

Análise de uma década de SEMISH: um breve panorama do precursor cinquentenário e ancião do CSBC

Luiz Paulo Carvalho¹, Silas Lima Filho¹, Michele A. Brandão²,
Jonice Oliveira¹, Flávia Maria Santoro³, Mônica Ferreira da Silva¹

¹UFRJ – Rio de Janeiro, RJ – Brasil

²IFMG/UFMG – Belo Horizonte, MG – Brasil

³UERJ – Rio de Janeiro, RJ – Brasil

luiz.paulo.carvalho@ppgi.ufrj.br, silaslfilho@ppgi.ufrj.br

michele.brandao@ifmg.edu.br, jonice@dcc.ufrj.br,

flavia@ime.uerj.br, mfsilvamail@gmail.com

Abstract. *CSBC is the biggest scientific event dedicated to computing in Latin America, integrating Brazilian computing and dealing with the direction of the area in the country. In one of its base events, SEMISH, there has been a significant Brazilian computational academic-scientific congregation for 50 years. The analysis of the last decade of SEMISH exposes a low recurrence of authorships, convergence between more prolific authors and with a greater number of collaborations and low participation of women. There are continuous and consistent connections. Hardware has almost negligible relevance, and software the opposite, with machine learning gaining prominence.*

Resumo. *O CSBC é o maior evento científico dedicado à computação da América Latina, integrando a computação brasileira e tratando dos rumos da área no país. Em um de seus eventos base, SEMISH, há congregação acadêmico-científica computacional brasileira significativa há 50 anos. A análise da última década de SEMISH expõe uma baixa reincidência de autorias, convergência entre autores mais prolíficos e com maior número de colaborações e baixa participação de mulheres. Notam-se conexões contínuas e consistentes. Hardware apresenta relevância quase insignificante, e software o contrário, com aprendizado de máquina ganhando destaque.*

1. Introdução

Em 2023 o Seminário Integrado de Hardware e Software (SEMISH) alcança sua quinquagésima edição, e sua importância histórica é ímpar ao cenário computacional brasileiro:

“Outro evento acadêmico histórico dessa época foi o Seminário Integrado de Software e Hardware (SEMISH). Em 1974, foi organizado internamente na UFRGS o primeiro SEMISH, criado para estimular o trabalho multidisciplinar. Em 1975, também na UFRGS foi organizado o segundo SEMISH como uma atividade de abrangência nacional. Em 1976, o SEMISH consolidou-se e transformou-se em evento permanente [...]

[...]

Ambos eventos e seus participantes, SECOMU desde 1971 e o SEMISH desde 1974, foram o embrião de uma sociedade científica de Computação no Brasil [...]. Uma das

principais motivações para criar sociedades científicas brasileiras foi o desejo de ampliar o contato e o conhecimento entre cientistas do Brasil e do exterior [...].

Com o objetivo de reunir a Comunidade Científica e Acadêmica de Computação do Brasil, a Sociedade Brasileira de Computação (SBC) foi fundada em 1978. A fundação ocorreu durante o 5º SEMISH, realizado na UFRJ, com a presença de cerca de 30 pessoas.

A SBC foi criada como sociedade científica para a organização de eventos e publicações da comunidade de pesquisa. A SBC gradualmente tornou-se a principal porta-voz da comunidade de Computação no Brasil [...]

Desde a década de 70, os eventos foram então espaços relevantes para as discussões de uma comunidade científica e profissional no desenvolvimento e divulgação da Computação no Brasil.” [Wazlawick e Silva Junior 2021]

O SEMISH é um evento de congregação acadêmico-científica computacional brasileira de enorme longevidade e significância, comparado com outros eventos, em 2023 se consagra mais antigo que outros simpósios renomados ¹, Simpósio Brasileiro de Engenharia de Software (37ª edição); Simpósio Brasileiro de Bando de Dados (38ª edição); e o Simpósio Brasileiro de Redes de Computadores (41ª edição). Ademais o caráter histórico que moldou as direções e caminhos da computação brasileira, apresenta contribuições acadêmico-científicas e tecnológicas tão significantes.

No âmbito científico tradicional, a ciência avança ancorada nas publicações e comunicações científicas formais [Agassi 2008], neste presente trabalho apresentamos uma visão materialista através de outras dimensões [Latour 1987, Lefèvre 2005], com uma ênfase meta-científica [Ioannidis et al. 2015, Ioannidis 2018] por uma abordagem analítico-descritiva [Marconi e Lakatos 2017, Wazlawick 2014]. Excedendo a limitação das publicações, analisamos outras sete dimensões: autoria, sexo, instituição/filiação, geolocalização institucional, idioma e resumos + palavras-chave.

