Brazilian Computer Science research: gender and regional distributions

Denis Arruda Fábio Bezerra Vânia Almeida Neris
Patricia Rocha de Toro Jacques Wainer

July 16, 2008

Abstract

This paper analysis the distribution of some characteristics of computer scientists in Brazil according to regions and gender. Computer scientist is defined as the faculty of a graduate level computer science department. Under this definition, there were 886 computer scientists in Brazil in November 2006.

1 Introduction

There are a couple of themes that are explored in the research on gender and academic careers. One of the themes is of the under-representation of females in Science, in general, and in particular in Engineering and Technology research. The other theme, called the "leaking pipeline", [1] which state that the under-representation increases as one moves up the prestige scale in science. The third theme of interest for this paper is the "productivity puzzle" [2] findings, which states that females have lower productivity than their male counterparts.

This research was inspired by our intuition that although the proportion of females in the Brazilian Computer Science research community is low, we did not have the feeling that there was a leaky pipeline in this community. At an impressionistic level, women were "common" in high prestige scientific positions in that community. Of course, that had to be balanced by the low proportion of females in that community, which we call the base line. Based on our experience, we had other intuitions that were investigated by this research. One of them was that there were computer science sub-areas with a much higher proportion of females than the baseline. That is verified by this research, although not all areas we though had a higher female proportion in fact did. Finally, we also tested another intuition that there was some regional specializations, that is, that research in some of the country regions tent to concentrate in some specific subareas of Computer Science.

1.1 Related research

There is an intense research regarding gender and computing. For example, GURER and CAMP [3] reviews the literature up to 2002 on gender and computer science, but with an emphasis on students, from preschoolers to graduate students. CAMP [4] also showed the decreasing proportion of females in CS from high school to graduate school to faculty positions, again an example of the “leaking pipeline”. But the main emphasis on gender and computing is the research on IT workplace and the barriers women face in accepting and keeping an IT job (for example [5, 6, 7]).

Regarding CS research, [8] examines the productivity of male and female researchers in the area of Information Systems (IS). Similar to our work, that work aims at comparing the proportion of published articles from female researchers in relation to male researchers, compared to the baseline - the proportion of female to male researchers. The authors select the top 251 most prolific researchers in 12 of the top journals in the area and within this group, the productivity of women and men are not significantly different.

Brazilian science has been the topic of many publications, although none regarding Computer Science. Brazilian science, in general, has been analyzed in [9, 10], for example. Brazilian science has also been studied in larger context of Latin-Caribbean countries [11]. Ibero-American countries [12], or third world countries [13]. In reference to particular disciplines, [14] analyses the distribution between domestic and international publication in Psychiatry for the period of 1981 to 1995. [15] discusses the increase in publication from Brazilian scientist in the area of Limnology (a sub area of Ecology) from 1970 to 2004, and the lower than average number of citation these article receive.

An analysis of scientific production of Brazilian men and women in astronomy, oceanography and immunology during 1997-2001 is presented in [16]. This study shows that there is effectively no difference between them in potential impact. The authors found evidence that Brazilian women tend to receive fewer “fellowships grants” (see section 2.1). Decreasing evidences of the a glass ceiling in Brazilian sciences was also found by LETA [17].

Finally, there are not much research on regional differences of science production within a single country. [18] discusses the patterns of inter-region scientific collaboration in China. [19] studies the distribution of R&D production, the flow of R&D personnel across different regions, and the patters of co-authorship in scientific publications in and between the different Swedish regions.

2 Background

2.1 The Lattes curriculum system

Brazil has developed an interesting system to record the production of its scientists. It is called the Lattes curricula system, or Lattes CV for short. The current version of the system allow researchers to update their Lattes CV and for other to consult the CVs using a Web system. We will discuss below that researchers have many incentives to keep their CV updated, and thus, the Lattes system is a very valuable resource regarding the Brazilian science.

