

Citation Disparity in Sub-Areas of Brazilian Computer Science

Fernando F. Druszcz, André L. Vignatti

Department of Computer Science, Federal University of Paraná, Curitiba, Brazil

Corresponding author: vignatti@inf.ufpr.br

Keywords: citation distribution, bibliometrics, social systems

It is an empirical known fact that some scientific areas receive more citations than others. The issue is twofold, the comparison between different areas and the use of the number of citations as a mean of comparison. Due to their unique natures, comparing different areas can be irrelevant or meaningless, but if they are subordinated to the same committee, comparisons are often made, and this leads to problems. A wide variety of metrics assess the quality of scientific output, one thing in common is that most revolve around the *number of citations*. Metrics that rely on the number of citations generally operate at the author level, but are often extrapolated to compare areas and sub-areas. Such extrapolation can be even more problematic than reducing the impact of articles and researchers to a single metric.

Our research: we analyze different sub-areas of Computer Science, with the data coming from the Brazilian community. We present the existing disparity between citations in sub-areas, which can cause problems if such sub-areas are submitted to the same evaluation criteria of scientific production. Also, we show how such issues are mitigated by a simple equation proposed in the literature, called *universal fit citation*. Let c be the total number of citations that an article has received since its publication date and μ be the average number of citations in the article's area in the year it was published. The *universal fit citation* \tilde{c} is defined as $\tilde{c} = \frac{c}{\mu}$. Focusing on the Brazilian community of computer science, we use CSIndexbr, to get a list of computer science authors in Brazil. From that list, we collect data about their publications on DBLP. Finally, with the DOI of these publications, we obtain their citation number in OpenCitations. To separate these by sub-areas, we used the venue-sub-area relation provided by CSIndexbr. This leads to 1,100 Brazilian computer science researchers with 46,000 publications, but only 11,000 could be related with a sub-area.

Preliminary results: our analysis showed that the distribution of citations follows a power law. Also, it is uneven for different areas of Brazilian computer science, e.g., publications in “computer vision” with at least 100 citations are 11 times more common than in “computer architecture”. Moreover, finding a paper with 20 or more citations is three times more likely in the sub-area of “security and cryptography” than in “formal methods and logic”. Finally, the proposed normalization factor appears to minimize the difference between citation curves, suggesting a fairer comparison between subareas of Brazilian computer science.