

BreSci na última década: Uma análise baseada em redes sociais

Silas Lima Filho¹, Luiz Paulo Carvalho¹, 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

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

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

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

Abstract. *BreSci is the scientific event dedicated to e-Science, which refers to computational science that supports scientific experimentation. In addition to discussing topics such as workflow modeling, ontologies, artificial intelligence, and machine learning, BreSci, which has been taking place since 2007, also supports the discussion of data science topics that can be transferred to the improvement of the sciences. A social network analysis of the last decade reveals an expanding scientific community across the editions, with a majority of male authors and a community characterized by rich institutional connections, concentrated in the Southeast region.*

Resumo. *O BreSci é o evento científico brasileiro dedicado à e-Ciência, ou seja, da ciência computacional que apoia a experimentação científica. Além de discutir tópicos como modelagem de workflows, ontologias, inteligência artificial e aprendizado de máquinas, o BreSci, que ocorre desde 2007, também apoia a discussão do tópicos em ciência de dados que possam ser transferidas para a melhoria das ciências. Em uma análise de redes sociais da última década se observa uma comunidade científica que se expande ao longo das edições, uma maioria masculina nas autorias e uma comunidade com rica conexão institucional, concentrada na região Sudeste.*

1. Introdução

“O Workshop Brasileiro de e-Ciência, conhecido como BreSci, é um fórum para discussão e disseminação de ideias sobre os temas envolvidos no desenvolvimento de infraestrutura de software em apoio às ciências como uma nova plataforma de pesquisa e experimentação científica.”¹. Assim se apresenta o BreSci, em sua décima sétima edição em 2023 e realizado desde 2007. Conta com dinâmicas e palestras com pesquisadores nacionais e internacionais renomados; despontando como um evento brasileiro focado em ciência de apoio computacional à experimentação científica, chamada de e-Ciência (do inglês e-Science).

¹<https://sbbd.org.br/2023/chamada-bresci/> [acesso 16-06-2023]

Este trabalho apresenta uma pesquisa meta-científica Ioannidis:2015, Ioannidis:2018 sobre o BreSci. O auto-exame sobre a própria rede ou comunidade é valiosa para entender, inclusive, os aspectos que contornam e compõem as pesquisas aplicadas nos temas específicos. O objetivo é **apresentar uma análise descritiva baseada em Análise de Redes Sociais (ARS) e estatística sobre os dados e os metadados da última década do BreSci, através de suas publicações, envolvendo dados diversos disponíveis, possíveis e válidos**. A 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 BreSci, como contribuição meta-científica descritiva do cenário de educação e computação brasileiro.

Trabalhos similares analisam comunidades e espaços acadêmico-científicos brasileiros meta-cientificamente [Ioannidis 2018], como [Lobato et al. 2021] e [Digiampietri et al. 2017] sobre o Workshop *Brazilian Social Network Analysis and Mining* (BraSNAM). 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]. Ao máximo do nosso conhecimento, este é um trabalho inédito com esta abordagem sobre desempenho computacional brasileiro, mais especificamente o BreSci.

A Seção 2 expõe o método, materiais e informações essenciais; a Seção 3 apresenta os resultados; e a Seção 4 encaminha as considerações finais.

2. Método, recursos, fontes de dados e conceitos

Foram coletados dados para análise presentes no *CSBCSet*², um conjunto de dados representando publicações de 2013 a 2022 do Congresso da Sociedade Brasileira de Computação (CSBC) [Lima Filho et al. 2023]. Partimos dos atributos: Ano, Evento, Título, Pessoa autora, Sexo, Instituição 1, UF Inst. 1, Idioma, filtrando pelo evento “BreSci”.

