

A scientific research overview at the Brazilian Computer Society Congress: a feminine perspective

Thalia Santos de Santana¹ and Adriano Honorato Braga¹

¹ Instituto Federal Goiano – Campus Ceres, Rodovia GO 154 Km 03, Ceres, Goiás, Brazil

Abstract

Disproportionality of gender in Information Technology field is well known. Women are a minority in the academic world, even in a scenario where there are efforts to value female models of success in computing. Reflecting on the historical bias, it can be seen that this fact causes impact on the scientific research development in this field. Scientometrics is the field of knowledge that recognizes indices of intellectual production, besides contributing to finding gaps and research trends. In this context, this fact can be an important tool to characterize gender diversity among scientists in Computer Science Research (CSR). This paper aimed to study the female presence at the Brazilian Computer Society Congress events (CSBC, Brazilian acronym), which is considered one of the main concerns in Latin America. Information was gathered by data collection, covering the period from 2017 to 2019, showing the following variables: name of the authors; title of the paper; and year of publication and events. The database results indicate male predominance in the audiences at CSBC events. However, considering works published, women stand out in almost 60% of them, even constituting only 31% of the relevant quantitative role of authors in development and scientific progress of computing.

Keywords

Computer Science Research, CSBC, Scientometrics, Scientific Production, Women in Computer Science

1. Introduction

In scientific records, usually inventions and other outstanding productions have omitted the female presence as promoters and collaborators of discoveries. Practices have been developed suppressing the role of women [1], which endorses the historical erasure of female figures, even with significant contributions in different field of knowledge, such as chemistry, physics, biological sciences, and even technology.

Taking scientific research in Computer Science (CS) in consideration, it is proven that only 21% of the total researchers in the field indexed by the Digital Bibliography & Library Project (DBLP) Computer Science Bibliography are women [2]. Brazil also follows this tendency with a noticeable small proportion of women in the scientific community in CS, even though in the beginning, more women were attracted to computer courses at universities [3].

However, indices of this gender imbalance are influenced by world historical elements, often reflected by the invisibility of women in Computer science. Pioneers as Ada Lovelace, the first person to build an algorithm and considered the first programmer; Grace Hopper, who had developed the first compiler and collaborated to the creation of the programming language Common Business Oriented Language (COBOL); and the girls of Electronic Numerical Integrator and Computer (ENIAC), programmers of one of the first digital/electronic computers in the world are examples of female presence in computing that had contributed to its conception as a science [4], and they are still not recognized by current generations, despite their valuable contributions [5].

Proceedings XIII Congress of Latin American Women in Computing 2021, October 25–29, 2021, San José, Costa Rica

EMAIL: thaliassantana15@gmail.com (T. S. Santana); adriano.braga@ifgoiano.edu.br (A. H. Braga)

ORCID: 0000-0003-1899-4819 (T. S. Santana); 0000-0002-1729-4039 (A. H. Braga)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

At the same time, the percentage of female students registered in technology-related courses has decreased over the years, increasing inequality in academic education and thus in scientific research. The most recent data gathered by the Brazilian Computer Society (SBC, Brazilian acronym) according to information from the National Institute of Educational Studies and Research Anísio Teixeira (INEP, Brazilian acronym) reports only 13.96% female students in courses related to computing at universities in Brazil, a total of 49,734 women and 306,392 men enrolled in 2019 [6].

Scientific publications have been studied to understand these issues related to gender inequality, in addition to promoting strategies for women's entry, permanence, and motivation [2]. Regardless women facing many barriers to promote female success at work sectors with masculine stereotypes, they show good results on fields such as technology [7], also shown in their scientific production.

However, there is still female underrepresentation created by direct discrimination and through subconscious behaviors that perpetuate stigmas in the field [8]. The same fact occurs in proportion to the increase in the scale of scientific prestige, defined as “leaking pipeline” [9] within Science, Technology, Engineering, and Mathematics (STEM) and in the focus of research in Computer Science Research (CSR), terminology used to differentiate studies focused on scientific research in the field.

Studies from the Organization of Ibero-American States (OEI) show that 72% of Brazilian papers are produced by women among all fields of scientific production [10]. However, even signing most of the them, the total number of researchers, about 49%, is still less than the male researchers. The assessment was based on the years from 2014 to 2017, analyzing the productions in which female researchers presented themselves either as first author or as co-author. Although this report highlights Brazil as having the highest participation of women when compared to the eighteen countries surveyed, field such as engineering show only 32% of women in publications, whereas there are 56% in Medicine.

Sciencimetry and Bibliometrics are indicators inherent to academic publications, and they are usually applied as a methodology for extracting measurable data in the literature for statistical analysis of scientific papers [2]. Scientometrics is defined as the study of quantitative aspects of science [11], enabling the assessment and quantification of scientific progress. In addition, it can provide subsidies in Academic field and society in general regarding topics of interest, contributing to find gaps, including regarding to gender diversity.

Thus, this paper had studied the scenario of Brazilian scientific productions published in the CS field and had carried out a quantitative analysis of the data registered under a gender perspective. This research had aimed to investigate the existence of possible gaps in events regarding the presence of female and male authors, in addition to collaborating with Brazilian studies on the scenario of gender diversity in Academic fields, presenting comparisons among events from one of the largest scientific conferences in this field. Data were gathered on the Web through publications from the proceedings of three editions of the Brazilian Computer Society Congress (CSBC, Brazilian acronym) (from 2017 to 2019).