In Brazil there are two main science funding agencies, CAPES (Coordination for the Improvement of Higher Education Personnel, a section of the Ministry of Education) and CNPq (National Council for the Development of Science and Technology, a part of the Ministry of Science and Technology). CNPq is the funding agency that evaluates and funds researchers, and CAPES evaluates and support graduate courses.

CNPq funds research based on the peer evaluation of the merits of the proponent and of their proposals. CNPq also provides a particular form of funding for researches, called *productivity fellowships* that pays a stipend to the researches, that is, a research fund that is discretionary and need not to be reported back to the CNPq. The fellowships are divided into classes 1A, 1B, 1C, 1D and 2, by decreasing order of value and prestige.

The number of fellowships, both level 1 and 2 are limited. Researchers submit a fellowship proposal to be evaluated by a committee of peers. A researcher with no fellowship can only be accepted into level 2, and a researcher who already receives the fellowship has to renew his fellowship every two or three years (depending on the level) and as the result may retain the fellowship at his current level, be promoted to the level above, or demoted to the level below.

The level of a researcher's fellowship is a recognized measure of scientific prestige. For example, some grant proposal can only be requested by level 1 researchers, and level 1 researchers are the only ones that can be members of both CAPES and CNPq evaluating committees, and that can vote for candidates for these committees.

CAPES evaluates all the graduate courses in Brazil where grade 3 and above grants the course a "recognition of quality". CAPES also distributes to each graduate course a number of student scholarships, based on the course evaluation. Most graduate level students receive these "institutional" scholarships, although some faculty may have research grants (from other sources) to support their students.

The evaluation of the courses is strongly based on the scientific production of the faculty, collected through the Lattes CV system.

Thus, the Lattes CV became the central repository of information regarding Brazilian researchers, and there are strong incentives, both personal and institutional, to keep one's Lattes CV updated.

2.1.1 Research topics

One of the fields in the Lattes CV is the researcher's self-declared research interests or research topics. The Lattes system, has a four level hierarchy to specify a research topic. The higher level, called *grand area*, is divided into Engineering, Exact Sciences, Social Sciences, and so on. The next level is called *area*; Computer Science, is one of the areas of Exact Sciences. The next level is called *sub-area*, and for Computer Science there are four areas: Computer

Sub-Area	Specialities
Computer Mathematics	Symbolic Mathematics Analytical and Simulation Models
Computer Methodology and Techniques	Database Software Engineering Programming Languages Graphics Processing Information Systems
Computer Systems	Computer Systems Architecture Hardware Basic Software Teleinformatics
Computer Theory	Algorithms and Complexity Analysis Computability and Computer Models Formal Languages and Automata Logic and Semantics of Programs

Table 1: Computer science knowledge areas (adapted from CNPq).

Mathematics, Computer Methodology and Techniques, Computer Systems, and Computer Theory, and each sub-area is further divided into *specialities*. The list of sub-areas and specialities is displayed in Table 1.

We could not find the source or the rationale for this particular division of CS, nor a clear definition of each of the categories. But the choices of subareas is clearly limited, and does not contemplate known subareas such as Artificial Intelligence, Human Computer Interfaces, among others. In the first versions of the system, the researcher would have to choose among those alternatives, but later versions allow the researcher to enter a set of keywords as their sub-areas, and specialities instead of the predefined choices. Furthermore, the researcher can declare many research interests, in different grand areas, if needed.

2.2 Brazilian geopolitical regions

Brazil is divided into five geopolitical regions. The North includes seven states and its most important feature is that it includes the Brazilian part of the Amazon rainforest. It is sparsely populated, and its economy is based on plant and mineral exploration, with a limited assembly industry in the city of Manaus, which is a import tax free zone. The Northeast includes nine states, an arid climate, and an economy based on agriculture, mainly sugarcane. The population is concentrated in a few large cities in the coast. The Midwest includes three states and Brasilia, the capital of Brazil. Sparsely populated, its economy is based on large farms agriculture and livestock. The Southeast includes four states, and it is the most developed region of Brazil. The population is distributed into very large metropolitan areas such as Sao Paulo and Rio, and in mid-size cities. The economy is based on a strong and diverse industry, services, agriculture and livestock. The South includes three states, and its economy is based on automobile and textile, livestock and small farm agriculture.