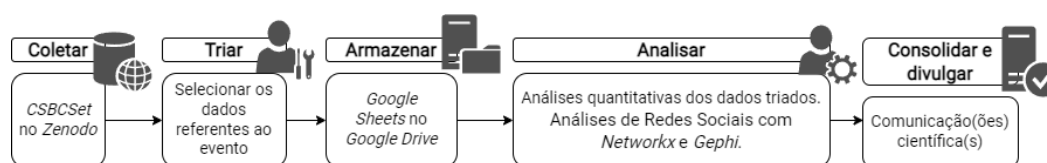


Figura 1. Etapas do método aplicado neste trabalho.

O método está ilustrado pela Figura 1, indicando as tecnologias e sistemas computacionais utilizados³. Foram realizadas análises quantitativas e estatísticas gerais e conduzidos métodos e abordagens em ARS [Tabassum et al. 2018]. Em ambos, o BreSci foi tratado como um todo e cada ano separadamente. Pelo excesso de informações geradas, e para melhor visualização, os artefatos gerados foram disponibilizados online.

Quanto a aspectos éticos, após uma deliberação sobre ponderações éticas ou morais, decidiu-se seguir o trabalho mantendo os autores identificados, ausente justificativa significativa o bastante para o anonimato, seja moral [Brasil 2016] ou ética

²<https://4658.short.gy/CSBC2023> [acesso 16-06-2023]

³<https://networkx.org/> - <https://gephi.org/> - <https://en.wikipedia.org/wiki/Spreadsheet> [acesso 16-06-2023]

[ANPEd 2019]. Tratam-se de dados abertos, dados de promoção da atividade-fim dos pesquisadores, estão ausentes juízos morais ou de valor moral, e sendo uma pesquisa secundária, é comum expor identidades profissionais dos envolvidos. Ademais, interessados podem utilizar destes dados ou informações em benefício, *e.g.*, em memoriais de promoção de carreira. Isto é, aqui são expostos **fatos acadêmico-científicos**.

3. Resultados e discussão

Esta seção expõe as análises combinando os dados e informações sobre o BreSci presentes no *CSBCSet*, de suas publicações entre 2013 – 2022, excluindo comunicações como abertura, prefácios, mensagens da coordenação, dentre outros. Na **análise de autorias e pessoas autoras/autores**, a Tabela 1 expõe os resultados das publicações e suas autorias. Considerando uma margem de aceites de 10% de todas as publicações (≈ 15), há uma estabilidade na variância de 10% deste valor (≈ 2 , entre 13 e 17), sem ultrapassar 20%, como em 2021 ou 2017. Entretanto, em 2013 este valor é extrapolado, sendo um valor atípico.

Em relação à quantidade de autorias por publicação e ano, 374 ($\approx 80\%$) apresentam apenas uma autoria (*i.e.*, publicaram apenas um trabalho, seja como autor individual ou em colaboração); 53 ($\approx 12\%$), duas autorias; 16 ($\approx 3,5\%$), três autorias, o somatório das demais é de $\approx 4,5\%$. A maior quantidade de autorias, em 18 publicações, é de Daniel de Oliveira; Kary Ocaña, em 11; Eduardo Ogasawara, em 10; Fábio Porto, em 8; Luciano Digiampietri e Marta Mattoso, em 7; e Sérgio Manuel Serra da Cruz, em 5. Sendo assim, podemos considerar estes os autores mais recorrentes no BreSci.

O grafo na Figura 2 expõe o componente gigante acumulado de 2013 – 2022 no BreSci. 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 [Tabassum et al. 2018]. Nem todos os autores mais prolíficos compõem o componente gigante. Dos 7 autores listados acima, 2 apenas não são conectados ao componente gigante. No entanto, os 5 demais autores dos mais prolíficos, tanto produzem mais, como colaboram mais e de maneira coesa e colaborativa. Ao longo de 10 anos, o evento contou com um total de 154 artigos. Na maioria das publicações, 38 ($\approx 25\%$), figuram 3 autores. 2 publicações ($\approx 1,3\%$) apresentam 12 autores, maior quantidade de autorias no evento, seguida de 2 publicações ($\approx 1,3\%$) com 10 autores. 3 publicações ($\approx 2\%$) apresentam autoria única.

