1,2%	2,4%
apoio	1,6%	3,4%	1,9%	2,5%	2,2%	2,0%	1,4%	1,3%	1,1%	1,6%	2,5%	2,6%	2,3%	1,4%	1,2%	2,0%	1,2%	1,1%
aprendizagem	7,7%	11,2%	8,8%	9,7%	8,7%	10,2%	10,5%	8,3%	10,3%	8,6%	9,8%	10,6%	8,4%	7,9%	5,8%	5,9%	9,2%	5,7%
avaliação	2,2%	3,4%	3,2%	3,6%	1,7%	2,6%	2,5%	2,7%	1,8%	1,4%	2,1%	2,0%	1,0%	1,1%	2,3%	3,1%	1,2%	2,8%
desenvolvimento	1,6%	2,0%	2,3%	3,2%	0,9%	1,0%	1,1%	3,0%	0,7%	1,4%	1,0%	2,0%	1,8%	1,4%	1,5%	1,6%	1,2%	1,4%
digitais	0,5%	1,0%	0,5%	0,7%	1,3%	4,3%	2,3%	4,3%	3,3%	3,7%	2,1%	1,2%	2,3%	1,3%	1,9%	1,0%	1,4%	1,9%
educação	2,7%	4,9%	2,3%	2,9%	1,7%	1,6%	3,6%	2,7%	2,6%	1,6%	0,8%	2,4%	2,8%	2,0%	1,5%	2,2%	1,4%	1,6%
educacionais	2,2%	2,4%	5,1%	2,2%	2,2%	3,3%	2,3%	2,3%	3,3%	3,2%	2,7%	2,2%	2,8%	3,9%	4,6%	4,5%	3,7%	3,2%
ensino	3,3%	5,9%	3,2%	2,5%	2,2%	3,6%	4,6%	2,3%	5,5%	5,1%	5,6%	4,5%	4,8%	5,7%	6,0%	6,3%	4,8%	6,0%
ferramenta	2,7%	2,0%	3,7%	3,2%	2,2%	0,7%	1,8%	1,7%	3,3%	3,0%	2,1%	2,0%	1,8%	1,8%	2,1%	1,8%	1,5%	1,0%
jogo	0,0%	1,5%	0,9%	0,4%	0,0%	1,6%	1,1%	2,0%	0,0%	3,2%	2,5%	2,6%	4,1%	3,9%	4,2%	3,9%	1,9%	3,3%
modelo	1,1%	2,0%	1,4%	0,7%	1,7%	2,0%	2,3%	2,0%	1,8%	1,2%	1,5%	2,2%	1,5%	1,6%	1,0%	1,2%	1,8%	1,1%
objeto	0,0%	0,0%	0,0%	3,6%	1,3%	1,6%	4,3%	2,3%	4,4%	4,6%	3,8%	4,7%	2,0%	2,9%	1,5%	1,2%	2,2%	0,9%
processos	0,0%	1,0%	0,9%	2,9%	2,2%	2,0%	0,9%	1,3%	1,5%	2,1%	2,1%	2,6%	2,0%	0,2%	1,0%	1,4%	2,1%	1,0%
programação	0,5%	1,0%	0,5%	0,7%	0,0%	0,7%	0,9%	0,0%	1,5%	1,2%	1,0%	1,8%	1,5%	2,5%	3,1%	3,5%	1,0%	2,8%
sistema	5,5%	3,4%	2,3%	5,0%	3,0%	3,3%	3,0%	2,7%	1,8%	2,8%	3,8%	3,1%	3,3%	1,4%	2,3%	1,8%	3,3%	2,8%
software	1,6%	2,9%	0,9%	1,4%	1,3%	2,0%	1,8%	1,3%	1,5%	2,1%	1,9%	2,0%	1,3%	0,4%	2,1%	1,8%	1,5%	1,1%
virtual	2,2%	6,3%	6,5%	2,2%	6,5%	4,6%	4,8%	5,3%	3,0%	2,1%	2,9%	3,5%	2,3%	3,8%	2,9%	1,6%	2,9%	1,0%

Figura 2. percentuais de artigos que possuem cada palavra entre as mais usadas por ano.

5. Redes de Colaboração

A produção de trabalhos no SBIE é predominantemente colaborativa e, de fato, a atuação de um pesquisador isolado faz parte de um passado distante. A produção científica relevante comumente requer efetiva colaboração e associação entre especialistas [PROCACI 2015] [MAGALHÃES 2013] [SILVA 2002][CRONIN 2005].