This study is structured as follows: Section 2 reports the related works; Section 3 reports the methodology used to gather the database and data analysis; Section 4 discusses the main results found; and finally, Section 5 presents the final considerations.

2. Related Papers

This section describes studies in the literature related to the CSR theme, checking the presence of women in computing context. The Agarwal et al. work [2] found out the gender imbalance by an exploratory analysis of data within the DBLP Computer Science Bibliography indexer, a repository of great relevance internationally in computing [2]. According to these authors, the CSR field is dominated by men, including subdomains, such as Software Engineering. In addition, Brazil has a larger number of female contributors to high-level conferences (37%) when compared to Japan and Germany, making the gender imbalance non-specific to developing countries.

When evaluating the curriculum Lattes of professors post graduated in CS in Brazil, Arruda et al. [9] ascertained the existence of 201 females and 685 males among the researchers in November 2006, totaling 23% women, value really close to that evaluated in this study [2]. Furthermore, it was statistically checked by research themes that there is a larger proportion of women in field which human is a component in common, such as artificial intelligence, computers in education, and man-machine

interface. However, the women move away from those themes of a technological component, such as computer and hardware architecture, as well as networks and distributed systems. In fact, in the more “feminine” research fields, women account for only 29.7% in this field and are more productive than men, with a quantity of 5.3 journal publications by man researchers and 6.0 by female researchers.

Regarding the CS subfields, Banshal et al. [12], Banshal et al. [13], Effendy and Yap [14], Noosrikong, Ngamsuriyaroj, and Ayudhya [15], and Yang et al. [16] investigated data from scientific productions with evaluations regarding identifying the most productive publication field in CS. However, it was ascertained that none of these studies covered gender connections when discussing about the inconsistency and under-representation of minority scientists, in addition to what Agarwal et al. had already mentioned in computing [2].

Consequently, Gallivan and Benbunan-Fich [17], and Vela, Caceres, and Cavero [18] studied the productivity among genders in Information Systems and Software Engineering, respectively. Gurer and Camp [19] carried out a systematic literature review, using the online database from the Association for Computing Machinery for Women (ACM-W). Cohoon, Nigai, and Kaye [20] had come close to previous authors when working within the ACM-W Digital Library, whereas Cavero et al. [21] evaluated DBLP, another large repository of publications on computing.

Thus, although there are many studies that assess specific fields and events and/or CSR in general in certain countries, studies regarding women authors in scientific works in Brazil focusing on different events and linked to a single conference were not found. The CSBC is a well-known congress, and it counts with a relevant segment of the active CS community; nevertheless, no comprehensive studies of the events linked to this conference have been made up to now, especially since it is one of the greatest achievements of the main Brazilian civil institution representative in this field [22].

3. Material and Methods

This study is exploratory in nature with a quantitative approach. This study had focused on information from the CSBC, aiming to diagnose the scientific production scenario in CS in Brazil. Considering that the event promoter (SBC) is the largest computer society in South America [23], it behaves as the main organizer and supporter of Brazilian technical-scientific events in computing. As an example, one of the most well-known fact is the CSBC itself, which has been carried out since 1980, providing opportunities for experience exchanges in the scientific, academic, and professional scope in CS with national and international coverage [24].

CSBC is the largest congress in academic field in a computing society in Latin America, expressing the research scenario in the country in a representative fashion, in addition to having great reputation within scientific communities, as well as institutions and development agencies. The event structure is composed of diverse base and satellite events. Currently, there are nine events regularly scheduled and other random events, which take place within the conference itself, after the responsible committee register and select the proposals.

Thus, it was decided to gather the works published on the CSBC website, as an input for scientometric analysis, considering the importance of the conference data as the object of this study. It is reassured that this approach had been necessary, considering that SBC does not have a privacy policy that allows exporting its conference data from its enrolling systems, such as the Journal and Event Management System (JEMS).

Regarding the time frame, the last three full years (2017 to 2019) available on the website at the time of data collection (2019) were taking for study by data scraping procedures, commonly known as Web Scraping. This technique allows the data to be collected automatically from web pages without the need to copy them manually. Thus, non-structured information, usually in HyperText Markup Language (HTML), are changed to structured data for analysis on databases or electronic spreadsheets [25].

Considering the information available at the CSBC editions electronic addresses, data were scraped and collected by the Python programming language, version 2.7.16. Therefore, the page structure was analyzed by inspection tools from the web browsers. After studying the blocks that contained the event information, it was possible to retrieve data from the internet that expresses CSBC publications in recent years, applying a script by Pandas, BeautifulSoup and URLOpen language libraries.

The BeautifulSoup library is specifically used for data collection by Web Scraping. Considering the web information format, editions prior to 2017 were not included in the scraping, as such data were not in HTML format or there were few events with the mentioned works, making automatic collection impossible.

A scraper-type script was necessary for each page because CSBC makes the congress proceedings available on different pages every year. On the scripts, the electronic addresses of each event and respective works were informed, as well as the object HTML to be mapped, providing as a result a file on Comma-separated-values (CSV) format, containing the event and its edition, title, and male/female authors, then the files were merged and, consequently, names and acronyms of the events were standardized.

Regarding the selection of the events, those works that did not have registration and acceptance were eliminated in the first instance, such as the University Computing Seminar (SECOMU, Brazilian acronym), as well as those that resulted in items other than papers, such as the Computer Update Conference (JAI, Brazilian acronym), which publish a book at the end of its event. Thus, eighteen events with some type of scientific production on CSBC, whether short papers or full papers, were returned in 2017, nineteen events in 2018, and sixteen events in 2019. In addition, the data collected was cleaned, eliminating records that contained opening prefaces, limiting only to specific publications and their male or female authors.