Figure 1 shows the regions, and Table 2 shows some relevant socio-economic data on each region. The GDP (Gross Domestic Product) and HDI (Human Development Index) values were obtained from IBGE 2004 [20] and PNUD [21].

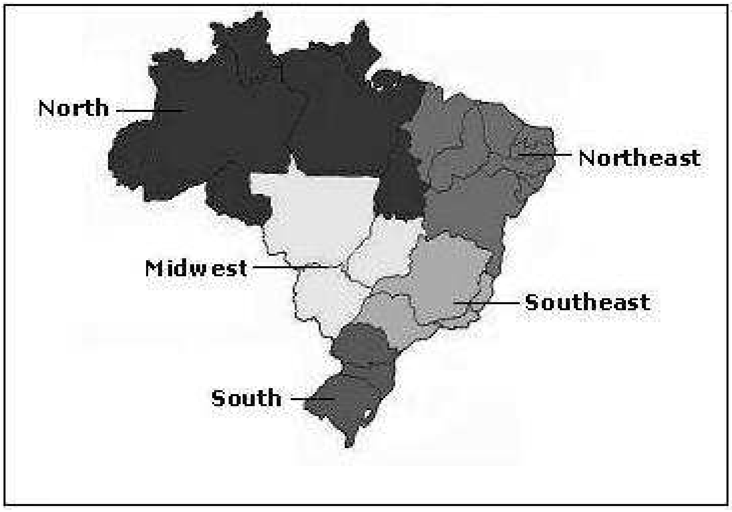

Figure 1: Brazilian Geopolitics Regions

Region	GDP	Average HDI 2000	Population (millions)
North	5%	0.725	12.9
Northeast	14%	0.676	47.7
Center-West	8%	0.792	11.6
Southeast	55%	0.791	72.4
South	18%	0.807	25.1

Table 2: Socio-economic characterization of the Brazilian geopolitical regions.

3 Methodology

We selected all graduate courses in Computer Science that were classified with a grade 3 or above by CAPES, which adds up to 44 courses (in October 2006), 2 from the North region, 10 from the Northeast, 3 from the Midwest, 19 from the Southeast, and 10 from the South.

The list of faculty for each of the 44 courses was found in each of the respective departments' websites. We considered as researcher all faculty listed in the website, including the faculty listed as external members. About 2% of the names were repeated, that is, the same faculty is listed in two or more courses. In that case, we classified the researcher in his self-declared primary department.

We accessed the Lattes CV of each of the names above. 1% of the researchers had no Lattes CV, and were discarded. In the end, we collected and classified the Lattes CV of 886 people, which we consider the set of CS researches in Brazil. The data collection regarding the production of each of the 886 researchers was performed from the middle of October to the middle of November, 2006.

We collected all publication mentioned in their Lattes CV from 2000 to 2006. The Lattes CV has different categories for journal and conference publications and we used these categories for our data collection. We made no distinction on whether the publication was in a national or an international journal, or if it was in a national, regional, or international conference. But some researchers did classify publications in Springer's Lecture Notes series as journal publications. In this case, we disconsidered the entry.

3.1 Research topic

Since the pre-defined subareas and specialities in computer science are uninformative, as discussed above, we defined 12 "research topics" which, we believe, better define the different sub-areas of Computer Science, as seen by the community. These research topics are: artificial intelligence, bioinformatics, collaborative systems, computer in education, data bases, hardware and computer architecture, human-computer interfaces, image processing and computer vision, networks and distributed systems, security, software engineering, and theory.

