Women in Computer Science Research
What is the Bibliography Data Telling Us?

Swati Agarwal$^1$  Nitish Mittal$^2$  Rohan Katyal$^1$
Ashish Sureka$^3$  Denzil Correa$^4$

$^1$IIIT Delhi, India - swatia@iiitd.ac.in
$^2$NSIT, India - nitishmittal94@gmail.com
$^3$ABB, India - ashish.sureka@in.abb.com
$^4$Max Planck, Germany - denzil@mpi-sws.org

ACM SIGCAS, 2016
Table of Contents

1. Research Motivation and Aim
   - Gender Imbalance
   - Scope and Objective

2. Related Work and Research Contributions
   - Literature Survey
   - Novelty and Uniqueness

3. Experimental Setup

4. Experimental Results
   - Author Distribution
   - Upward or Downward Trend
   - Domain Specific Analysis
   - Conference Leadership Positions
   - Prolific Authors
   - Geographical Location Analysis

5. Conclusion

6. References
Table of Contents

1 Research Motivation and Aim
   - Gender Imbalance
   - Scope and Objective

2 Related Work and Research Contributions
   - Literature Survey
   - Novelty and Uniqueness

3 Experimental Setup

4 Experimental Results
   - Author Distribution
   - Upward or Downward Trend
   - Domain Specific Analysis
   - Conference Leadership Positions
   - Prolific Authors
   - Geographical Location Analysis

5 Conclusion

6 References
Gender Imbalance in Computer Science

Women Participation as Authors in Research Paper and Scientific Publications

Studies on women participation as authors in research paper are conducted to understand gender inequality and reasons for gender differences [3][4][13].

DBLP\(^a\) is a well-known Computer Science bibliography database indexing more than four million publications (as of November 2015).

**Motivation:** Need to measure scientific output and contributions of women, gender gap and imbalances in the field of Computer Science by conducting a bibliometric study on DBLP bibliographic database.

\(^a\)http://DBLP.uni-trier.de/
# Table of Contents

1. **Research Motivation and Aim**
   - Gender Imbalance
   - Scope and Objective

2. **Related Work and Research Contributions**
   - Literature Survey
   - Novelty and Uniqueness

3. **Experimental Setup**

4. **Experimental Results**
   - Author Distribution
   - Upward or Downward Trend
   - Domain Specific Analysis
   - Conference Leadership Positions
   - Prolific Authors
   - Geographical Location Analysis

5. **Conclusion**

6. **References**
Research Questions

Facets for Analysis

1. Number of articles published by all women authors

Swati Agarwal, Nitish Mittal, Rohan Katyal, Ashish Sureka, Denz
Women in Computer Science Research
Research Questions

Facets for Analysis

1. Number of articles published by all women authors
2. Contributions of women authors in CSR conferences as an editor
Facets for Analysis

1. Number of articles published by all women authors
2. Contributions of women authors in CSR conferences as an editor
3. Percentage of women authors making among top 25 most prolific authors in CSR
Research Questions

Facets for Analysis

1. Number of articles published by all women authors
2. Contributions of women authors in CSR conferences as an editor
3. Percentage of women authors making among top 25 most prolific authors in CSR
4. Research productivity of women authors across various sub-fields within CSR
Research Questions

Facets for Analysis

1. Number of articles published by all women authors
2. Contributions of women authors in CSR conferences as an editor
3. Percentage of women authors making among top 25 most prolific authors in CSR
4. Research productivity of women authors across various sub-fields within CSR
5. Research productivity of women authors across various world regions
Research Motivation and Aim
- Gender Imbalance
- Scope and Objective

Related Work and Research Contributions
- Literature Survey
- Novelty and Uniqueness

Experimental Setup

Experimental Results
- Author Distribution
- Upward or Downward Trend
- Domain Specific Analysis
- Conference Leadership Positions
- Prolific Authors
- Geographical Location Analysis

Conclusion

References
Closely Related Work

Vela et al. [13]
Conduct an empirical study for female participation in 12 leading Software Engineering (SE) journals [13]. They conduct an analysis of the gender of the authors, editorial board members, associate editors and editors-in-chief 12 journals over a two-year period (2007 and 2008) [13].