A Tabela 1 indica que há uma maior acumulação distribuída de autorias por publicações se comparado com outros eventos similares [Carvalho et al. 2023]. Em 4 dos 10 anos há uma proporção geral de 5 autorias por publicação, refletindo no quantitativo total com mais publicações apresentando cinco autores (23 – $\approx 15\%$) do que quatro autores (21 – $\approx 13,5\%$). A quantidade de autores é um tópico relevante na epistemologia científica [Shaffer 2014], faz parte da cultura de certos espaços e comunidades científicas.

Com 140 autorias primárias no total, Luciano Pacifico, Kary Ocaña e Luciano Digiampietri apresentam 3 autorias primárias em publicações; 8 autores ($\approx 5\%$) com 2 autorias primárias; 129 autores ($\approx 92\%$) com apenas uma, sem reincidência. Sendo assim, nota-se uma baixa ocorrência de reincidências de autorias primárias.

A Figura 3 expõe o grafo completo com as colaborações e co-autorias de todo BreSci, além do componente gigante. A configuração do grafo concentra na parte mais

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

	(a) Publicações			(d) Idioma das publicações				(b) Autores (todos) por sexo **				(c) Autores (primários) por sexo **			
	Total (autores)	Total (primários)	Pub./autor	pt-br	Qtd. %	en	Qtd. %	M	Qtd. %	F	Qtd. %	M	Qtd. %	F	Qtd. %
2013	102	24	4	18	75%	6	25%	74	73%	22	22%	17	71%	6	25%
2014	74	16	5	9	56%	7	44%	47	64%	26	35%	9	56%	6	38%
2015	65	14	5	11	79%	3	21%	49	75%	13	20%	9	64%	3	21%
2016	52	12	4	8	67%	4	33%	39	75%	11	21%	10	83%	2	17%
2017	45	11	4	7	64%	4	36%	29	64%	16	36%	9	82%	2	18%
2018	69	15	5	7	47%	8	53%	42	61%	27	39%	10	67%	5	33%
2019	39	12	3	8	67%	4	33%	28	72%	10	26%	9	75%	3	25%
2020	63	17	4	16	94%	1	6%	43	68%	20	32%	14	82%	3	18%
2021	88	19	5	12	63%	7	37%	60	68%	26	30%	13	68%	6	32%
2022	53	14	4	7	50%	7	50%	44	83%	8	15%	13	93%	0	0%
Total	650	154	4	103	67%	51	33%	455	70%	179	28%	113	73%	36	23%

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

central o componente gigante da Figura 2. Os 2 autores mais prolíficos do evento, e que não pertenciam ao componente gigante, agora se fazem presente no grafo completo. No entanto, dadas as suas conexões, tais autores permanecem na periferia do grafo. O grafo da Figura 3 também apresenta autores não citados na lista dos que possuem maior número de artigos, mas que conectam-se à algum nó do componente gigante. Nesse grafo completo, de modo a facilitar a leitura, exibimos apenas os rótulos dos nós que possuem grau de conexão com o valor mínimo 14.

