

A Preliminary Analysis of the Scientific Production of Latin American Computer Science Research Groups

Juan F. Delgado-Garcia, Alberto H.F. Laender and Wagner Meira Jr.

Computer Science Department, Federal University of Minas Gerais
31270-901 - Belo Horizonte - Brazil
{jfdgarcia, laender, meira}@dcc.ufmg.br

Abstract. In this paper, we present a preliminary analysis of the scientific production of Latin American Computer Science research groups. Our analysis is based on data over a period of 20 years collected from DBLP, and addresses 24 groups from academic institutions in Argentina, Chile, Colombia, Cuba, Mexico, Peru, Uruguay and Venezuela. Our results show a clear improvement in the publication output of these groups in the last 10 years, particularly in Argentina, Chile and Mexico.

Keywords: Latin America, Computer Science, Scientific Production, Coauthorship Analysis, Bibliometrics

1 Introduction

According to SCIMago Journal and Country Rank (JCR)¹, recent years have witnessed a tremendous increase in the scientific production in Computer Science (CS) all over the world. Considering data from 2002 to 2012, for instance, in North America (Canada and USA) the number of publications increased 59.73%, in Western Europe (considering only the top-5 countries, UK, Germany, France, Italy and Spain) 184%, and in Latin American (also considering only the top-5 countries, Brazil, Mexico, Argentina, Chile and Colombia) 319%. In other regions, countries like Australia, China, Korea, India and Poland have achieved figures even higher. Although SCIMago numbers reflect only publications that appeared in selected journals, they show that CS is a very productive research area with many active research groups spread around the world.

In view of this scenario, in this paper we present a preliminary analysis of the scientific production of Latin American CS research groups. Our analysis is based on data over a period of 20 years collected from DBLP², and addresses 24 groups from academic institutions in Argentina, Chile, Colombia, Cuba, Mexico, Peru, Uruguay and Venezuela. Despite being the country in Latin America with the highest productivity in the area [4], we have not included Brazil in our analysis for two reasons. First, Brazil is by far the Latin American country

¹ <http://www.scimagojr.com/countryrank.php>

² <http://www.informatik.uni-trier.de/~ley/db>

with the largest number of CS research groups. According to a recent report from CAPES³, Brazil's Ministry of Education agency in charge of graduate programs, there are in the country 69 Computer Science graduate programs spread through 56 academic institutions. This number of programs would make the comparison with groups from other Latin American countries unbalanced. Second, to the best of our knowledge, this is the first work that focuses on a comparison involving CS research groups only from Latin American, whereas the literature includes some other studies that compare CS research groups from Brazil with those of other countries [2,3,4].

Related Work. Although there are some previous works in the literature that analyze and compare the research production of several countries, here we focus on three specific ones that involve Computer Science research groups in Latin America. Laender *et al.* [2] analyzed the quality of the top Computer Science graduate programs in Brazil and found that they are comparable to programs in North America and Europe w.r.t. publication and graduation rates. In particular, the study showed that the ratio between conference and journal papers in Brazilian programs, around 2.5, was close to the ratio presented by European and North American programs, which ranges from 2.3 to 2.8. Wainer *et al.* [4] presented a comparative study of the Brazilian CS scientific production with some Latin American (Argentina, Chile, Mexico), European, BRIC (Russia, India, China), and other relevant countries such as South Korea, Australia, and USA from 2001 to 2005. The findings show that Brazil's scientific production is the largest in Latin America, getting close to that of European countries such as Spain and Italy, and almost the same as India and Russia. Menezes *et al.* [3] analyzed the characteristics of three coauthorship networks in CS communities formed, respectively, by researchers from Brazil, North America (Canada and USA) and Europe (France, Switzerland and UK). They provided several statistics of the three networks and performed a temporal analysis of them over a span of 12 years, from 1994 to 2006. In this paper we not only focus on the overall production of the eight Latin American countries considered, but also analyze the performance of the research groups of the main institutions in each country.

The rest of this paper is organized as follows. Section 2 describes the data gathered for our analysis, Section 3 discusses our preliminary results, and Section 4 presents some conclusions and insights for future work.

2 Data Gathered

In this paper we base our analysis on publications from the last 20 years that were authored by researchers from institutions in Latin America that offer graduate programs in Computer Science, Informatics and System Engineering [1]. In particular, we consider the following institutions:

³ <http://avaliacaotrienal2013.capes.gov.br/relatorios-de-avaliacao>

- **Argentina:** Universidad de Buenos Aires (UBA), Universidad Nacional de la Plata (UNLP), Universidad Nacional del Centro de la Provincia de Buenos Aires (UNICEN) and Universidad Nacional del Sur (UNS).
- **Chile:** Pontificia Universidad Católica de Chile (PUC-Chile), Universidad de Chile (UCHILE) and Universidad de Concepción (UDEC).
- **Colombia:** Universidad ICESI (ICESI), Pontificia Universidad Javeriana - Cali (PUJ-Cali), Universidad de los Andes (ANDES), Universidad del Valle (UNIVALLE) and Universidad Nacional de Colombia (UNAL).
- **Cuba:** Universidad de La Habana (UH), Universidad de las Ciencias Informáticas (UCI) and Universidad de Oriente (UO).
- **Mexico:** Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV), Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM), Universidad Autónoma del Estado de México (UAEMEX) and Universidad Nacional Autónoma de México (UNAM).
- **Peru:** Universidad Católica San Pablo (USCP).
- **Uruguay:** Universidad de la República (UDELAR).
- **Venezuela:** Universidad Central de Venezuela (UCV), Universidad de Carabobo (UC) and Universidad Simón Bolívar (USB).