Muito antes das tecnologias de comunicação modernas, a noção de centralidade em redes sociais foi introduzida por [BAVELAS 1950]. Conforme seu trabalho, “*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*”. Devido a sua localização privilegiada, “*esse indivíduo deve ser responsável por transmitir, modificar, ou reter a informação entre membros do grupo*”. Dessa maneira pode-se afirmar que a centralidade na rede está relacionada à *influência*.

O uso de modelagens baseadas em conceitos de teorias de grafos é comumente consideradas para a resolução de problemas. Para analisar redes de colaboração

deve-se realizar um mapeamento entre as coautorias em trabalhos. [PROCACI 2015] [MAGALHÃES 2013]. Formalmente, um grafo G consiste em um conjunto de vértices V e arestas A , em que cada aresta conecta dois vértices, e é denotado pelo conjunto $G = (V, A)$. No presente trabalho, um vértice representa um autor e cada aresta representa uma parceria (coautoria) em um trabalho.

O grau de um vértice é a quantidade de arestas conectadas a ele. A distância entre dois vértices em um grafo é a quantidade mínima de arestas que devem ser percorridas para sair de um dado vértice v_i e chegar em um outro vértice v_j . Define-se um grafo como desconexo quando existe ao menos um par de vértices que não está ligado a nenhum caminho.

Embora tenham sido apresentadas dados e estatísticas gerais para embasarem a análise das comunidades por meio das colaborações entre autores, assim como em [PROCACI 2015], acredita-se que a abordagem, mesmo sendo de grande interesse e importância, seja insuficiente para uma visão holística da colaboração das comunidades.

A centralidade de grau considera a quantidade de arestas incidentes a um vértice, e pode ser interpretada como a probabilidade que ele possui de receber alguma informação da rede (ou grafo). Em outras palavras, trata-se da influência direta que um vértice tem em relação a seus vizinhos. Assim, é possível destacar autores que possuem mais relações com outros autores em trabalhos (coautoria) [BORGATTI 2006].

Já a centralidade de proximidade é relacionada à distância total de um vértice a todos os demais vértices do grafo. Ou seja, a distância de cada autor a todos os demais autores pode sugerir seu poder de influência. Por exemplo, um dado autor pode não conhecer (de maneira pessoal ou virtual) um outro autor, mas eles podem ter um amigo em comum. Assim, a distância entre eles é de duas arestas.

Por fim, a centralidade de intermediação considera a quantidade de vezes que um vértice age como ponte ao longo do caminho mais curto entre dois outros vértices. Pode ser utilizado para mensurar o controle sobre a comunicação entre autores em redes de colaboração (ou mesmo numa rede social).

De maneira a analisar os critérios de centralidade de grau e centralidade de proximidade, um autor pode ser muito influente em sua instituição, com diversos trabalhos e parcerias com autores da mesma instituição (alto grau). Porém, pode ter poucas parcerias com autores de outras instituições e ser distante na rede de colaboração, com baixo valor de centralidade de proximidade. Em contrapartida, um autor pode ter poucos trabalhos com parceiros de sua instituição, mas atuar com pesquisadores de diversas instituições, o que tende a maximizar seu valor de centralidade de proximidade.

No contexto do presente trabalho, o uso da modelagem em grafos e de medidas de centralidade podem identificar e quantificar a influência de autores por meio de diferentes critérios, conforme sua posição estrutural no grafo [BORGATTI 2006]. A ferramenta SocNetV 2.5 (*Free and Open-Source Tool for Social Network Analysis*) foi utilizada nesse trabalho tanto para os cálculos das centralidades nos grafos quanto para as suas representações gráficas. Para investigar os autores em destaque do SBIE no período considerado, o trabalho tomou como base: a frequência de publicação, a quantidade de colaboradores (grau), as centralidade de grau, de proximidade e de intermediação.

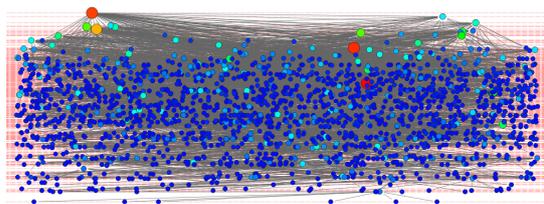


Figura 3. Autores com grau 10 ou superior e seus colaboradores.