Nossa motivação central e primária é analítico-descritiva e tem o objetivo de **apresentar uma análise descritiva baseada em Análise de Redes Sociais (ARS) e estatística sobre os dados e os metadados da última década do SEMISH, através de suas publicações, envolvendo dados diversos disponíveis, possíveis e válidos.** À esta categoria de pesquisas é facultada a ausência de perguntas de pesquisa ou hipótese [Marconi e Lakatos 2017], com ênfase nas informações e conhecimentos gerados; expomos o panorama de uma década de SEMISH, compondo uma contribuição meta-científica descritiva do cenário de educação e computação brasileiro.

Alguns trabalhos similares analisam comunidades e espaços acadêmico-científicos brasileiros sob uma perspectiva meta-científica [Ioannidis 2018], como [Lobato et al. 2021] e [Digiampietri et al. 2017] sobre o Workshop *Brazilian Social Network Analysis and Mining* (BraSNAM); [Santana e Braga 2020] sobre o Workshop *Women in Information Technology* (WIT); [Barbosa et al. 2017] sobre o Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC-Br); ou [Kauer e Moreira 2013] sobre a evolução dos tópicos de interesse no Simpósio Brasileiro de Banco de Dados (SBBD). O exame sobre si, envolvendo análises, avaliações e percepções formais e estruturadas; permite que rede(s) ou comunidade(s) amadureçam, tomem decisões baseadas em evidências e se identifiquem objetivamente [Ioannidis et al. 2015].

¹<https://cbsoft2023.ufms.br/sbes> – <https://sbbd.org.br/2023/> – <https://sbrc.sbc.org.br/2023/> [acesso 04-04-2023]

O trabalho está estruturado da seguinte forma, a Seção 2 expõe o método, materiais e conceitos essenciais; a Seção 3 apresenta os resultados; e a Seção 4 encaminha as considerações finais.

2. Método, recursos, fontes de dados e conceitos

A Figura 1 expõe o método deste trabalho. **Coletamos dados** disponíveis e possíveis para análise presentes no *CSBCSet*, um conjunto de dados representando publicações de 2013 a 2022 do Congresso da Sociedade Brasileira de Computação (CSBC), com diversos dados e metadados [Lima Filho et al. 2023]. As análises, individuais ou combinadas, partem de dados dos atributos: Ano, Evento, Edição, Título, Pessoa autora, Sexo, Instituição 1, UF Inst. 1, Idioma, Resumo, Palavras-chave². O *CSBCSet* traz dados sobre diversos eventos, para as análises neste presente trabalho utilizamos o sub-conjunto de dados “SEMISH” no atributo Evento.

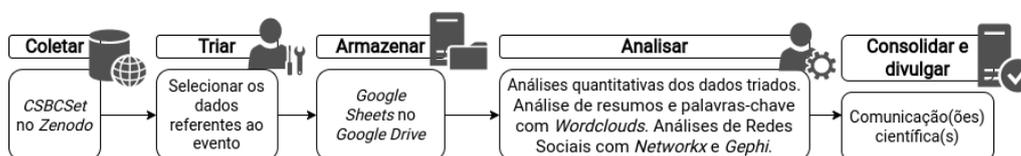


Figura 1. Etapas da metodologia aplicada neste trabalho.

A partir do método, utilizamos os seguintes sistemas computacionais³: *NetworkX* e *Gephi*, gratuitos e abertos, usados para criação, manipulação e estudo de estruturas, dinâmicas e funções de redes complexas; (iii) *Microsoft Excel*, versão comercial, e *Google Sheets*, gratuito, usados para tabelas e planilhas para estruturação, armazenagem e compartilhamento de dados, com funcionalidades aritméticas e estatísticas, representação gráfica de dados e combinações complexas. Utilizamos as funcionalidades gratuitas do sistema online *WordClouds* para geração das nuvens de palavras.

Em relação à etapa de **análise**, objetivo primário deste trabalho, realizamos análises quantitativas e estatísticas gerais e conduzimos métodos e abordagens em ARS [Tabassum et al. 2018]. Em ambos os casos, o SEMISH foi tratado como um todo e cada ano separadamente. Pelo excesso de informações geradas, e para melhor visualização, disponibilizamos os artefatos gerados online⁴.

Algumas definições e conceitos de ARS são fundamentais ao decorrer deste trabalho [Tabassum et al. 2018]: Componente gigante é um sub-grafo da rede principal com a maior quantidade de nós conectados e sua medida é importante para mostrar o quão coesiva ou separada uma rede é; coeficiente de agrupamento representa a tendência de um nó formar uma comunidade, ou seja, a tendência dos vizinhos de um nó se conectarem uns aos outros; o diâmetro é o menor caminho entre dois nós com a maior distância na rede; densidade é a proporção de conexões possíveis na rede em relação às existentes; e, finalmente, um clique consiste de um sub-grafo completo na rede, no qual todos os nós estão conectados uns aos outros.