Cohoon et al. [4]
Cohoon et al. conduct a study on measuring trends and influences on women authorship of computing conference papers [4]. They analyze data from over 3000 ACM-affiliated conferences, workshops, symposia and forums held between 1966 and 2009 [4].
Closely Related Work

Cavero et al. [3]
Cavero et al. analyze the evolution of women in computing research by analyzing Computer Science publications from year 1936 to 2010. They use DBLP dataset for their study and present insights on various aspects of the evolution of female authorship in computing research [3]

Aspray et al. [2]
Aspray investigates the relationship between gender and information technology among preteens and adolescents, with each study considering what could lead girls’ interest in computing to diverge from boys [2]
## Table of Contents

1. Research Motivation and Aim
   - Gender Imbalance
   - Scope and Objective

2. Related Work and Research Contributions
   - Literature Survey
   - Novelty and Uniqueness

3. Experimental Setup

4. Experimental Results
   - Author Distribution
   - Upward or Downward Trend
   - Domain Specific Analysis
   - Conference Leadership Positions
   - Prolific Authors
   - Geographical Location Analysis

5. Conclusion

6. References
Contribution to Body of Knowledge

Novel Research Contributions

First in-depth and focused work on mining 4.7 million records of DBLP database spanning 16 years on the topic of women contribution and gender imbalance in Computer Science scholarly publications.

We conduct a series of experiments to answer several research questions and present empirical results and insights.
Experimental Dataset and Statistics

**DBLP Data**

We download September 17, 2015 snapshot of DBLP dataset\(^a\)

We choose Software Engineering (SE), Data Engineering (DE) and Theory (TH) as three distinct sub-fields of CSR and select 81 unique conferences (CSE: 29, DE: 20, SE: 11, TH: 30) based on the parameters used in previous literature\(^b\) [9][12][14].

We restrict our analysis to 16 years of data.

\(^a\)http://dblp.uni-trier.de/xml/
\(^b\)http://projects.csail.mit.edu/dnd/ranking/
We extract the first name of all the authors and apply Genderize API\(^a\) to determine the gender of authors.

For a data sample of 81 conferences, we were able to determine the gender of 3,57,436 authors (2,87,936 as Male and 69,500 as Female) while 59,009 authors’ gender can not be determined using the API.

We obtain a total of 2,14,865 unique affiliations and extract country information also

\(^a\)https://genderize.io
## Table of Contents

1. **Research Motivation and Aim**
   - Gender Imbalance
   - Scope and Objective

2. **Related Work and Research Contributions**
   - Literature Survey
   - Novelty and Uniqueness

3. **Experimental Setup**

4. **Experimental Results**
   - Author Distribution
     - Upward or Downward Trend
     - Domain Specific Analysis
     - Conference Leadership Positions
     - Prolific Authors
     - Geographical Location Analysis

5. **Conclusion**

6. **References**
Only 21% of the researchers (actively publishing in CSR conferences) are female authors and rest 79% are male authors.
Distribution of Male and Female Authors

Software Engineering has the least number of authors publishing papers in SE conferences (7.85% of male authors and 1.97% of female authors)

Both female and male authors are publishing actively in both Data Engineering and Theory (difference of approximate 1%). However, women participation is relatively lower than men authors

For example, in DE, there is a participation gap of 8.29% and similarly, in TH domain women participation is 10.63% lower than men participation
Table of Contents

1 Research Motivation and Aim
   • Gender Imbalance
   • Scope and Objective

2 Related Work and Research Contributions
   • Literature Survey
   • Novelty and Uniqueness

3 Experimental Setup

4 Experimental Results
   • Author Distribution
   • Upward or Downward Trend
   • Domain Specific Analysis
   • Conference Leadership Positions
   • Prolific Authors
   • Geographical Location Analysis

5 Conclusion

6 References
### Trend of Gender Imbalance

**Figure:** Trend of Gender Imbalance in Computer Science Research over a Period of Time from 2000 to 2015

Figure reveals an increase in the percentage of female authors from 17% in 2000 to 23% in 2013.

