

ERAD/RS 20 anos: Quem fez esta história?

Sandro da Silva Camargo¹, Leonardo Bidese de Pinho¹

¹Programa de Pós-Graduação em Computação Aplicada
Universidade Federal do Pampa (UNIPAMPA)
Caixa Postal 007 – 96.400-590 – Bagé – RS – Brasil

{sandro.camargo, leonardo.pinho}@unipampa.edu.br

Resumo. *A Escola Regional de Alto Desempenho da Região Sul (ERAD/RS), evento itinerante criado em 2001 com foco no RS, expandiu-se ao longo dos anos almejando divulgar a produção científica e qualificar profissionais da região sul em Processamento de Alto Desempenho (PAD). Na sua vigésima edição, atingiu quase 1200 trabalhos que contam a evolução da pesquisa Sul-brasileira em PAD. Este estudo explora, com técnicas de Análise de Redes Sociais e de Genealogia Acadêmica, as publicações das vinte edições, identificando os autores e grupos que se destacaram nesta história. Seus resultados fornecem subsídios à Comissão Regional de Alto Desempenho para o reconhecimento dos pesquisadores e relações que construíram os vinte anos da ERAD/RS.*

1. Introdução

As Tecnologias de Informação e Comunicação (TIC) desempenham um papel fundamental na sociedade moderna, sendo uma das bases para o desenvolvimento socioeconômico [Roztocki et al. 2019]. Ciente desta importância, a Sociedade Brasileira de Computação (SBC) tem incentivado e financiado a oferta de eventos com o objetivo de disseminar temas atuais e relevantes relacionados à Computação em todo o país. Dentre estes eventos, destacam-se as Escolas Regionais, que visam a divulgação e o fortalecimento da relação da SBC com os cursos da área de TIC.

Inserida neste contexto, a Escola Regional de Alto Desempenho da Região Sul (ERAD/RS) é um evento anual, que teve sua primeira edição no ano de 2001, sendo promovida pela SBC, por meio da Comissão Regional de Alto Desempenho da Região Sul (CRAD/RS). Os objetivos desta escola são: a) qualificar profissionais da região nas áreas relacionadas a PAD; b) prover um fórum regular onde se possa apresentar os avanços recentes nessas áreas; e c) discutir as formas de ensino de PAD nas universidades do sul do Brasil [Schnorr and Pillon 2018].

A partir da análise dos anais das vinte edições da ERAD/RS, este trabalho buscou identificar os autores que construíram o evento ao longo de sua história, assim como identificar as redes de coautorias. O foco nas redes é relevante pois, hoje em dia, a colaboração em pesquisa tem se tornado uma atividade fundamental para o progresso da ciência por vários motivos, estando entre eles a melhoria da comunicação entre grupos de pesquisadores e o intercâmbio de experiências e competências, além de facilitar a produção e disseminação de novos conhecimentos científicos [Camargo and Camargo 2018]. Medir o nível de colaboração entre membros de comunidades de pesquisadores é uma tarefa complexa, que se pode fazer de diferentes formas [Su 2010]. No entanto, a forma mais evidente de colaboração é a coautoria em publicações [Newman 2004], que tem

sido um instrumento para a análise de colaborações, associações científicas e associações tecnológicas, tornando possível a obtenção da compreensão dos padrões de cooperação entre autores [Sampaio et al. 2015].

Na literatura, podem ser encontrados diversos trabalhos com a aplicação da abordagem de Análise de Redes Sociais, ou *Social Network Analysis (SNA)*, para identificar autores e redes de colaboração mais importantes em eventos e periódicos. Como exemplo mais próximo, [Camargo et al. 2019a] identificaram os autores e comunidades mais importantes nos anais dos primeiros dezoito anos da ERAD/RS. Já [Camargo and Camargo 2018] construíram e analisaram a rede de colaboração entre co-autores dos mais de 400 artigos publicados na Revista do CCEI, ao longo dos 20 anos de existência deste periódico científico. Já [Camargo et al. 2019b] também aplicaram da mesma abordagem no âmbito da Agroinformática na Argentina, envolvendo a análise de 270 trabalhos publicados nos Anais de dez edições do Congresso Argentino de Agroinformática, ocorridos nos anos de 2008 a 2018. Expandindo a análise do primeiro trabalho relacionado citado, de [Camargo et al. 2019a], este estudo busca, a partir de autorias dos artigos publicados nos anais das vinte edições da ERAD/RS artigos, identificar que autores e grupos desempenharam um papel de destaque na história deste evento. Este estudo foi posteriormente ampliado com ênfase nas instituições e em identificações de tendências [Camargo et al. 2020].

De forma complementar, mostrou-se relevante correlacionar os dados dos principais autores e grupos com indicadores de Genealogia Acadêmica (GA). GA envolve métodos computacionais para criação de grafos que permitam o estudo de parentesco acadêmico [Damaceno et al. 2017]. No contexto brasileiro, a Plataforma Acácia¹, disponibilizada em 2018, destaca-se como fonte de dados para estudos de GA, provendo indicadores relevantes a partir dos registros na Plataforma Lattes de orientações ou supervisões formais concluídas nos níveis de mestrado, doutorado e pós-doutorado. Portanto, trata-se de uma base de dados relevante, passível de consulta por meio de técnicas de *Web Scraping*², não apenas para indicadores individuais de genealogia, mas também para identificação de vínculos institucionais dos pesquisadores e, por conseguinte, da verificação das instituições cujos pesquisadores mais publicaram em determinada área.

O restante deste trabalho está organizado conforme segue. A Seção 2 detalha a metodologia utilizada no desenvolvimento do trabalho. Na Seção 3 são apresentados e discutidos os resultados obtidos. Por fim, são elencadas, na Seção 4, as principais conclusões e perspectivas de trabalhos futuros.

2. Material e Métodos

De acordo com a abordagem utilizada em [Camargo et al. 2019a], esta pesquisa, de natureza quantitativa, envolveu uma pesquisa documental na Biblioteca Digital Brasileira de Computação (BDBComp), por meio da página da ERAD³. Para implementação de

¹<http://plataforma-acacia.org>

²Prática que envolve a coleta de dados sem o uso de uma API, sendo mais comumente realizada por meio da escrita de um programa automatizado que consulta servidores web, requisita dados (usualmente na forma de HTML e outros arquivos que compõem páginas web) e então analisa os dados para extrair as informações necessárias [Mitchell 2018]

³<http://amplus.ufpel.edu.br/erad/doku.php?id=anais>

scripts de análise e processamento de dados, foi utilizado o pacote estatístico R, versão 3.4.4 [R Core Team 2018].

