

Uma Análise Cienciométrica das Publicações do Congresso da Sociedade Brasileira de Computação na Perspectiva das Mulheres na Computação

Thalia S. Santana¹, Adriano H. Braga¹

¹Instituto Federal Goiano – Campus Ceres (IF GOIANO)
Caixa Postal 51 – 76.300-000 – Ceres – GO – Brasil

thaliassantana15@gmail.com, adriano.braga@ifgoiano.edu.br

Abstract. *The present work carries out a study about the female presence in the scientific productions of the events of the Congress of the Brazilian Computer Society (CSBC), that is considered one of the major concerns of relevance in the country. The data were obtained through Web Scraping techniques, between the years 2017 to 2019, through the creation of a database for investigations regarding the panorama of women in computing, making use of scientometric analyzes in the characterization of gender diversity in computer science research (CSR).*

Resumo. *O presente trabalho realiza um estudo acerca da presença feminina nas produções científicas dos eventos do Congresso da Sociedade Brasileira de Computação (CSBC), o qual é considerado uma das conferências de maior relevância em computação no país. Os dados foram obtidos por meio de técnicas de Web Scraping, entre os anos de 2017 a 2019, mediante a criação de uma base de dados para investigações quanto ao panorama das mulheres na computação, fazendo uso de análises cienciométricas na caracterização da diversidade de gênero na pesquisa em ciência da computação (CSR – em inglês, computer science research).*

1. Introdução

Análises sobre diversidade e desequilíbrio de gênero na pesquisa e a consequente participação de mulheres na popularização científica, são uma importante questão para a equidade e entendimento sobre a sub-representação de minorias em Ciência da Computação (CC) (AGARWAL; MITTAL; SUREKA, 2017). Levando em conta as produções na área, somente cerca de 21% dos pesquisadores que são ativos em conferências de renome são mulheres, considerando o indexador internacional DBLP *Computer Science Bibliography* (AGARWAL et al., 2016).

Avaliando-se todas as áreas de produção científica, as mulheres brasileiras são responsáveis por 72% dos artigos publicados. Contudo, mesmo assinando a maior parte destes, o quantitativo total de mulheres cientistas (49%) é menor do que do público masculino (ALBORNOZ et al., 2018). Em computação, os estudos de Arruda et al. (2009) demonstraram que no Brasil o percentual de pesquisadoras totaliza apenas 23%, evidenciando o cenário da baixa proporção feminina na comunidade científica em CC.

Neste contexto, a SBC é a maior sociedade de computação de toda a América do Sul (CSBC, 2019) e reflete-se como a principal realizadora e apoiadora da execução de

eventos técnico-científicos brasileiros na área. A exemplo disso, um de seus feitos mais reconhecidos trata do Congresso da Sociedade Brasileira de Computação (CSBC). Com uma grade de eventos fixos (base) e demais eventos esporádicos (satélites), é realizado desde o ano de 1980, oportunizando troca de experiências no âmbito científico, acadêmico e profissional, com prestígio nacional e internacional (SBC, 2019).

Sendo assim, objetivou-se diagnosticar produções científicas publicadas na área de CC, em específico nos anais do CSBC, a fim de realizar uma análise quantitativa por meio dos dados catalogados sob a perspectiva de gênero em pesquisa em ciência da computação (CSR – em inglês, *computer science research*). Além da introdução, o estudo subdivide-se da seguinte maneira: a Seção 2 apresenta a metodologia empregada para coleta e análise dos dados; na Seção 3 são descritos os resultados e discussões encontrados; e por fim, as Considerações Finais na Seção 4.

2. Material e Métodos

Foram obtidas as publicações presentes nas páginas eletrônicas do CSBC como material para análises cienciométricas. Considerou-se artigos publicados entre 2017 e 2019 por meio de procedimentos de raspagem de dados, conhecidos como *Web Scraping*. A partir de informações disponíveis nos endereços eletrônicos das três últimas edições do CSBC, realizou-se a raspagem e coleta de dados. Contudo, não foi possível obter dados anteriores, a exemplo da edição de 2016 no mesmo formato que os anos mais atuais. Foi utilizada a linguagem de programação Python, versão 2.7.16, no mês de agosto de 2019, fazendo uso de um *script* com auxílio das bibliotecas *Pandas*, *BeautifulSoup* e *URLOpen*, gerando um arquivo no formato CSV (*Comma-separated-values* – em português, valores separados por vírgula), com o evento e sua edição, título e autoras/es.