²<https://zenodo.org/record/7977462> [acesso 04-04-2023]

³<https://networkx.org/> - <https://gephi.org/> - <https://en.wikipedia.org/wiki/Spreadsheet> - <https://www.wordclouds.com/> [acesso 04-04-2023]

⁴https://4658.short.gy/CSBC_ARS_2023 [acesso 04-04-2023]

2.1. Aspectos éticos

Extraímos esta discussão de [Carvalho et al. 2023], pela semelhança e aplicação do *CSBCSet*. Manter os nomes das pessoas autoras ou seguir com anonimato foi um dilema ético que nos acometeu. **Moralmente**, é ausente uma justificativa, fundamentação ou base que objetivamente **determine** um anonimato, pelas diretrizes institucionais governamentais que regem a ética em pesquisa brasileira [Brasil 2016] ou questões éticas gerais de pesquisa [ANPEd 2019, Bos 2020, London 2022]. Pois lidamos com dados abertos; dados de promoção e publicidade de atividade-fim; sem qualquer dado pessoal sensível ou potencialmente prejudicial ou consequencialmente negativo às partes; consiste de uma pesquisa secundária que trata especificamente de dados meta-científicos; por fim, é ausente de qualquer juízo moral ou de valor primário associado aos dados.

Eticamente debatemos uma série de possíveis consequências negativas ou prejudiciais materiais e concretas, além da ética relativista (e.g., “não quero” ou “não gosto”), e fomos incapazes de pontuar justificativas morais significativas o bastante para seguir com o anonimato. Adicionalmente, pessoas autoras podem utilizar destes dados ou informações para sua vantagem, e.g., em memoriais de promoção de carreira.

3. Resultados e discussão

Nesta seção apresentamos as análises quantitativas (Seção 3.1), gerais e centralizadas em cada um dos atributos-chave; e as análises baseadas em ARS (Seção 3.2) da colaboração entre pesquisadores e instituições.

3.1. Análises quantitativas gerais

Nesta seção apresentamos as análises quantitativas gerais, aritméticas ou estatísticas, combinando os dados e informações sobre o SEMISH presentes no *CSBCSet*, de suas publicações entre 2013 – 2022, excluindo comunicações como abertura, prefácios, mensagens da coordenação, dentre outros.

Como primeiro atributo, iniciamos pela **análise de autorias e pessoas autoras/autores**, a Tabela 1 expõe os resultados das publicações e suas autorias no conjunto (a). Há uma distribuição esparsa da quantidade de autores por publicação em 2013 e 2015, i.e., proporcionalmente mais autores para menos publicações; em uma relação de ≈ 5 para 1; em 2014 este valor é de ≈ 3 para 1; de 2016 em diante, de ≈ 4 para 1.

Se considerarmos 10% de todas as publicações (≈ 18) como um valor estável para os anos, e 10% deste valor (≈ 2) como um valor estável de variância, variando entre 16 e 20, percebemos que o SEMISH é um evento instável no quantitativo de publicações por ano. Percebemos anos nos quais o valor é muito abaixo (8 em 2017) ou muito acima (35 em 2021). Fomos incapazes de encontrar os dados de submissão e aceite, o que nos limita a esta análise simples.

Em relação à quantidade de autorias por publicação e ano, 472 ($\approx 82\%$) apresentaram apenas uma autoria (i.e., publicaram apenas um trabalho, seja como autor individual ou em colaboração); 72 ($\approx 13\%$), duas autorias; 18 ($\approx 3\%$), três autorias; o somatório das demais é de $\approx 2\%$. Isolado como autor envolvido na maior quantidade de autorias está Adenauer Yamin, com 15 autorias. Com 6 autorias, Marcos César Seruffo, Ana Marilza Pernas, João Ladislau Lopes; com 5, Thais Batista e Cinthyan de Barbosa. Sendo assim, podemos considerar estes os autores mais recorrentes no SEMISH.

Tabela 1. Análises quantitativas gerais de publicações, sexo e idioma

Ano	(a) Publicações		(d) Idioma das publicações				(b) Autores (todos) por sexo **				(c) Autores (únicos) por sexo **			
	Por autor (todos)	Por publicação	pt-br	Qtd. %	en	Qtd. %	M	Qtd. %	F	Qtd. %	M	Qtd. %	F	Qtd. %
2013	61	12	10	83%	2	17%	47	77%	12	20%	9	75%	2	17%
2014	42	13	10	77%	3	23%	29	69%	9	21%	10	77%	2	15%
2015	105	21	20	95%	1	5%	84	80%	17	16%	18	86%	2	10%
2016	74	19	16	84%	3	16%	55	74%	17	23%	16	84%	2	11%
2017	30	8	3	38%	5	63%	20	67%	9	30%	6	75%	1	13%
2018	40	10	8	80%	2	20%	30	75%	10	25%	8	80%	2	20%
2019	63	15	14	93%	1	7%	49	78%	9	14%	11	73%	4	27%
2020	93	24	18	75%	6	25%	67	72%	23	25%	21	88%	2	8%
2021	135	35	23	66%	12	34%	100	74%	30	22%	27	77%	8	23%
2022	82	22	17	77%	5	23%	59	72%	16	20%	17	77%	4	18%
Geral	725	179	139	78%	40	22%	540	74%	152	21%	143	80%	29	16%

