

# Participação Feminina em Comitês de Programa da Computação: Análises de Eventos Nacionais e Internacionais

Breno C. S. Pimenta, Mirella M. Moro  
*Departamento de Ciência da Computação*  
*Universidade Federal de Minas Gerais*  
Belo Horizonte, Brasil  
brenopimenta@ufmg.br, mirella@dcc.ufmg.br

**Resumo**—Esta pesquisa analisa a participação feminina nos comitês de programas dos principais eventos internacionais na área da Computação, comparando-a com dados de eventos nacionais. Observou-se que, embora ainda haja sub-representação feminina em ambos os cenários, os eventos brasileiros apresentam índices ligeiramente mais altos, com 25,09% de mulheres contra 20,05% nos internacionais. Conferências voltadas à interação humano-computador, educação em computação e sistemas colaborativos tendem a apresentar maiores índices de representatividade feminina, enquanto subáreas mais técnicas, como arquitetura de computadores, circuitos integrados e segurança, exibem taxas mais baixas.

**Palavras-chave**—Participação feminina, comitês de programas, diversidade de gênero, análise de dados.

## I. INTRODUÇÃO

A igualdade de gênero desempenha um papel fundamental na sociedade, incluindo na área de Ciência da Computação e Tecnologia da Informação. A inclusão de pessoas de perfis diversos é essencial para garantir múltiplas perspectivas e contínua inovação. Contudo, apesar de muitos avanços, a representatividade feminina e de outras minorias nas áreas de computação ainda enfrenta desafios.

A Sociedade Brasileira de Computação (SBC) é uma entidade científica sem fins lucrativos que desempenha um papel crucial na promoção da pesquisa em computação no Brasil. Através de publicações, conferências e congressos, a SBC reúne pessoas da academia e da indústria que atuam na área para promover a disseminação do conhecimento científico [1]. Em âmbito internacional, diversas organizações desempenham papel semelhante na promoção do avanço da computação.

Dentre essas organizações, destacam-se a *Association for Computing Machinery* (ACM), o *Institute of Electrical and Electronics Engineers* (IEEE), a *International Joint Conference on Artificial Intelligence* (IJCAI), a *European Design Automation Conference* (EDAC) e a *Association for Information Systems* (AIS). A ACM é reconhecida por sua ampla gama de publicações científicas e pela realização de conferências de grande impacto global [2]. O IEEE, por sua vez, se destaca como uma das maiores organizações técnicas do mundo, promovendo o desenvolvimento da ciência, engenharia elétrica e computacional [3]. A IJCAI é uma referência na área de

inteligência artificial, promovendo um dos congressos mais prestigiados do setor. A EDAC tem papel central na área de automação de design eletrônico, enquanto a AIS se dedica ao estudo de sistemas de informação e sua influência nas organizações e na sociedade.

Essas instituições foram selecionadas para análise devido à relevância de seus eventos no cenário da computação. Seus Comitês de Programa (CPs) são essenciais para garantir a qualidade dos trabalhos apresentados, ajudando a definir as principais tendências e direcionamentos da pesquisa em computação. O estudo da composição desses comitês pode fornecer *insights* muito relevantes sobre a diversidade e inclusão no meio acadêmico e profissional da computação.

### A. Objetivos

Este estudo visa continuar a análise da participação feminina nos CPs, agora focando nos simpósios promovidos por sociedades científicas internacionais. A representatividade de gênero não se limita apenas à igualdade numérica, mas também à voz e influência que as pessoas têm em decisões acadêmicas e científicas. A participação nos CPs não é apenas um indicador da diversidade de participação, mas também de diversidade na influência na comunidade científica, visto que a participação em comitês influencia a aceitação de artigos científicos, bem como a orientação de pesquisas futuras.

Assim, este projeto busca identificar desafios e oportunidades para promover uma maior inclusão de gênero na área de computação. Através da continuação de estudos e da monografia anterior, serão analisados dados quantitativos e qualitativos para avaliar a participação feminina em eventos a partir de 2011. As questões de pesquisa que norteiam o estudo são as seguintes: Como é a presença feminina por evento e como ela evolui ao longo do tempo? Qual é a participação feminina nos núcleos de tais eventos? E como os eventos nacionais se comparam aos internacionais?

Esta pesquisa visa contribuir para o debate acadêmico sobre diversidade de gênero na computação, e também fornecer *insights* que podem informar políticas e práticas futuras na SBC e em outras instituições acadêmicas e científicas. Os resultados das análises têm um potencial significativo de

utilização. Por exemplo, ao identificar as áreas com maior atuação feminina nos CPs, pode-se analisar que aspectos atraem mulheres para essas áreas e enfatizá-las em ações de atração de talentos. Além disso, outras iniciativas brasileiras e internacionais podem também se beneficiar dos resultados aqui apresentados, promovendo uma maior inclusão de gênero na área de computação.

### B. Estrutura do Texto

Este trabalho está organizado em cinco seções, além das referências bibliográficas. Nesta seção, é apresentada a introdução ao tema, incluindo a caracterização do problema, a motivação para o estudo e os objetivos da pesquisa. A Seção II é dedicada ao referencial teórico, no qual são discutidos os principais conceitos e revisada a literatura correlata, abordando estudos sobre representatividade de gênero na computação e metodologias e resultados já alcançados, além de um breve resumo das conclusões obtidas pela Monografia em Sistemas de Informação 1 (MSI 1).

A Seção III detalha as atividades desempenhadas, incluindo a coleta e tratamento de dados, a definição dos critérios de análise, as ferramentas estatísticas utilizadas e o processo de validação dos dados. Os resultados obtidos e sua discussão são apresentados na Seção IV, na qual é analisada a participação feminina nos CPs dos simpósios da SBC e examinando a evolução temporal dessa participação. Finalmente, a Seção V apresenta as conclusões do estudo, destacando as principais contribuições para a área e sugerindo caminhos para pesquisas futuras e ações práticas que possam promover uma maior diversidade nos comitês científicos.

## II. REFERENCIAL TEÓRICO

A seguir, serão apresentados os principais conceitos e estudos já existentes que estão relacionados à temática.

### A. Participação feminina e participação masculina

O foco da pesquisa é analisar a participação feminina nos Comitês de Programa dos simpósios promovidos pela Sociedade Brasileira de Computação (SBC). É importante esclarecer que essa análise compara as representações de gênero masculino e feminino em diversos momentos. No entanto, essa dicotomia não abrange todas as identidades de gênero.

É amplamente aceito que o gênero é uma construção social e cultural, em oposição a uma característica inata ou biologicamente determinada [4]. As sociedades atribuem significados e expectativas específicas a diferentes categorias de gênero, e essas construções sociais exercem uma influência profunda na vida das pessoas. Além disso, a compreensão tradicional do sexo como estritamente binário (masculino e feminino) é inadequada, considerando as variações biológicas e anatômicas que não se encaixam facilmente nessa estrutura binária [5].

No entanto, é importante observar que essas questões estão fora do escopo de análise deste trabalho. O estudo se concentra na categorização por nome para criar um perfil da comunidade com base em gênero. Isso é feito com o objetivo de evitar

ambiguidades nos termos utilizados, e optou-se por usar as expressões “participação feminina” ou “participação masculina” para fins de análise.

### B. Representação Feminina na Computação

A igualdade de gênero é um objetivo global, mas os dados mais recentes do Objetivo de Desenvolvimento Sustentável (ODS) número cinco indicam que ainda não estamos progredindo o suficiente para alcançá-lo até 2030. A área da Ciência da Computação no Brasil reflete essa desigualdade, com uma notável sub-representação das mulheres. Estudos revelam que as mulheres representavam apenas 23% das pessoas pesquisadoras nesse campo em [6].

De acordo com relatório de 2024 da UNESCO, a proporção de mulheres entre os graduados em STEM (Ciência, Tecnologia, Engenharia e Matemática) é de 35% e não mudou tanto nos últimos 10 anos. Mulheres são sub-representadas no processo de design e implantação tecnológica. Em 2022, as mulheres ocupavam menos de 25% dos empregos em ciência, engenharia e tecnologia da informação e comunicação. Hoje, elas representam apenas 26% dos funcionários em dados e inteligência artificial [7].

A cultura da Ciência da Computação, em muitos aspectos, é percebida como excludente para as mulheres. Estereótipos de gênero, falta de modelos femininos e um ambiente muitas vezes hostil podem desencorajar as mulheres a ingressar nesse campo. É fundamental compreender que a sub-representação das mulheres na Ciência da Computação é uma questão cultural e social, não uma diferença inata de habilidades ou interesses. Percepções de que a computação é um domínio masculino contribuem para afastar as mulheres [8].

Segundo Misa [9], a trajetória da participação feminina pode ser dividida em três fases: (1) um período inicial de baixa representação, mas com algumas pioneiras, (2) um crescimento acelerado entre os anos 1970 e 1980, e (3) um declínio após os anos 1980 devido a mudanças culturais e estruturais, como a popularização dos computadores pessoais como produtos direcionados ao público masculino. Estudos como o de Wang et al. [10] demonstram que, embora a participação feminina na autoria de artigos científicos de computação tenha crescido, esse crescimento desacelerou recentemente, indicando uma possível estabilização abaixo da paridade de gênero.

Além disso, conforme apontado por Tives et al. [11], a inclusão de mulheres enfrenta muitos obstáculos. Conforme destacado pelo Relatório do Fórum Econômico Mundial, a pandemia de COVID-19 exacerbou as disparidades de gênero, projetando que a igualdade econômica entre homens e mulheres só será alcançada em 135 anos. Diante desse cenário, torna-se ainda mais urgente e necessário promover iniciativas que aumentem a participação feminina no campo da Computação.

### C. Participação Feminina na Computação Acadêmica

Um dos problemas enfrentados pelas mulheres na área de computação é a sub-representação na pesquisa científica. Segundo Dal Pizzol et al. [12], apesar do aumento do número

de publicações femininas nas últimas duas décadas, as autoras ainda são minoria na produção de artigos de computação. Esse fenômeno não se deve a uma menor qualidade ou quantidade dos trabalhos das mulheres, mas sim à baixa participação delas na área. No Brasil, a SBC reportou que apenas 21,67% de seus associados eram do gênero feminino em novembro de 2018 [13].

Um fator que pode influenciar a participação das mulheres na pesquisa em computação é a colaboração internacional. Aksnes et al. [14] analisaram as diferenças de gênero na colaboração internacional em diferentes campos de pesquisa, incluindo tecnologia e ciências sociais. Eles constataram que os homens tendem a ter uma taxa de colaboração internacional mais alta do que as mulheres, especialmente entre as pessoas com poucas publicações. No entanto, entre as pessoas com mais publicações científicas, as mulheres superam os homens na colaboração internacional.