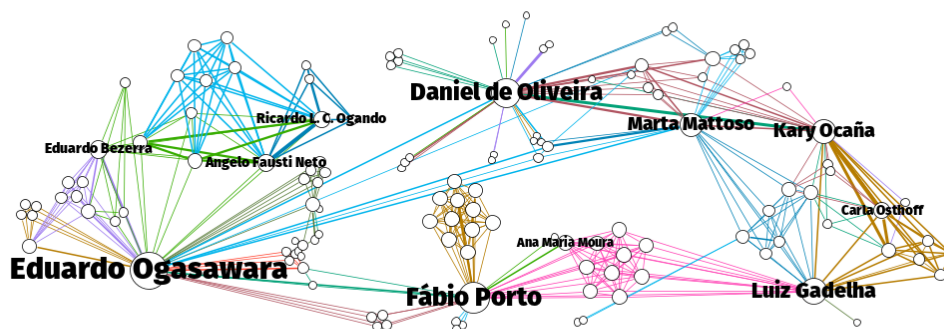


Figura 2. Componente gigante da colaboração de 10 anos do BreSci., 2013 – 2022

Sobre a **análise de sexo**, a quantidade de homens supera a de mulheres em todos os anos. 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. Das 10 autorias mais prolíficas, 9 são homens; sendo que na décima posição a distribuição de mulheres é por empate. Em 2021 a quantidade de autorias primárias de homens e mulheres empata (6 e 6), sendo um fenômeno significativamente atípico em relação aos demais. Sobre a **análise de idiomas**, português brasileiro é o idioma dominante do BreSci, mesmo que sua imanência técnica abstrai regionalismos ou âncoras a fenômenos locais, percebe-se uma predominância do português, mesmo que acidental e não proposital. Em nenhum ano a quantidade de publicações em inglês supera em português.

As relações da **análise de instituições/filiações** estão representadas 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. A Tabela 2 complementa o grafo com

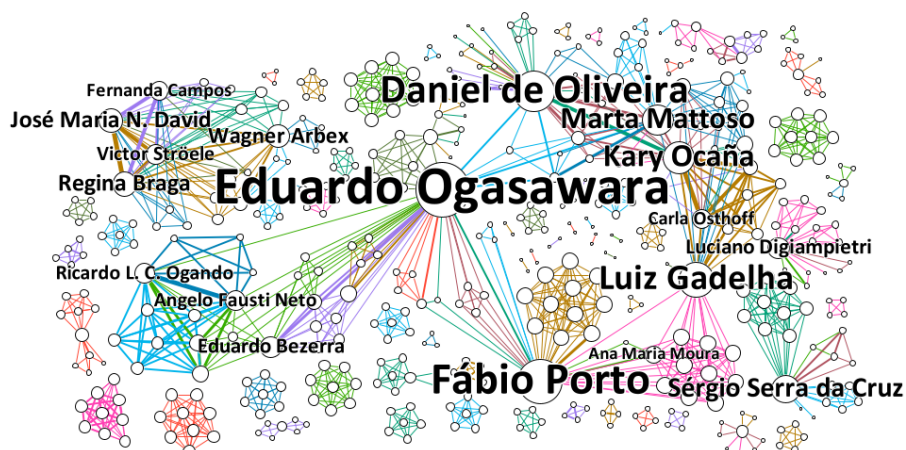


Figura 3. Grafo completo acumulativo dos autores durante os anos 2013-2022.

o quantitativo de autorias por instituição.

O LNCC apresenta maior relevância de colaborações e publicações ao longo dos anos. A quantidade de publicações do LNCC (83) é bem superior ao segundo instituto listado, CEFET-RJ (58), que se aproxima dos valores do terceiro ranqueado, UFF (52). Das instituições listadas na Tabela 2, com a exceção da UFPB, que é penúltima no ranking da quantidade de publicações, todas as instituições listadas fazem parte do grafo de colaborações de instituições. Das colaborações apresentadas no grafo, ressaltam as colaborações recorrentes (pela largura das arestas) entre LNCC, CEFET-RJ, UFF e UFRJ. O BreSci apresenta instituições bastante conectadas, onde a maioria e as mais relevantes estão no componente gigante.