After cleaning, there were 919 works published in 2017, 2018, and 2019, that are, 308, 348, and 263 papers, respectively. Searching for analysis throughout the time, only the congresses in common were kept within the years analyzed. Finally, 14 events were listed (Table 1) and 759 works published.

Table 1
CSBC events selected by acronym, name, and type

Acronymus	Event name	Type
BraSNAM	Brazilian Workshop on Social Network Analysis and Mining	Satellite
BreSci	Brazilian e-Science Workshop	Satellite
CTD	SBC Thesis and Dissertation Contest (Concurso de Teses e Dissertações da SBC)	Base
CTIC	SBC Scientific Initiation Papers Contest (Concurso de Trabalhos de Iniciação Científica da SBC)	Base
ENCompIF	National Computing Meeting of Federal Institutes (Encontro Nacional de Computação dos Institutos Federais)	Satellite
ETC	Computation Theory Meeting (Encontro de Teoria da Computação)	Satellite
SBCUP	Brazilian Symposium on Ubiquitous and Pervasive Computing (Simpósio Brasileiro de Computação Ubíqua e Pervasiva)	Satellite
SEMISH	Integrated Software and Hardware Seminar (Seminário Integrado de Software e Hardware)	Base
WCAMA	Workshop on Computing Applied to Management of Environment and Natural Resources (Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais)	Satellite
WEI	Computer Education Workshop (Workshop sobre Educação em Computação)	Base
WIT	Women in Information Technology	Base
WPerformance	Workshop on Computer and Communication Systems Performance (Workshop em Desempenho de Sistemas Computacionais e de Comunicação)	Satellite
WPIETF	Pre-Internet Engineering Task Force (IETF) Workshop (Workshop Pré-IETF)	Satellite
WTrans	Systems Transparency Workshop (Workshop de Transparência em Sistemas)	Satellite

It is noted that none of the events provides information regarding the gender of the works' authors. Therefore, it was necessary to use an automatic classification library to identify the gender of the individuals surveyed and, consequently, provide analysis regarding female participation in the CSBC. After tests with diverse systems that perform such prediction, the genderBR package version 3.4.4 was chosen for this paper; it was developed considering public information from the 2010 Census of the Brazilian Institute of Geography and Statistics (IBGE, Brazilian acronym) [26]. Most publications in Brazil are authored by Brazilians, which is an important factor to be considered for their classification, which is done using the personal name. Thus, genderBR is connected to the R, an environment to implement statistical functions widely used to create views and perform data analysis [27]. At the time of this study, this package was not available for using along with the Python programming language.

It is important to highlight that SBC has its own paper submission model, which is widespread by all events that compose it. Unlike most conferences and even the Lattes platform, one of the most used by Brazilian researchers, the names of the authors are written in full without abbreviations instead of the last surname and initial letter of the first name, which contributed to identify the authors' gender.

However, even if there would be an SBC standard for identifying authors in the publication of papers, abbreviated names were still found. Thus, these identifications had been searched manually online and the complete nomenclature was found through other works already published, or even through networks of researchers, such as Lattes, Research Gate, and/or citations by a person from Google Scholar. Having the base prepared, the classification was done along with the use of the function clipboard, which takes the data, creating a data frame, a type of matrix with the possibility of different data type values among them.

The result returns to ranking into "Female", "Male" or NA (Unidentified). The genderBR considers the names with proportions below 90% as NA, generally when the name can be used for both genders and even for those that were not found in the IBGE database. About 4.4% of individuals were not identified and underwent a manual classification. Finally, 2,723 records comprised the total set of authors.

A new spot was created for storing corrected authors' names, and this process is called disambiguation, as described by Duarte, Moura, and Moro [28]. This was done by grouping records of authors, which referred to the same individual (with or without a middle surname), keeping the name as a rule in the most complete form, also performed manually.

After such procedures, the data were organized according to relational modeling of an Entity-Relationship Diagram (DER, Brazilian acronym) in a MySQL database version 8.0.16, through the Database Management System (DBMS) MySQL Workbench version 8.0.15. The data underwent analysis through queries in Structured Query Language (SQL), a standard search format for databases. Later, the Python 3.7.3 programming language was used within the Google Collaboratory platform, making possible to study collaboratively and in the cloud. Numpy, Scipy, and Matplotlib libraries were imported, being possible to create visualizations and statistical analyses.

Taking this panorama into account, the following research questions (RQ) were listed:

- RQ1 - What is the gender distribution in the period studied?
- RQ2 - How does female participation vary over the years in relation to the event?
- RQ3 - How does the number of works accepted per event and year vary?
- RQ4 - How is the difference between female authors and male authors configured in order of authorship?
- RQ5 - How many women are among the three most productive authors?
- RQ6 - Considering the authors, what is the female presence among scientific productions? and
- RQ7 - What are the different trends in events in relation to articles published by women?

Therefore, among the selected works and events, the following variables were found: authors' name and gender; title; edition and year of conferences; and authorship order. This study focused on articles published in the event proceedings, and the RQs guided this work, which will be answered in the next sessions.

4. Results and Discussion

Among the 14 CSBC events from the total of records by authorship, 1,878 researchers are male (69%) and 845 are female (31%). The three people with a larger number of publications are male, being the man in the first place with twelve works, and the second and third men were in the ranking tied with eleven works in the three years studied. Assessing the number of authors by gender in relation to the order of authorship, 220 women are first authors, and 539 men are main authors. The disproportionality condition is repeated at all levels of co-authorship (Figure 1) checked in the publications, since the number of male authors is greater than the female authors in any of the order of authors.