The data gathering process consisted of three major steps. In the first step, we identified the researchers, i.e., faculty, from each institution of interest. In the second step, we collected data from the DBLP entry of each identified researcher. Finally, we parsed the resulting XML files and populated a relational database with 6126 publications (conference papers and journal articles) spread over 1643 publication venues. The data analyzed in this paper reflects the DBLP repository on March 27, 2014.

3 Results

In this section we present some results of our analysis. Table 1 summarizes the publication statistics of each research group by country from 1994 to 2013. The publication rate of each institution λ_i is given by Equation 1

$$\lambda_i = \frac{\sum_{y=1994}^{2013} \frac{P_{iy}}{R_{iy}}}{20} \quad (1)$$

where P_{iy} is the total number of publications from institution i in year y , and R_{iy} is the total number of researchers affiliated to institution i in year y . The overall mean \bar{X}_c of the scientific production per country was calculated according to Equation 2

$$\bar{X}_c = \frac{\sum_{i \in U_c} \lambda_i}{|U_c|} \quad (2)$$

where U_c is the set of institutions from country c .

As we can see from Table 1, the overall mean of the scientific production per country is 2.59 for Argentina, 3.18 for Chile, 1.02 for Colombia, 0.92 for Cuba,

2.28 for Mexico, 0.87 for Peru, 1.41 for Uruguay and 1.58 for Venezuela. These numbers show a clear predominance of the research groups in Argentina, Chile and Mexico, with an average of more than two publications per year over the 20-year time period. Moreover, these groups contribute with, respectively, 28.80%, 28.72% and 25.30% of the overall production of the 24 institutions considered. They are also the groups associated with graduate programs that are more than 10 years old and have a more consolidated structure with the majority of the faculty members holding a PhD degree. A second group of countries comprises Colombia, Venezuela and Uruguay with an average publication rate by year between 1.0 and 1.58. Finally, we have Cuba with an average of 0.92 publication by year and Peru, with a young graduate program at the Universidad Católica de San Pablo, showing an average of 0.87 publication by year.

| Country | Institution | # of Pub. | Pub. Rate λ_i | Overall Mean \bar{X}_c |
|-----------|-------------|-----------|-----------------------|--------------------------|
| Argentina | UBA | 888 | 2.21 | 2.59 |
| | UNLP | 492 | 2.76 | |
| | UNICEN | 385 | 2.84 | |
| | UNS | 336 | 2.55 | |
| Chile | PUC-Chile | 408 | 2.51 | 3.18 |
| | UCHILE | 1530 | 5.49 | |
| | UDEC | 155 | 1.53 | |
| | ICESI | 36 | 0.75 | |
| Colombia | PUJ-Cali | 38 | 1.36 | 1.02 |
| | ANDES | 73 | 1.10 | |
| | UNIVALLE | 29 | 0.66 | |
| | UNAL | 204 | 1.23 | |
| Cuba | UH | 38 | 0.72 | 0.92 |
| | UCI | 27 | 0.69 | |
| | UO | 127 | 1.35 | |
| Mexico | CINVESTAV | 655 | 3.79 | 2.28 |
| | ITESM | 179 | 1.66 | |
| | UAMEX | 31 | 0.94 | |
| | UNAM | 979 | 2.73 | |
| Peru | UCSP | 35 | 0.87 | 0.87 |
| Uruguay | UDELAR | 286 | 1.41 | 1.41 |
| | UCV | 221 | 1.56 | |
| Venezuela | UC | 44 | 0.97 | 1.58 |
| | USB | 285 | 2.21 | |

Table 1: Production of the research groups in the period 1994-2013.

We also observed that the average number of coauthors during the period between 1994 and 2013 is 2.26 for journal articles and 2.41 for conference papers, which shows that Latin American CS groups follow a coauthorship pattern similar to North American and European ones [2]. It is also worth mentioning that the Latin American groups publish more in conferences than in journals following again a similar pattern of many groups in North America, Europe and Brazil [2]. We can also analyze in more detail the temporal evolution of the region scientific