Este trabalho utilizou como base os 1172 artigos publicados nas edições de 2001 a 2020 da ERAD/RS, o que inclui os trabalhos aceitos para apresentação na vigésima edição do evento. Foram coletados manualmente os nomes dos autores dos artigos, com exceção dos artigos de 2020 cujos dados foram disponibilizados em formato BibTex. Em uma análise preliminar, foram encontradas situações onde o nome de um mesmo autor estava escrito de maneira distinta em diferentes artigos. Para resolver este problema, foi utilizada a função *Adist*⁴, do pacote R, que implementa o algoritmo de cálculo de distância de Levenshtein para identificar os nomes de autores escritos de forma similar. Como resultado, após a unificação dos nomes similares, foram identificados 1118 autores distintos. Foram divididos os dados de autorias dos artigos em dois arquivos: autores e colaborações. O arquivo de autores continha o número sequencial de identificação de cada autor e seu respectivo nome. O arquivo de colaborações apresentava, em cada linha, um número sequencial de identificação da colaboração, e os números de identificação de dois autores que participaram da colaboração. Foi criada uma linha neste arquivo para cada combinação de autores, tomados dois a dois, para cada artigo com mais de dois autores. Esta abordagem já foi utilizada em trabalhos correlatos realizados sobre outras bases de dados [Newman 2004, Camargo and Camargo 2018, Brum et al. 2019, Camargo et al. 2019a].

Para representar as colaborações na forma de uma rede social, assim como para o cálculo das métricas da rede e dos nodos, foi utilizada a ferramenta *Gephi* 0.9.2 [Bastian et al. 2009]. Neste trabalho, foram consideradas as seguintes métricas:

- Quantidade de Publicações (Publ): quantidade absoluta de trabalhos nos quais este autor participou;
- Grau (Gr): indica com quantos outros autores diferentes houve colaboração na coautoria de trabalhos;
- Centralidade de Intermediação (CI): destaca a importância do autor como elemento de conexão entre diferentes grupos de pesquisa [Umadevi 2013];
- *Pagerank* (PR): ressalta os autores que têm um papel de liderança na rede [Wang et al. 2013].

A identificação das comunidades de coautoria (Com) foi realizada por meio do método Louvain [Blondel et al. 2008]. As instituições de cada um dos autores mais relevantes foram identificadas por meio da aplicação de uma abordagem manual de consulta aos seus currículos na Plataforma Lattes⁵, tendo sido considerado o vínculo atual declarado no resumo do currículo.

Para a coleta de dados da Plataforma Acácia, atualizados em Julho de 2019, foi criada uma ferramenta de *Web Scraping*, em Python 3, com base nas bibliotecas Requests⁶ e BeautifulSoup⁷, capaz de recuperar em lote as informações genealógicas dos pesquisadores a partir de um arquivo de entrada com os seus nomes completos. Como resultado é gerado um arquivo de saída contendo, para cada pesquisador, os seguintes dados:

⁴<https://www.rdocumentation.org/packages/utlils/versions/3.5.1/topics/adist>

⁵<http://lattes.cnpq.br/>

⁶<https://requests.readthedocs.io/en/master>

⁷<https://www.crummy.com/software/BeautifulSoup>

- Pesquisador: nome completo do pesquisador. *Ex.: Philippe Olivier Alexandre Navaux;*
- Grande Área: primeira grande área registrada no Lattes. *Ex.: Ciências Exatas E Da Terra;*
- Área: primeira área registrada no Lattes. *Ex.: Ciência Da Computação;*
- Instituição: instituição em que o pesquisador trabalho, conforme endereço profissional registrado no Lattes. *Ex.: Universidade Federal Do Rio Grande Do Sul;*
- Descendência (Ds): número de descendentes diretos e indiretos. *Ex.: 547;*
- Índice Genealógico (IG): maior número g de descendentes que possuem, no mínimo, g descendentes cada. *Ex.: 10;*
- Fecundidade (Fc): número de descendentes diretos. *Ex.: 83;*
- Fertilidade (Ft): número de descendentes diretos que possuem, no mínimo, um descendente. *Ex.: 25;*
- Gerações (G): número de gerações de acadêmicos, isto é, maior distância hierárquica de um acadêmico em relação a sua descendência). *Ex.: 4;*
- Relações (R): número de conexões (arestas) na descendência de um acadêmico. *Ex.: 553;*
- Primos (Pr): número de acadêmicos que possuem avô(ós) em comum e pai(s) distintos. *Ex.: 0;*
- Data dos dados: data da atualização do Lattes pelo pesquisador. *Ex.: 6/6/2019;*
- Quantidade de orientadores: número de orientações ou supervisões recebidas em nível de mestrado, doutorado ou pós-doutorado registradas no Lattes. *Ex.: 2;*
- Nomes dos orientadores: nomes dos orientadores ou supervisores registrados no Lattes. *Ex.: Francois Anceau - John Rogers.*

3. Resultados e Discussões

Os resultados e discussões aqui apresentados foram divididos em três subseções: Análise de Autores, Análise de Colaborações e Análise de Genealogia.

3.1. Análise de Autores

A Tabela 1 mostra a estatística descritiva das métricas dos autores. A métrica de quantidade de publicações, mostra que mais de 50% (Mediana) dos autores mostra publicou uma única vez, e que mais de 75% (Q_3) dos autores tem no máximo tres publicações. A baixa quantidade de publicações da maioria dos autores também é reforçada pela baixa média (3,210). Já em relação à métrica de Grau, é apresentada uma média de cinco (5,016) colaborações por autor, sendo que mais de 75% (Q_3) dos autores estão dentro desta faixa de valor. As métricas de Centralidade de Intermediação e Pagerank, a fim de aumentar seus baixos valores, foram multiplicadas por constantes, respectivamente 100 e 1000. A Figura 1 mostra, através de um diagrama de caixa, as variações das métricas de Redes Sociais analisadas.

Sob uma perspectiva mais detalhada, a Tabela 2 mostra os autores que tiveram ao menos onze publicações na história da ERAD/RS, o que representa uma média de mais de um trabalho publicado a cada duas edições. Este limite permitiu filtrar os 47 autores com mais publicações entre os 1118 identificados, o que representa 4,2% do total. A primeira parte desta tabela, até o autor 23, apresenta os pesquisadores que publicaram, em média, ao menos um trabalho por edição. Além destes contidos no ranking, foram

Tabela 1. Estatísticas descritivas das métricas dos autores.