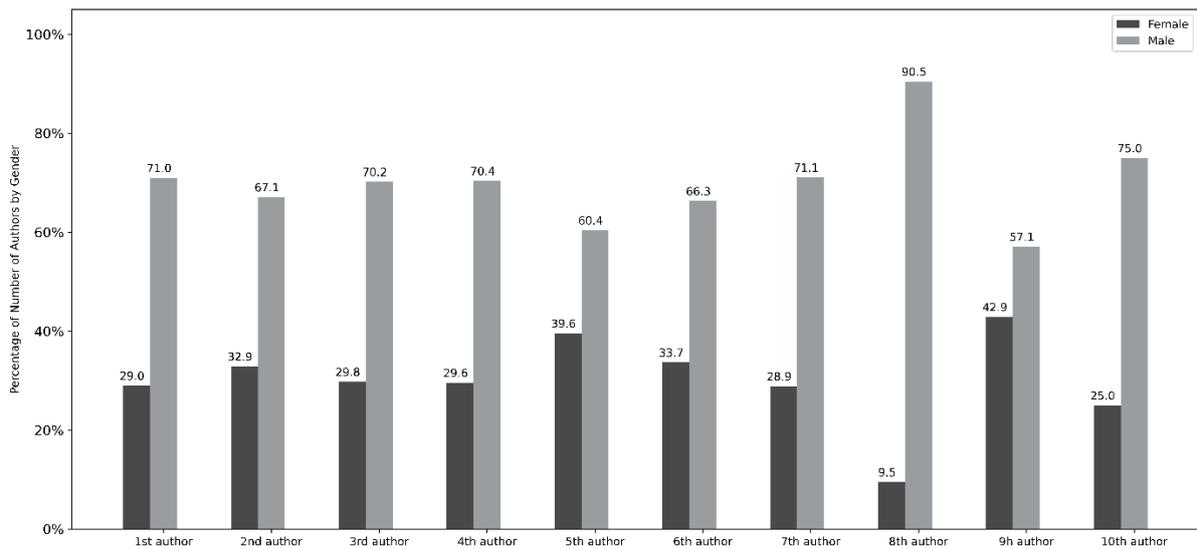


Figure 1: Bar graph regarding the gender of authors and authorship order

As for the events, the Computer Education Workshop (WEI, Brazilian acronym) is the one with the largest number of works with 122 publications, next, Computation Theory Meeting (ETC, Brazilian acronym) with 94 productions, and, lastly, Women in Information Technology (WIT) with 87 works. It is noteworthy that the number can be considered as expected, in the case of the WEI, as it is a consolidated event, occurring for almost thirty editions and is included in the Qualis/Capes B3 index (journal classification by the Brazilian Coordination for the Improvement of Higher Education Personnel), in addition to having a comprehensive field. However, the events following WEI are new; the ETC in 2019 was in its 4th edition and is still only a satellite event. WIT, in its 13th edition in 2019, despite being a base event, is the youngest among those in this category, whose events have occurred for almost half a century, such as Integrated Software and Hardware Seminar (SEMISH, Brazilian acronym), demonstrating that the scientific community has commonly participated in new events, which meet the research theme, even if not yet recognized by the Qualis/Capes extract. Studying year by year, the WEI stands out in the years 2018 and 2019, with the greatest number of publications, as seen in Figure 2.

Among all the events promoted at CSBC, WIT is the one with the largest number of women as main authors with eighty women and seven men. This event aims exactly to address the theme of women in computing, and the high rate of female researchers is expected. However, in the Thesis and Dissertation Competition (CTD, Brazilian acronym) event, in which computer theses and dissertations are awarded, only one woman was found as the first author, what reflexes the low percentage of masters and doctors in CS [29].

Still within the scope of authorship, in 2017, 2018, and 2019, there were, respectively, 58, 89, and 73 works by women and 172, 198, and 169 by men, corresponding to a percentage of 25.2%, 31%, and 30.2% of female authorship compared to male authorship. Taking the total number of works studied into account, there were 449 publications by women among the authors, what represents a percentage of about 59.1%. Thus, even in a smaller number (31% of the total number of authors), there are female

scientists in more than half of the works (among the 759 works studied) published in CSBC. There were 699 works by male scientists.