** Alguns dos dados deste atributo constam como “-” (desconhecidos); sendo assim, o somatório pode não alcançar 100%.

O grafo na Figura 2 expõe o componente gigante acumulado de 2013 – 2022 no SEMISH. Percebemos que há uma semelhança entre os autores mais prolíficos e o componente gigante, indicando que tanto apresentam mais autorias, como uma rede fechada, estável e centralizada neles, com ênfase em Adenauer Yamin. Há uma concentração da região Sul, especificamente envolvendo a UCPel, UFPel, IFSul. As cores diferentes na Figura 2 indicam anos de publicação distintos, há proporcional excesso de interações, colaborações e produções dos autores envolvidos nesta rede, assim como em anos distintos, demonstrando relevância, importância e coesão com consistência temporal.

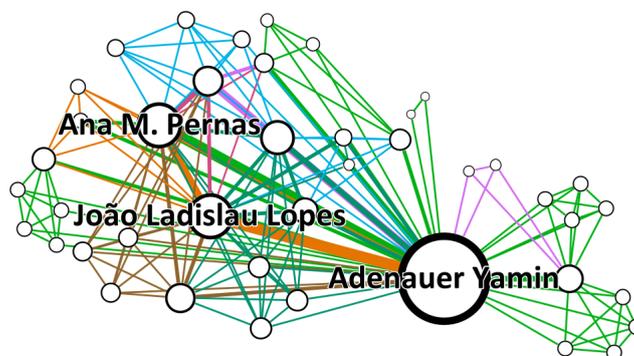


Figura 2. Componente gigante de 10 anos de co-autorias no SEMISH

A maioria das publicações do SEMISH apresenta três autores, 48 ($\approx 27\%$). Há uma variância entre um extremo com as autorias individuais ($\approx 2\%$), até 9 autores, em três publicações ($\approx 2\%$), maior quantidade observada. Comparado com o CSBC como um todo [Carvalho et al. 2023], o SEMISH é um evento enxuto de autorias em publicações, variando de um autor até nove; com uma característica inusitada, tem tantas publicações com autoria única, quanto com 8 ou 9 autores, 3 em cada uma destas categorias.

O SEMISH tem como característica a pouca concentração de autorias primárias. De 179 publicações analisadas, apenas um autor apresenta 3 autorias primárias, Rogério Albendes, da UFPel. Além dele, 6 autores apresentam duas autorias primárias ($\approx 3\%$), $\approx 97\%$ apresentam apenas uma autoria primária. Pode indicar mais interesse em pesquisas consolidadas ou finalizadas, menos em pesquisas em andamento ou particionadas.