No âmbito nacional, há diferenças significativas no número de trabalhos científicos publicados por homens e mulheres. Em um estudo que utilizou dados da plataforma Lattes, Soares et al. [15] analisaram os currículos de pessoas pesquisadoras entre os anos de 1966 e 2015. Seu objetivo consistiu em avaliar a produção e participação das mulheres na área de Ciências Exatas e da Terra. Os resultados indicaram que 86% dos currículos compreenderam homens e somente 14% foram mulheres.

Laberge et al. [16] identificaram que áreas de maior prestígio acadêmico, como Teoria da Computação, apresentam menor participação feminina (cerca de 13,1%), enquanto subáreas mais aplicadas, como Interação Humano-Computador (HCI), possuem cerca de 20% de pesquisadoras. Além disso, o estudo mostrou que áreas com mais mulheres tendem a estar associadas a instituições de menor prestígio, o que pode impactar o avanço de carreira e a visibilidade das pesquisadoras na área.

Esses resultados indicam que há uma desigualdade de gênero na área de tecnologia, com os homens ocupando uma posição superior em todas as categorias de produtividade. Além disso, sugerem que a colaboração internacional pode ser uma estratégia para aumentar a visibilidade e o reconhecimento das mulheres na pesquisa em computação.

#### D. Gênero na Sociedade Brasileira de Computação

A diversidade de gênero nos comitês de programa de eventos de computação é um tema que tem recebido atenção da comunidade acadêmica. Lorens et al. [17] realizaram um estudo sobre a presença de autoras mulheres na área de computação, considerando eventos nacionais e internacionais. Elas observaram que as áreas interdisciplinares, aplicadas ou com um aspecto social tendem a ter uma maior participação feminina do que as áreas mais teóricas ou fundamentais. Essa tendência se reflete também nos comitês de programa, que são formados em grande parte pelas pessoas autoras dos trabalhos aceitos nos eventos.

No entanto, o percentual de mulheres em comitês de programa ainda é baixo, tanto em eventos nacionais quanto

internacionais, indicando uma sub-representação de mulheres na ciência da computação global. O estudo propõe algumas ações para atrair mais mulheres para a área, como destacar as características das áreas com maior participação feminina, desenvolver materiais e atividades que inspirem meninas a seguir carreiras na computação e promover iniciativas de inclusão e retenção de mulheres na área.

Outro estudo que aborda a questão da diversidade de gênero na computação é o de Soares et al. [15], que analisou os eventos promovidos pela SBC entre 2018 e 2022. Elas constataram que o evento com maior participação feminina é o *Women in Information Technology* (WIT), uma iniciativa para discutir os assuntos relacionados às questões de gênero em Tecnologia da Informação. Em termos de quantidade de artigos publicados, o WIT possui destaque na quantidade, perdendo apenas para o Workshop sobre Educação em Computação (WEI) nos anos de 2018 e 2021. As autoras argumentam que a possibilidade de ver outras mulheres em uma área majoritariamente masculina pode trazer motivação e acolhimento para outras que pensam em seguir na área.

Esses estudos mostram que a diversidade de gênero na ciência da computação é um desafio que precisa ser enfrentado por meio de políticas e práticas que valorizem e incentivem a participação feminina. Além disso, eles evidenciam a importância de considerar as especificidades das diferentes áreas da computação, bem como os fatores culturais e sociais que influenciam as escolhas profissionais das mulheres.

#### E. Métricas para Análise de Comitês de Programa

Os artigos principais de Lorens et al. [17] e Duarte et al. [18] explicaram que, dado a variação no tamanho dos comitês e o fato de muitos indivíduos participarem de várias edições dos eventos a análise precisou ser estruturada em três aspectos principais: número de mulheres em termos de pessoas, participação feminina ao longo do tempo e definição do núcleo dos eventos.

1) *Número de Pessoas e Participação*: O número de pessoas desconsidera quantas vezes a mesma pessoa participou em diferentes edições do comitê. Por exemplo, no comitê do evento  $X$ , a mulher A participou das edições  $\{x_1, x_2\}$ , o homem B participou das edições  $\{x_1, x_2, x_3, x_4\}$ , e a mulher C participou da edição  $\{x_3\}$ . Isso significa que o número de mulheres neste evento foi 2, representando  $\frac{2}{3} \times 100 \approx 66.7\%$  do total de pessoas. Para o mesmo exemplo, ao considerar as participações, nota-se que a mulher A participou de 2 edições, o homem B de 4 edições, e a mulher C de 1 edição. Ou seja, a participação feminina foi de 3, enquanto a masculina foi de 4. Em termos percentuais, a participação feminina foi de aproximadamente  $\frac{3}{7} \times 100 \approx 42.9\%$ .

2) *Definição do Núcleo*: Considera-se que uma pessoa faz parte do núcleo de um evento se sua participação nos comitês estiver entre as 25% maiores, ou seja, no terceiro quartil das participações daquele evento. Por exemplo, em um comitê do evento  $Y$  com cinco membros A, B, C, D, E e F e somatória de suas participações respectivas de  $\{1, 3, 5, 7, 9, 11\}$ , o valor do terceiro quartil é 8.50. Assim, os membros E e F, com

participações acima de 8.50, são considerados integrantes do núcleo do evento Y.

#### F. Classificação de Gênero a partir do Nome

Durante a pesquisa, os dados coletados não continham o atributo “gênero” ou “sexo” das pessoas. Por isso, foi necessário inferir o gênero a partir do nome da pessoa. Por isso, realizou-se uma investigação mais aprofundada sobre as abordagens utilizadas em trabalhos anteriores para lidar com classificações similares a fim de auxiliar na melhoria da tarefa de classificar os nomes.

1) *Desafios da classificação*: A tarefa de inferir o gênero a partir de nomes apresenta várias dificuldades. Primeiramente, a ambiguidade dos nomes é um grande desafio. De acordo com Karimi et. al. [19], nomes unissex, que podem ser atribuídos a ambos os gêneros, dificultam a classificação. Exemplos incluem nomes como Alex e Taylor. Além disso, a variação cultural adiciona outra camada de complexidade. Um nome que é tipicamente masculino em uma cultura pode ser feminino em outra, complicando a inferência de gênero em um contexto de dados de pessoas de muitos países. Por exemplo, Michele no Brasil é geralmente feminino, enquanto na Itália é frequentemente masculino. Assim sendo, inferir o gênero usando o nome completo pode ser mais eficaz, usando o exemplo anterior, se o nome Michele está seguido de sobrenomes italianos, é mais provável que seja masculino do que se estiver seguido de sobrenomes tipicamente brasileiros.

Métodos baseados em bancos de dados de nomes podem apresentar um viés significativo em relação a determinadas regiões ou culturas. Nomes comuns em países ocidentais podem ser bem classificados, enquanto nomes de países emergentes podem não ser, e como exemplo o estudo cita países como Brasil e China, os quais representam uma parcela significativa dos dados coletados nesta pesquisa. Além disso, alguns métodos podem ter uma precisão desigual entre os gêneros, afetando negativamente a representatividade de gênero nas análises [19].

A precisão dos métodos de inferência de gênero é outra questão crítica. Métodos que utilizam apenas nomes têm uma precisão limitada e podem não lidar bem com nomes raros ou pouco comuns, como apontado por Burger et al. [20]. A combinação de métodos baseados em nomes com reconhecimento de imagens pode melhorar a precisão, mas também requer acesso a dados de imagem, o que pode não ser viável ou ético, além disso essa abordagem reforça estereótipos físicos e estéticos. Por fim, serviços de inferência de gênero precisam ser constantemente atualizados para lidar com novas tendências de nomes e mudanças culturais.

2) *Serviços de Classificação Automática*: O trabalho base conduzido por Lorens et al. [17] utilizou uma abordagem de classificação automática com a biblioteca de processamento de linguagem natural Natural Language Toolkit (NLTK). Para casos sem resultado, utilizou-se outra biblioteca específica para identificar gênero, conhecida como *Gender Guesser*. Os resultados de ambas foram verificados manualmente por amostragem. Após essa verificação, cerca de 30% das pessoas

ainda estavam com gênero não identificado, como no caso de nomes ambíguos como Ariel. Tais nomes foram, então, manualmente classificados por meio da busca das respectivas pessoas online.

Outro trabalho [21] comparou e avaliou cinco serviços de inferência de gênero: Gender API, genderize.io, NameAPI, NamSor e o pacote Python gender-guesser. A análise revelou que os serviços Gender API e NamSor geralmente apresentam melhor desempenho, embora sejam serviços pagos a precisão variou bastante dependendo da origem geográfica dos nomes. O artigo também destaca que, embora o *gender-guesser* tenha uma coleção de nomes presumidamente de alta qualidade, o banco de dados não foi atualizado desde sua publicação em 2008. Mesmo assim, o *gender-guesser* é mencionado por fornecer atribuições de gênero de alta qualidade, principalmente se ignoradas as classificações não encontradas. Em termos de desempenho, o *gender-guesser* apresentou baixa taxa de erros, indicando que é eficaz em minimizar o número de classificações erradas quando comparado com outros serviços avaliados.

Wang et al. [10] utilizaram a *Gender API*, associando uma probabilidade a cada classificação. No entanto, esse método apresenta limitações, especialmente para nomes ambíguos e de origem asiática, que frequentemente são classificados com menor precisão. Laberge et al. [16] complementaram essa abordagem com pesquisas diretas para validar a precisão dos algoritmos de classificação, identificando uma taxa de acerto de 97%.

#### G. Conclusões obtidas em MSI 1

O trabalho desenvolvido em MSI 1 possibilitou a análise da participação feminina nos CPs de diversos eventos nacionais na área de Computação. A análise revelou que, apesar de avanços em alguns eventos, a porcentagem total de mulheres nos CPs foi de 25,09%, valor similar à participação feminina anual média de 26,28% entre 2011 e 2023.

No núcleo dos eventos, onde estão as 25% maiores participações, as mulheres representam 25,94% do total de participantes, o que indica uma média relativamente baixa. Observou-se que eventos como o IHC (Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais), o SBIE (Simpósio Brasileiro de Informática na Educação), o SBSC (Simpósio Brasileiro de Sistemas Colaborativos) e o SBSI (Simpósio Brasileiro de Sistemas de Informação) apresentam os maiores índices de participação feminina. Esses eventos estão majoritariamente ligados a áreas que envolvem interação humano-computador, educação e colaboração.

Por outro lado, eventos como o SBSeg (Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais), o SIBGRAPI (Conference on Graphics, Patterns and Images), o SBESC (Simpósio Brasileiro de Engenharia de Sistemas Computacionais) e o SBCCI (Symposium on Integrated Circuits and Systems Design) registram os menores índices de participação feminina. Essas conferências estão concentradas em áreas como segurança da informação, visão computacional, engenharia de software e design de circuitos integrados. Esse

padrão sugere que a temática dos eventos e os fatores culturais e estruturais das subáreas da Computação podem ter relação com a presença feminina.