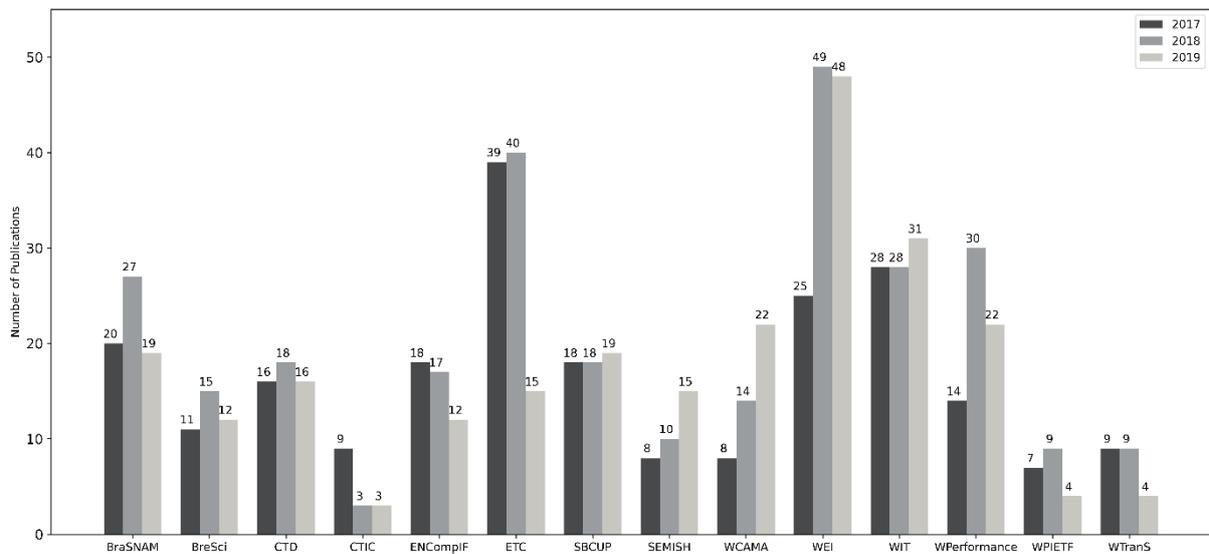


Figure 2: Bar graph of the number of publications by year and event

In addition to analyzing the total and percentage by gender, the female/male ratio (F/M) of people was calculated to analyze the proportion of authorship by gender of each event (Table 2), as done by Ribeiro et al. [22]. In absolute number and proportion, the “most female” events were WIT, WEI and BreSci. All events with a multidisciplinary approach deal, in general, with issues involving women in computing with incentive projects, informatics in education, and e-Science aiming to computational support applications for the processes of scientific experimentation. With a smaller proportion, CTD, WPIETF, and WPerformance had lower F/M ratios, which are, in order, related to final postgraduate works, Internet standards, and computer system performances.

Table 2
Events, number of authors by gender and F/M ratio

Event	Geral	Female (F)	Male (M)	%F	%M	Ratio F/M
WIT	339	282	57	83.2	16.8	4.9474
WEI	452	158	294	35.0	65.0	0.5374
BreSci	153	53	100	34.6	65.4	0.5300
WTranS	66	22	44	33.3	66.7	0.5000
BraSNAM	241	66	175	27.4	72.6	0.3771
ENCompIF	185	49	136	26.5	73.5	0.3603
WCAMA	192	49	143	25.5	74.5	0.3427
SEMISH	133	33	100	24.8	75.2	0.3300
CTIC	44	10	34	22.7	77.3	0.2941
ETC	260	45	215	17.3	82.7	0.2093
SBCUP	232	34	198	14.7	85.3	0.1717
CTD	112	13	99	11.6	88.4	0.1313
WPIETF	71	8	63	11.3	88.7	0.1270
WPerformance	243	23	220	9.5	90.5	0.1045

The single-factor ANOVA method was used to check whether or not there was a statistically significant difference among the data presented in Table II, which allows comparison among means of different populations, being carried out among the five largest groups (WEI, BreSci, WTranS, BraSNAM, and ENCompIF) and a smaller proportion of authors (ETC, SBCUP, CTD, WPIETF, and WPerformance). Even the WIT having the largest ratio among the female participants, it was necessary to disregard it to have a more realistic view among the events and their fields. Thus, f-ratio=53.809631

and $p\text{-value} = 0.000081$, significant for $p < 0.05$, were found, showing that there is a significant difference (alternative hypothesis) in the distribution by gender in CSBC events in relation to the female/male proportion, rejecting the null hypothesis of equal means.

More specifically, the percentage of female authors in each of the fourteen events in the three years studied was also verified (Figure 3). From 2017 to 2018, a considerable increase in female scientists was found, and therefore there was a decrease in female authors percentage in nine events in 2019. In five events, in which the percentage was higher compared to the previous year (2018), such as CTD, ETC, WCAMA, WIT and WTranS, the rate of increase was minimal, which was, respectively, 0.3%, 1.2 %, 3.7%, 4.1%, and 5.3%.