We collected all sub-areas and specialities under the Computer Science area, as entered by the researchers, and grouped in the 12 research topics. Bellow the keywords used to define each topic:

artificial intelligence : computational intelligence, artificial intelligence, knowledge representation, neural networks, data mining, fuzzy logic, machine learning, automatic reasoning, knowledge based system, natural language processing.

bioinformatics bioinformatics, computational biology, computational molecular biology, biotechnology

collaborative systems computer supported cooperative work, computer supported collaborative learning, groupware, workflow, CSCW.

computer and education distance education, educational informatics, informatics applied to education, artificial intelligence applied to education

data base data base, distributed data base, temporal data base, data integration, data integration in the Web, information integration, XML and semi-structured data bases

hardware and computer architecture computer architecture, computer systems architecture, tool for integrated circuit design, hardware, microelectronics, instruction level parallelism, parallel processing

human computer interfaces interface design, human-computer interface, human machine interface, usability and accessibility, user interface.

image processing and vision graphic computing, image processing, image analysis, computer vision.

network and distributed systems computer networks, management of computer networks, grid computing, mobile computing, computer protocols, Middleware, distributed systems, parallel and distributed processing, wireless networks, sensor networks, teleinformatics, fault tolerant systems.

security biometry, cryptography, computer security, information security, systems security, computer systems security, computer network security

software engineering : software development, formal methods, object oriented programming/development, aspect oriented programming/development, software components

theory algorithms, distributed algorithms, algorithm complexity, cryptography, quantum computing, computational geometry, formal languages and automata, optimization, combinatorial optimization, continuous optimization, non-linear optimization, graphs.

3.2 Statistical analysis

When comparing some data regarding the regional distribution we used a Chi-squared goodness-of-fit test, as implemented in the `chisq.test` function in R[1]. The distribution of researchers for each region is displayed in table 3 is considered as the target distribution. The fitness of any other regional data to the target distribution is verified with the Chi-squared test, and if the resulting p-value is smaller than 0.05 (95% of confidence), the regional data is considered not to fit the target distribution.

When comparing gender data, we use the exact binomial test for goodness of fit of a binomial distribution, as implemented by the function `binom.test` in R. Given any gender data, the exact binomial test will verify its fitness to a target distribution, which we take to be the total proportion of male and females taken from table 4. The exact binomial test is more precise than a chi-squared test for binomial distributions, but again the p-value of less than 0.05 indicates that the data does not fit the target distribution (with 95% of confidence).

[1] www.r-project.org

4 Results: regional differences

Table 3 shows the distribution of CS graduate courses and researchers by the geopolitical regions.

Region	Graduate Courses	Researchers	Researchers per course
North	2	28	14.0
Northeast	10	155	15.5
Midwest	3	41	13.7
Southeast	19	454	23.9
South	10	208	20.8

Table 3: Distribution of CS graduate courses and number of researchers by region.

Table 4 shows the distribution of CS researchers by region and by gender.

Region	Female Researchers	Male Researchers
North	3	25
Northeast	27	128
Midwest	11	30
Southeast	116	338
South	44	164
Total	201	685

Table 4: Number of researchers by gender and region

The distribution of courses and researchers by region (table 3) cannot be considered both data from the same distribution (χ^2 = 34.1981, df = 4, p-value = 6.786e-07) and thus the ratios of researchers per course are significantly different. There is a linear correlation between the number of courses and the number of researchers per course - the more graduate courses there are in a region, the higher the number of researchers each course has.

The distribution of male and female researchers by the regions (table 4) is a standard contingency table. The Chi-squared test for independence reveals that the one cannot reject the hypothesis that the table is independent (χ^2 = 7.543, df = 4, p-value = 0.1098), that is, that there are no significant differences in the gender distribution across the regions.

4.1 Research Areas

Table 5 presents the distribution of the declared research topics by gender and by the geopolitical regions.