A análise da evolução da participação feminina ao longo dos anos, utilizando mapas de calor e regressões lineares, revelou padrões distintos entre os eventos. Alguns congressos, como o IHC, o SBSC e o GEOINFO (Simpósio Brasileiro de Geoinformática), demonstram um crescimento gradual na participação feminina. Por outro lado, eventos como o SBIE, o SBSI, o BRACIS (Brazilian Conference on Intelligent Systems) e o WSCAD (Workshop em Sistemas Computacionais de Alto Desempenho) apresentaram uma redução na participação feminina ao longo do tempo, mesmo em casos em que o número total de participantes aumentou.

A análise dos núcleos dos CPs indicou que a participação feminina dentro desses núcleos é estável e, em alguns eventos, proporcionalmente maior do que no total de participantes. Eventos como SBSC e SVR (Simpósio sobre Realidade Virtual e Aumentada) apresentaram uma diminuição no crescimento dos CPs, mas um aumento na participação feminina, destacando a importância de criar ambientes inclusivos mesmo em contextos menores.

### III. METODOLOGIA

Nesta seção, são apresentadas as principais atividades realizadas para o avanço da temática.

#### A. Preparação e Estudos

O primeiro passo consistiu em uma revisão da literatura sobre a participação feminina na Computação, pesquisa científica e comitês de programas. Foram analisados os artigos principais publicados pelo grupo de pesquisa da professora Mirella M. Moro (e.g., [22], [17], [18], [23], [24]), juntamente com todas as suas referências bibliográficas. Esse estudo inicial foi fundamental para a consolidação do tema e permitiu a identificação de referências mais atuais no assunto.

Novas referências foram obtidas a partir de diversas fontes, incluindo a SBC OpenLib<sup>1</sup> (SOL) - Biblioteca Digital da Sociedade Brasileira de Computação, Google Scholar<sup>2</sup>, IEEE Xplore<sup>3</sup> e ACM Digital Library<sup>4</sup>. Essa pesquisa foi essencial para a construção do referencial teórico apresentado e forneceu a base necessária para a definição da direção da pesquisa, que é detalhada nas seções seguintes.

Os dados já coletados pelos artigos principais foram integrados ao estudo, permitindo a continuidade da pesquisa com informações de anos anteriores. Isso facilitou a análise longitudinal e a comparação temporal dos dados, enriquecendo a qualidade e a abrangência do estudo.

#### B. Seleção de Eventos

A seleção dos eventos nacionais de MSI 1 foi baseada na continuidade da pesquisa de Lorens et al. [17]. Foram considerados os mesmos eventos nacionais utilizados na pesquisa

<sup>1</sup><https://sol.sbc.org.br/>

<sup>2</sup><https://scholar.google.com/>

<sup>3</sup><https://ieeexplore.ieee.org/Xplore/home.jsp>

<sup>4</sup><https://dl.acm.org/>

original, com a adição de alguns novos eventos. A seleção buscou incluir diversas áreas da Computação, com o objetivo de proporcionar diferentes perspectivas sobre a participação feminina.

Além disso, para este projeto, foram selecionados eventos internacionais em áreas similares aos eventos nacionais, visando uma comparação entre a participação feminina em contextos nacionais e internacionais. O planejamento incluiu a coleta de dados de ambos os tipos de eventos para possibilitar uma análise comparativa da participação feminina. Os eventos selecionados e a relação entre eles podem ser observados nas Tabelas: I para eventos nacionais; II para eventos internacionais; e III para a equiparação entre tais eventos.

Tabela I  
EVENTOS NACIONAIS

Evento	Simpósio Brasileiro em
BRACIS	Sistemas Inteligentes
BSB	Bioinformática
GEOINFO	Geoinformática
IHC	Fatores Humanos em Sistemas Computacionais
SBBDD	Bancos de Dados
SBCCI	Circuitos Integrados e Sistemas
SBES	Engenharia de Software
SBESC	Engenharia de Sistemas Computacionais
SBIE	Informática na Educação
SBMF	Métodos Formais
SBRC	Redes de Computadores e Sistemas Distribuídos
SBSC	Sistemas Colaborativos
SBSeg	Segurança da Informação e de Sistemas Computacionais
SBSI	Sistemas de Informação
SIBGRAPI	Gráficos, Padrões e Imagens
SVR	Realidade Virtual e Aumentada
WebMedia	Sistemas Multimídia e Web
WSCAD	Arq. de Computadores e Computação de Alto Desempenho

#### C. Coleta de dados

Para cada evento internacional selecionado, foi realizada uma coleta manual dos membros integrantes do comitê de programa (CP) de cada edição realizada entre 2011 e 2023. A coleta seguiu uma lista de prioridades de fontes:

- 1) Anais do evento, físicos digitalizados ou digitais. Para os anais digitalizados, foram utilizadas ferramentas de extração de texto para obter a lista de membros dos CPs.
- 2) Site oficial da edição do evento. O acesso a esses sites era mais fácil que aos anais, mas a informação poderia estar desatualizada.
- 3) Caso o site do evento estivesse fora do ar, foi utilizado o *WebArchive*<sup>5</sup> para recuperar a página enquanto estava ativa. O *WebArchive* é um serviço digital que arquiva páginas da web ao longo do tempo, permitindo o acesso a versões antigas de sites que podem não estar mais disponíveis online.

<sup>5</sup><https://web.archive.org/>

Tabela II  
EVENTOS INTERNACIONAIS

Evento	International Conference / Symposium on
IJCAI	Artificial Intelligence
SIGSPATIAL	Advances in Geographic Information Systems
CHI	Human Factors in Computing Systems
SIGMOD	Management of Data
ICCAD	Computer-Aided Design
ICSE	Software Engineering
DAC	Design Automation
SIGCSE	Computer Science Education
MEMOCODE	Formal Methods and Models for System Design
SIGCOMM	Applications, Technologies, Architectures, and Protocols for Computer Communication
CSCW	Computer-Supported Cooperative Work and Social Computing
CCS	Computer and Communications Security
ICIS	Information Systems
SIGGRAPH	Computer Graphics and Interactive Techniques
VRST	Virtual Reality Software and Technology
SIGMM	Multimedia
MICRO	Microarchitecture
BCB	Bioinformatics, Computational Biology, and Health Informatics

Tabela III  
EVENTOS NACIONAIS E INTERNACIONAIS EQUIPARÁVEIS

Evento nacional	Evento internacional	Temática principal
BRACIS	IJCAI	Inteligência Artificial
BSB	BCB	Biologia
GEOINFO	SIGSPATIAL	Geoinformática
IHC	CHI	Interação Humano-Computador
SBBB	SIGMOD	Bancos de Dados
SBCCI	ICCAD	Circuitos e CAD
SBES	ICSE	Engenharia de Software
SBESC	DAC	Sistemas Computacionais
SBIE	SIGCSE	Educação
SBMF	MEMOCODE	Métodos Formais
SBRC	SIGCOMM	Redes de Computadores
SBSC	CSCW	Sistemas Colaborativos
SBSeg	CCS	Segurança
SBSI	ICIS	Sistemas de Informação
SIBGRAPI	SIGGRAPH	Computação Gráfica
SVR	VRST	Realidade Virtual
WebMedia	SIGMM	Multimídia
WSCAD	MICRO	Arquitetura e Desempenho

- 4) Contato com as pessoas coordenadoras dos comitês. Para os poucos casos faltantes, foi realizado contato com as pessoas que coordenaram os comitês da edição, embora sem sucesso, diferentemente do retorno obtido para algumas edições faltantes de eventos nacionais, em MSI 1.

Foram criadas planilhas para cada evento e abas de ano referente a cada coleta realizada. Para cada coleta, a informação mínima obtida era: sigla do evento e nome da pessoa.

Quando disponíveis, também foram coletados dados de país e instituição de ensino da pessoa. Também foram coletados os nomes das pessoas coordenadoras do evento (*General Chair*) e das pessoas coordenadoras dos CPs (*Program Chair*). Os dados brutos coletados foram salvos no Google Drive em arquivos de texto e planilhas, com referência à fonte de onde foram coletados e com as informações citadas anteriormente. Um exemplo de um arquivo gerado pela coleta pode ser visto na Figura 1.

Para a coleta dos dados, foram incluídos os integrantes do *Technical Program Committee* (TPC) na maioria dos eventos. Nos casos em que o TPC era subdividido em áreas ou trilhas temáticas, considerou-se apenas a trilha principal de artigos. Quando a divisão em áreas servia apenas para detalhar a temática do evento, todas as áreas foram incluídas, com a exclusão de indivíduos que estavam listados em mais de uma delas para evitar duplicações.

Para os eventos SIGCSE (edições de 2011 a 2017 e 2023) e ICIS (edição de 2011), foram coletados os membros da categoria *Reviewers*. Quando havia trilhas distintas, foi considerada apenas a trilha principal de artigos (*papers*). Embora em muitos eventos os *Reviewers* não desempenhem a mesma função que os membros do TPC, essa inclusão foi necessária nesses casos específicos, pois essas edições não possuíam uma categoria formal equivalente ao TPC.

A coleta, conciliação e tratamento de dados dos eventos internacionais foram demorados devido à natureza manual desse processo. Cada evento apresenta características muito específicas e a estrutura dos dados não segue um padrão consistente. Isso exigiu um esforço considerável para localizar, extrair e organizar as informações de forma adequada, para as mais de 200 coletas realizadas manualmente.

```
Source:
link.springer.com/book/10.1007/978-3-031-45368-7

General Chairs:
Gisele Lobo Pappa, UFMG, Brazil
Wagner Meira Jr., UFMG, Brazil

Program Committee Chairs:
Murilo Coelho Naldi, UFSCar, Brazil
Reinaldo A. C. Bianchi, FEI, Brazil

Program Committee:
Adenilton da Silva, UFP, Brazil
Adriane Serapiao, Unesp, Brazil
...
```

Figura 1. Exemplo de arquivo gerado pela etapa de coleta de dados para a edição de 2023 do evento BRACIS.

#### D. Conciliação dos Dados

A partir desta etapa, os dados registrados manualmente em planilhas do Google Sheets foram processados de forma automatizada por meio de notebooks em Python. Para acessar e manipular os dados, foram utilizadas as APIs do Google Sheets<sup>6</sup>. Esse processo permitiu a extração, o tratamento e a

<sup>6</sup><https://developers.google.com/sheets/api/reference/rest>

análise das informações de cada evento de maneira eficiente e sistemática.

A conciliação dos dados consistiu na unificação de todas as informações coletadas ao longo dos anos em uma única tabela por evento. Cada tabela continha as colunas Nome, Instituição, País, Gênero e Ano. Inicialmente, os dados de cada ano foram simplesmente transferidos para essa tabela geral. Para simplificar a fusão das informações, adotou-se o critério de que indivíduos com o mesmo nome, dentro do mesmo ano, seriam considerados a mesma pessoa. Essa decisão foi tomada devido à inviabilidade de uma revisão manual completa, considerando o elevado volume de dados nas etapas seguintes.