Quanto a seleção dos eventos, foram eliminados aqueles que não possuíam submissão de trabalhos como o Seminário de Computação (SECOMU), bem como aqueles que produziram itens diferentes de artigos como resultado, a exemplo da Jornada de Atualização de Informática (JAI), a qual publica um livro ao final da conferência. Além disso, foi realizada a limpeza dos dados coletados, eliminando registros que contiveram prefácios de abertura, limitando-se somente às publicações em específico e seus autores/as. Em busca de uma análise ao longo do tempo, manteve-se somente os congressos em comum entre os anos analisados. Ao todo, 14 eventos foram elencados (Tabela 1), com 759 trabalhos publicados.

Cabe destacar que nenhum dos eventos supracitados fornecem informações quanto ao gênero dos componentes da autoria dos artigos. A SBC possui um modelo próprio de submissão de artigos, sendo utilizado por toda a conferência, onde os nomes das/os autoras/es são descritos de forma literal e completa, sem abreviações pelo último sobrenome e inicial da letra do primeiro nome, o que contribuiu para identificar o gênero das/os autoras/es. Para isso, foi necessário o uso da biblioteca de classificação automática *genderBR*, na versão 3.4.4, para identificação do gênero dos pesquisados. Mesmo após, cerca de 4,4% dos indivíduos não foram identificados e assim, também passaram por uma classificação manual, com busca das/os pesquisadoras/es *online*.

Como última etapa de construção da base, foi criado um novo campo para armazenamento de nomes de autores corrigidos, sendo este processo denominado como desambiguação, como descrito por Duarte, Moura e Moro (2019). Isso foi feito para

agrupamento de alguns registros de autores que referiam-se ao mesmo indivíduo (com ou sem sobrenome do meio), mantendo-se como regra o nome em sua forma mais completa, etapa também realizada de modo manual.

Por fim, os dados foram organizados de acordo com a modelagem relacional de um Diagrama Entidade-Relacionamento (DER), em um banco de dados MySQL, versão 8.0.16, por meio do Sistema Gerenciador de Banco de Dados (SGBD) MySQL Workbench, versão 8.0.15. A análise ocorreu por intermédio de consultas em linguagem SQL (*Structured Query Language* – em português, Linguagem de Consulta Estruturada). Posteriormente, utilizou-se da linguagem Python 3.7.3 via plataforma Google *Colaboratory*, juntamente das bibliotecas *Numpy*, *Scipy* e *Matplotlib*, onde foi possível gerar visualizações e análises estatísticas.

Tabela 1. Eventos do CSBC selecionados por sigla, nome e tipo. Fonte: Própria (2019).

Sigla	Nome do evento	Tipo
BraSNAM	<i>Brazilian Workshop on Social Network Analysis and Mining</i>	Satélite
BreSci	<i>Brazilian e-Science Workshop</i>	Satélite
CTD	Concurso de Teses e Dissertações da SBC	Base
CTIC	Concurso de Trabalhos de Iniciação Científica da SBC	Base
ENCompIF	Encontro Nacional de Computação dos Institutos Federais	Satélite
ETC	Encontro de Teoria da Computação	Satélite
SBCUP	Simpósio Brasileiro de Computação Ubíqua e Pervasiva	Satélite
SEMISH	Seminário Integrado de <i>Software</i> e <i>Hardware</i>	Base
WCAMA	Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais	Satélite
WEI	Workshop sobre Educação em Computação	Base
WIT	<i>Women in Information Technology</i>	Base
WPerformance	Workshop em Desempenho de Sistemas Computacionais e de Comunicação	Satélite
WPIETF	Workshop Pré-IETF	Satélite
WTrans	Workshop de Transparência em Sistemas	Satélite

3. Resultados e Discussão

Considerando os eventos base e satélite do CSBC, do total de registros da análise por autor, 1.878 pesquisadores são do gênero masculino (69%), em contraponto à 845 autoras do gênero feminino, que compõem um percentual de cerca de 31%. Os três autores com maior número de publicações são todos do gênero masculino, sendo o primeiro com 12 artigos, e o segundo e terceiro do *ranking* empatados com 11 trabalhos nos três anos de análise temporal da base do CSBC.