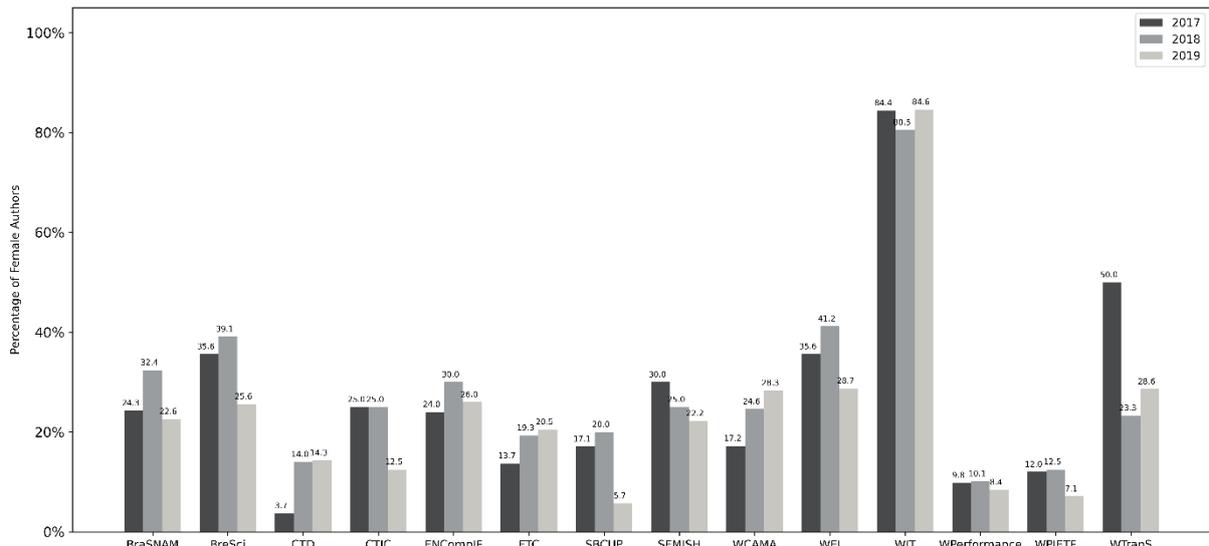


Figure 3: Bar graph of percentage of authors by year and event

Considering the data presented, it is worth highlighting the importance of the activities that promote gender equity, which can mitigate this gap in scientific research in computing, aiming to encourage young women to continue in research. Digital Girls Program [30] is an example of attracting the interest of girls and women to the technology field endorsed by the SBC; since its introduction in 2011, it has partnered projects spread across many Brazilian States with the view of being a reference in Latin America on gender equity issues in technological careers. From this perspective, it is understood that the Digital Girls Program supports minimizing this imbalance in the academic field, by attracting more girls who, consequently, will be immersed in the field of computing and, therefore, will have contact with scientific production at some point.

Regarding the permanence of women researchers, actions to support women in the scientific career are mentioned in works such as Parent in Science, which has discussed issues related to maternity and science in Brazil since 2017, and how this fact has a direct impact on indicators of scientific production. This movement (Parent in Science) has also been providing financial aid to help mother students in the final stages of postgraduate courses [31] through the “Amanhã” (Tomorrow) program, launched in 2020.

In addition, recently in 2020, the National Council for Scientific and Technological Development (CNPq, Brazilian acronym), a Brazilian body responsible for fostering scientific research and development, updated the curriculum available through Lattes platform to allow the registration of maternity leave periods for scientists. The discussions fostered in this panorama have had an impact on the way academic productivity indices are gathered, in which some research funding announcements from agencies and universities started to take the period of maternity leave in consideration by extending the productivity evaluation period, since this productivity decreases due to this career break.

Complementarily, to ensure a greater presence of women in scientific events, there are actions promoting the presence of children, providing a children's space that allows them and their mothers to participate in scientific events. This demand was pointed out by the scientific community participating

in the SBC events and discussed by Machado et al. [32], highlighting the relevance of the event and its great importance for contact and collaboration networks among scientists.

5. Final Considerations

The references should be formatted according to the following guidelines: A paginated journal article [2], an enumerated journal article [3], a reference to an entire issue [4], a monograph (whole book) [5], a monograph/whole book in a series (see 2a in spec. document) [6], a divisible-book such as an anthology or compilation [7] followed by the same example, however we only output the series if the volume number is given [8] Scientometrics allows to perform more detailed analysis about scientific knowledge, which trends the path of science production and is shared in journals and conferences. It also contributes to the knowledge consolidation, which is considered an important mechanism to measure reflexes, especially regarding gender bias approached in this study, thus enabling individuals to understand and make an impact on the historical record of science.

This study contributes to map Brazilian research, focusing on gender in computing field, even if the available data are dependent on isolated and insufficient research for comprehensive views [22]. Thus, no study had been found that specifically investigates all events of this conference, especially under the perspective of minority groups within the context of CSR, although there are studies that specifically assess some of the events that occur at the CSBC under a scientometric perspective.

Considering the RQs above mentioned regarding RQ1, women are still in a smaller number (31%) in the item of scientific production, categorizing the research in CS as a male field. According to Elsevier's "Gender in the Global Research Landscape" report [33], there is only gender balance when women express from 40% to 60% of the total number of researchers. The authors also pointed out that, among 27 knowledge field, computing in Brazil accounts for 7% of scientists with only 2% women (almost 1/3 of the male gender), making the CSBC's projection slightly above from that analyzed internationally by Elsevier.

Furthermore, compared to recent studies by Ribeiro et al. [22], this value exceeds the 21.87% of female SBC members, and remains almost similar to that found by Duarte, Moura, and Moro [28], about 30% of program committee participants in Brazilian computing events. However, it almost doubles the percentage of students enrolled in this field in university education (about 14%), showing that, in short, even in smaller numbers, a considerable percentage of women has been present in publications and outlining scientific knowledge at CSBC, especially when compared to other indicators.

In relation to RQ2, the number of men and women per event is still unequal and there were no numbers of parity close to 50% for both genders in any of those events analyzed. The same fact occurs when seen year after year in a specific fashion, in which the index dropped sharply in most in the following year, despite a growing increase in the female percentage in 2018, making a gradual and linear growth impossible. In RQ3, therefore, eight events of the conference also decreased in number of papers approved for 2019; regarding to the order of authorship (RQ4), there is no parity (F/M ratio of 0.4081), considering the main authors.

Taking the authors of the most active works (RQ5) into account and evaluating the three authors with the highest number of articles published, it was found that none of the authors was women. The most productive author shared his publications to SBCUP (ubiquitous and pervasive computing) and SEMISH (general aspects of software and hardware), highlighting the publications of the first author, which is an event with little more than a decade of existence. The author with the largest number of publications is in fourth position with articles shared in four different events (BraSNAM, WIT, CTIC, and WEI), tied in quantitative with another male author. However, regarding the female presence in RQ6, it is clear that women are a representative part of most of the cataloged productions, even in a smaller number, showing the prominent role of women in the spreading of scientific knowledge.