Em seguida, a tabela de participantes foi processada para gerar uma nova tabela, com as colunas Nome, Instituição, País, Sexo e colunas de 2011 a 2023. Caso uma pessoa A tenha participado em 2023, ela recebe 1 na coluna desse ano ou 0 caso contrário. Contudo, apenas migrar os dados para essa aba apresenta um problema: havia repetição de nomes, pois uma pessoa pode ter participado de mais de um evento.

Para resolver esse problema, foi necessário unir linhas que se referiam à mesma pessoa. Dessa forma, a presença de uma pessoa poderia ser verificada em cada coluna de ano. Essa tarefa de conciliação não poderia ser feita de forma manual, visto que era muito demorada e não era escalável para a quantidade de eventos e edições restantes. Logo, foi desenvolvido um algoritmo com o propósito de automatizar a consolidação dos dados. O algoritmo desenvolvido funciona da seguinte forma:

- 1) **Normalização de Nomes:** O algoritmo começa normalizando os nomes, padronizando-os para garantir que pequenas variações na escrita não resultem em entradas duplicadas. Isso inclui remover caracteres especiais, unificar diferentes formas de escrita (como transformar "Junior" em "Jr"), e padronizar o formato de capitalização.
- 2) **Estratégia de Agrupamento por Similaridade:** Para comparar nomes e identificar registros que possam se referir à mesma pessoa, o algoritmo utiliza uma abordagem simples baseada em grupos de bloqueio, ou *blocking*<sup>7</sup>. Inicialmente, os nomes são organizados em blocos conforme a primeira letra, reduzindo a quantidade de comparações e aumentando a eficiência do processo.
- 3) **Cálculo de Similaridade:** Dentro de cada bloco, o algoritmo calcula a similaridade entre pares de nomes utilizando a métrica *Jaro-Winkler*<sup>8</sup> e a distância de *Levenshtein*<sup>9</sup>. Essas técnicas permitem medir o grau de semelhança entre duas cadeias de caracteres, levando em conta tanto a proximidade ortográfica quanto pequenas variações na escrita. Além disso, o algoritmo emprega

<sup>7</sup>Técnica que divide os dados em subconjuntos menores para reduzir o número de comparações necessárias, melhorando a eficiência da análise.

<sup>8</sup>Algoritmo utilizado para medir a similaridade entre duas cadeias de caracteres, atribuindo maior peso para caracteres iniciais correspondentes.

<sup>9</sup>Métrica que quantifica a diferença entre duas palavras pelo número mínimo de edições (inserções, deleções ou substituições) necessárias para transformá-las uma na outra.

a codificação fonética *Metaphone*<sup>10</sup>, assegurando que nomes com pronúncias similares também sejam reconhecidos como equivalentes. Nomes que atingem um limiar mínimo de similaridade e apresentam baixa distância de edição são considerados como sendo da mesma pessoa ao longo dos anos. Em MSI 1, foi utilizada a biblioteca *diffli*<sup>11</sup>, que fornece ferramentas para comparar sequências, contudo o uso dessa biblioteca para os dados intencionais se mostrou pouco eficiente, unindo dados de forma muito agressiva, o que levava a união errônea de pessoas.

- 4) **Construção de Grupos por Grafos:** A relação entre os nomes identificados como semelhantes é representada em um grafo, onde cada nó corresponde a um nome e cada aresta indica uma correspondência acima do limiar estabelecido. O algoritmo então encontra os componentes conectados para formar grupos de nomes que compartilham similaridade, garantindo que todas as variantes de um mesmo nome sejam associadas corretamente.
- 5) **Consolidação dos Dados:** Após a identificação dos grupos, os registros pertencentes a uma mesma pessoa são consolidados. A consolidação dos dados preserva informações relevantes, como instituição, país e gênero, e agrega a participação de cada pessoa ao longo das diferentes edições do evento.

Essas etapas permitem a automatização do processo de conciliação de dados, tornando-o mais eficiente e escalável para grandes volumes de dados. Após a consolidação dos dados, é realizada uma leitura e verificação por amostra para garantir a precisão da união de nomes. Essa verificação inclui a conferência do total de pessoas, assegurando que todas as entradas duplicadas foram corretamente agrupadas e que cada pessoa é representada apenas uma vez na tabela final.

A precisão desse método foi avaliada verificando-se a consistência do número de participantes por evento ao longo dos anos. Para quantificar possíveis erros na conciliação, foi calculada uma métrica baseada na variação do total de participações antes e depois da unificação dos registros. O erro percentual foi determinado conforme a Equação (1). O resultado desse cálculo de erro pode ser visto na Tabela IV.

$$\text{Erro(\%)} = \frac{\text{Total coletado} - \text{Total após conciliação}}{\text{Total coletado}} \times 100 \quad (1)$$

Eventos sem taxa de erro indicam que não houve registro de perda na participação ao longo do tempo. No entanto, isso não garante que todos os registros são de fato de pessoas distintas. Nos eventos com maiores taxas de erro, a principal causa foi a presença de pessoas com nomes muito semelhantes ou idênticos, tornando a distinção entre elas uma tarefa desafiadora. Essa ambiguidade pode resultar na consolidação inadequada de registros, impactando a precisão dos dados finais.

<sup>10</sup>Método que converte palavras em códigos fonéticos para facilitar a comparação de palavras com pronúncias semelhantes.

<sup>11</sup><https://pypi.org/project/cdiffli/>

Tabela IV  
ERRO NO ALGORITMO DE CONCILIAÇÃO POR EVENTO

Evento	Erro na conciliação (%)	Total de participações
SIGMM	6,76%	9.114
IJCAI	2,66%	22.448
CCS	1,18%	2.195
SIGGRAPH	0,26%	763
SIGCSE	0,24%	8.647
ICIS	0,19%	1.050
BCB	0,15%	1.312
SIGSPATIAL	0,14%	1.476
DAC	0,08%	2.643
CHI	0,06%	3.212
CCSW	0,00%	1.340
ICCAD	0,00%	1.543
ICSE	0,00%	1.313
MEMOCODE	0,00%	478
MICRO	0,00%	919
SIGCOMM	0,00%	660
SIGMOD	0,00%	1.942
VRST	0,00%	749
<b>Total</b>	<b>2,06%</b>	<b>61.804</b>

### E. Classificação do Gênero pelo Nome

Nesta etapa, foram testados vários classificadores para identificar se um nome é feminino ou masculino. Seguindo o método utilizado em MSI 1, todos os nomes foram classificados com as seguintes ferramentas, além do último classificador que não estava incluído em MSI 1 (GPT-4o).

- **Gender Guesser:** Utiliza listas de nomes comuns para inferir o gênero com base em ocorrências conhecidas.
- **NLTK:** Uma biblioteca de processamento de linguagem natural que inclui funcionalidades para classificação de texto.
- **Zero-Shot Learning:** É uma técnica de aprendizado de máquina em que o modelo pode fazer previsões para categorias nas quais não foi treinado diretamente. Ao invés de iniciar com um treinamento específico para cada categoria, o modelo usa uma compreensão geral da linguagem para “adivinhar” a classe correta. Foram utilizados três modelos de zero-shot:
  - **Zero-Shot BART:** O BART (*Bidirectional and Auto-Regressive Transformers*) é um modelo de linguagem que combina características de análise bidirecional (para entender o contexto de palavras que vêm antes e depois) com a geração de texto (auto-regressivo). Isso o torna útil tanto para tarefas de classificação quanto de geração de texto.
  - **Zero-Shot RoBERTa:** O RoBERTa (*Robustly optimized BERT approach*) é uma versão otimizada do BERT, com ajustes nos hiperparâmetros e mais dados de treinamento. Por ser bidirecional, ele entende o contexto das palavras em ambas as direções e se destaca em tarefas de compreensão de linguagem

natural.

- **Zero-Shot XLM-RoBERTa:** É a variante multilíngue do RoBERTa, treinada em diversos idiomas. Isso permite que o modelo lide bem com tarefas de processamento de linguagem natural em vários contextos linguísticos.

- **GPT-4o:** Um modelo de linguagem da *OpenAI* que também é capaz de realizar tarefas de classificação de gênero com base no nome. O GPT-4o pode atuar em regime zero-shot, isto é, sem necessitar de um treinamento específico para cada classe, pois é treinado em uma grande quantidade de dados e tem uma forte capacidade de compreensão contextual da linguagem.

Inicialmente, foram aplicados os classificadores *Gender Guesser*, *NLTK* e os modelos baseados em *Zero-Shot Learning* para atribuir gênero aos nomes do conjunto de dados dos CPs internacionais. No entanto, o desempenho desses classificadores mostrou-se insatisfatório: muitos nomes eram classificados como gênero indefinido e, quando conseguiam atribuir gêneros, a proporção entre nomes masculinos e femininos variava de forma inconsistente. Diferentemente do que acontecia no cenário nacional, uma verificação manual de cada nome seria impraticável, pois muitos são ambíguos em relação à língua portuguesa e a conferência de todos poderia levar muito tempo.

Diante disso, foi necessário um classificador que fosse capaz de lidar com nomes provenientes de diversos países de maneira mais precisa. Por isso, optou-se pelo uso do GPT-4o, que demonstrou maior eficiência e assertividade em comparação aos demais classificadores avaliados. Os resultados dos testes são apresentados na Tabela V, na qual a acurácia representa a porcentagem de acertos em relação ao total de registros classificados, excluindo os resultados indefinidos. Para avaliar o desempenho do GPT-4o, utilizou-se a base de dados de nomes de MSI 1, previamente conferida manualmente, garantindo uma referência para a avaliação dos resultados.

Uma verificação dos resultados foi realizada por amostragem para garantir a precisão das classificações. Para confirmar o gênero, foram pesquisadas páginas que faziam referência à pessoa, observando-se os pronomes utilizados para se referir a ela. Verificou-se que o GPT-4o conseguiu lidar com nomes bastante variados, apresentando classificações mais assertivas mesmo em casos ambíguos. Por exemplo, nomes tipicamente masculinos de origem italiana, como *Daniele P.* e *Gabriele S.*, bem como nomes reconhecidos como femininos de origem do Oriente Médio, como *Nassim P.* e *Hazar H.*. Além disso, o classificador demonstrou bom desempenho diante de nomes totalmente ambíguos para os outros classificadores, como *Chetan T.*, *Anshul K.* e *Jian P.*, provenientes de países da Ásia.

Esses nomes causam incertezas entre falantes de português, pois não seguem os padrões ortográficos ou fonéticos da língua. Além disso, muitos classificadores existentes são treinados com nomes de origem ocidental e predominantemente masculinos. Ainda assim, o GPT-4o mostrou-se robusto o suficiente para classificar o gênero de forma consistente,



destacando-se como a melhor solução para este cenário de alta diversidade de nomes.