Métrica	Min	Q_1	Mediana	Média	Q_3	Max	DP	CV
Pub	1,00	1,00	1,00	3,210	3,00	109	7,92	2,46
Gr	0,00	2,00	3,00	5,016	5,00	111	8,20	1,63
CI*100	0,00	0,00	0,00	1,977	0,00	183	0,01	5,44
PR*1000	0,14	0,47	0,69	0,894	0,91	14,6	1,07	1,27

Obs.: Min: Valor mínimo; Q_1 : Primeiro quartil; Q_3 : Terceiro quartil; Max: Valor máximo; DP: Desvio Padrão; CV: Coeficiente de Variação.

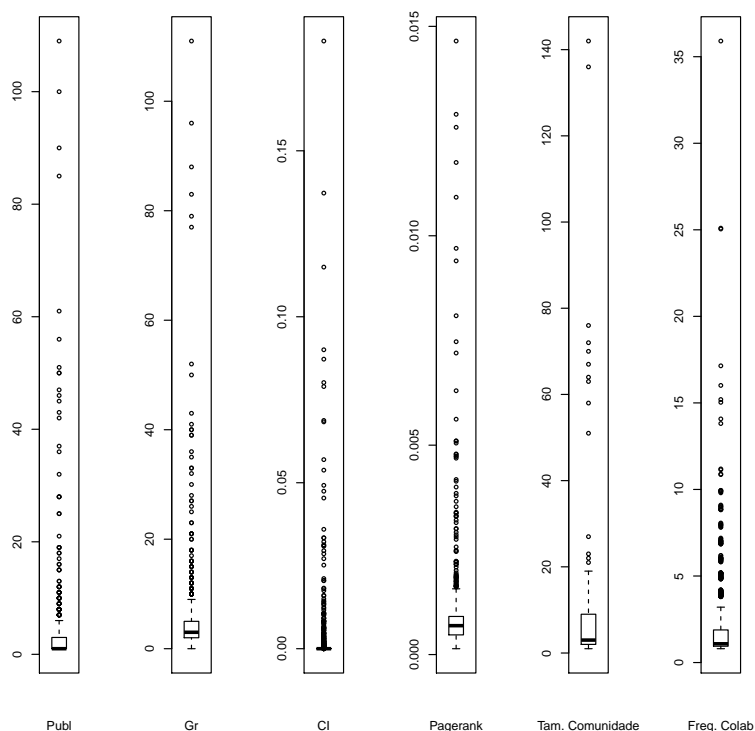


Figura 1. Diagramas de caixa com os valores das métricas dos autores e tamanho das comunidades e frequências das colaborações.

identificados 59 autores que tiveram de seis a dez artigos publicados, 181 autores com três a cinco artigos, 187 autores com dois artigos e 644 autores que participaram de um único artigo, mostrando que 57,55% dos autores publicaram uma única vez no evento. Estes números mostram que uma ampla maioria dos autores participa uma ou poucas vezes no evento. A tabela está ordenada em forma decrescente de Publicações, Grau, Centralidade de Intermediação e *Pagerank*, indicando as respectivas comunidades de coautoria. As estatísticas descritivas das métricas analisadas são apresentadas na Tabela 1.

Embora haja uma quantidade expressiva de *outliers* nos diagramas da Figura 1, optou-se por restringir a análise somente aos maiores valores de cada métrica. Assim, dada esta restrição, e de acordo com as métricas apresentadas, fica evidente a importância dos quatro autores de maior ranking (1, 2 e 3) na história da ERAD/RS. Já em relação ao grau, os autores (1, 2 e 4) apresentam um papel de destaque em relação às suas quantidades

Tabela 2. Ranking, métricas e comunidades dos autores com mais trabalhos publicados na história da ERAD/RS.

N	Nome	Instituição	Publ	Gr	CI	PR	Com
1	Philippe Olivier Alexandre Navaux	UFRGS	109	111	0,18307	0,01465	2
2	Gerson Geraldo Homrich Cavalheiro	UFPEL	100	88	0,11496	0,01290	5
3	Maurício Lima Pilla	UFPEL	90	79	0,07898	0,00940	3
4	Adenauer Corrêa Yamin	UFPEL	85	96	0,06876	0,01092	7
5	Andrea Schwertner Charão	UFSM	61	77	0,13726	0,01259	1
6	Edson Luiz Padoin	UNIJUI	56	83	0,09006	0,01175	2
7	André Rauber Du Bois	UFPEL	51	40	0,05696	0,00510	3
8	Rodrigo da Rosa Righi	UNISINOS	50	52	0,06834	0,00970	10
9	Renata Hax Sander Reiser	UFPEL	50	41	0,00820	0,00478	3
10	César Augusto FonticIELha De Rose	PUCRS	47	39	0,08723	0,00720	4
11	Cláudio Schepke	UNIPAMPA	46	40	0,04922	0,00630	9
12	Nicolas Bruno Maillard	UFRGS	45	33	0,03131	0,00473	4
13	Cláudio Fernando Resin Geyer	UFRGS	43	50	0,05379	0,00747	8
14	Luiz Gustavo Leao Fernandes	PUCRS	42	36	0,01502	0,00562	9
15	Dalvan Griebler	PUCRS	37	32	0,03597	0,00479	9
16	Guilherme Piegas Koslovski	UDESC	36	43	0,08018	0,00809	1
17	Carlos Amaral Hölbig	UPF	32	27	0,03089	0,00474	6
18	Laércio Lima Pilla	UFSC	28	39	0,03341	0,00510	3
19	Tiaraju Asmuz Diverio	UFRGS	28	35	0,04745	0,00505	6
20	Márcia Cristina Cera	UNIPAMPA	28	26	0,04539	0,00418	4
21	Lucas Mello Schnorr	UFRGS	25	25	0,02048	0,00400	2
22	João Vicente Ferreira Lima	UFSM	25	21	0,03339	0,00338	1
23	Marcelo Trindade Rebonatto	UPF	21	21	0,02059	0,00379	6
24	Cristiano André da Costa	UNISINOS	19	28	0,03104	0,00469	10
25	Matheus da Silva Serpa	UFRGS	19	23	0,00401	0,00292	2
26	Benhur Stein	UFSM	19	20	0,03222	0,00301	1
27	Ana Marilza S. Pernas	UFPEL	18	33	0,02858	0,00414	7
28	Francieli Zanon Boito	INRIA	18	27	0,00838	0,00332	2
29	Arthur Francisco Lorenzon	UNIPAMPA	17	21	0,01750	0,00340	4
30	João Ladislau Barbará Lopes	IFSul	16	30	0,00425	0,00330	7
31	Márcio Bastos Castro	UFSC	16	23	0,02113	0,00330	15
32	Rafael dos Santos	CEITEC	15	21	0,01808	0,00300	13
33	Marcelo Pasin	UFSM	15	18	0,02518	0,00286	1
34	Maurício Aronne Pillon	UDESC	15	18	0,01439	0,00366	1
35	Alexandre da Silva Carissimi	UFRGS	13	16	0,01010	0,00223	2
36	Valderi Reis Quietinho Leithardt	UNIVALI	13	14	0,02952	0,00315	8
37	Rafael Bohrer Ávila	IFSul	12	18	0,00226	0,00249	2
38	Mário Dantas	UFJF	12	16	0,02712	0,00385	12
39	Daniela Saccol Peranconi	UNISC	12	9	0,00778	0,00153	5
40	Patrícia Pitthan de Araújo Barcelos	UFSM	12	9	0,00308	0,00183	1
41	Alan Schlindvein de Araújo	UFPEL	12	9	0,00017	0,00117	5
42	Luthiano Rodrigues Venecian	UCPEL	11	20	0,00027	0,00216	7
43	Haroldo Fraga de Campos Velho	INPE	11	14	0,00072	0,00221	1
44	Fábio Diniz Rossi	IFFar	11	13	0,01593	0,00222	4
45	Jorge Luis Victoria Barbosa	UNISINOS	11	13	0,01321	0,00188	5
46	Adriano Fiorese	UDESC	11	12	0,01825	0,00338	16
47	Charles Christian Miers	UDESC	11	11	0,00843	0,00281	1