For RQ7, women participating in CSBC tend to concentrate on events that have multidisciplinary characteristics. Informatics in Education is considered one of the areas typically defined as "feminine" in this country [28]. According to Duarte, Moura, and Moro [28], the interdisciplinary factor can be seen as favorable, considering the possibility of acting in diverse fields related to computing. Under a quantitative perspective, the events with fewer women decreased in 2018 compared to 2019, showing a trend of female distancing from fields that have the technological component and/or higher degrees

(in the case of CTD) as a factor, ensuring greater participation of women in events aimed at the female contingent or those related to human and social factors.

During the development of this study, some obstacles were overcome, such as the limited literature focusing on its subject, the lack of data availability from article submission platforms and/or any similar sources for the respective purpose, which makes it difficult to understand a more trustworthy perspective related to the field. Nevertheless, even though the proceedings of the CSBC were at different addresses when performing this work, later, the SBC itself launched this study on the SBC OpenLib (SOL) library as a repository for insertion of publications within the scope of society and supported events as response to the demand of its members and interested parties [34], and keeping all the publications of the events held by it just in one place that will be properly populated in the future and help those interested in this process.

Furthermore, it is expected that the data studied can provide subsidies for the SBC and the entire scientific community, regarding the scenario of Brazilian bibliographic research and production. This fact is in line with the promotion of activities in favor of gender equity, maintaining and highlighting supported actions, such as the Digital Girls Program aimed to encourage girls from elementary school onwards in the area, institutionalized by the aforementioned computer society, seen as of interest by the CC-linked community.

For future studies, it is suggested to increase the analysis period to five years allowing to check the evolution of female participation in CSBC by event, including online events that have occurred due to the COVID-19 pandemic. Furthermore, it also emphasizes the need of easy access to proceedings and abstracts for the execution of algorithms classification to find out their research trends, compared to genre and event. In this way, it will be possible to correlate the evolution among subfields where there is the presence of women.

With the increase in number of records and variables, it becomes possible to submit artificial intelligence and machine learning algorithms to analysis as in the case of association rules [35], aiming to assess how much the presence of certain items in the database correlates with different records, finding out the CS sub-fields that have evolved the most through the publications of female scientists.

6. References

- [1] L. Schiebinger, “Mais mulheres na ciência: questões de conhecimento,” *História, Ciências, Saúde*, vol. 15, pp. 269-281, 2008. Available: doi: 10.1590/s0104-59702008000500015.
- [2] S. Agarwal, N. Mittal, R. Katyal, A. Sureka, and D. Correa, “Women in computer science: What is the Bibliography Data Telling Us?,” *ACM Comput. Soc.*, vol. 46, no. 1, pp. 7-19, 2016. Available: doi: 10.1145/1118178.1118185.
- [3] M. C. Monard and R. P. de M. Fortes, “Uma visão da Participação Feminina nos Cursos de Ciência da Computação no Brasil,” in *V Congresso de la Mujer Latinoamericana en La Computacion, LAWCC*, 2013, pp. 6-12.
- [4] J. Schwartz, L. S. Casagrande, S. A. C. Leszczynski, and M. G. de Carvalho, “Mulheres na informática: quais foram as pioneiras?,” *Cad. Pagu*, no. 27, pp. 255–278, 2006. Available: doi: 10.1590/s0104-83332006000200010.
- [5] A. J. Hussain, L. Connell, H. Francis, D. Al-Jumeily, P. Fergus, and N. Radi, “An Investigation into Gender Disparities in the Field of Computing,” *Proc. - 2015 Int. Conf. Dev. eSystems Eng. DeSE 2015*, pp. 20-25, 2015. Available: doi: 10.1109/DeSE.2015.17.
- [6] SBC, “Educação Superior em Computação Estatísticas-2019,” 2021. Available: <https://www.sbc.org.br/documentos-da-sbc/summary/133-estatisticas/1354-educacao-superior-em-computacao-estatisticas-2019>.
- [7] A. Smeding, “Women in Science, Technology, Engineering, and Mathematics (STEM): An Investigation of Their Implicit Gender Stereotypes and Stereotypes’ Connectedness to Math Performance,” *Sex Roles*, vol. 67, no. 11-12, pp. 617–629, 2012. Available: doi: 10.1007/s11199-012-0209-4.
- [8] T. Beaubouef and W. Zhang, “Where are the women computer science students?,” *J. Comput. Sci. Coll.*, pp. 14-20, 2011.