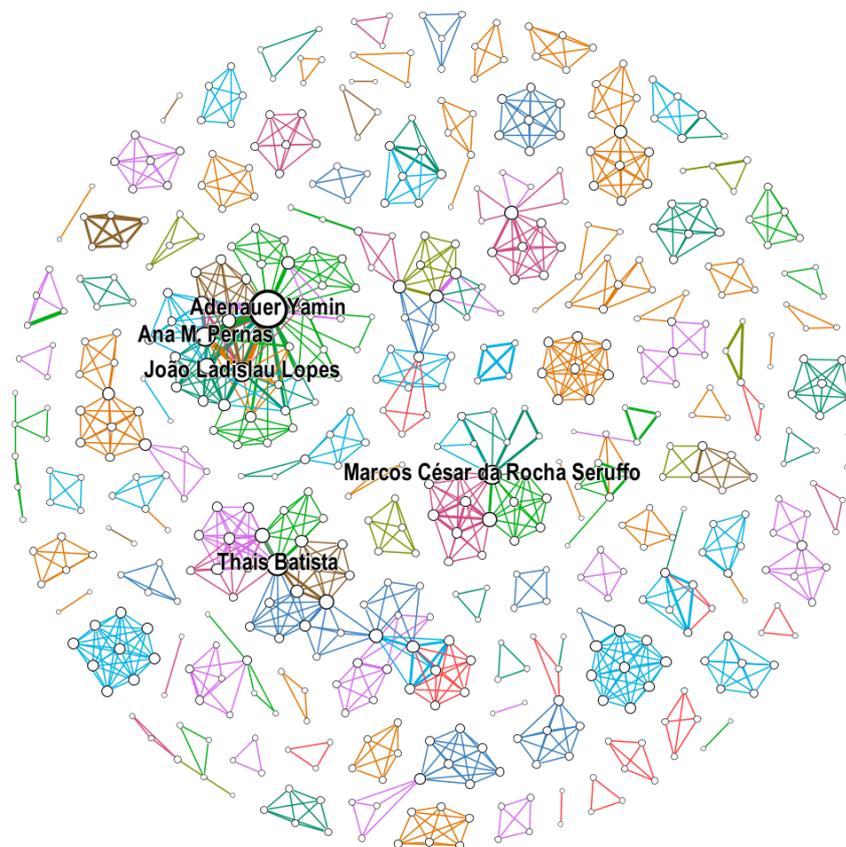


Figura 3. Grafo de colaboração e co-autorias em 10 anos de SEMISH, 2013 – 2022

Sobre a **análise de sexo**⁵ (seções c e d da Tabela 1), em todos os anos a quantidade de homens superou a quantidade de mulheres. Considerando a agenda de inclusão, incentivo e manutenção das mulheres na computação [Ribeiro et al. 2020], também no âmbito acadêmico-científico, este é um dado valioso. Em nenhum ano, sejam de autorias totais ou primárias, o quantitativo de mulheres alcançou um terço ($\approx 33\%$) dos homens. O pior resultado ocorre em 2020, no qual 88% de autorias primárias são de homens (21), apenas 8% de mulheres (2). E no total, o pior resultado ocorre em 2019, com 78% de homens (49) e 14% de mulheres (9).

Apesar de indicarmos a ausência de juízo moral, precisamos anunciar um complemento ético sobre este tópico. Reforçamos a agenda de inclusão, incentivo e manutenção de mulheres na computação, sendo nossa intenção central e única descrever e analisar a realidade já concretizada, material. Um aumento da presença de mulheres em edições futuras deve ser uma consequência primária do incentivo e promoção das mulheres no âmbito científico da computação, **consequencial**.

Em suma, pelos valores basilares da ética científica [London 2022,

⁵Estamos conscientes sobre a diferença entre sexo e gênero. Esta é uma limitação de abordagens envolvendo dados mecanicamente e estatisticamente rotulados devido ao volume como do *CSBCSet* i.e., sem que os autores envolvidos anunciem seu próprio sexo. Ademais, este é um dado significativo à investigação do envolvimento de mulheres na computação brasileira [Ribeiro et al. 2020, Santana e Braga 2020].

Marconi e Lakatos 2017], publicações não devem ser aceitas com base no sexo dos autores envolvidos, em primeira autoria ou colaboração, para “fazer um malabarismo” com estes dados. Devemos promover um aumento concreto da quantidade de mulheres agindo neste contexto, atualmente bastante deficitária, para potencialmente alavancar, através de uma abordagem moralmente positiva, o aumento deste quantitativo, como consequência e não princípio.

Apesar da maioria significativa e expressiva de homens, mulheres ocupam quantidades expressivas de produção e colaboração científica no SEMISH, como expostas na Figura 2 e Figura 3; os resultados de Ana Marilza Pernas e Thais Batista se sobressaem. Iniciativas de inclusão, incentivo e manutenção de mulheres na computação não devem estar restritas ou limitadas ao WIT, para que haja mais engajamento e presença de mulheres também no SEMISH.

Sobre a **análise de idiomas** (seção b da Tabela 1), português brasileiro é o idioma de maior ocorrência no SEMISH. O idioma das publicações, e seu desdobramento político, é objeto de discussão acadêmico-científica na computação brasileira [Carvalho et al. 2022, Mendonça 2021]. Não há um padrão constante entre a proporção de publicações em inglês ou português, indicando uma insignificância do idioma enquanto característica determinante de avaliação.

Representamos as relações da **análise de instituições/filiações** como um grafo, exposto na Figura 4. Os nós representam as instituições, e o tamanho o respectivo grau, i.e., relevância quantitativa; e as arestas representam as relações entre instituições, e a largura a relevância quantitativa dessa interação. Representa o grau de colaborações entre pessoas autoras.

Aqui notamos dois componentes principais, desconexos; e três componentes menores. Um dos componentes principais é concentrado na região Sul, pela UFPel, UCPel, IFSul e UFRGS; que reflete as informações na Figura 2. Além da significância quantitativa, que pode ser percebida pelo tamanho dos nós, as arestas indicam uma relação contínua, consistente e quantitativamente relevante.

O segundo componente tem uma representação regional dupla, entre UFRJ e USP (Sudeste) e UFPA (Norte). A UFPA apresenta mais colaborações se comparada com UFPel ou UCPel, entretanto as mesmas são menos relevantes, indicada pela largura da aresta. A UFPA colaborou com diversas instituições das quais os autores publicaram em menor quantidade, enquanto na UCPel e UFPel os autores publicaram em maior quantidade e recorrentemente. Os três componentes menores são isolados e conduzidos por universidades federais, UFMG, UFAM e UFPE. São fortemente influenciados pelo regionalismo.