Tabela V  
ACURÁCIA DOS CLASSIFICADORES PARA NOMES MULTICULTURAIIS

Classificador	Indefinidos (%)	Acurácia (%)
GPT-4o	0,01	99,99
Gender Guesser	40,43	96,58
Zero-Shot RoBERTa	39,21	90,29
Zero-Shot XLM-RoBERTa	38,59	87,62
NLTK	49,78	81,34
Zero-Shot BART	25,78	73,92

#### F. Tratamento dos Dados

O primeiro passo do tratamento dos dados foi corrigir as edições sem dados. Para essas edições, foram utilizados os dados do anterior mais próximo. Esses casos foram corrigidos da seguinte forma, conforme a Tabela VI.

Tabela VI  
CORREÇÕES DAS EDIÇÕES FALTANTES

Evento	Edições faltantes corrigidas
IJCAI	2012 ← 2011, 2014 ← 2013
CHI	2011 ← 2010, 2012, ← 2010, 2013 ← 2010, 2015 ← 2014, 2016 ← 2014
ICCAD	2023 ← 2022
ICIS	2020 ← 2019
VRST	2020 ← 2019

Os dados faltantes para o IJCAI (2012, 2014) e ICIS (2020) ocorreram visto que essas edições não foram realizadas. As demais edições faltantes referem-se a eventos que não tiveram os dados dos membros do comitê disponibilizados nos anais ou em outros meios. Após essa limpeza e normalização, o conjunto de dados totalizou 28.931 pessoas.

#### G. Análise dos Dados

Nesta etapa, foi desenvolvido um notebook em Python para realizar uma análise abrangente dos eventos coletados. O objetivo principal dessas análises foi criar uma visão clara e detalhada dos dados dos eventos, seguindo a metodologia descrita nos artigos de base. Para permitir comparações uniformes, alguns dados foram convertidos em porcentagens. A análise foi estruturada em três aspectos principais: número de mulheres em termos de pessoas, participação feminina ao longo do tempo e definição do núcleo dos eventos. Esses conceitos foram explicados no referencial teórico.

Foram extraídas diversas informações do conjunto de dados. Primeiro, foi feita uma análise da composição geral dos comitês, gerando gráficos sobre a presença e participação feminina para cada evento em cada ano. Em seguida, foram gerados gráficos mostrando a mudança da presença feminina ao longo dos anos.

Diversas visualizações foram criadas na tentativa de mostrar essa variação temporal, mas como as escalas eram diversas e a variabilidade dos dados era alta, muitas delas falharam em comunicar as informações principais. Para solucionar esse problema, foram criadas duas visualizações qualitativas: um mapa de calor que mostra a presença feminina em cada ano de cada evento e um gráfico de bolhas mostrando a tendência de crescimento seguindo uma regressão linear para cada evento. Além disso, para comparar os eventos nacionais e os internacionais, foram gerados gráficos para mostrar as diferenças e semelhanças entre ambos os contextos. Cada uma dessas visualizações será apresentada e analisada no capítulo de resultados e conclusões.

## IV. RESULTADOS

A primeira informação extraída do conjunto de dados foi a porcentagem de mulheres, que totalizam 20.05% das pessoas. Esse valor é bem similar ao da participação feminina ao longo dos anos, já que no período de 2011 a 2023, as mulheres representaram 19.23% das participações dos comitês de programas dos eventos nacionais. Dentro dos núcleos dos eventos, as mulheres representam 18.20% das pessoas. As mulheres dentro do núcleo representam 27.33% do total de mulheres, enquanto, no caso dos homens, 30.81% deles estão nos núcleos dos eventos.

A comparação entre os dados nacionais e internacionais revelou diferenças na participação feminina nos comitês de programas dos eventos analisados. Nos eventos internacionais, as mulheres representam 20,05% do total de participantes, enquanto nos eventos nacionais esse percentual é 25,09%, indicando uma participação feminina maior no cenário brasileiro. Essa tendência se mantém também para as outras métricas, com as mulheres correspondendo a 19,23% das participações nos comitês de programas internacionais, contra 26,28% nos nacionais.

Esses dados sugerem que, apesar de a representatividade feminina ainda ser inferior à masculina tanto no Brasil quanto no exterior, os eventos nacionais apresentam uma participação feminina mais expressiva nas métricas analisadas, incluindo uma maior presença das mulheres nos núcleos dos eventos.

#### A. Análises da participação feminina geral

O primeiro gráfico gerado refere-se ao percentual de participação feminina nos comitês de programas dos eventos analisados. Nenhum evento apresentou uma participação feminina superior a 50%. O gráfico mostrado na Figura 2 evidencia a disparidade existente na participação feminina entre os diferentes eventos.

Eventos como o CHI, CSCW, SIGCSE e ICSE destacam-se por possuírem os maiores índices de participação feminina nos CPs. Estes eventos compartilham características temáticas centradas em interação humano-computador, trabalho colaborativo, educação em computação e engenharia de software, que são áreas com uma maior presença feminina, tradicionalmente.

Por outro lado, eventos como o ICIS, MICRO, DAC e ICCAD apresentam os menores índices de participação feminina

em seus comitês de programa. Esses eventos são predominantemente focados em áreas como sistemas de informação, arquitetura de computadores, automação de design eletrônico e computação embarcada.

Ao comparar esses resultados com os dados dos eventos nacionais analisados em MSI 1, observa-se um padrão semelhante. No Brasil, os eventos com maior participação feminina incluem IHC, SBIE, SBSC e SBSI, que abordam tecnologias colaborativas, comunicação, interação humano-computador e educação em informática, áreas que, assim como nos eventos internacionais, possuem uma presença feminina relativamente maior.

Por outro lado, eventos como SBSeg, SIBGRAPI, SBESC e SBCCI registram os menores índices de participação feminina. Essas conferências são predominantemente voltadas para segurança de sistemas, processamento de imagens, engenharia de sistemas computacionais e circuitos integrados, subáreas tradicionalmente dominadas por homens e com menor diversidade de gênero.

A diferença nas taxas de participação feminina pode estar relacionada não apenas à temática dos eventos, mas também a fortes fatores culturais e estruturais dentro das sub-áreas da Computação. Eventos com maior enfoque em aspectos sociais, colaborativos e educacionais tendem a atrair mais mulheres, enquanto áreas mais técnicas e tradicionalmente dominadas por homens ainda falham em alcançar uma maior diversidade de gênero.

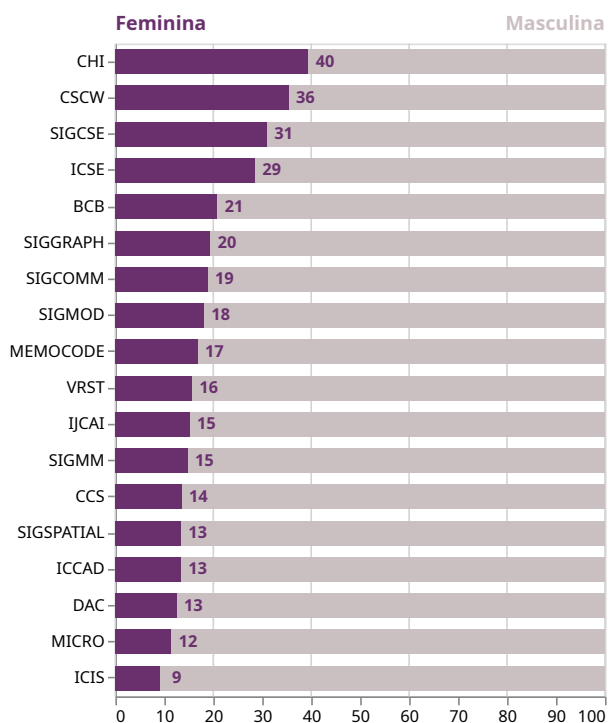


Figura 2. Percentual da participação feminina e masculina nos comitês de programas dos eventos internacionais.

Ao analisar as Figuras 3 e 4, observa-se que não há uma relação direta entre o tamanho do comitê e a participação

feminina. Eventos como IJCAI e SIGMM possuem grandes comitês, mas apresentam baixa representatividade feminina. No entanto, verifica-se que eventos com comitês menores tendem a registrar índices ainda mais baixos de participação feminina, como é o caso do MEMOCODE, ICIS e SIGSPATIAL.

As Figuras 5 e 6 reforçam essa observação, indicando que eventos com um número absoluto reduzido de mulheres também tendem a apresentar menores percentuais de participação feminina em relação ao total de membros. Entretanto, para eventos de porte médio e grande, não há uma correlação clara entre o número total de participantes e a representatividade feminina.

Essa análise foi conduzida com o objetivo de identificar possíveis *outliers* na participação feminina, ou seja, comitês que inicialmente apresentam um número significativo de mulheres, mas que não conseguem manter essa participação ao longo dos anos. Esse padrão pode sugerir que comitês menores oferecem um ambiente menos inclusivo ou adotam políticas menos eficazes de incentivo à diversidade de gênero.

Por outro lado, a baixa presença feminina em eventos de grande porte, como IJCAI e SIGMM, indica que fatores além do tamanho do comitê podem influenciar a participação, incluindo a temática do evento, a existência de redes de apoio entre as participantes e as iniciativas específicas de inclusão promovidas por cada conferência. Esses resultados destacam a importância não apenas de ampliar a participação feminina nos comitês de programas, mas também de implementar políticas que garantam condições favoráveis para a permanência e crescimento contínuo nessas posições ao longo dos anos.

### B. Participação Feminina nos Núcleos

Ao analisar o núcleo dos comitês de programas (CPs) pela Figura 7, percebe-se que a quantidade de mulheres no núcleo é semelhante à quantidade de mulheres participando no CP como um todo. Em outras palavras, eventos que apresentam as maiores porcentagens de presença feminina são aqueles que também têm as maiores participações femininas entre as pessoas com as 25% maiores participações. Eventos como o CHI, CSCW, ICSE e SIGCSE destacam-se por possuírem os maiores índices de participação feminina nos CPs. Em contraste, eventos como o ICIS, MICRO, DAC e VRST apresentam os menores índices de participação feminina nos núcleos de seus comitês de programa.

Essa observação sugere que analisar o núcleo dos comitês pode ser um bom método para entender a diversidade de gênero dentro dos CPs de eventos internacionais, assim como foi possível concluir em MSI 1 sobre os eventos nacionais. A porcentagem de mulheres dentro dos núcleos não se manteve tão estável do que a participação de mulheres no geral ao comparar diferentes eventos, conforme ilustrado na Figura 8. No entanto, é interessante notar que os eventos que possuem menores porcentagens de mulheres dentro do núcleo, em comparação com o total de mulheres, tendem a ser aqueles com menor participação feminina geral.