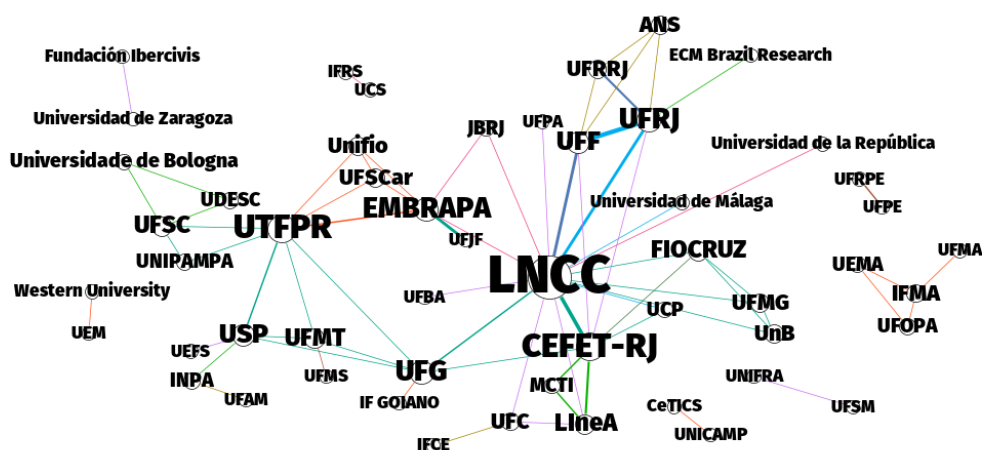


Figura 4. Grafo de colaboração entre instituições dos autores

Pela análise de geolocalização institucional, há uma concentração na região Sudeste no Rio de Janeiro (RJ) ($\approx 44\%$), São Paulo (SP) ($\approx 13\%$), e Minas Gerais (MG) ($\approx 9\%$) acumulando juntos $\approx 66\%$ (428) das autorias. Paraná (PR) tem menos da metade de ocorrências ($\approx 4\%$) se comparado com o terceiro maior, MG. A UTFPR é, inclusive,

Tabela 2. Análises de instituições com publicações com 31 ou mais autorias

	Qtd. anos	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	Total
LNCC	9	6	18	3	7	12	7	5	9	16		83
CEFET-RJ	9	1		9	2	4	8	9	8	12	5	58
UFF	8	6		8	6	9	9	5	6	3		52
USP	9	2	16	3	2	3	10		2	6	4	48
UFJF	6	2			10		7	5	5	6		35
UFRJ	6	9	2	6	3	5	2					27
UFRRJ	4	10		4		4	8					26
LIneA	3	19		4	2							25
UFPB	1		22									22
UFMG	3	6		10						4		20

uma das instituições com maior grau de colaboração, vide Figura 4, e neste aspecto ultrapassa instituições de SP e MG. Observando a Tabela 2, das 10 instituições com maior quantidade de autorias, 9 se encontram na região Sudeste, e 6 especificamente no RJ.

3.1. Análise da Rede Social do BreSci

A Tabela 3 apresenta 15 métricas de ARS para as coautorias no período entre 2013 e 2022. Como os resultados apontam, a rede social da comunidade do BreSci possui diâmetro de certa forma estável, variando ao longo dos anos de 1 a 3. Sendo os anos de 2019, 2020 e 2021 caracterizados por uma diminuição dos valores de algumas métricas. Por exemplo, em 2019 houve a menor quantidade de nós (quantidade de autores) no grafo total, em 2020 a menor quantidade de nós pertencentes ao componente gigante. Houve também em 2019 um queda significativa na quantidade de colaborações entre autores (número de arestas), mas com uma crescente nos dois seguintes anos.

Uma importante análise se dá em comparar a evolução do componente gigante ao longo dos anos. A Figura 5 mostra a evolução dos principais autores e suas conexões, sob o ponto de vista do maior subgrafo de cada ano, bem como certos autores foram criando novas comunidades científicas ao longo das edições. Comparando com a Figura 3, percebe-se a presença de autores como Eduardo Ogasawara, Marta Mattoso e Daniel de Oliveira no componente de 2013. Em 2015, Ogasawara já possui sua comunidade com nós até então ausentes. O mesmo acontece com Daniel de Oliveira no componente de 2018, onde novos autores surgem, e sem a presença de Marta Mattoso. Em 2021 e 2022 novas comunidades surgem como do LNCC em torno de Fábio Porto, e Priscila Saito da UTFPR.