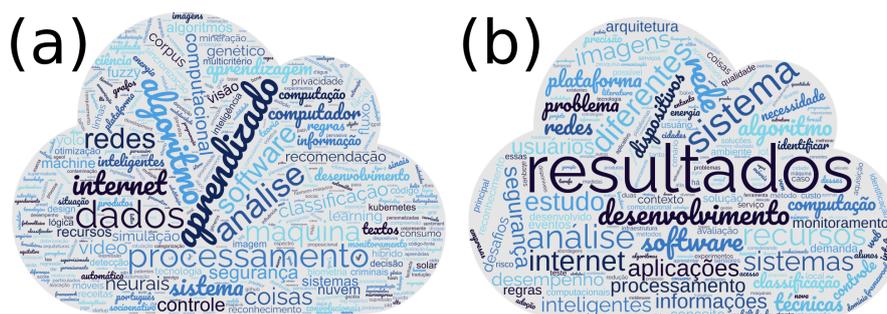
O primeiro componente é conduzido pelos autores nas Figuras 2 e 3, Adenauer Yamin, Ana Marilza Pernas e João Ladislau Lopes. O segundo, tanto por Thais Batista quanto Marcos César Seruffo. Há uma característica curiosa distinta entre as duas figuras combinadas com a Figura 4, mesmo que Thais esteja associada à UFRN e Marcos César Seruffo à UFPA e ambas apresentem colaboração, os dois autores nunca colaboraram entre si, o que podemos notar na Figura 3 com dois componentes significativamente expressivos, embora destacados. Ainda combinando estas figuras, a relevância de colaboração de Thais Batista é significativa, ao contrário da UFRN, indicando que as colaborações da UFRN são concentradas nela.

Tabela 2. Análises de instituições com 16 ou mais autorias

	Qtd.	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
UFPA	6			7	9			15	10	25	4	70
UFRJ	7	14	2	19	3		5		6		5	54
UFRN	6	3			3		7	6	18		6	43
UFPeI	7			7	5		4	3	4	2	10	35
UCPeI	7			2	3		3	1	3	2	13	27
UFOP	6	4			4	6		4	3		3	24
UTFPR	4		3		5	3	10					21
UFAM	4					5		3	6	6		20
UFRGS	8	5		2	2		1	1	4	2	2	19
UNICAMP	6		3	5	2				2	3	3	18
USP	4		4		6	4				3		17
UEL	3		2					12	3			17
UFPE	3			10					3	3		16
IFCE	4			8		2		3		3		16

Para contextualização semântica do conteúdo das publicações, **analisamos os resumos e palavras-chave**. A Figura 5 expõe a nuvem de palavras após tradução para português brasileiro e tratamento de termos vazios (e.g., artigos), insignificantes (e.g., verbos vagos) ou descartáveis (e.g., termos que isolados perdem significado). A maior parte das palavras-chave disponíveis ocorrem nos anos de 2020 a 2022, e este dado já nos faz perceber um fenômeno interessante, a ascensão de importância do aprendizado de máquina (*machine learning*) e seus termos correlatos, os termos de maior ocorrência estão associados com contextos desta técnica específica, e.g., algoritmo, aprendizado, análise, processamento, dados.

A análise de resumos ilustra uma realidade já esperada no contexto computacional brasileiro [da Costa Marques 2015], a total ausência de “hardware” (apesar do título do evento), enquanto software permanece um tópico recorrente. O SEMISH apresenta uma natureza generalista em suas pesquisas, com tópicos e temas de alto nível de abstração na computação, e.g., dispositivos, software, internet, sistema.



**Figura 5. (a) Nuvem de palavras das palavras-chave do SEMISH (disponíveis)
(b) Nuvem de palavras dos resumos do SEMISH entre 2013 – 2022**

3.2. Análise da Rede Social do SEMISH

A Tabela 3 apresenta 15 métricas de ARS para as coautorias no período entre 2013 e 2022. É possível perceber que o SEMISH apresenta um baixo diâmetro, variando entre 1 e 2, ao longo dos anos. A exceção é o ano de 2013, que tem a rede de coautoria com maior diâmetro, valor igual a 3. Esse resultado dá indícios que o SEMISH tem uma comunidade

Tabela 3. Métricas da rede de coautoria de uma década do SEMISH.

Métricas	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Diametro	3	1	2	2	1	1	2	2	2	2
DensidadeGrafo	0,081	0,078	0,051	0,052	0,121	0,133	0,079	0,041	0,028	0,0525
DensidadeCompGigante	0,439	1	0,6363	0,745	1	1	0,49	0,571	0,563	0,237
TamAbsolutoCompGigante	14	7	11	11	6	8	11	8	11	23
TamRelativoCompGigante	0,245	0,179	0,115	0,157	0,222	0,242	0,207	0,094	0,085	0,319
NumMaxCliquesCompGigante	3	1	2	2	1	1	3	2	2	6
TamMaxCliquesCompGigante	7	7	7	9	6	8	7	5	7	6
NumMaxCliquesGrafo	12	11	19	19	8	8	14	24	35	22
TamMaxCliquesGrafo	7	7	9	9	6	8	7	8	8	7
CoefAgrupamentoCompGigante	0,899	1	0,919	0,943	1	1	0,943	0,928	0,951	0,944
CoefAgrupamentoGrafo	0,975	0,95	0,963	0,926	0,904	0,823	0,898	0,891	0,867	0,861
NumNosCompGigante	14	7	11	11	6	8	11	8	11	23
NumNosGrafo	58	40	96	71	28	34	54	86	129	73
NumArestasCompGigante	40	21	35	41	15	28	27	16	31	60
NumArestasGrafo	134	61	235	131	46	75	114	151	236	138