- [9] D. Arruda, F. Bezerra, V. A. Neris, P. R. de Toro, and J. Wainera, "Brazilian computer science research: Gender and regional distributions," *Scientometrics*, vol. 79, no. 3, pp. 651-665, 2009. Available: doi: 10.1007/s11192-007-1944-0.
- [10] M. Albornoz, R. Barrere, L. Matas, L. Osorio, and J. Sokil, "Las brechas de género en la producción científica Iberoamericana," 2018.
- [11] J. Tague-Sutcliffe, "An introduction to informetrics," *Inf. Process. Manag.*, vol. 28, no. 1, pp. 1-13, 1992.
- [12] S. K. Banshal, K. Singhal, A. Uddin, and V. K. Singh, "Mapping Computer Science research in Bangladesh," *Ski. 2014 - 8th Int. Conf. Software, Knowledge, Inf. Manag. Appl.*, 2014. Available: doi: 10.1109/SKIMA.2014.7083526.
- [13] S. K. Banshal, A. Uddin, K. Singhal, and V. K. Singh, "Computer science research in India: A scientometric study," *12th IEEE Int. Conf. Electron. Energy, Environ. Commun. Comput. Control (E3-C3), INDICON 2015*, 2015. Available: doi:10.1109/INDICON.2015.7443320.
- [14] S. Effendy and R. H. C. Yap, "Analysing Trends in Computer Science Research: A Preliminary Study Using The Microsoft Academic Graph.," *WWW '17 Companion*, pp. 1245-1250, 2017. Available: doi: 10.1145/3041021.3053064.
- [15] C. Noosrikong, S. Ngamsuriyaroj, and S. P. N. Ayudhya, "Identifying focus research areas of computer science researchers from publications," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, pp. 811-816, 2017. Available: doi: 10.1109/TENCON.2017.8227970.
- [16] L. Yang, J. Zhou, J. Shi, and Y. Li, "Big data analysis for SCIE research output of Chinese computer science," *6th Int. Conf. Comput. Sci. Netw. Technol. ICCSNT 2017*, pp. 261-265, 2017. Available: doi: 10.1109/ICCSNT.2017.8343699.
- [17] M. J. Gallivan and R. Benbunan-Fich, "Examining the relationship between gender and the research productivity of IS faculty," *SIGMIS-CPR'06*, pp. 103-113, 2006. Available: doi: 10.1145/1125170.1125201.
- [18] B. Vela, P. Cáceres, and J. M. Cavero, "Participation of women in software engineering publications," *Scientometrics*, vol. 93, no. 3, pp. 661-679, 2012. Available: doi: 10.1007/s11192-012-0774-x.
- [19] D. Gürer and T. Camp, "An ACM-W literature review on women in computing," *ACM SIGCSE Bull.*, vol. 34, no. 2, p. 121, 2002. Available: doi: 10.1145/543812.543844.
- [20] J. M. Cohoon, S. Nigai, and J. Kaye, "Gender and computing conference papers," *Commun. ACM*, vol. 54, no. 8, p. 72, 2011. Available: doi: 10.1145/1978542.1978561.
- [21] J. M. Cavero, B. Vela, P. Cáceres, C. Cuesta, and A. Sierra-Alonso, "The evolution of female authorship in computing research," *Scientometrics*, vol. 103, no. 1, pp. 85-100, 2015. Available: doi: 10.1007/s11192-014-1520-3.
- [22] K. da S. F. M. Ribeiro, J. Azevedo, C. Maciel, and S. A. Bim, "Uma análise de gênero a partir de dados da Sociedade Brasileira de Computação," in *Women in Information Technology (WIT) - CSBC*, 2019, p. 5.
- [23] CSBC, "Inscrições," 2019. Available: <http://csbc2019.sbc.org.br/inscricoes/> (accessed Oct. 11, 2019).
- [24] SBC, "CSBC - Congresso da Sociedade Brasileira de Computação," 2019. Available: <https://www.sbc.org.br/eventos/csbc> (accessed Oct. 11, 2019).
- [25] E. Vargiu and M. Urru, "Exploiting web scraping in a collaborative filtering-based approach to web advertising.," *Artif. Intell. Res.*, vol. 2, pp. 44-54, 2013.
- [26] F. Meireles, "Predizendo sexo a partir de nomes próprios com o genderBR," 2018. Available: <https://fmeireles.com/blog/rstats/predizendo-sexo-nomes-proprios-genderbr/> (accessed Oct. 10, 2019).
- [27] L. A. da Silva, S. M. Peres, and C. Boscarioli, *Introdução à Mineração de Dados com aplicações em R*, 1. ed. 2016.
- [28] B. D. Duarte, A. F. C. Moura, and M. M. Moro, "Mulheres na Computação: Análises por Sub-Áreas," in *Women in Information Technology (WIT) - CSBC*, 2019, p. 5.
- [29] R. Almeida and G. Zanlorensi, "Qual o gênero e a idade de mestres e doutores no Brasil," 2018. <https://www.nexojournal.com.br/grafico/2018/05/23/Qual-o-gênero-e-a-idade-de-mestres-e-doutores-no-Brasil> (accessed Apr. 20, 2019).

- [30] Meninas Digitais, “Sobre o Programa Meninas Digitais,” 2019. Available: <http://meninas.sbc.org.br/index.php/sobre/> (accessed Nov. 04, 2019).
- [31] Parent in Science, “Parent in Science,” 2021. Available: <https://www.parentinscience.com/> (accessed Aug. 20, 2021).
- [32] L. S. Machado, S. A. Bim, K. D. S. F. M. Ribeiro, L. B. Frigo, and C. Maciel, “Espaço infantil em eventos da Sociedade Brasileira de Computação: uma necessidade ou não?,” *Women Inf. Technol. - CSBC*, pp. 30-39, 2020. Available: doi: 10.5753/wit.2020.11273.
- [33] L. Allagnat, S. Berghmans, H. Falk-Krzesinski, S. Hanafi, R. Harbert, S. Hiuggett, and S. Tobin. “Gender in the Global Research Landscape,” 2017. Amsterdam (Netherlands): Elsevier. Available: https://www.elsevier.com/_data/assets/pdf_file/0008/265661/ElsevierGenderReport_final_for-web.pdf
- [34] SBC, “Conheça a SBC Open Lib, a biblioteca digital da SBC,” 2019. Available: <https://www.sbc.org.br/noticias/10-slideshow-noticias/2134-conheca-a-sbc-open-lib-a-biblioteca-digital-da-sbc> (accessed Oct. 12, 2019).
- [35] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases,” *ACM SIGMOD Rec.*, vol. 22, no. 2, pp. 207-216, 1993.