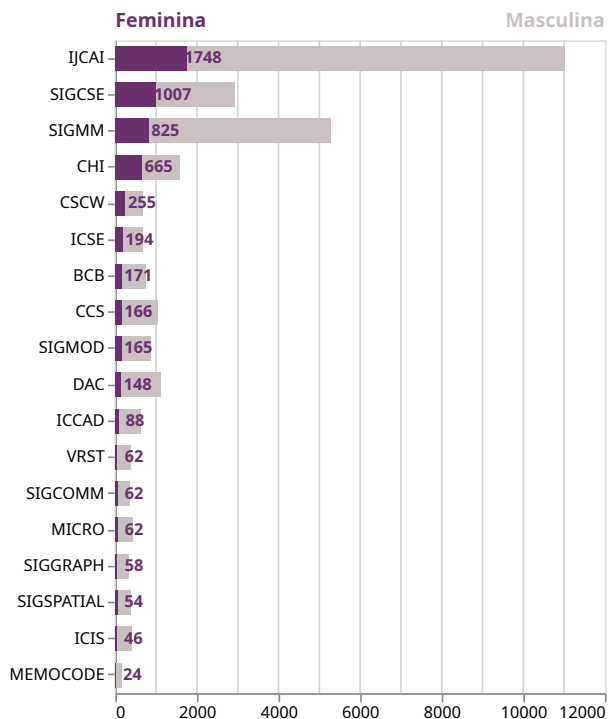


Figura 3. Número absoluto de pessoas nos comitês de programas dos eventos internacionais.

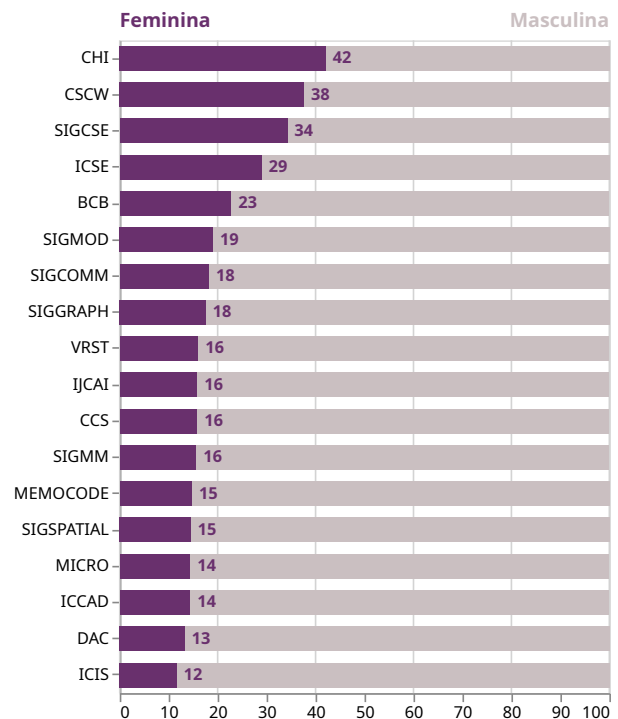


Figura 5. Percentual de pessoas comitês de programas dos eventos internacionais.

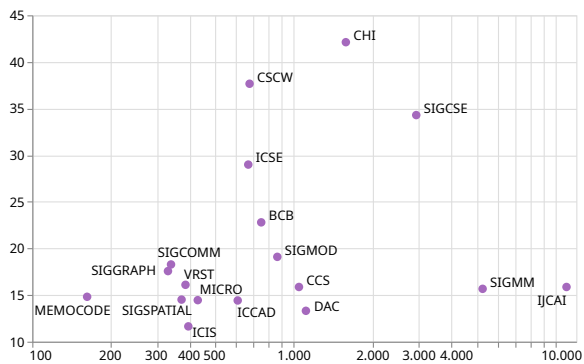


Figura 4. Relação entre tamanho do comitê de programa e a participação feminina nos eventos internacionais.

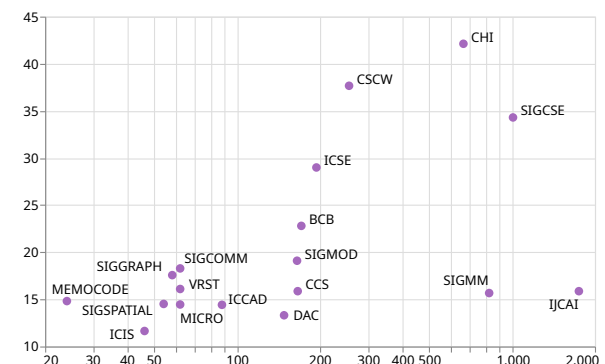


Figura 6. Relação entre o número absoluto de mulheres no comitê de programa e a participação feminina nos eventos internacionais.

### C. Evolução da Participação Feminina

Para analisar a participação feminina ao longo dos anos, foi desenvolvido um mapa de calor que permite uma visualização mais clara e qualitativa das mudanças ao longo do tempo, conforme mostrado na Figura 9. Este gráfico revela a discrepância entre eventos com maior e menor participação feminina. A maioria dos eventos apresenta uma baixa participação, resultando em um gráfico predominantemente frio em termos de cores. No entanto, pode-se observar que alguns eventos, como o CHI, ICSE, CCS, SIGCOMM, MEMOCODE e SIGGRAPH, tiveram um crescimento na participação feminina em seus CPs ao longo dos anos. Por outro lado, eventos como o CSCW, BCB, MICRO e ICIS mostraram uma diminuição

na participação feminina, principalmente nos últimos 4 anos analisados.

A análise do mapa de calor revela que apenas uma edição apresentou uma participação feminina superior à masculina: o IHC de 2018, no qual 54,2% dos membros do comitê de programa eram mulheres. Nos eventos internacionais, não foi identificado nenhum caso semelhante.

Para analisar a evolução da participação feminina ao longo do tempo, foram ajustadas regressões lineares para as séries temporais de cada evento. A regressão linear foi selecionada por ser um modelo estatístico de fácil interpretação e simples. No entanto, devido à natureza dos dados, essa abordagem não apresenta um ajuste ideal, pois a participação feminina ao

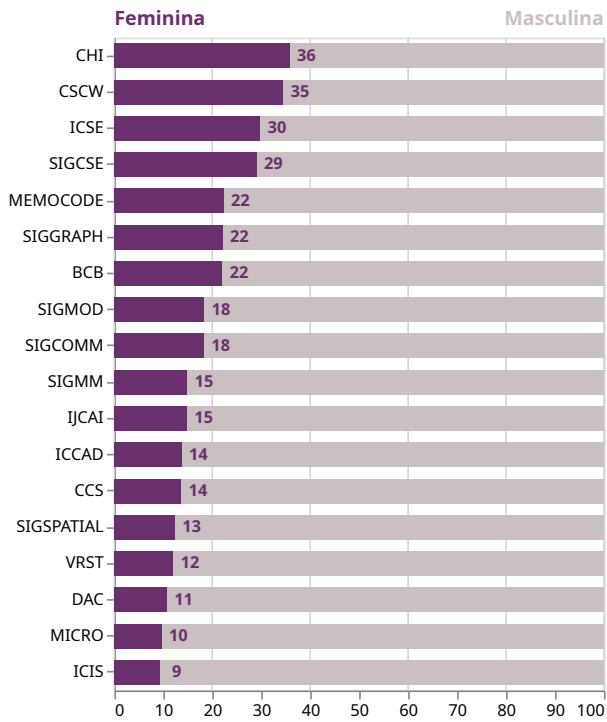


Figura 7. Percentual de pessoas no núcleo dos comitês de programa dos eventos internacionais.

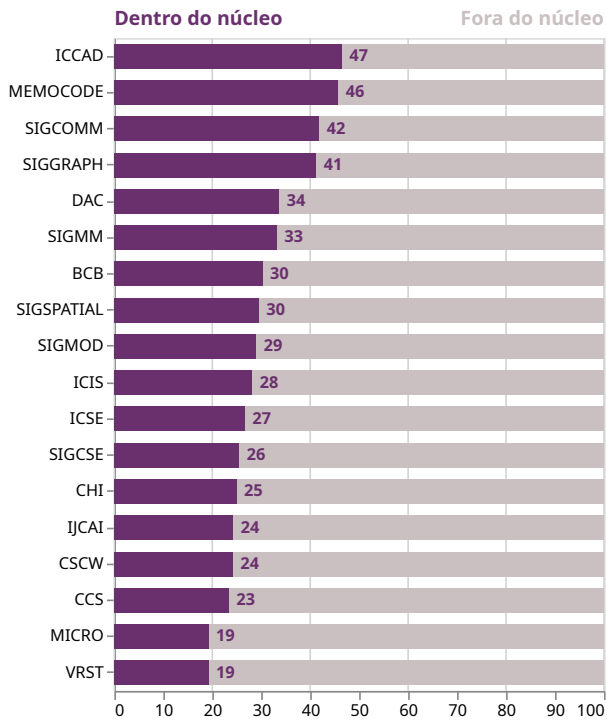


Figura 8. Percentual de mulheres dentro e fora do núcleo dos comitês de programa dos eventos internacionais.

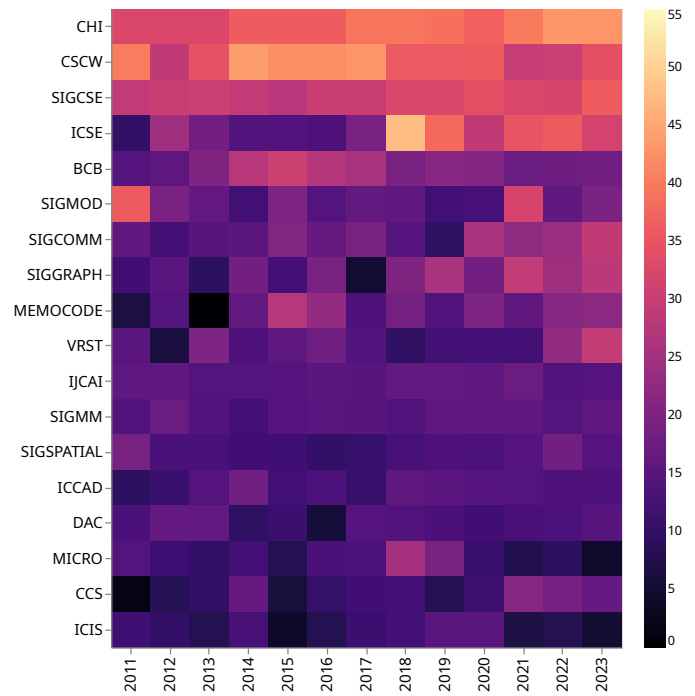


Figura 9. Mapa de calor da porcentagem da participação feminina nos comitês de programas dos eventos internacionais. Ordenado do evento com maior média de participação feminina para o menor.

longo dos anos não segue um crescimento linear contínuo, mas passa por fases distintas de aumento, estagnação e até declínio.