com muitas publicações de autores que colaboram entre si. Além disso, observa-se que o ano de 2021 possui uma rede de coautoria com maior quantidade de nós (autores) e arestas (relacionamentos de coautoria) quando comparada aos outros anos, valores iguais a 129 e 236, respectivamente. O componente gigante da rede desse ano também possui um dos mais altos números de nós e arestas, valores iguais a 11 e 3, respectivamente. Entretanto, esse aumento no número de nós e arestas não contribuiu para reduzir o diâmetro da rede. Isso reforça a proximidade entre as coautorias no SEMISH.

Pela densidade do grafo percebemos que a rede é bastante esparsa ao longo dos anos, pois o valor da densidade não chega nem a 0,15. É possível a formação de muitas conexões na rede de coautoria. Entretanto, a densidade do componente gigante é bem mais alta, principalmente para os anos 2014, 2017 e 2018, pois o valor é igual a 1.

É importante ressaltar que o tamanho absoluto trata apenas da quantidade total de nós pertencente ao componente gigante. Já o tamanho relativo contabiliza a proporção do número de nós citados anteriormente para a quantidade total de nós do grafo completo. Para as métricas tamanho absoluto e relativo do componente gigante, é possível observar que o ano de 2022 se destaca com maior componente, um valor de 23 e 0,31, respectivamente. Não obstante, apesar de ter um maior tamanho do componente gigante, a rede de coautoria de 2022 não possui maior quantidade de cliques, ou seja, grupos de autores diretamente conectados. Para essa métrica, o ano de 2021 que se destaca, com 35 cliques.

Finalmente, o alto coeficiente de agrupamento das redes de coautoria do componente gigante e do grafo como um todo revela que os autores do SEMISH tendem a se agrupar e publicar em conjunto. O menor coeficiente de agrupamento é 0,82 para o ano de 2018, e o grafo desse ano possui menor quantidade de autores em relação aos outros anos. Ou seja, um aumento no número de autores e publicações não contribuiu para reduzir o coeficiente de agrupamento.

4. Considerações finais

Neste trabalho apresentamos análises quantitativas gerais e ARS sobre uma década do SEMISH, através de suas publicações, dados e metadados. O SEMISH é um dos eventos mais antigos e longevos da ciência computacional brasileira, completando 50 anos em 2023, e através de suas dinâmicas podemos montar parte do panorama história da computação brasileira, gerando informações e conhecimentos e contribuindo meta-cientificamente com o amadurecimento, memória e cultura do cenário. Desconhecemos

trabalhos similares ou fielmente próximos a este, em magnitude de abrangência, complexidade ou profundidade.

Além das contribuições analítico descritivas aqui apresentadas, algumas limitações e ameaças à validade são pertinentes [Wazlawick 2014, Marconi e Lakatos 2017]. Estamos limitados ao SEMISH e ao intervalo de tempo cobertos, 2013 – 2022, sem que isto restrinja sua complementação, o que é bem-vindo. Devido à limitação de espaço, alguns dos objetos resultantes de análises gráficas foram omitidos e podem ser acessados e apreciados com melhor qualidade no repositório complementar, online (indicado na Seção 2). Este trabalho se limita às análises e métricas realizadas, sendo que outras constituem potenciais trabalhos futuros, quantitativas ou qualitativas, por dimensões ou perspectivas variadas, e.g., outras abordagens ou métricas de ARS.

5. Agradecimentos

Este trabalho foi realizado com apoio da CAPES – Brasil – Código de Financiamento 001; e parcialmente pelo programa *Oracle for Research* (número do prêmio CPQ-2160239).