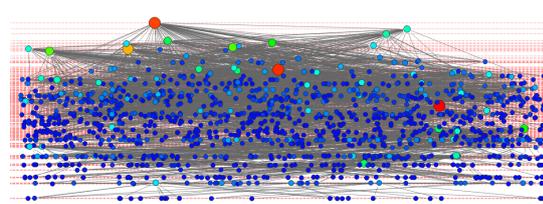


Figura 4. Top 16 autores mais frequentes e autores com distância máxima 2.

A Tabela 2 apresenta os top 10 autores mais influentes do SBIE conforme quantidade de publicações, colaborações e centralidade de grau, de proximidade e de intermediação. As colunas Grafo 1 (Figura 3) e Grafo 2 (Figura 4) referem-se a colocação com base no valor de centralidade de proximidade dentre os 1312 e 2032 autores considerados nos grafos, respectivamente. O layout considerado está no formato em nível pelo critério proximidade dos vértices. Assim, quanto mais próximo ao limite superior da imagem, maior sua importância de acordo com o critério adotado. Os nomes foram relacionados em ordem alfabética e, embora não seja o objetivo principal do trabalho, esses resultados prestam uma homenagem pela contribuição à comunidade de Informática na Educação.

Tabela 2. 10 autores mais influentes no SBIE em quantidade de artigos, grau de colaboração e centralidades de proximidade e de intermediação.

Nome	Artigos	Grau	C.Proximidade		C.Intermediação
			Grafo 1	Grafo 2	
Alex Sandro Gomes	42	90	412 ^a	460 ^a	3 ^a
Crediné Silva de Menezes	73	84	1 ^a	1 ^a	6 ^a
Davidson Cury	27	39	4 ^a	9 ^a	68 ^a
Elaine Oliveira	19	30	3 ^a	6 ^a	8 ^a
Evandro Costa	40	86	38 ^a	29 ^a	18 ^a
Ig Ibert Bittencourt	28	74	11 ^a	7 ^a	7 ^a
Magda Bercht	19	35	40 ^a	16 ^a	9 ^a
Orivaldo de Lira Tavares	25	45	5 ^a	15 ^a	32 ^a
Rosa Maria Vicari	27	52	9 ^a	5 ^a	5 ^a
Seiji Isotani	25	52	14 ^a	11 ^a	4 ^a

Na busca por tornar a comunidade do SBIE mais coesa, com menos grupos isolados e mais interações, em [PROCACI 2015] foram realizados experimentos empíricos com o objetivo de minimizar o coeficiente de agrupamento médio por meio de adições de arestas entre autores, simulando colaborações de maneira artificial. Com base nos resultados obtidos, deve-se incentivar a atuação em conjunto entre autores que colaboram muito (alto grau de colaboração) e autores que colaboram pouco (baixo grau). Além disso, a adição de arestas entre autores que colaboram muito ou entre pessoas que colaboram pouco não torna a comunidade mais coesa.

6. Conclusões

O presente trabalho teve como objetivo principal brindar o SBIE em sua 30^a edição com a apresentação de um panorama geral sobre suas publicações, com base em todo acervo de edições anteriores disponível no site, em uma apresentação neutra, de maneira impessoal, apolítica e restrita aos dados públicos. Para isso foram destacados os tópicos: estatísticas gerais, termos em destaque, autores e redes de colaboração.

O acervo do SBIE nas últimas 18 edições possui 2.147 trabalhos de 4.286 autores em 9.596 colaborações entre eles. Informações importantes podem ser observadas, como: (i) o crescimento do evento em quantitativos de artigos e também de autores; (ii) média de 3 autores por artigo, com apenas 4% artigos feitos por apenas um autor; (iii) cerca de 74% dos autores possui apenas um artigo no evento, e 88% dos autores possuem até dois artigos; (iii) comunidade fiel, com 223 autores que possuem ao menos 5 artigos, 62 autores possuem ao menos 10 artigos; (iv); devido ao grande número de autores com apenas um artigo publicado foi investigado os estreantes por ano, que obteve pico de 70% em 2010 e média superior a 50% por evento; (v) termos em destaque presentes nos títulos dos trabalhos foram organizados e apresentados em função do tempo; (vi) com o objetivo de representar redes de colaboração grafos foram construídos e cálculos de centralidade foram considerados para identificar e destacar autores mais influentes.

Devido a restrições de espaço algumas análises e figuras foram suprimidas deste texto. São exemplos destas análise: gráficos para apresentar as Estatísticas Gerais (resumidos na Tabela 1), as nuvens de palavras com autores e termos mais frequentes por período (a cada cinco anos e por década), grafos gerados a partir de outras configurações que podem apoiar a análise de influência e as colaborações entre autores conforme diferentes configurações e seleções de vértices.

