Female Semantic Web Researchers: Does Collaboration with Male Researchers Influence their Network Status?

Tamy Chambers, Staša Milojević, Ying Ding

School of Informatics & Computing, Department of Information & Library Science

Indiana University, Bloomington, Indiana

tischt@indiana.edu, smilojev@indiana.edu, dingying@indiana.edu

ABSTRACT

The movement to increase gender diversity in computing and computer science is a well-funded and well researched mission. However, despite this, female representation within the field tends to lag behind that of other disciplines within the academy, both among students and faculty. This study sets out to correlate bibliometric and network measures to identify successful and influential women in the computer science subdomain of semantic web research and identify the impact of their collaboration with both men and other women as part of the their success.

Categories and Subject Descriptors

K.7 THE COMPUTING PROFESSION

1. INTRODUCTION

Gender diversity in computing has been addressed by the highest of organizations though both research and financial support. Since 2001, The United States National Science Foundation (NSF) has funded more than $130M in ADVANCE grant projects to increase the representation of women in the STEM sciences, of which Computer Science is one. The Association of Computing Machinery (ACM), the largest and most influential scientific computing society in the world, maintains a *"Women in Computing"* committee (ACM-W) to advocate on behalf of women in all computing fields. Based on findings form a European Commission report by the ETAN Expert Working Group on Women and Science (2002), the European Union has initiated "gender mainstreaming" to increase female participation in all activities relating to science and technology.

Research in the area of gender diversity in computing is considerably broad as it addresses female representation both within the educational and employment process [2]. The current study however, focuses on academic faculty in the computer sciences; in particular the collaboration patterns of female researchers within the Semantic Web area. Given that collaboration and co-authorship is very prevalent in this domain, this study seeks to understand how gender impacts collaboration. The study first identifies the most successful and influential female Semantic Web researchers based on the co-authorship network centrality measures (degree, betweeness, and Eigenvector), and second analyzes their collaboration patterns

based on gender to see if those who are most influential have more collaborations with men than women.

Other studies of gender in academic publishing have included, a survey study in 2002 [4] that concluded "rank, years since PhD, type of university, discipline and department, amount of research time, and marital status are better associated with publication than gender" (p.172), a more recent study in 2013, where researchers assigned *h-indices* to faculty based on their publication record, noted that "men had significantly higher h-indices than women" [1] (p.215), and a study that analyzed paper counts and citation counts in a far reaching study of all women in science across the world, to conclude that "despite many good intentions and initiatives, gender inequality is still rife in science" [3] (p. 211).

2. METHODOLOGY

Bibliographic data for this study was obtained from Arnetminer and supplemented with data mined to identify gender in a process described below. Arnetminer (arnetminer.org) is search and mining service which automatically creates semantic based researcher profiles by extracting information distributed throughout the Web. The data captured by this system is primarily from the computer science domain and includes more than 6000 conferences and 3,200,000 publications integrated into 700,000 researcher profiles from more than 200 countries. Our dataset included 8,193 publications from 11,290 authors published between 1983 and 2009.

Identification of the gender of each author was based on analysis of their first name. This means of gender identification has been used in other studies including a study commissioned by the European Commission [5] which created a database from multiple sources including dictionaries, name lists from foreign consulates, academies, and government organizations. More recently a study analyzing gender disparities in scholarly publishing also used a similar gender assignment method from sources such as the US census, WikiName, and Wikipedia [3].

In this study we used the US census list and lists from two international baby name websites (Baby Names Wizard and Babynames World) to obtain gender. The US census list of first names is drawn from the 2010 census and each name is identified with the percentage of its use as either a male or female. This, however, identified only 20% of the names used in our study. Lists of names and gender identification for European, Indian, African, and Asian names were run against the remaining names, after which the gender of less than 30% of the authors remained unknown.

A co-authorship network was created and analyzed based on the following three centrality measures: degree centrality, which identifies those nodes most connected to the community and which are thus the most influential, betweeness centrality which

identifies nodes which serve as bridges in the network, and Eigenvector centrality which identifies nodes most connected to influential nodes.

The collaboration patterns of the top female authors for each centrality measure were identified based on the gender of all their co-authorship connections. The percentage of both male and female collaboration for each was found by dividing each by the total collaborations for that author. The average collaboration and standard deviation are based on the totals of the top 24 women identified using the three centrality measures.

3. RESULTS

The first goal of this study was to identify the most successful and influential female researchers in the Semantic Web domain given co-authorship. Table 1 lists the top researchers identified by each centrality measure. The following three tables identify the collaboration percentage of each researcher by centrality measure.