---

<table>
<thead>
<tr>
<th>Year</th>
<th>Female (%)</th>
<th>Male (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>17.2%</td>
<td>82.8%</td>
</tr>
<tr>
<td>2001</td>
<td>17.4%</td>
<td>82.6%</td>
</tr>
<tr>
<td>2002</td>
<td>17.7%</td>
<td>82.3%</td>
</tr>
<tr>
<td>2003</td>
<td>18.1%</td>
<td>81.9%</td>
</tr>
<tr>
<td>2004</td>
<td>18.5%</td>
<td>81.5%</td>
</tr>
<tr>
<td>2005</td>
<td>18.9%</td>
<td>81.1%</td>
</tr>
<tr>
<td>2006</td>
<td>19.2%</td>
<td>80.8%</td>
</tr>
<tr>
<td>2007</td>
<td>19.5%</td>
<td>80.5%</td>
</tr>
<tr>
<td>2008</td>
<td>19.8%</td>
<td>80.2%</td>
</tr>
<tr>
<td>2009</td>
<td>20.0%</td>
<td>80.0%</td>
</tr>
<tr>
<td>2010</td>
<td>20.2%</td>
<td>79.8%</td>
</tr>
<tr>
<td>2011</td>
<td>20.4%</td>
<td>79.6%</td>
</tr>
<tr>
<td>2012</td>
<td>20.6%</td>
<td>79.4%</td>
</tr>
<tr>
<td>2013</td>
<td>20.8%</td>
<td>79.2%</td>
</tr>
<tr>
<td>2014</td>
<td>21.0%</td>
<td>79.0%</td>
</tr>
<tr>
<td>2015</td>
<td>21.2%</td>
<td>78.8%</td>
</tr>
</tbody>
</table>

---

Swati Agarwal, Nitish Mittal, Rohan Katyal, Ashish Sureka, Denzil Correa

**Women in Computer Science Research**
The percentage of female authors from 2000 to 2013 is not monotonically increasing (as there is a slight decrease from 2004 to 2005 as well as 2006 to 2007) but in general has an upward trend.

Results for year 2013 shows a significant gender gap in authorship as the women representation is still less than 25%.

The rate of women as a percentage of the rate for men for past 3 years is $23/77 = 29.87\%$. 

---

Swati Agarwal, Nitish Mittal, Rohan Katyal, Ashish Sureka, Denzil Correa

Women in Computer Science Research
Table of Contents

1 Research Motivation and Aim
   - Gender Imbalance
   - Scope and Objective

2 Related Work and Research Contributions
   - Literature Survey
   - Novelty and Uniqueness

3 Experimental Setup

4 Experimental Results
   - Author Distribution
   - Upward or Downward Trend
   - Domain Specific Analysis
     - Conference Leadership Positions
     - Prolific Authors
     - Geographical Location Analysis

5 Conclusion

6 References
**Figure**: Domain Specific Trend of Gender Imbalance Over a Period of 16 Years (2000 to 2015)

Upward trend for both male and female authors in all four domains
In first 10 years the rate of increment in female authors is relatively low in comparison to male authors.

From 2005 to 2007 male authors are increasing by 2.36, 1.25, 1.54 percentage while increment in female authors is always less than 1%.

In 2015, there is a large decrement in number of authors across all sub-fields of CSR.
Figure: A Timeline Based Review of Female Authors Holding a Leadership Position in CSR Conferences

Data Engineering conferences, number of male editors are 71.1% more than female editors
Conference Leadership Positions

There is a strong evidence of gender imbalance on leadership positions in CSR.

Despite the variation in numbers, DE and TH domains have same patterns till 2010 except two major gaps in 2005 (70%) and 2007 (64.3%).

We observe a relatively large percentage of female editors in SE in comparison to Theory domain.
# Table of Contents

1. **Research Motivation and Aim**
   - Gender Imbalance
   - Scope and Objective

2. **Related Work and Research Contributions**
   - Literature Survey
   - Novelty and Uniqueness

3. **Experimental Setup**

4. **Experimental Results**
   - Author Distribution
   - Upward or Downward Trend
   - Domain Specific Analysis
   - Conference Leadership Positions
   - Prolific Authors
   - Geographical Location Analysis

5. **Conclusion**

6. **References**
**Female Authors in Top 25 Most Prolific Authors**

**Figure**: A Donut Presentation of Domain Specific Distribution of Top 25 Authors Publishing in CSR Conferences

Number of females (the ones who are determined) in top 25 authors are always below 5
The number of top female researchers are always lower than number of male researchers.
Despite several sudden decrements over 16 years, Theory domain has the maximum number of male authors making among top 25 researchers of year