Referências

- [Agassi 2008] Agassi, J. (2008). *Science and its History: A Reassessment of the Historiography of Science*. Boston Studies in the Philosophy and History of Science Nº253. Springer, 1st edition.
- [ANPEd 2019] ANPEd (2019). *Ética e pesquisa em educação: subsídios – volume 1*. volume 1. ANPEd, Rio de Janeiro, RJ.
- [Barbosa et al. 2017] Barbosa, S., Silveira, M., e Gasparini, I. (2017). What publications metadata tell us about the evolution of a scientific community: the case of the brazilian human–computer interaction conference series. *Scientometrics*, 110(14):275–300.
- [Bos 2020] Bos, J. (2020). *Research Ethics for Students in the Social Sciences*. Springer Cham, 1st edition.
- [Brasil 2016] Brasil (2016). Ministério da saúde. RESOLUÇÃO Nº 510, DE 07 DE ABRIL DE 2016.
- [Carvalho et al. 2023] Carvalho, L. P., Lima Filho, S., Brandão, M. A., Oliveira, J., Santoro, F. M., e da Silva, M. F. (2023). Uma década de interações entre eventos e pesquisadores do CSBC: Um estudo meta-científico. In *Anais do XII BraSNAM*, Porto Alegre, RS, Brasil.
- [Carvalho et al. 2022] Carvalho, L. P., Suzano, J. A., Santoro, F. M., e Oliveira, J. (2022). A meta-scientific broad panorama of ethical aspects in the Brazilian IHC. *Journal on Interactive Systems*, 13(1):105–126.
- [da Costa Marques 2015] da Costa Marques, I. (2015). THE COMPUTER MARKET RESERVE IN BRAZIL: DEMOCRACY, AUTHORITARIANISM, RUPTURES, AND CONTINUITIES. *IEEE Annals of the History of Computing*, 37:1–1.
- [Digiampietri et al. 2017] Digiampietri, L., Mugnaini, R., Pérez-Alcázar, J., Delgado, K., Tuesta, E., e Mena-Chalco, J. (2017). Análise da evolução, impacto e formação de redes nos cinco anos do brasnam. In *Anais do VI BraSNAM*, Porto Alegre, RS, Brasil. SBC.

- [Ioannidis et al. 2015] Ioannidis, J., Fanelli, D., Dunne, D., e Goodman, S. (2015). Meta-research: Evaluation and improvement of research methods and practices. *PLoS biology*, 13(10).
- [Ioannidis 2018] Ioannidis, J. P. A. (2018). Meta-research: Why research on research matters. *PLoS Biol*, 16(3).
- [Kauer e Moreira 2013] Kauer, A. U. e Moreira, V. P. (2013). Evolução dos temas de interesse do sbbd ao longo dos anos. In *SBBB (Short Papers)*, pp. 25–1.
- [Latour 1987] Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard university press.
- [Lefèvre 2005] Lefèvre, W. (2005). Science as Labor. *Perspectives on Science*, 13(2):194–225.
- [Lima Filho et al. 2023] Lima Filho, S., Carvalho, L. P., Suzano, J., Brandão, M., Oliveira, J., e Santoro, F. M. (2023). CSBCSet: Um conjunto de dados para uma década de CSBC, seus eventos e publicações. In *Anais do XII BraSNAM*, Porto Alegre, RS, Brasil.
- [Lobato et al. 2021] Lobato, F., Sousa, G., e Jr., A. J. (2021). Brasnam em perspectiva: uma análise da sua trajetória até os 10 anos de existência. In *Anais do X BraSNAM*, pp. 217–228, Porto Alegre, RS, Brasil. SBC.
- [London 2022] London, A. J. (2022). *For the Common Good: Philosophical Foundations of Research Ethics*. Oxford University Press, 1st edition.
- [Marconi e Lakatos 2017] Marconi, M. d. A. e Lakatos, E. M. (2017). *Fundamentos de Metodologia Científica*. Atlas, São Paulo, SP, 8ª edition.
- [Mendonça 2021] Mendonça, N. (2021). Abertura e Internacionalização do SBES: Um Oxímoro Irreconciliável? In *Anais do I Workshop de Práticas de Ciência Aberta para Engenharia de Software*, pp. 49–51, Porto Alegre, RS, Brasil. SBC.
- [Ribeiro et al. 2020] Ribeiro, K. S. F. M., Maciel, C., e S. Bim, M. A. (2020). Gênero e tecnologias. In Maciel, C. e Viterbo, J., editors, *Computação e sociedade: a profissão - volume 1*, pp. 104–140. EdUFMT Digital, Mato Grosso, Brasil.
- [Santana e Braga 2020] Santana, T. e Braga, A. (2020). Uma análise cienciométrica das publicações do congresso da sociedade brasileira de computação na perspectiva das mulheres na computação. In *Anais do XIV WIT*, pp. 279–283, Porto Alegre, RS, Brasil. SBC.
- [Tabassum et al. 2018] Tabassum, S., Pereira, F. S. F., Fernandes, S., e Gama, J. (2018). Social network analysis: An overview. *WIREs Data Mining and Knowledge Discovery*, 8(5):e1256.
- [Wazlawick 2014] Wazlawick, R. S. (2014). *Metodologia de Pesquisa para Ciência da Computação*. Elsevier, São Paulo, SP, 2ª edition.
- [Wazlawick e Silva Junior 2021] Wazlawick, R. S. e Silva Junior, D. P. (2021). Histórico de Eventos da SBC no Brasil. *Em SBC Horizontes*. <http://horizontes.sbc.org.br/?p=5493> [acesso 04-04-2023].