The distribution of researchers in most research areas is not significantly different than the distribution of researchers in the political regions, as shown in table 6. The exceptions are the areas of software engineering, which has a

	North	Northeast	Midwest	Southeast	South
Artificial intelligence	6	20	5	66	36
Bioinformatics	0	1	2	12	4
Collaborative systems	1	2	0	10	1
Computers in education	2	6	2	16	12
Data base	6	19	5	104	43
Hardware and computer architecture	2	21	6	63	58
Human-computer interfaces	1	3	0	9	9
Image processing and vision	1	15	3	52	29
Networks and distributed systems	8	47	5	79	48
Security	0	3	1	8	8
Software engineering	8	48	8	91	29
Theory	3	38	15	104	43

Table 5: Number of researchers for each Ad hoc research interests by region.

lower than expected concentration of researchers in the South, and a higher than expected in the North and Northeast; network and distributed systems, with a higher concentration on the North and Northeast, and hardware and computer architecture with a much higher than expected concentration of researchers in the South.

Research topics	p-value
Artificial intelligence	0.7515
Bioinformatics	0.3878
Collaborative systems	0.3952
Computers in education	0.7006
Data base	0.1424
Hardware and computer architecture	0.0027
Human-computer interfaces	0.3414
Image processing and vision	0.4880
Networks and distributed systems	0.0485
Security	0.4883
Software engineering	0.0291
Theory	0.3299

Table 6: Significance of the differences in the distribution of research interests by region.

4.2 Publications

In Brazil for the period from 2000 to 2006, for each paper published in Journals there are 4.27 papers published in Conference proceedings. From 2000 to 2006, the Brazilian researchers considered in this paper published 4470 papers in Journals and 19081 papers in Conference proceedings. Table 7 shows the distribution of publications according to region, and table 10, the distribution according to gender.

	Journals	Conference	Journal productivity	Conference productivity
North	68	511	2.43	18.25
Northeast	547	3305	3.53	21.32
Midwest	142	357	3.46	8.71
Southeast	2637	8922	**5.81**	19.65
South	1076	5986	5.17	**28.78**
Total	4470	19081	5.05	21.54

Table 7: Number of publications and productivity in journals and conferences by political regions between the years 2000 and 2006.

The productivity of researchers in both journal and proceedings are significantly different in the regions, resulting in χ^2 = 39.0387, df = 4, p-value = 6.84e-08, for the goodness-of-fit of the journal production, and χ^2 = 53.6579, df = 4, p-value = 6.206e-11 for the goodness of fit test for the proceeding production. Researchers in the Southeast seems to be concentrating their production in journal articles, at the expense of their production in conference proceedings, where as researchers in the South seems to have achieved a more balanced distribution of their efforts. The lower productivity for the North, Northeast and Midwest researchers may be explained by two factors, the lower supply of graduate students in these regions, and, possibly, a lower critical mass of researchers in each department. Some of the Southern and Southeastern universities and computer science departments are more prestigious and thus attract a lot of graduate students, some from the North, Northeast and Midwest regions. It is likely that researchers from these two regions have more graduate students than average. A second explanation is that, as seen in table 3, there are more researchers per department in the South and Southeast; with less colleagues, the researchers in the lower productivity regions have less chances to collaborate and have a bigger share of the administrative burden.

4.3 Productivity Grants

Among the 886 researchers considered in the paper, 206 received a productivity grant on 2006. Tables 8 and 11 display the distribution of CNPq fellowships by region and gender, respectively.

Finally, the distribution of fellowships (total of both levels) by regions does not follow the distribution of researchers in the regions (p-value = 0.0003396) but it seems to follow the distribution of journal publications (p-value = 0.149). That seems to be consistent with the direction that the CNPq committee have been following of assigning more importance to publications in journals than in conferences.

Fellowship	North	Northeast	Midwest	Southeast	South	Total
1A	0	1	0	11	0	12
1B	0	0	0	13	1	14
1C	0	4	0	11	8	23
1D	0	3	0	29	5	37
2	0	12	6	76	24	120
Total	2	20	6	140	38	206

Table 8: Number of fellowship grants per region.