Table 1: Top 5 ranked women by centrality measure

Degree Centrality	Betweeness Centrality	Eigenvector Centrality
C Goble	C Goble	C Goble
A Gomez-Perez	K Sycara	M Sabou
D McGuinness	A Gomez-Perez	A Gomez-Perez
M Sabou	D McGuinness	T Catarci
Y Gil	M Sabou	A Illarramendi

Table 2: Female researchers ranked by degree centrality and compared by collaboration percentages

Degree	Male Coll.	SD	Female Coll.	SD
C Goble	69.5%	0.77	18.9%	-0.31
A Gomez-Perez	64.5%	0.30	22.4%	0.41
D McGuinness	64.0%	0.26	22.7%	0.47
M Sabou	59.2%	-0.19	30.6%	2.14
Y Gil	56.3%	-0.46	20.8%	0.08

Avg. male collaboration = .61 (σ =.107784).   Avg. female collaboration = .20 (σ =.047582)

Table 3: Female researchers ranked by betweeness centrality and compared by collaboration percentages.

Betweeness	Male Coll.	SD	Female Coll.	SD
C Goble	69.5%	0.77	18.9%	-0.31
K Sycara	55.8%	-0.50	16.3%	-0.87
A Gomez-Perez	64.5%	0.30	22.4%	0.41
D McGuinness	64.0%	0.26	22.7%	0.47
M Sabou	59.2%	-0.19	30.6%	2.14

Avg. male collaboration = .61 (σ =.107784).   Avg. female collaboration = .20 (σ =.047582)

Table 4: Female researchers ranked by Eigenvector centrality and compared by collaboration percentages.

Eigenvector	Male Coll.	SD	Female Coll.	SD
C Goble	69.5%	0.77	18.9%	-0.31
M Sabou	59.2%	-0.19	30.6%	2.14
A Gomez-Perez	64.5%	0.30	22.4%	0.41
T Catarci	68.3%	0.66	22.0%	0.32
A Illarramendi	69.2%	0.74	17.9%	-0.52

Avg. male collaboration =. 61 (σ =.107784). Avg. female collaboration = .20 (σ =.047582)
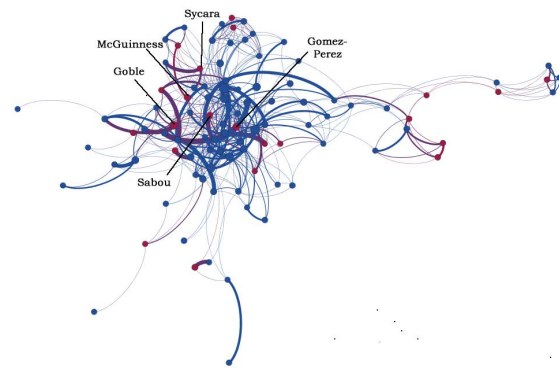

Figure 1: Top 99 researchers based on degree centrality

4. CONCLUSIONS

The results of the current study, show a small correlation between the network measures of women in the Semantic Web domain and their higher than average collaboration with male researchers. Figure 1 shows that most collaboration is either between men (blue line) or men and women (purple line), but that there is little collaboration between women (red line).

C. Goble ranked first in all three network measures, she also had a higher than average collaboration rate with male researchers and a lower than average collaboration rate with female researchers. **A Illarramendi** shows a similar collaboration pattern. **A. Gomez-Perez, T Catarci,** and **D.McGuinness** have similar higher than average collaboration with men, but all also have a higher than average collaboration with women. The collaboration pattern of **M Sabou and Y Gil** are complete opposite to both **C. Goble** and **A Illarramendi** in that they both have lower than average collaboration with men and higher than average collaboration with women.

Later stages of this research will review results past the top five women for each measure and increase the gender identification process to identify the gender of more researchers.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] Eloy, J.A. et al. 2013. Gender disparities in scholarly productivity within academic otolaryngology departments. *OTOLARYNG HEAD NECK* 148, 2 (Feb. 2013), 215–22.

[2] Franklin, D. 2013. *A Practical Guide to Gender Diversity for Computer Science Faculty*. Morgan Claypool.

[3] Lariviere, V. et al. 2013. Bibliometrics: Global gender disparities in science. *Nature*. 504, 7479 (2013), 211–213.

[4] Nakhaie, M.R. 2002. Gender Differences in Publication among University Professors in Canada. *CAN REV SOCIOL* 39, 2 (2002), 151–179.

[5] The ETAN Expert Working Group on Women in Science 2002. *Science policies in the European Union: Promoting excellence through mainstreaming gender equity*. European Commission.