Avaliando a quantidade de autores por gênero juntamente da ordem de autoria, 220 mulheres são primeira autora enquanto 539 homens se encontram como autor principal. A situação de desproporcionalidade se repete em todos os níveis de coautoria verificados nas publicações, visto que em qualquer uma das ordens, o público

masculino é maior que o feminino, mesmo não havendo regras de limite de autores por artigo. Quanto aos eventos, o WEI é o que possui maior número de trabalhos (122 publicações) considerando os anos analisados. Em seguida, vem o ETC, com 94 produções e por último, o WIT, com 87 trabalhos, o qual é o mais novo evento base dentre os desta categoria e com elevado número de trabalhos.

Ainda no âmbito de autoras/es principais, em 2017, 2018 e 2019, respectivamente, publicaram 58, 89 e 73 mulheres, em relação à 172, 198 e 169 homens, o que corresponde a 25,2%, 31% e 30,2%. Quanto à presença feminina nos artigos, 449 publicações analisadas contêm mulheres dentre as/os autoras/es, representando cerca de 59,1%. Assim, mesmo em menor número, as cientistas estão contidas em mais da metade dos trabalhos que são publicados no CSBC. Já em relação ao público masculino, estão presentes em 699 artigos. Ademais, também foi verificado o percentual de autoras em cada um dos 14 eventos durante os três anos, conforme apresentado na Figura 1.

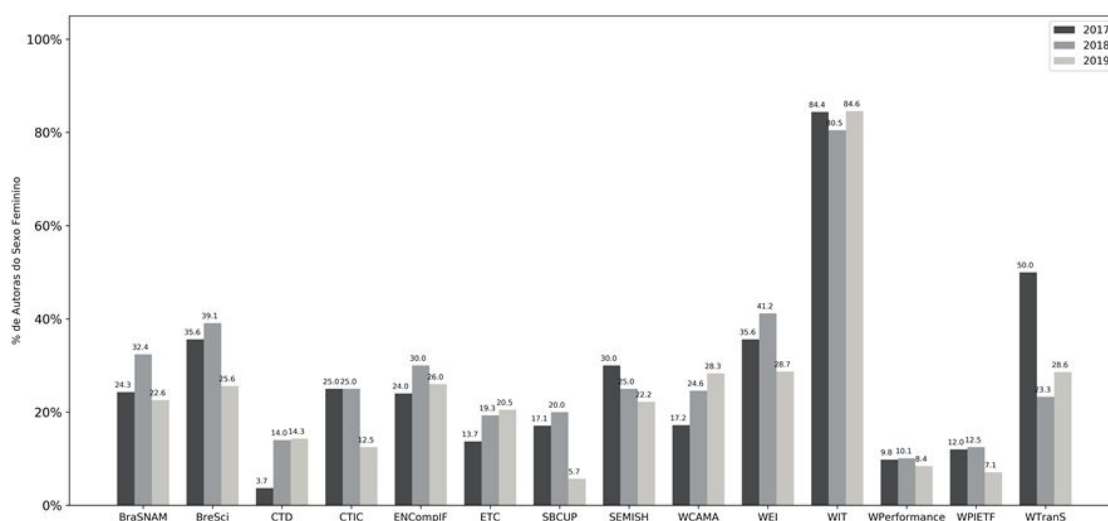


Figura 1. Percentual de autoras por ano e evento. Fonte: Própria (2019).