O panorama apresentado além de destacar autores considerados influentes conforme conceitos de centralidade e palavras que podem indicar temas de interesse para novas pesquisas, constitui uma sólida referência. Dentre os trabalhos futuros propostos pelos autores podem ser destacados: busca por correlações entre publicações, áreas de pesquisa e as palavras mais frequentes por período; a busca por associações entre os termos dos títulos; o uso de outras modelagens com grafos ponderados ou mesmo dígrafos para análise de colaborações; análise das colaborações entre instituições por meio de grupos de pesquisas; analisar a relevância e as citações dos artigos publicados; contemplar também bases de dados de outros eventos importantes no cenário nacional.

Referências

- [ARANHA 2007] ARANHA, C. N. (2007). *Uma Abordagem de Pré-Processamento Automático para Mineração de Textos em Português: Sob o Enfoque da Inteligência Computacional*. PhD thesis, Pontifícia Universidade Católica do Rio de Janeiro.
- [BAVELAS 1950] BAVELAS, A. (1950). Communication patterns in task oriented groups. *Journal of the Acoustical Society of America*, 22.
- [BORGATTI 2006] BORGATTI, S. P.; EVERETT, M. G. (2006). *A Graph-Theoretic perspective on centrality*. Social Networks.
- [CAVALCANTI 2011] CAVALCANTI, T. R; da SILVA, F. Q. B. (2011). Historical, conceptual, and methodological aspects of the publications of the brazilian symposium on software engineering: A systematic mapping study. In *25th Brazilian Symposium on Software Engineering (SBES)*.
- [CBIE 2019] CBIE (2019). Portal da comissão especial de informática na educação. Disponível em <http://www.br-ie.org>. Acesso: 01/05/2019.
- [CRONIN 2005] CRONIN, B. (2005). *The hand of science: academic writing and reward*. Oxford: Scarecrow Press.

- [FERNANDES 1999] FERNANDES, C. T.; SANTOS, N. (1999). Pesquisa e desenvolvimento em informática na educação no brasil – parte i. *Revista Brasileira de Informática na Educação*, (4).
- [FREEMAN 1978] FREEMAN, L. C. (1978). Centrality in social networks: Conceptual clarification. In *Social Networks*.
- [FREITAS 2010] FREITAS, I. Q. (2010). Medidas de centralidade em grafos.
- [GLANZEL 2001] GLANZEL, W. (2001). National characteristics in international scientific coauthorship relations. *Scientometrics*, 51(1).
- [GOLDSCHMIDT 2005] GOLDSCHMIDT, R.; PASSOS, E. (2005). *Data Mining: Um Guia Prático*. Elsevier.
- [LEE and BOZEMAN 2005] LEE, S. and BOZEMAN, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5).
- [LIMA 2017] LIMA, L.H.C.; PENHA, G.; ROCHA, L. M. M.; S. A. L. A. H. F.; O. J. M. (2017). The collaboration network of the Brazilian symposium on databases - 30 editions of history. In *Journal of the Brazilian Computer Society*.
- [MAGALHÃES 2013] MAGALHÃES, C.V.C.; SANTOS, R. S. F. G. A. (2013). Caracterizando a pesquisa em informática na educação no Brasil: Um mapeamento sistemático das publicações do SBIE. In *XXIV Simpósio Brasileiro de Informática na Educação - SBIE*.
- [NAVARRO 2001] NAVARRO, G. (2001). A guided tour to approximate string matching. *ACM Computing Surveys*, 33.
- [OLIVEIRA 2012] OLIVEIRA, E. A. (2012). Sobre a colaboração na comunidade de sistemas de informação através dos simpósios SBIE.
- [PROCACI 2015] PROCACI, T.B.; SIQUEIRA, S. P. F. N. B. (2015). 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. In *Simpósio Brasileiro de Informática na Educação - SBIE*.
- [SBIE 2019] SBIE (2019). Portal do simpósio brasileiro de informática na educação. Disponível em <http://www.br-ie.org/pub/index.php/sbie/index>. Acesso: 01/05/2019.
- [SILVA 2002] SILVA, E. L. (2002). Rede científica e a construção do conhecimento. *Informação e Sociedade: Estudos*, 12(1).
- [VANZ 2009] VANZ, S. A. S. (2009). *As redes de colaboração científica no Brasil (2004-2006)*. PhD thesis, Universidade Federal do Rio Grande do Sul.