5 Results: gender differences

Let us repeat the totals for all researchers in Brazil (table4). There are 201 female and 685 male researchers. Thus the baseline female to male proportion is 0.29, or females are 23% of the Computer Science researchers in Brazil.

5.1 Research areas

	Female	Male	Proportion Female/Male	p-value
Artificial intelligence	43	90	0.48	0.01248
Bioinformatics	5	14	0.36	0.7836
Collaborative systems	7	7	1.0	0.02305
Computers in education	19	19	1.0	0.0002661
Data base	25	69	0.36	0.3882
Hardware and computer architecture	12	138	0.09	3.262e-06
Human-computer interfaces	11	11	1.0	0.0045
Image processing and vision	22	78	0.28	1
Networks and distributed systems	27	160	0.17	0.006574
Security	4	16	0.25	1
Software engineering	50	140	0.36	0.2263
Theory	42	161	0.15	0.5575

Table 9: Number of researchers for each research topics by gender, ratio female/male, and statistical significance.

Table 9 list number of male and female researchers according to each ad hoc research area, the proportion of female to male, and the p-value of the goodness-of-fit of the proportion of female to male researchers in the research topics with the baseline of the total number of female and male researchers. The table shows that there is no significant differences between the distribution of the gender of the researchers in the areas software engineering, image processing, theory, security, bioinformatics, and data base. The areas of artificial intelligence, computers in education, collaborative systems, and human computer interfaces have a statistically significant higher proportion of females than the general ratio 0.29, and the areas of network and distributed systems, hardware and computer architectures are "male" research areas.

5.2 Publications

	Female	Male productivity	Female productivity	Male
Journal	976	3494	4.85	5.10
Conferences	4445	14636	21.11	21.37

Table 10: Number of publications in journals and conferences by gender, from 2000 to 2006.

Regarding production, there is no significant difference in the distribution of the journal article production given the proportion of the gender of the researchers (p-value = 0.1805), but there is a small but significant difference in conference proceedings (p-value = 0.04495). That is, there is no significant difference in productivity between male and female researchers regarding journal papers, but females are slightly less productive than males in conference publications.

But the distribution of females are not evenly distributed when we consider the researchers ranked by their production (articles in journals). For example, among the 30 most prolific researchers, only two are females (or 6.7%). Among the 100 most prolific researchers, 17 are females; among the top 250, 47 (or 18.9%) are females. Only among the top 400 is the female proportion similar to the proportion on the whole population of researchers.

5.3 Fellowships

Fellowship	Female	Male
1A	0	12
1B	3	11
1C	2	21
1D	9	28
Sub Total 1	14	72
2	38	82
Total	52	144

Table 11: Number of fellowship grants by gender.

Finally regarding the fellowships, there is no significant differences in the distribution of the fellowships of level 1 or the total number of fellowships across gender (p-value = 0.1969 and p-value = 0.2011 respectively), but for fellowships of level 2 the differences are significant (p-value = 0.02199). That is, while 12%

of the male researchers have a level 2 fellowship, 19% of the female researchers have them, and the difference is statistically significant.

6 Conclusions

The field of computer science in Brazil seems to be reasonable egalitarian regarding gender, given the low number of female computer scientists. Female scientists tend to concentrate in the areas of artificial intelligence, collaborative systems, computer in education, and human-computer interfaces, areas in which the “human component” is more salient. And females tend to avoid areas such as hardware and networks, in which the “technological component” is more relevant. That seems consistent with research such as [22] which show that men prefer more technological activities, although that work is not about research choices.