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

Métricas	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Diametro	2	2	2	2	3	3	2	1	1	2
DensidadeGrafo	0,0516	0,0894	0,0918	0,0980	0,1096	0,0774	0,1136	0,0713	0,0623	0,0860
DensidadeCompGigante	0,5273	0,3718	0,6364	0,6889	0,3524	0,3636	0,6667	1,0000	1,0000	0,5111
TamAbsolutoCompGigante	11	13	11	10	15	12	9	8	12	10
TamRelativoCompGigante	0,1325	0,2131	0,1833	0,2273	0,4167	0,1875	0,2813	0,1509	0,1500	0,2128
NumMaxCliquesCompGigante	3	4	2	2	5	4	2	1	1	3
TamMaxCliquesCompGigante	6	6	7	8	6	5	7	8	12	6
NumMaxCliquesGrafo	22	16	14	11	11	15	9	14	18	14
TamMaxCliquesGrafo	9	12	10	8	6	10	7	8	12	9
CoefAgrupamentoCompGigante	0,9202	0,8660	0,9192	0,9611	0,8950	0,8167	0,9524	1,0000	1,0000	0,9056
CoefAgrupamentoGrafo	0,8682	0,9098	0,8980	0,9136	0,9394	0,8973	0,8052	0,8654	0,9012	0,8081
NumNosCompGigante	11	13	11	10	15	12	9	8	12	10
NumNosGrafo	84	62	61	45	37	65	33	54	81	48
NumArestasCompGigante	29	29	35	31	37	24	24	28	66	23
NumArestasGrafo	180	169	168	97	73	161	60	102	202	97