Obs.: Os três maiores valores para cada métrica estão enfatizados em negrito.

de colaborações. As mais altas centralidades de intermediação mostram que os autores 1, 2 e 5 desempenharam um papel significativamente relevante na comunicação entre dife-

rentes comunidades de pesquisa, sendo importantes na integração destas comunidades no contexto da ERAD/RS. Os maiores valores de *Pagerank* dos autores 1, 2 e 5 evidenciam que eles possuem um papel de liderança no contexto do evento, provavelmente decorrente de suas atividades como formadores de novos pesquisadores nesta área.

Em relação aos estados, das dezenove diferentes instituições representadas na Tabela 2, três são de Santa Catarina (UDESC, UFSC e UNIVALI), uma de Minas Gerais (UFJF)⁸ e uma estrangeira (INRIA). Todas as demais catorze instituições são do Rio Grande do Sul. Neste sentido, fica evidente a possibilidade de expansão do evento nos estados de Santa Catarina e, principalmente, Paraná.

Neste trabalho, através da aplicação do método Louvain [Blondel et al. 2008], foram identificadas 76 comunidades de pesquisa das quais algumas continham apenas um autor. Conforme a Tabela 2, os autores com maior quantidade de publicações fazem parte das maiores comunidades de pesquisa, ficando evidente o papel das comunidades de pesquisa na produtividade dos pesquisadores. O tamanho médio das comunidades é $1,645 \pm 1,703$ autores, mostrando um potencial para integração de autores isolados ou de comunidades com poucos autores.

Na Tabela 2, as comunidades 1, 2 e 3 englobam mais de 50% dos autores presentes no ranking. A maior comunidade (Com 1), composta por 142 autores, representada pela cor magenta, é composta principalmente por membros da UFSM e UDESC, também com participação do INPE. A principal autora desta comunidade, Andrea Schwertner Charão (N 5), está entre as maiores métricas de CI e PR. A segunda maior comunidade (Com 2), com 136 autores, indicada pela cor verde, engloba principalmente membros da UFRGS, com participação da UNIJUI, INRIA e IFSul. O principal autor desta comunidade, Philippe Olivier Alexandre Navaux (N 1), tem o maior valor para todas as métricas analisadas. A terceira maior comunidade (Com 3), composta por 76 autores, ilustrada pela cor azul, tem membros da UFPEL, com participação de membros da UFSC.

A Figura 2 mostra a rede de colaboração da ERAD/RS. Nodos representam os autores que participaram de artigos publicados. Arestas representam colaborações em coautoria. O tamanho dos nodos é proporcional à quantidade publicações do autor. A espessura das arestas é proporcional à quantidade de colaborações entre os autores. As cores dos nodos representam as principais comunidades de coautoria. Podem ser identificadas as cores das nove maiores comunidades. A partir da décima maior comunidade, a cor utilizada para representação foi cinza. A fim de permitir a visualização com foco exclusivo nos autores mais representativos, e suas respectivas colaborações, foram aplicados filtros para autores com mais de dez publicações, segundo a Figura 3.

3.2. Análise de Colaborações

A Tabela 3 mostra as treze colaborações mais frequentes entre as 2804 encontradas nos vinte anos da ERAD/RS, tendo sido selecionadas aquelas que se repetiram onze vezes ou mais, representando 0,46% do total. Foi dada ênfase às três primeiras colaborações, que se repetiram com frequência superior à quantidade de edições da Escola. Além destas, 64 colaborações entre autores se repetiram entre seis e dez trabalhos publicados, 296 colaborações que ocorreram entre três e cinco vezes, 423 colaborações em dois trabalhos

⁸Presente em função do pesquisador Mario Dantas que até 2017 atuava na UFSC.