There is also an interesting phenomena regarding these research choices. Within the “feminine” research areas, females are more productive than males. There are 142 male and 60 female researchers that declare themselves to work in the artificial intelligence, computers in education, and human-computer interfaces areas. They published 755 and 362 journal papers, and 3370 and 1855 conference papers from 2000 to 2006 respectively. That results in a 5.3 journal publications per male researchers and and 6.0 per female researcher, and the difference is statistically significant (p-value = 0.04947). And for conferences, the productivity is is 23,73 and 30,92 for males and females respectively (p-value $< 2.2e\text{-}16$). In the “male” areas of hardware, network, distributed systems, and theory the number researchers are 314 males and 77 females, the total number of journal publications are 1782 for males and 416 for females, and the total number of conference publications are 5890 and 1457. There is no significant differences in productivity between males and females in these research areas (p-value = 0.3762 for journal productivity and p-value = 0.769 for conference).

Women are as recognized as “major researchers” as men, at least using the fellowship measure - in fact women slightly but significantly more likely to receive such a grant than man, contrasting an opposite result for astronomy, oceanography and immunology reported in [16].

The “productivity puzzle” also is not apparent if we take the whole CS community - women are at least as productive than men. But if we take the top 30 or the top 100 most prolific CS researchers in Brazil, women are not as present in these groups as their proportion in the whole CS community. It would be interesting to find out reasons for such disparity, beyond the ones raised by [2] regarding the demands of family versus work life. A possible explanation is that the most prolific researchers are more likely the older ones, people that had the time to create large research groups, and maybe women are newcomers to the Brazilian CS community. This hypothesis can be evaluated using the Lattes systems.

Regarding regional differences, there are some statistically significant differences in productivity in the different regions, and some differences in the concentration of researchers in a few research topics. We do not believe that the two disparities are related - that the research topics of favor in the less productive regions are "more difficult" and thus result in less publications. We believe that there are other, more likely explanations. One hypothesis that the Southern and Southeastern departments attract the more motivated graduate students, reducing the availability of graduate students in the other regions. That would reduce the productivity of the researchers in the other regions. A second hypotheses is that because of a lower number of researchers per department, faculty in the lower productivity regions have a higher teaching and administrative burden, and thus have less time to spend in research as their Southeastern and Southern counterparts. Finally, a third hypothesis is that older researchers are in general more productive, as we discussed above, and some of the computer science departments in the most productive regions are among the oldest in Brazil, and thus more likely to harbor the older researchers. As we mentioned, this hypothesis can be evaluated using the Lattes system, and is a important future continuation of this work.

Finally, the regional differences regarding emphasis in some research topics is an interesting phenomena, which deserves some further research. We propose two competing hypothesis. The first is that once a Computer Science department or a region achieves a certain research status in a particular subarea it attracts new faculty in that area. The alternative hypothesis is that a strong research group in one of the subareas, produces a lot of PhD that are either hired by the department as faculty, or move to a same region university.

6.1 Limits of this research

In this research we used a limited definition of "computer science researcher", as faculty associated to departments that grant a graduate degree in computer science and which are accredited by CAPES with a grade 3 or more. There are also computer science researchers in other departments, for example, computer engineering, electronic engineering, applied mathematics and maybe mecatronics. Despite working on these other departments some of these researchers would classify themselves as computer scientists and would be recognized as such by the CS community. It is also possible that researchers in Computer Science are also present in non-graduate granting departments, or in government and industry research centers. Expanding our research to encompass them would be an important future step of this work.

A second potential limitation of this research is the overcounting of papers. Since we count all papers in a researcher CV, a paper that is co-authored with another Brazilian CS researcher is doubly counted, because that publication will also be added to the other researcher's CV. The overcounting of papers may be an issue if co-authorship between researchers is correlated with one of the independent variables used in this paper (gender, region, research topic).

References

[1] J ALPER. The pipeline is leaking women all the way along. *Science*, 260(16):409–411, 1993.

[2] J. R. COLE and H. ZUCKERMAN. The productivity puzzle: persistence and change in patterns of publication of men and women scientists. In P. Maehr and M. W. Steinkamp, editors, *Advances in Motivation and Achievement*, volume 2, pages 217–258. JAI Press, 1984.