A revisão da literatura indica a existência de modelos alternativos para análise temporal, como o uso de modelos autorregressivos [10]. No entanto, considerando que cada evento possui apenas 13 pontos de dados, a aplicação de técnicas mais complexas não se mostrou vantajosa, uma vez que métodos estatísticos mais sofisticados requerem uma maior quantidade de observações para fornecer estimativas robustas.

Com os coeficientes das regressões lineares, foi gerada uma visualização de bolhas, ilustrada na Figura 10, permitindo uma representação gráfica das tendências de crescimento ou declínio da participação feminina nos comitês de programa dos eventos analisados. Nessa visualização, o eixo  $x$  representa a tendência de crescimento do total de participantes do evento, enquanto o eixo  $y$  representa a tendência de crescimento da participação feminina.

1) *Primeiro quadrante*: Observa-se que a maioria dos eventos internacionais está localizada no primeiro quadrante, o que indica que o aumento da participação feminina acompanha o crescimento do tamanho dos comitês. Destacam-se eventos como ICSE e CHI, reconhecidos por apresentarem uma das mais altas taxas de participação feminina. Esse quadrante também inclui eventos com baixa representatividade feminina, como SIGGRAPH, CCS, SIGCOMM e VRST.

No cenário nacional, encontram-se no primeiro quadrante eventos como GEOINFO, IHC e SBSeg. O IHC, assim como o CHI, está relacionado à interação humano-computador, uma

área que tende a ter maior participação feminina. Já o SBSeg, semelhante ao CCS, está ligado à segurança da informação, um campo que é tradicionalmente dominado por homens, mas que possui uma tendência positiva da participação feminina.

2) *Segundo quadrante:* Os eventos localizados no segundo quadrante apresentaram um aumento na participação feminina, apesar da redução no tamanho dos comitês. Apenas dois eventos internacionais se encaixam nesse perfil: SIGCSE e SIGSPATIAL. O SIGCSE, voltado para a educação em computação, já possuía uma participação feminina elevada em comparação com os demais eventos, enquanto o SIGSPATIAL já possuía participação feminina reduzida. No contexto nacional, essa mesma tendência foi observada nos eventos SBSC e SVR. O SBSC e o SIGCSE abrangem áreas que tradicionalmente atraem mais pesquisadoras, mas que possuem uma tendência negativa na participação feminina nos últimos anos.

3) *Terceiro quadrante:* No terceiro quadrante estão os eventos que apresentaram queda tanto na participação feminina quanto no crescimento do comitê. No cenário internacional, apenas o ICIS se enquadra nessa categoria, demonstrando uma tendência negativa em eventos voltados para sistemas de informação. No Brasil, destacam-se SBFM e WSCAD, que seguiram um padrão semelhante ao ICIS, com uma redução mais acentuada na participação feminina em comparação à diminuição do tamanho de seus comitês. Isso pode indicar que, nesses eventos, a retração dos CPs tem impactado de maneira desproporcional a presença de mulheres, agravando a desigualdade de gênero.

4) *Quarto quadrante:* O quarto quadrante reúne eventos nos quais o número total de participantes nos CPs aumentou, mas a participação feminina diminuiu. Entre os eventos internacionais, destaca-se o CSCW, que, apesar de estar entre aqueles com as maiores taxas de participação feminina, sofreu uma retração nesse índice. Outros eventos desse grupo incluem MICRO, SIGMOD e BCB.

No cenário nacional, a maioria dos eventos encontra-se nesse quadrante. Entre os exemplos notáveis estão SBIE e SBSI, que, assim como o CSCW, possuem uma das mais altas taxas de participação feminina, mas apresentaram um declínio nesse índice. Da mesma forma, eventos como SIBGRAPI, SBRC e WebMedia, que já possuíam baixa participação feminina, seguiram a mesma tendência dos eventos internacionais de áreas mais técnicas, como MICRO e SIGMOD.

5) *Tendências Futuras:* Muitos eventos internacionais localizados nos primeiro e segundo quadrantes encontram-se próximos ao eixo  $x$ . Isso indica que, ao longo do tempo, podem migrar para um cenário de declínio na participação feminina. Essa possibilidade se torna mais preocupante em eventos que já apresentam uma baixa presença feminina e que podem sofrer uma piora sem a implementação de iniciativas ou políticas voltadas à inclusão. Entre os eventos internacionais que merecem atenção nesse sentido, destacam-se SIGMM, DAC, IJCAI, ICCAD e SIGSPATIAL, cujos padrões de crescimento indicam riscos de queda na participação feminina caso não sejam adotadas estratégias para entender e reverter essa tendência.

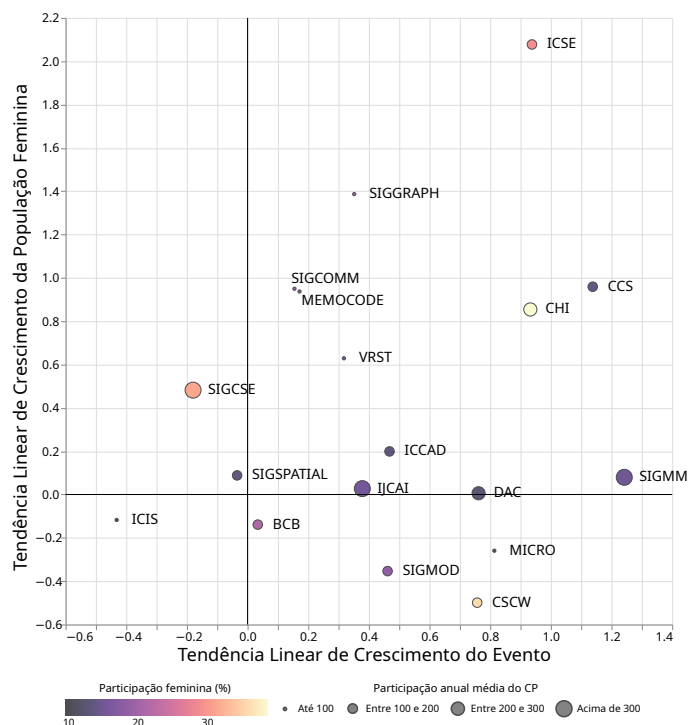


Figura 10. Tendência linear de crescimento da participação feminina nos comitês de programa de eventos internacionais pelo crescimento total desses comitês.

#### D. Comparativo dos Eventos Nacionais e Internacionais por Temática

A Tabela 11 apresenta diagramas de caixa (*boxplot*) que evidenciam as semelhanças e diferenças na participação feminina entre eventos nacionais e internacionais, agrupados por temática. Observa-se que, nos eventos nacionais, os primeiro e terceiro quartis estão mais próximos entre si, indicando uma menor dispersão nos índices de participação feminina. Já nos eventos internacionais, os quartis são mais afastados, o que sugere uma maior variabilidade ao longo dos anos. Esse comportamento pode estar associado ao fato de que os comitês internacionais costumam ser mais amplos e compostos por pesquisadoras de diversos países, o que pode influenciar a oscilação na representatividade feminina.

As medianas das participações femininas nos eventos nacionais tendem a ser superiores às dos eventos internacionais, reforçando a tendência de maior inclusão feminina no Brasil. Além disso, algumas temáticas que apresentam baixa participação feminina no cenário internacional registram índices relativamente mais altos no Brasil. Exemplos incluem Sistemas de Informação (SBSI e ICIS), Bancos de Dados (SBBD e SIGMOD), Geoinformática (GEOINFO e SIGSPATIAL) e Inteligência Artificial (BRACIS e IJCAI). Isso sugere que algumas subáreas tradicionalmente dominadas por homens em nível global podem estar em um estágio mais avançado de inclusão feminina no Brasil. Outra possibilidade é que, no contexto nacional, a abordagem temática desses eventos seja mais atrativa para mulheres em comparação com seus

equivalentes internacionais.

Por outro lado, verifica-se que, em diversas temáticas, os índices de participação feminina são semelhantes nos contextos nacional e internacional. Áreas com baixa representatividade feminina em eventos brasileiros também apresentam baixos índices nos eventos internacionais. Esse padrão é evidente em áreas como Circuitos e CAD (SBCCI e ICCAD), Sistemas Computacionais (SBESC e DAC), Computação Gráfica (SIBGRAPI e SIGGRAPH), Segurança (SBSEg e CCS), Arquitetura e Desempenho (WSCAD e MICRO), Redes de Computadores (SBRC e SIGCOMM) e Realidade Virtual (SVR e VRST). Para esses eventos com baixa participação feminina no Brasil, a participação feminina no âmbito internacional tende a ser maior, o que indica um maior interesse por parte das mulheres pela área ou temática dos eventos.

Já as áreas com maior participação feminina estão concentradas em temáticas como Interação Humano-Computador (IHC e CHI), Educação (SBIE e SIGCSE) e Sistemas Colaborativos (SBSC e CSCW). Essa tendência sugere que eventos voltados para pesquisa interdisciplinar, que combinam Computação com áreas como Ciências Sociais e Educação, tendem a atrair maior diversidade de gênero, tanto no Brasil quanto no cenário internacional.

## V. CONCLUSÕES E TRABALHOS FUTUROS

Esta pesquisa comparou a participação feminina nos comitês de programas de eventos internacionais em Computação com dados de eventos nacionais obtidos em MSI 1. Observou-se que, de modo geral, as mulheres representam 20,05% do total de participantes nos eventos internacionais, contra 25,09% no cenário brasileiro. Além disso, as participações anuais médias também se mantiveram em patamares ligeiramente mais altos nos eventos nacionais (26,28%) em comparação aos internacionais (19,23%). Esse resultado reforça a persistência de um cenário de sub-representação feminina, embora os eventos brasileiros apresentem índices ligeiramente mais altos em todas as métricas analisadas.

A análise de séries temporais, reforçada por regressões lineares e mapas de calor, evidenciou que alguns eventos internacionais, como CHI, ICSE, CCS, SIGCOMM, MEMOCODE e SIGGRAPH, apresentaram crescimento na participação feminina, enquanto outros, a exemplo de CSCW, BCB, MICRO e ICIS, registraram declínio, sobretudo nos últimos anos avaliados. Esse mesmo comportamento também foi observado em eventos nacionais estudados em MSI 1: há conferências com aumento constante da presença feminina (IHC, GEOINFO, SBSC) e outras em que esse índice decaiu (SBIE, SBSI, BRACIS).