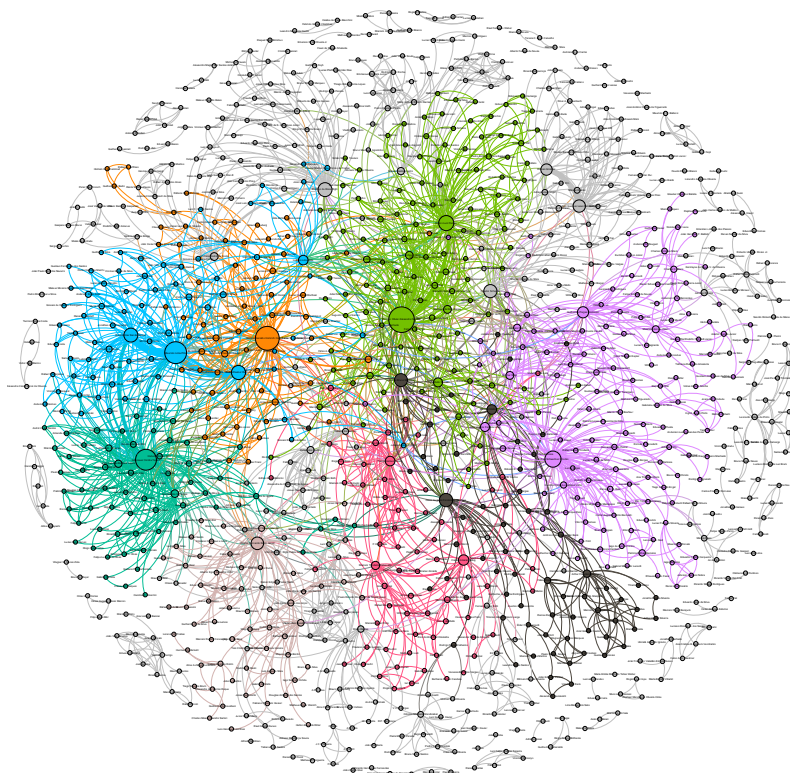


Figura 2. Rede de colaboração histórica dos 20 anos da ERAD/RS.

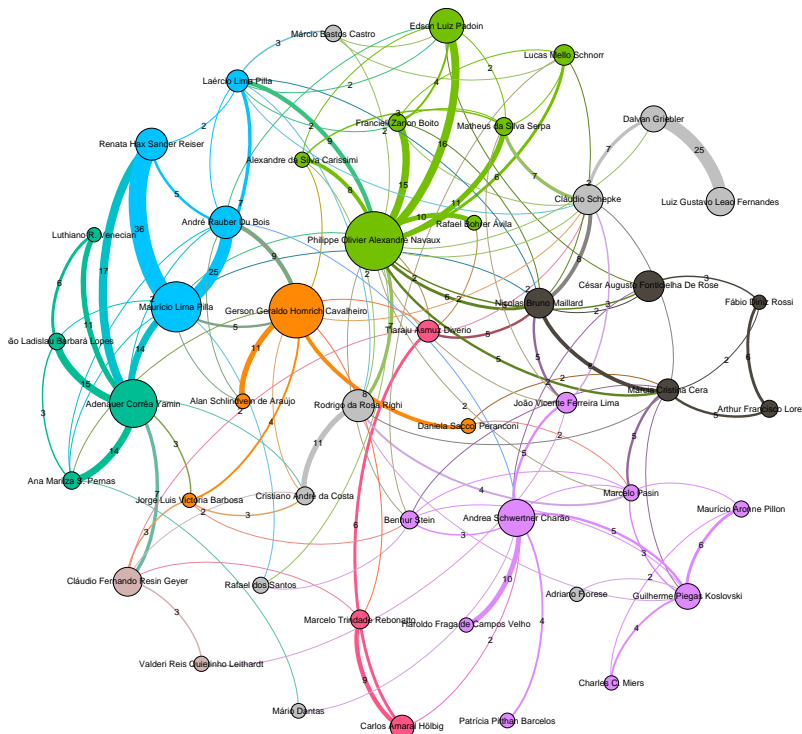


Figura 3. Rede de colaboração histórica dos 20 anos da ERAD/RS para autores com 11 ou mais publicações.

e 2008 colaborações que ocorreram uma única vez, representando 71,6% do total. Pode ser evidenciado que as treze colaborações que mais se repetem ocorreram com os autores previamente apresentados na Tabela 2, e que a maioria destas ocorreu uma única vez.

Tabela 3. Maiores quantidades de colaborações de coautoria.

N	Autores	Freq
1	Maurício Lima Pilla (N 3) e Renata Hax Sander Reiser (N 9)	36
2	Maurício Lima Pilla (N 3) e André Rauber Du Bois (N 7)	25
3	Luiz Gustavo Leao Fernandes (N 14) e Dalvan Griebler (N 15)	25
4	Adenauer Corrêa Yamin (N 4) e Renata Hax Sander Reiser (N 9)	17
5	Philippe Olivier Alexandre Navaux (N 1) e Edson Luiz Padoin (N 6)	16
6	Adenauer Corrêa Yamin (N 4) e João Ladislau Barbará Lopes (N 30)	15
7	Philippe Olivier Alexandre Navaux (N 1) e Francieli Zanon Boito (N 28)	15
8	Maurício Lima Pilla (N 3) e Adenauer Corrêa Yamin (N 4)	14
9	Adenauer Corrêa Yamin (N 4) e Ana Marilza S. Pernas (N 27)	14
10	Adenauer Corrêa Yamin (N 4) e Luthiano Rodrigues Venecian (N 42)	11
11	Gerson Geraldo H. Cavalheiro (N 2) e Alan Schlindvein de Araújo (N 41)	11
12	Rodrigo da Rosa Righi (N 8) e Cristiano André da Costa (N 24)	11
13	Philippe Olivier Alexandre Navaux (N 1) e Matheus da Silva Serpa (N 25)	11

Obs.: Freq: Frequência de colaborações entre os autores.

3.3. Análise de Genealogia

Por fim, utilizou-se a ferramenta de *Web Scraping* desenvolvida para coleta dos indicadores individuais de genealogia dos 46⁹ principais autores, bem como para a verificação dos vínculos institucionais destes pesquisadores e, por conseguinte, da identificação das instituições cujos pesquisadores mais publicaram em determinado contexto. Todos os autores estão cadastrados na grande área Ciências Exatas e da Terra (43 na área de Ciência da Computação e 1 na Matemática), com exceção de dois: um em Outros (área de Microeletrônica) e outro em Não identificada (área não identificada).

A Figura 4 contempla os resultados observados para os seis primeiros indicadores disponibilizados na Plataforma Acácia. Para cada indicador, podem ser vistos os dez autores com maior destaque. Conforme pode ser observado, os pesquisadores Philippe Olivier Alexandre Navaux (N 1) e Cláudio Fernando Resin Geyer (N 13) se destacam em todos estes indicadores, o que os coloca na condição de principais precursores das pesquisas em PAD na Região Sul, com contribuições diretas para a formação inicial e sustentação das redes de colaboração que se consolidaram ao longo dos últimos vinte anos, compostas por pesquisadores produtivos, tanto no que se refere à quantidade de produções científicas na área, como na quantidade de orientações de novos pesquisadores. Estas redes podem ser observadas na Figura 5, onde os nodos são os autores e as arestas representam as relações de orientação, com seta apontando para o orientador.