[3] D. GURER and T. CAMP. An acm-w literature review on women in computing. *SIGCSE Bull.*, 34(2):121–127, 2002.

[4] T. CAMP. The incredible shrinking pipeline. *Commun. ACM*, 40(10):103–110, 1997.

[5] M. K. AHUJA. Women in the information technology profession: a literature review, synthesis and research agenda. *European Journal of Information Systems*, 11(1):20–34, 2002.

[6] M. SUMNER and K. WERNER. The impact of gender differences on the career experiences of information systems professionals. In *SIGCPR '01: Proceedings of the 2001 ACM SIGCPR conference on Computer personnel research*, pages 125–131, New York, NY, USA, 2001. ACM Press.

[7] C. K. RIEMENSCHNEIDER, D. J. ARMSTRONG, M. W. ALLEN, and M. F. REID. Barriers facing women in the IT work force. *SIGMIS Database*, 37(4):58–78, 2006.

[8] M. J. GALLIVAN and R. BENBUNAN-FICH. Examining the relationship between gender and the research productivity of is faculty. In *SIGMIS CPR '06: Proceedings of the 2006 ACM SIGMIS CPR conference on computer personnel research*, pages 103–113, New York, NY, USA, 2006. ACM Press.

[9] W. GLANZEL, J. LETA, and B. THIJS. Science in Brazil. Part 1: A macro-level comparative study. *Scientometrics*, 67(1):67–86, 2006.

[10] J. LETA, W. GLANZEL, and B. THIJS. Science in Brazil. Part 2: Sectoral and institutional research profiles. *Scientometrics*, 67(1):87–105, 2006.

[11] M. KRAUSKOPF, M. I. VERA, V. KRAUSKOPF, and A. WELLJAMS-DOROF. A citationist perspective on science in Latin America and the Caribbean, 1981-1993. *Scientometrics*, 34(1):3–25, 1995.

[12] A. M. CALVINO. Assessment of research performance in food science and technology: Publication behavior of five Iberian-American countries (1992-2003). *Scientometrics*, 69(1):103–116, 2006.

[13] F. OSAREH and C. S. WILSON. Third World Countries (TWC) research publications by disciplines: A country-by-country citation analysis. *Scientometrics*, 39(3):253–266, 1997.

[14] I. FIGUEIRA, R. JACQUES, and J. LETA. A comparison between domestic and international publications in Brazilian psychiatry. *Scientometrics*, 56(3):317–327, 2003.

[15] A. S. MELO, L. M. BINI, and P. CARVALHO. Brazilian articles in international journals on Limnology. *Scientometrics*, 67(2):187–199, 2006.

[16] J. LETA and G. LEWISON. The contribution of women in brazilian science: A case study in astronomy, immunology and oceanography. *Scientometrics*, V. 57(n. 3):p. 339–353, 2003.

[17] J. LETA. As mulheres na ciência brasileira: crescimento, contrastes e um perfil de sucesso. *Estudos Avançados*, 17:271 – 284, 12 2003. in Portuguese.

[18] L. LIANG and L. ZHU. Major factors affecting China's inter-regional research collaboration: Regional scientific productivity and geographical proximity. *Scientometrics*, 55(2):287–316, 2002.

[19] R. DANELL and O. PERSSON. Regional R&D activities and interactions in the Swedish Triple Helix. *Scientometrics*, 58(2):203–218, 2003.

[20] IBGE. Brazilian institute of geography and statistics. http://www.ibge.gov.br/home/, 2007.

[21] PNUD. United nations development programme. http://www.pnud.org.br/, 2007.

[22] T. KLIEF and W. FAULKNER. Boys and their toys: Men's pleasures in technology. `http://www.rcss.ed.ac.uk/sigis/public/backgrounddocs/BOYS_AND_THEIR_TOYS-ZIF8.rtf`, 2002.