Assim como nos eventos nacionais, a participação feminina varia de acordo com a temática dos eventos internacionais. Conferências focadas em áreas como interação humano-computador, colaboração e educação em computação (por exemplo, CHI, CSCW, SIGCSE) destacam-se por apresentar maiores índices de participação feminina, enquanto eventos mais técnicos, como CCS, SIGCOMM, MICRO, DAC e ICCAD, registram percentuais menores. Além disso, eventos com

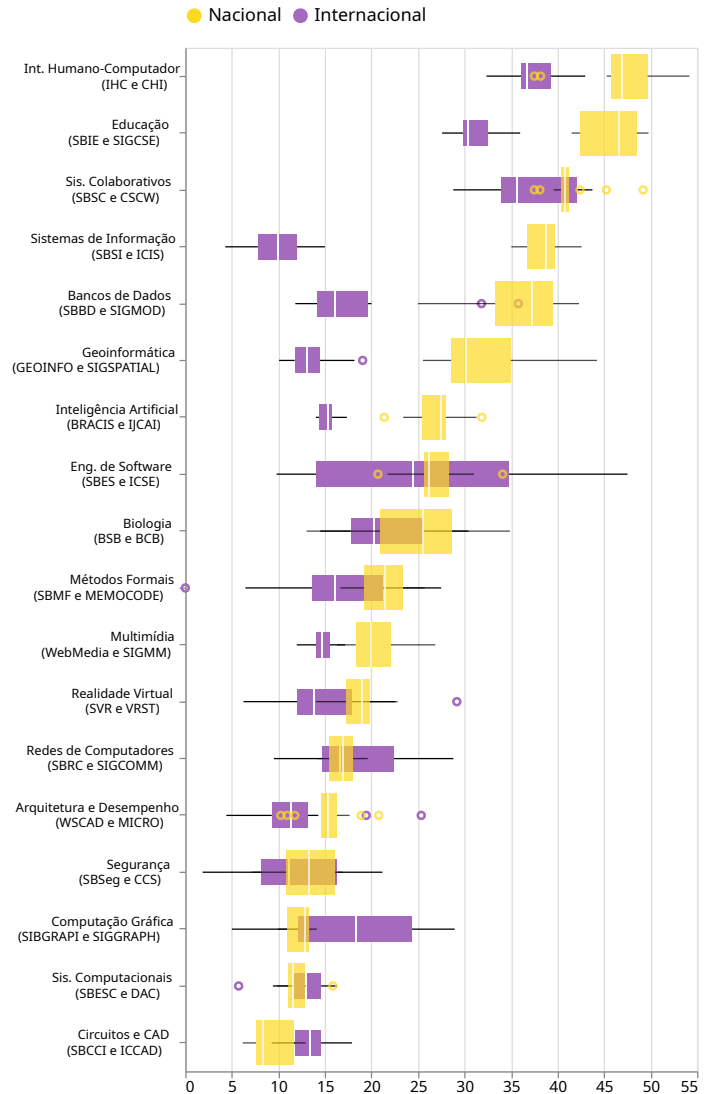


Figura 11. Diagramas de caixa (*boxplot*) dos eventos nacionais e internacionais por temática dos comitês, ordenados da maior para menor média de participação feminina dos eventos nacionais

grandes CPs podem ter tanto alta quanto baixa participação de mulheres, sugerindo que fatores culturais, estruturais e de políticas de inclusão podem exercer um papel mais significativo do que a dimensão do comitê.

A comparação temática dos eventos internacionais e nacionais por meio de *boxplots* confirmou o padrão destacado anteriormente. Áreas como interação humano-computador, educação em computação e sistemas colaborativos (por exemplo, CHI, CSCW, SIGCSE internacionalmente; IHC, SBSC, SBIE nacionalmente) apresentam maior participação feminina. Em contrapartida, conferências com foco em subáreas mais técnicas (arquitetura de computadores, CAD, circuitos integrados e segurança) exibem índices menores tanto no cenário internacional (MICRO, DAC, ICCAD, CCS) quanto no nacional (SBESC, SBCCI, WSCAD, SBSEg). Isso indica a influência de fatores culturais, estruturais e temáticos na

atratividade das conferências para pesquisadoras.

Em relação aos núcleos dos comitês (as 25% maiores participações), verificou-se que a porcentagem de mulheres nesses grupos tende a acompanhar o índice geral de cada evento, porém com oscilações mais acentuadas. Observou-se ainda que eventos com menor presença feminina no total do CP são justamente aqueles que mantêm baixa representatividade também em seus núcleos. Outro ponto importante foi a evolução temporal, na qual apenas uma edição no âmbito nacional (IHC em 2018) ultrapassou 50% de participação feminina, destacando-se como um caso isolado, já que a maioria dos eventos nacionais e internacionais ainda se encontra abaixo e longe desse patamar.

O trabalho desenvolvido evidenciou um cenário que é ainda muito desigual, mas mostra algumas tendências positivas de crescimento da participação feminina em certos eventos. No entanto, a presença de quedas em vários eventos, mesmo em alguns tradicionalmente mais inclusivos, ressalta a necessidade de políticas contínuas e efetivas de incentivo à diversidade de gênero.

Como trabalhos futuros, sugere-se a adoção de modelos estatísticos que capturem variações não lineares, o que pode oferecer maior clareza sobre tendências de crescimento, estagnação ou declínio na participação feminina, especialmente ao considerar séries de dados mais antigas e mais recentes do que as analisadas. Além disso, coletar depoimentos por meio de entrevistas ou questionários pode ajudar a compreender os desafios enfrentados pelas mulheres nos comitês, contribuindo para direcionar políticas de inclusão mais assertivas. Por fim, avaliar a duração da permanência das mulheres nos CPs, suas funções específicas e o impacto de suas contribuições pode fornecer uma visão mais completa sobre a dinâmica de participação feminina.

Essas ações podem fortalecer a compreensão dos fatores que favorecem ou inibem a diversidade de gênero em eventos científicos na área de Computação, tanto em âmbito nacional quanto internacional, contribuindo efetivamente para reduzir as disparidades de gênero e a promoção de ambientes mais equitativos, inclusivos e inovadores.

## REFERÊNCIAS

- [1] Sociedade Brasileira de Computação, “Sobre a sbc,” 2022. [Acessado em: Jan. 2025].
- [2] Association for Computing Machinery, “About acm,” 2024. [Acessado em: Jan. 2025].
- [3] Institute of Electrical and Electronics Engineers, “About ieee,” 2024. [Acessado em: Jan. 2025].
- [4] J. Lorber, *The Social Construction of Gender*. SAGE Publications, 1991.
- [5] A. Fausto-Sterling, “The five sexes: Why male and female are not enough,” *SCIENCES*, pp. 20–254, Mar. 1993.
- [6] D. Arruda, F. Bezerra, V. Neris, P. Toro, and J. Wainer, “Brazilian computer science research: Gender and regional distributions,” *Scientometrics*, vol. 79, pp. 651–665, 06 2009.
- [7] Global Education Monitoring Report Team, *Global Education Monitoring Report 2024, Gender Report: Technology on Her Terms*. Paris: UNESCO, 2024.
- [8] J. Margolis and A. Fisher, *Unlocking the Clubhouse: Women in Computing*. MIT Press, revised ed., 2003.
- [9] T. J. Misa, “Dynamics of gender bias in computing,” *Commun. ACM*, vol. 64, p. 76–83, May 2021.

- [10] L. L. Wang, G. Stanovsky, L. Weihs, and O. Etzioni, “Gender trends in computer science authorship,” *Commun. ACM*, vol. 64, p. 78–84, Feb. 2021.
- [11] H. Tives and R. Pereira, “Diagnóstico sobre a presença das mulheres na distribuição de prêmios na área da computação,” in *Anais do XVI Women in Information Technology*, (Porto Alegre, RS, Brasil), pp. 45–55, SBC, 2022.
- [12] N. D. Pizzol, E. Barbosa, and S. Musse, “Gender representation in Brazilian computer science conferences,” in *Anais do XVI Women in Information Technology*, (Porto Alegre, RS, Brasil), pp. 67–76, SBC, 2022.
- [13] M. Moraes, M. Moraes, and U. Biondo, “Ciência livre: contribuições das mulheres para a computação,” in *Anais do XIX Congresso Latino-Americano de Software Livre e Tecnologias Abertas*, (Porto Alegre, RS, Brasil), pp. 153–156, SBC, 2022.
- [14] D. Aksnes, F. Piro, and K. Rørstad, “Gender gaps in international research collaboration: a bibliometric approach,” *Scientometrics*, vol. 120, 06 2019.
- [15] M. Soares, A. Silva, Érica Carneiro, L. Varjolo, and G. Guedes, “Análise do gênero de pesquisadores em eventos do congresso da sociedade brasileira de computação,” in *Anais do XVII Women in Information Technology*, (Porto Alegre, RS, Brasil), pp. 57–67, SBC, 2023.
- [16] N. Laberge, K. H. Wapman, A. C. Morgan, S. Zhang, D. B. Larremore, and A. Clauset, “Subfield prestige and gender inequality among u.s. computing faculty,” *Commun. ACM*, vol. 65, p. 46–55, Nov. 2022.
- [17] A. L. Lorens, J. Botelho, A. F. Moura, B. Duarte, and M. Moro, “Participação feminina em comitês de programa de simpósios da computação,” in *Anais do XIV Women in Information Technology*, (Porto Alegre, RS, Brasil), pp. 90–99, SBC, 2020.
- [18] B. Duarte, A. Moura, and M. Moro, “Mulheres na computação: Análises por sub-Áreas,” in *Anais do XIII Women in Information Technology*, (Porto Alegre, RS, Brasil), pp. 174–178, SBC, 2019.
- [19] F. Karimi, C. Wagner, F. Lemmerich, M. Jadidi, and M. Strohmaier, “Inferring gender from names on the web: A comparative evaluation of gender detection methods,” in *Proceedings of the 25th International Conference Companion on World Wide Web*, p. 53–54, 2016.
- [20] J. Burger, J. Henderson, G. Kim, and G. Zarrella, “Discriminating gender on twitter,” in *EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1301–1309, 01 2011.
- [21] L. Santamaría and H. Mihaljević, “Comparison and benchmark of name-to-gender inference services,” *PeerJ Computer Science*, vol. 4, p. e156, 2018.
- [22] A. Araújo and M. M. Moro, “Mulheres digitais: Desafios (a serem) vencidos na academia para equidade de fato,” *Revista de Educação Pública*, vol. 31, no. jan/dez, pp. 1–20, 2022.
- [23] M. Moro, L. Salgado, and A. Araujo, “Wit 18 anos: A evolução de seus comitês de programa,” in *Anais do XVIII Women in Information Technology*, (Porto Alegre, RS, Brasil), pp. 207–218, SBC, 2024.
- [24] L. H. C. Lima, G. Penha, L. M. de Alencar Rocha, M. M. Moro, A. P. C. da Silva, A. H. F. Laender, and J. P. M. de Oliveira, “The collaboration network of the Brazilian symposium on databases - 30 editions of history,” *J. Braz. Comput. Soc.*, vol. 23, no. 1, pp. 10:1–10:16, 2017.