A Figura 6 apresenta os resultados dos dez autores com maiores quantidades de Primos (Pr). A quantidade expressiva observada sugere que estes autores possuem grande

⁹Um dos 47 autores – Alan Schlindvein de Araujo – não consta na Plataforma Acácia pois possui apenas graduação registrada no Lattes

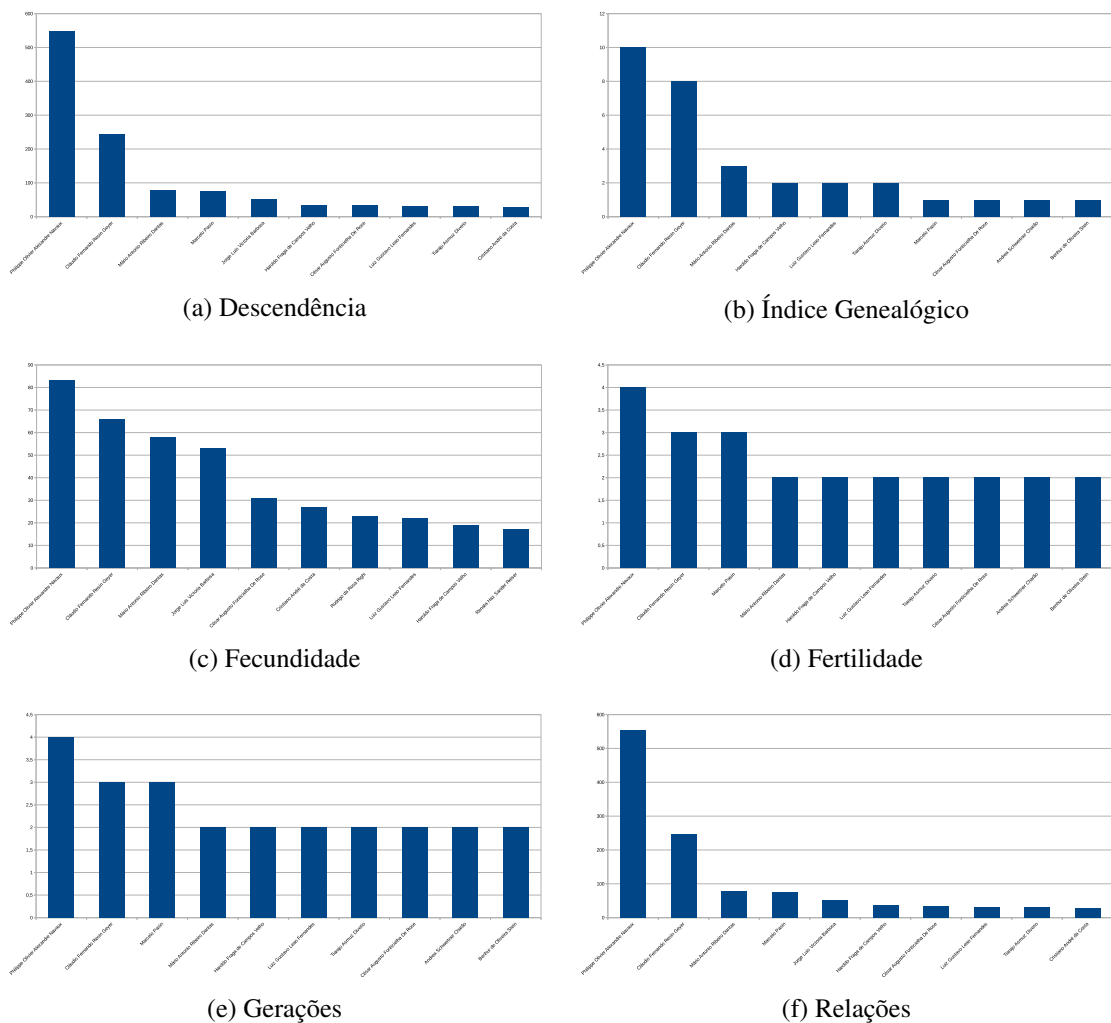


Figura 4. Autores com maiores indicadores genealógicos.

potencial para desenvolvimento de projetos em colaboração nos próximos anos devido ao facilitador causado pela presença de vínculo genealógico com diversos pesquisadores.

Além destes indicadores, outro dado relevante trazido pela ferramenta de *Web Scraping* é a quantidade de orientadores, a qual pode ser contrastada com os demais indicadores para identificar possíveis correlações. Por exemplo, no contexto composto pelo conjunto de pesquisadores que colaborou como autor de mais trabalhos na ERAD/RS nos últimos vinte anos, surgem duas questões interessantes de pesquisa, passíveis de análise preliminar com os resultados obtidos:

- 1) Existe correlação entre ter recebido orientação de diferentes pesquisadores e ter sido capaz de fomentar uma robusta rede de pesquisa a partir dos pesquisadores que formou em nível de mestrado, doutorado ou pós-doutorado?
- 2) Existe correlação entre ter recebido orientação de diferentes pesquisadores e ter sido capaz de formar mais pesquisadores nestes níveis?

De acordo com os resultados da Figura 7, apresentados em escala logarítmica, é possível perceber duas leves tendências: quantidade de descendentes (indicada por Ds)

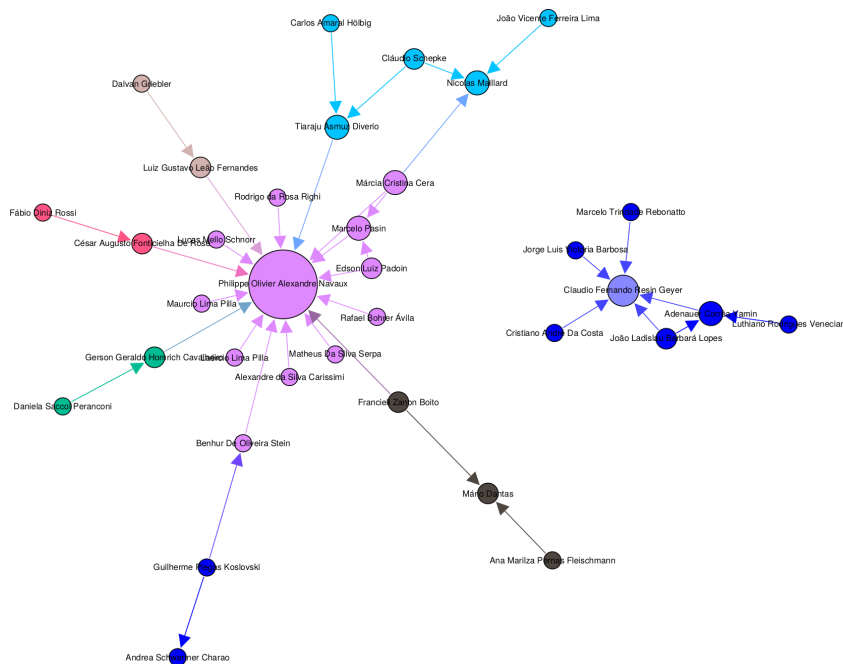


Figura 5. Redes hierárquicas de orientação.