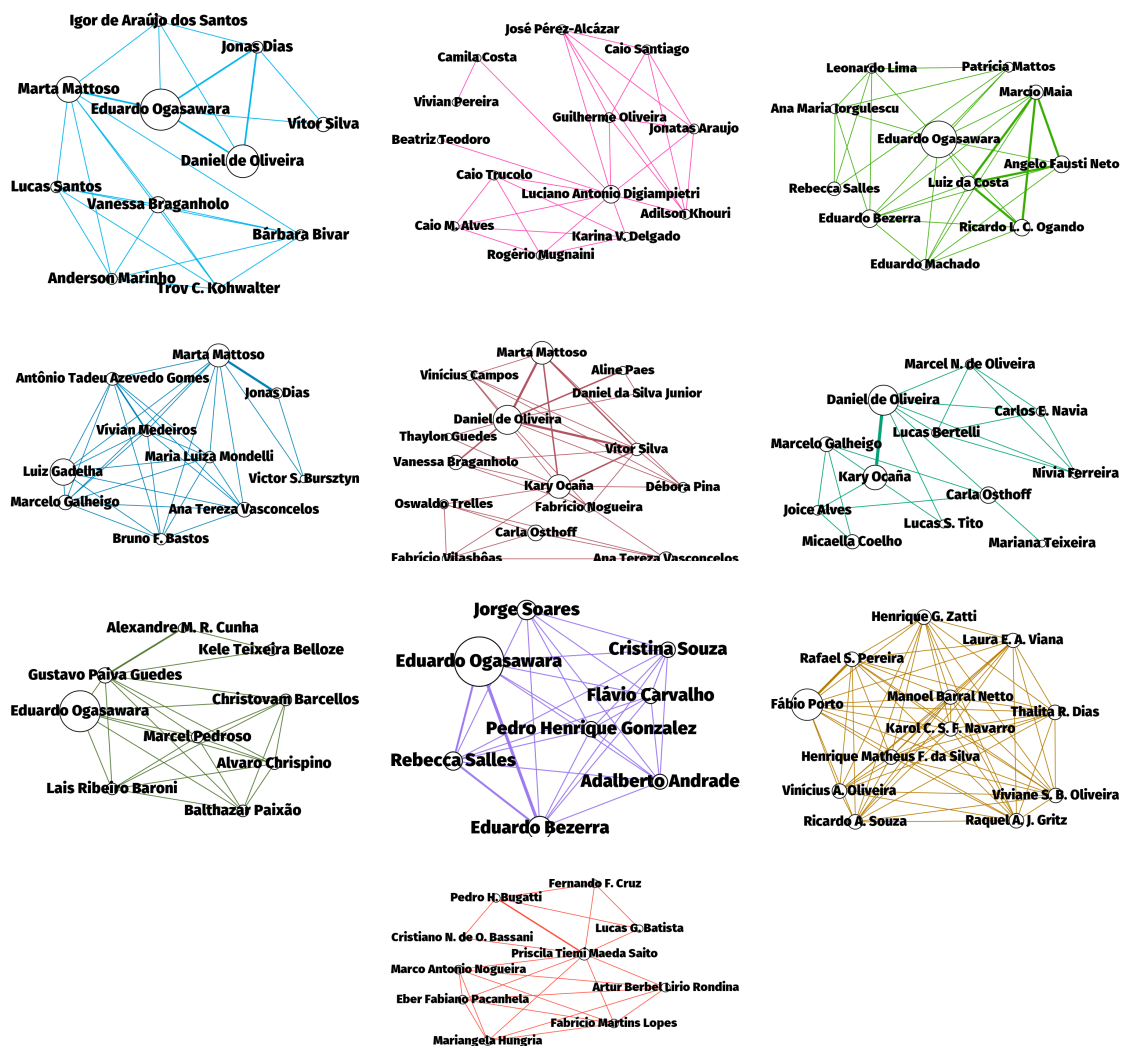


Figura 5. Evolução do Componente Gigante ao longo das edições (2013-2022).

4. Considerações finais

Neste trabalho apresentamos análises quantitativas gerais e ARS sobre uma década do BreSci, através de suas publicações, dados e metadados. Analisar o BreSci meta-cientificamente compõe a pesquisa de desempenho computacional e comunicacional brasileira. Pelas suas dinâmicas montamos parte do quebra-cabeça através dos dados, gerando informações e conhecimentos e contribuindo meta-cientificamente com o amadurecimento, memória e cultura do cenário [Ioannidis et al. 2015, Ioannidis 2018]. Há ausência de trabalhos similares ou fielmente próximos a este, em magnitude de abrangência, completude ou profundidade.

Além das contribuições analítico descritivas aqui apresentadas, algumas limitações e ameaças à validade são pertinentes [Wazlawick 2014, Marconi e Lakatos 2017]. Este trabalho está limitado ao BreSci, enquanto espaço de desempenho computacional e comunicacional no Brasil; e intervalo de tempo cobertos, 2013 – 2022, sem que isto

restringa 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.

Referências

- ANPEd (2019). *Ética e pesquisa em educação: subsídios – volume 1*. volume 1. ANPEd, Rio de Janeiro, RJ.
- Brasil (2016). Ministério da saúde. RESOLUÇÃO Nº 510, DE 07 DE ABRIL DE 2016.
- 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.
- 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, J., Fanelli, D., Dunne, D., e Goodman, S. (2015). Meta-research: Evaluation and improvement of research methods and practices. *PLoS biology*, 13(10).
- Ioannidis, J. P. A. (2018). Meta-research: Why research on research matters. *PLoS Biol*, 16(3).
- Lima Filho, S., Carvalho, L. P., Suzano, J., Brandão, M. A., Oliveira, J., e Santoro, F. M. (2023). Csbcsset: Um conjunto de dados para uma década de csbc, seus eventos e publicações. In *Anais do XII Brazilian Workshop on Social Network Analysis and Mining*, pp. 240–245, Porto Alegre, RS, Brasil. SBC.
- 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.
- Marconi, M. d. A. e Lakatos, E. M. (2017). *Fundamentos de Metodologia Científica*. Atlas, São Paulo, SP, 8ª edition.
- 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.
- Shaffer, E. (2014). Too many authors spoil the credit. *Canadian journal of gastroenterology hepatology*, 28:605.
- 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, R. S. (2014). *Metodologia de Pesquisa para Ciência da Computação*. Elsevier, São Paulo, SP, 2ª edition.