In 2011, there are only male authors in top 25 prolific researchers in Theory domain

Even after taking all the numbers collectively, maximum number of female authors (8) never made it up to the minimum value of male authors (11) which is below 50%
Table of Contents

1. Research Motivation and Aim
   - Gender Imbalance
   - Scope and Objective

2. Related Work and Research Contributions
   - Literature Survey
   - Novelty and Uniqueness

3. Experimental Setup

4. Experimental Results
   - Author Distribution
   - Upward or Downward Trend
   - Domain Specific Analysis
   - Conference Leadership Positions
   - Prolific Authors
   - Geographical Location Analysis

5. Conclusion

6. References
Developing and Developed Countries

Figure: Choropleth Map

Choropleth map showing 2,131 women authors publishing from 109 countries
Developing and Developed Countries

Figure: Choropleth Map

Distribution of 3,494 men authors publishing from 137 countries across the world.

Swati Agarwal, Nitish Mittal, Rohan Katyal, Ashish Sureka, Denzil Correa

Women in Computer Science Research
USA has the maximum number of male and female authors (1920 and 1162 respectively)

France and Germany have the second highest number of male and female authors. While Canada, Australia, Spain also have a large number of authors publishing in top rank CSR publishers

China has maximum rate of women participation in top rank CSR publishers i.e. 53%. While USA has only 37% women participation. Similarly, Japan, France and South Korea have women research participation below 35%
Figure: Choropleth Map for the Number of Unique Women Author

1828 women and 3253 men authors publishing articles in top CSR conferences from a total of 81 and 93 countries
Figure: Choropleth Map for the Number of Unique Men Author

Among all advanced developed countries, Japan, Germany and USA have the least women participation (between 33 and 35%)
Brazil has higher number of unique women contributing in top rank CSR conferences in comparison to Japan and Germany (37%-31 out of 84 authors)

China has 61 unique women authors publishing papers in top rank CSR conferences while only 22 women authors from South Korea have papers in top CSR conferences in last 6 years

India being a less developed country, it has 29% women research participation while France- an advanced developed country has 0% women authors publishing in prestigious conferences of CSR
Experimental results shows that CSR is a male-dominated field including all selected sub-domains
despite having an upward trend in number of female authors, the rate of increment is always below 1%
The trend of number of articles co-authored by only women authors is increasing only with 0.05%
Number of articles co-authored and primarily authored by at-least one female are least in SE conferences
Gender Imbalance or Gap in Computer Science Research

The number of male authors holding a leadership position in CSR conference is significantly higher than female authors (between 55 and 80%) including the least women participation in Theory domain.

Despite having the maximum number of publications and being an advanced developed country, USA (33%) has significantly low women participation in top rank CSR conferences than other less developed countries; such as China (48.4%) and Brazil (37%).

Canada (42%) and Australia (45%) have higher women research contribution than USA, Japan and Germany (between 33 and 35%).
References I

Denis Arruda, Fábio Bezerra, Vânia Neris, Patricia Rocha De Toro, and Jacques Wainera.
Brazilian computer science research: Gender and regional distributions.

William Aspray et al.
*Women and information technology: Research on underrepresentation*, volume 1.

José María Cavero, Belén Vela, and Paloma Cáceres.
The evolution of female authorship in computing research.

J. McGrath Cohoon, Sergey Nigai, and Joseph "Jofish" Kaye.
Gender and computing conference papers.
Michael J Gallivan and Raquel Benbunan-Fich.
Examining the relationship between gender and the research productivity of is faculty.

KC Garg and Suresh Kumar.
Scientometric profile of indian scientific output in life sciences with a focus on the contributions of women scientists.

Michael Ley.
The dblp computer science bibliography: Evolution, research issues, perspectives.
Is there a gender gap in authorship of the main brazilian psychiatric journals at the beginning of the 21st century?

Antonis Sidiropoulos and Yannis Manolopoulos.
A citation-based system to assist prize awarding.

Hajar Sotudeh and Nahid Khoshian.

Pleun van Arensbergen, Inge van der Weijden, and Peter Van den Besselaar.
Gender differences in scientific productivity: a persisting phenomenon?
References IV