Foi possível perceber que o WIT dentre todos os eventos promovidos no CSBC, é o que possui maior número de mulheres como autoras principais, com 80 mulheres e 7 homens. Sendo o evento justamente destinado a temática de mulheres na computação, é esperado o elevado índice de pesquisadoras. Entretanto, eventos como o CTD, onde premia-se teses e dissertações da computação consideradas como de relevância, somente uma mulher foi encontrada como primeira autora – reflexos do baixo percentual de mestres e doutoras em CC (ALMEIDA; ZANLORENSSI, 2018).

4. Considerações Finais

É notório que mesmo em menor número, as mulheres são parte representativa da maioria das produções catalogadas (presentes em quase 60% das publicações), demonstrando o proeminente papel feminino na difusão e efetivação de saberes científicos. Desta maneira, apesar de na literatura possuírem pesquisas que avaliem especificamente alguns dos eventos presentes no CSBC sob a ótica cienciométrica, não foi encontrado nenhum estudo que investigue em específico todos os eventos desta conferência, ainda mais na perspectiva de grupos minoritários no contexto de CSR.

Além disso, destaca-se o próprio WIT como evento de relevância para aumento no número de autoras e conseqüentemente, participantes mulheres no CSBC.

Por meio deste estudo, acredita-se que haja contribuição com o corpo de pesquisas brasileiras com recorte de gênero na computação, já que os dados disponíveis frequentemente são dependentes de pesquisas isoladas e insuficientes para visualizações abrangentes (RIBEIRO et al., 2019). Outrossim, espera-se que os dados analisados possam vir a fornecer subsídios para a SBC e de encontro quanto ao fomento de atividades em prol de equidade de gênero, bem como manutenção e destaque de ações apoiadas como o já chancelado Programa Meninas Digitais (MENINAS DIGITAIS, 2019). Para trabalhos futuros, é importante que haja a criação de uma política de privacidade que permita exportar os dados das conferências ligadas à SBC, a fim de possibilitar a disponibilização pública da base de dados própria, e assim, seja possível aumentar o período de tempo da análise a fim de verificar efetivamente a evolução da participação feminina no CSBC por evento.

5. Referências

- Agarwal, S., Mittal, N., & Sureka, A. (2017). Minority ethnic groups in computer science research: what is the bibliography data telling us?. *ACM SIGCAS Computers and Society*, 47(2), 5-15.
- Agarwal, S., Mittal, N., Katyal, R., Sureka, A., & Correa, D. (2016). Women in computer science research: What is the bibliography data telling us?. *ACM SIGCAS Computers and Society*, 46(1), 7-19.
- Albornoz, M., Barrere, R., Matas, L., Osorio, L., & Sokil, J. (2018). Las brechas de género en la producción científica Iberoamericana. *Papeles Obs*, 9, 30.
- Almeida, R., Zanlorenssi, G. (2018). Qual o gênero e a idade de mestres e doutores no Brasil. <https://www.nexojornal.com.br/grafico/2018/05/23/Qualo-genero-e-a-idade-de-mestres-e-doutores-no-Brasil>, Março de 2019.
- Arruda, D., Bezerra, F., Neris, V. A., De Toro, P. R., & Wainera, J. (2009). Brazilian computer science research: Gender and regional distributions. *Scientometrics*, 79(3), 651-665.
- CSBC (2019). Inscrições. <http://csbc2019.sbc.org.br/inscricoes/>, Outubro de 2019.
- Duarte, B., Moura, A., & Moro, M. (2019). Mulheres na Computação: Análises por Sub-Áreas. In *Anais do XIII Women in Information Technology*, (pp. 174-178). Porto Alegre: SBC. doi:10.5753/wit.2019.6732.
- Meninas Digitais (2019). Sobre o Programa Meninas Digitais. <http://meninas.sbc.org.br/index.php/sobre/>, Novembro de 2019.
- Ribeiro, K., Azevedo, J., Maciel, C., & Bim, S. (2019). Uma análise de gênero a partir de dados da Sociedade Brasileira de Computação. In *Anais do XIII Women in Information Technology*, (pp. 159-163). Porto Alegre: SBC. doi:10.5753/wit.2019.6729.
- SBC (2019). CSBC - Congresso da Sociedade Brasileira de Computação, <https://www.sbc.org.br/eventos/csbc>, Outubro de 2019.