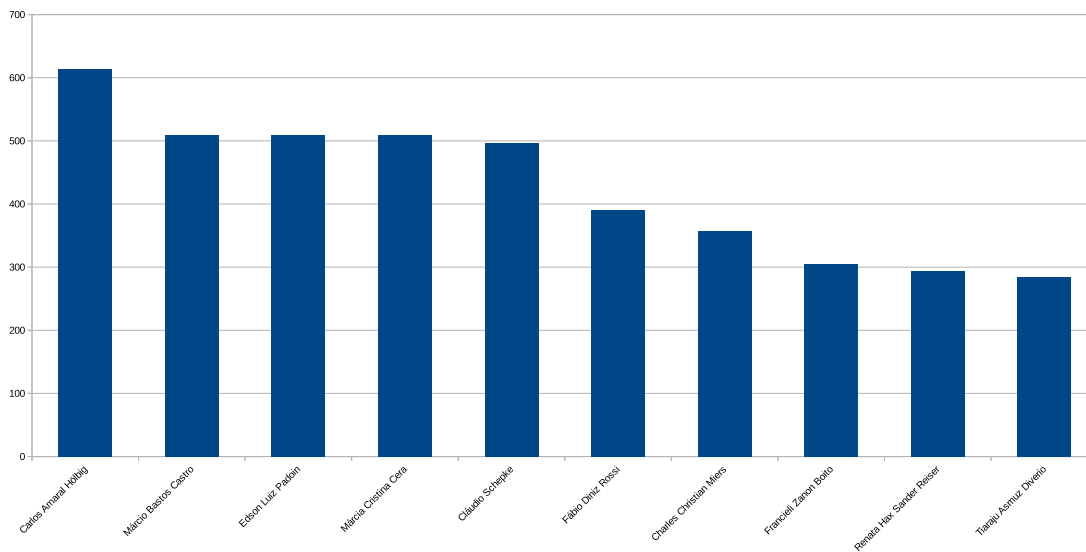


Figura 6. Dez autores com quantidade de Primos.

tendeu a crescer com o aumento no número de orientações diferentes recebidas, enquanto que a quantidade de pesquisadores formados (indicada por F_c) tendeu a cair. Contudo, como o R é baixo, respectivamente 0,002 e 0,007, é possível afirmar que, para este conjunto de pesquisadores, que a experiência com uma quantidade maior de orientadores não é determinante para um aumento significativo na capacidade de constituir redes de pesquisa robustas, tampouco de ampliar o número de pesquisadores diretamente formados.

A Tabela 4 sintetiza as instituições às quais estavam vinculados os autores principais da ERAD, conforme registro presente no Lattes em Julho de 2019. É possível notar

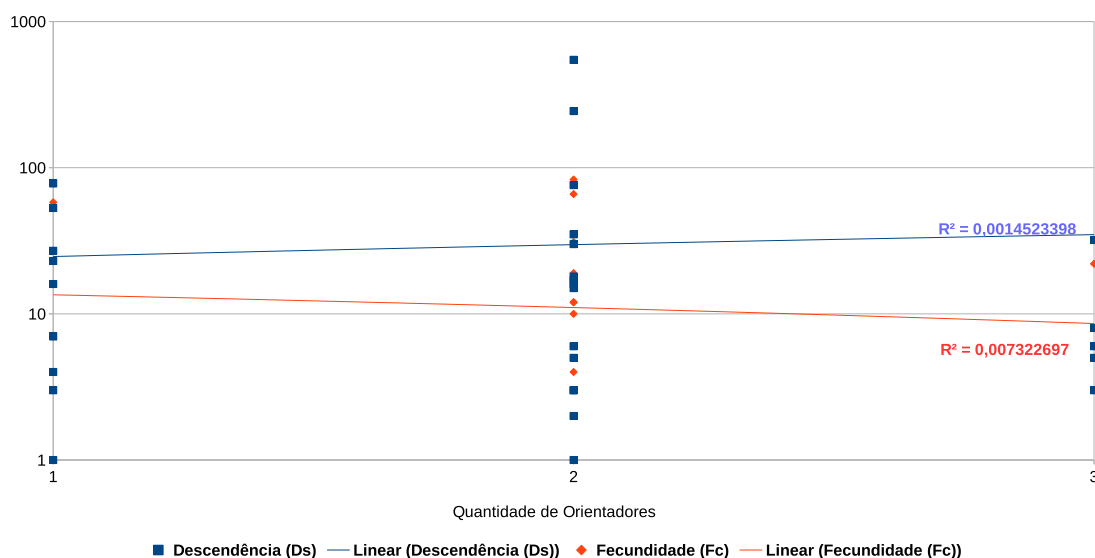


Figura 7. Correlação entre quantidade de orientadores e indicadores genealógicos.

que cerca de um terço dos principais autores possuíam vínculo com a UFRGS, UFPEL ou UFSM. Além disso, percebe-se que as cinco Universidades Federais da Região Sul presentes na lista – UFRGS (8), UFPEL (6), UFSM (5), UNIPAMPA (2) e UFSC (2) – concentravam a metade destes autores (23).

Tabela 4. Quantidade de autores principais por instituição conforme Plataforma Acácia.