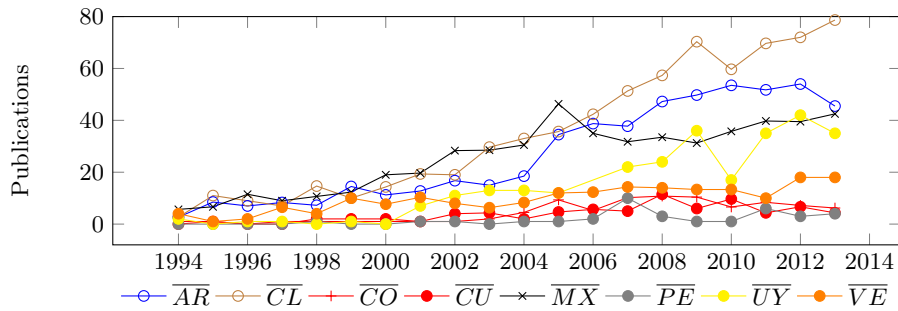


Fig. 1: Average number of publications per country over time.

production in Figure 1, which shows the average number of publications per country over time. Notice that there has been an increase w.r.t. publications in all countries, in particular Chile, which presented the largest increase, followed by Argentina and Mexico, where we observe a consistent increase from 1998 to 2008 (probably as a consequence of the Internet access and availability of scientific works on the Web).

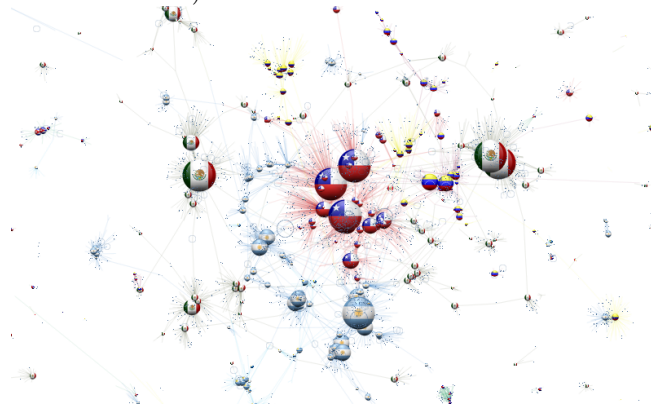


Fig. 2: LACCompNet - Latin American Computer Science Network.⁴

Finally, we present an analysis of the research networks originated from the coauthorships (see Figure 2) and discuss how they evolved over time. Here, we define a research network as a group of authors who have published together at least five papers in a decade. We determined research networks by mining maximal sets of authors, that is, the largest set of authors whose subsets are also significant in terms of number of publications [5]. Table 2 shows both the number of groups and their average size per country and decade. We can see that all countries presented an increase w.r.t both indicators. The increase in the number of groups demonstrate that there is an increasing research density

⁴ <http://tortuga.lbd.dcc.ufmg.br/LACCompNet>

in the region, while the increase in the average size of the groups shows that researchers are cooperating more and there is a growing critical mass in the CS area.

| Country | 1994 - 2003 | | 2004 - 2013 | | Total |
|-----------|-------------|----------|-------------|----------|-------|
| | Groups | Avg Size | Groups | Avg Size | |
| Argentina | 35 | 2.57 | 160 | 3.04 | 195 |
| Chile | 19 | 2.53 | 137 | 2.70 | 156 |
| Colombia | 0 | 0.00 | 30 | 2.43 | 30 |
| Cuba | 2 | 2.50 | 12 | 3.83 | 14 |
| Mexico | 25 | 2.48 | 101 | 2.79 | 126 |
| Uruguay | 4 | 2.25 | 25 | 2.96 | 29 |
| Venezuela | 7 | 2.43 | 33 | 2.64 | 40 |

Table 2: Group analysis by country.

4 Conclusions and Future Work

In this paper, we assessed the CS scientific production in Latin American, identifying research groups in each country. A temporal study was also performed, using a period of 20 years, from 1994 to 2013. We have compared the production among some research groups from eight countries in Latin American, and found a fast increase in publications from some of them, possibly due to cooperation work with groups of North America and Europe, spotted when analyzing the research groups. We also found other groups that were structured through offering of Master (MSc) and Doctorate (PhD) programs. As future work we would like to explore how these groups are formed through their coauthorship networks. In addition, we plan to use other sources of bibliographic information for exploring coauthorship networks in more detail.

Acknowledgments. This work is partially funded by CAPES, CNPq, FAPEMIG and InWeb, Brazil.

References

1. Cuadros-Vargas, E., Silva-Sprock, A., Delgado-Castillo, D., Hernandez-Bieliukas, Y., Collazos, C. Evolution of the Computing Curricula for Computer Science in Latin America. *Proc. of CLEI*, pp. 1-10, 2013
2. Laender, A.H.F., Lucena, C.J.P., Maldonado, J.C., Souza e Silva, E., Ziviani, N. Assessing the research and education quality of the top Brazilian Computer Science graduate programs. *ACM SIGCSE Bulletin*, 40(2):135-145, 2008.
3. Menezes, G.V., Ziviani, N., Laender, A.H.F., Almeida, V.F.A. A Geographical Analysis of Knowledge Production in Computer Science. *Proc. of WWW*, pp.1041-1050, 2009.
4. Wainer, J., Xavier, E.C., Bezerra, F. Scientific production in computer science: A comparative study of Brazil and other countries. *Scientom.*, 81(2):535-547, 2009.
5. Zaki, M., Meira Jr., W. *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge, 2014.