Instituição	Autores
Universidade Federal Do Rio Grande Do Sul	8
Universidade Federal De Pelotas	6
Universidade Federal De Santa Maria	5
Universidade Do Estado De Santa Catarina	4
Universidade Do Vale Do Rio Dos Sinos	3
Pontifícia Universidade Católica Do Rio Grande Do Sul	3
Universidade Federal Do Pampa	2
Universidade Federal De Santa Catarina	2
Universidade De Passo Fundo	2
Não identificada	1
Instituto Federal Sul Rio Grandense	2
Universidade Federal De Juiz De Fora	1
Universidade De Santa Cruz Do Sul	1
Universidade Da Beira Interior	1
Instituto Nacional De Pesquisas Espaciais	1
Instituto Federal Farroupilha Campus Alegrete	1
Fundação De Integração Desenvolvimento E Educação Do Noroeste Do RS	1
Centro Nacional De Tecnologia Eletrônica Avançada	1
Centre De Recherche INRIA Grenoble Rhône Alpes	1

4. Conclusões e Trabalhos Futuros

Esta pesquisa documental analisou quantitativamente os trabalhos publicados na ERAD/RS de 2001 a 2019, e os trabalhos aceitos em 2020. Foram encontrados os principais autores do evento, com base na análise das métricas de quantidade de publicações, grau de colaboração, centralidade de intermediação e *Pagerank*. Foram identificadas as quatro principais comunidades de coautoria, assim como suas quantidades de autores e principais elementos. Espera-se que os resultados aqui apresentados contribuam para o reconhecimento dos autores que fizeram a história do evento e auxiliem ações de expansão das comunidades de pesquisa em PAD no sul do país. Salienta-se que a execução desta pesquisa foi totalmente dependente da acesso aos anais de todas as edições do evento. Neste sentido, faz-se necessário parabenizar a CRAD/RS pela publicação destes na Biblioteca Digital Brasileira de Computação. Além disso, a partir da lista de principais autores e de técnicas de *Web Scraping* usadas para coleta de dados de genealogia acadêmica disponibilizados na Plataforma Acácia, foi possível (i) evidenciar os pesquisadores precursores na área de PAD na Região Sul, (ii) ressaltar o potencial para o crescimento de pesquisas colaborativas na área nos próximos anos, (iii) concluir preliminarmente que a quantidade de orientações diferentes não está diretamente relacionada a capacidade de constituir redes de pesquisa robustas, tampouco de ampliar o número de pesquisadores diretamente formados e (iv) destacar tanto a expressiva participação de pesquisadores vinculados a Universidades Federais localizadas no RS e SC, como a ausência de pesquisadores de instituições do Paraná, sugerindo a necessidade de ações específicas para estimular a sua participação no futuro, as quais podem incluir a adoção de edições itinerantes da ERAD, até então realizadas sempre no RS, nos demais estados da região. Como trabalho futuro, almeja-se estudar a relevância dos pesquisadores da Região Sul na comunidade nacional, por meio da aplicação de metodologia similar à do presente trabalho nos artigos publicados nos principais eventos da área no contexto brasileiro: *International Symposium on Computer Architecture and High Performance Computing* (SBAC-PAD) e Simpósio de Sistemas Computacionais de Alto Desempenho (WSCAD).

Agradecimentos

O presente trabalho foi realizado com suporte da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, tendo em vista o vínculo dos autores ao Programa de Pós-Graduação em Computação Aplicada (PPGCAP), resultado da cooperação entre UNIPAMPA e EMBRAPA. Destaca-se também o apoio recebido da coordenação da comissão executiva da CRAD/RS e dos organizadores da 20^a edição da ERAD/RS, em especial no acesso aos dados dos trabalhos aceitos para apresentação. Adicionalmente, os autores gostariam de agradecer, *in memoriam*, à pesquisadora Márcia Cristina Cera, pelas suas relevantes contribuições à área de PAD no RS, sobretudo pelo seu engajamento na organização de edições da ERAD/RS e na formação de novos pesquisadores na UNIPAMPA.

Referências

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media*, volume 8, pages 361–362, Menlo Park, California. The AAAI Press.

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):1–12.
- Brum, L. M. L., Lampert, V. N., and Camargo, S. S. (2019). Business intelligence and data warehouse in agrarian sector: A bibliometric study. *Journal of Agricultural Science*, 11(2):353 – 368.
- Camargo, S. S., Brum, L. M. L., and Pinho, L. B. (2019a). ERAD/RS: Uma visão histórica de autores e redes de colaboração. In *Anais da 19ª Escola Regional de Alto Desempenho da Região Sul*, volume 1, pages 87–94, Três de Maio, RS.
- Camargo, S. S. and Camargo, S. L. T. (2018). 20 anos da revista do CCEI: Um panorama de autores e comunidades de coautoria. *Revista do CCEI*, 23(38):1 – 15.
- Camargo, S. S., Pinho, L. B., Bosch, M., Machado, C., and Saibene, Y. B. (2020). Diez años del congreso argentino de agroinformática: Un análisis histórico del alcance geográfico y redes de colaboración. *Electronic Journal of SADIO (EJS)*, 19(1):118–139.
- Camargo, S. S., Pinho, L. B., and Saibene, Y. B. (2019b). Quién es quién en la agroinformática argentina. In *Anales del XI Congreso Argentino de Agroinformática (CAI)*, pages 59–72, Salta, Argentina.
- Damaceno, R., Rossi, L., and Mena-Chalco, J. (2017). Identificação do grafo de genealogia acadêmica de pesquisadores: Uma abordagem baseada na plataforma lattes. In *Proceedings of the 32nd Brazilian Symposium on Databases*, pages 76–87.
- Mitchell, R. (2018). *Web Scraping with Python: Collecting Data from the Modern Web*. O’Reilly Media, Inc., 2nd edition.
- Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5200–5205.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Roztocki, N., Soja, P., and Weistroffer, H. R. (2019). The role of information and communication technologies in socioeconomic development: towards a multi-dimensional framework. *Information Technology for Development*, 25(2):171–183.
- Sampaio, R. B., Sacerdote, H. C. S., Fonseca, B. P. F., and Fernandes, J. H. C. (2015). A Colaboração científica na pesquisa sobre coautoria: um método baseado na análise de redes. *Perspectivas em Ciência da Informação*, 20:79 – 92.
- Schnorr, L. M. and Pillon, M. A. (2018). Apresentação. In Schnorr, L. M. and Pillon, M. A., editors, *Anais da ERAD/2018*. Sociedade Brasileira de Computação.
- Su, Y. F. (2010). *Study on the Cooperation Network of Sport Research Papers*. Shanghai University of Sport, Shanghai.
- Umadevi, V. (2013). Case study - centrality measure analysis on co-authorship network. *Journal of Global Research in Computer Science*, 4(1):67 – 70.
- Wang, R., Zhang, W., Deng, H., Wang, N., Miao, Q., and Zhao, X. (2013). Discover community leader in social network with pagerank. In Tan, Y., Shi, Y., and Mo, H., editors, *Advances in Swarm Intelligence*, pages 154–162, Berlin, Heidelberg. Springer.