

## Introduction

Google Scholar is one of the most eminent web search engines and contains literature from almost all publishing formats and disciplines .

It also allows authors to create their profiles and link articles and journals to the profile based on which the profile is given a total number of citations and various other indexed rankings.

This project aims to check the accuracy of these Google Scholar profiles as it was observed that some of the profiles include articles not written by the author, but Google Scholar still included the citations for those articles in calculating the total number of articles for the scholar profile.

Previously, a lot of research has been conducted on Google Scholar and Google has always modified its website to cater to those faults the research presented in this report is novel and does not derive from any previous research.

## Methods and Materials

- Initially names of 1000 computer science professors and researchers from top-ranked universities were collected.
- Then a scraper coded in python3 using the selenium library, was used to scrape the data from Google Scholar and store it in text files.
- Out of these 1000 professors 736 had public profiles, their total citation counts and titles of the first 500 articles were recorded.
- Using this data and excel sheet was created that contained the information about all the authors and the number of fake articles found
- Another field was added to the dataset with attribute name False citation ratio where:
 
$$\text{False Citation Ratio} = \frac{\text{False Citation Count}}{\text{Total Citations (Copied)}}$$
- Later another data set was collected which categorized the authors according to their region to find any relation region and fake articles on profiles.
- The process of creating a profile was also studied by creating a fake profile and adding articles suggested by Google. To try and identify who was at fault for these fake articles, Google or the authors themselves.

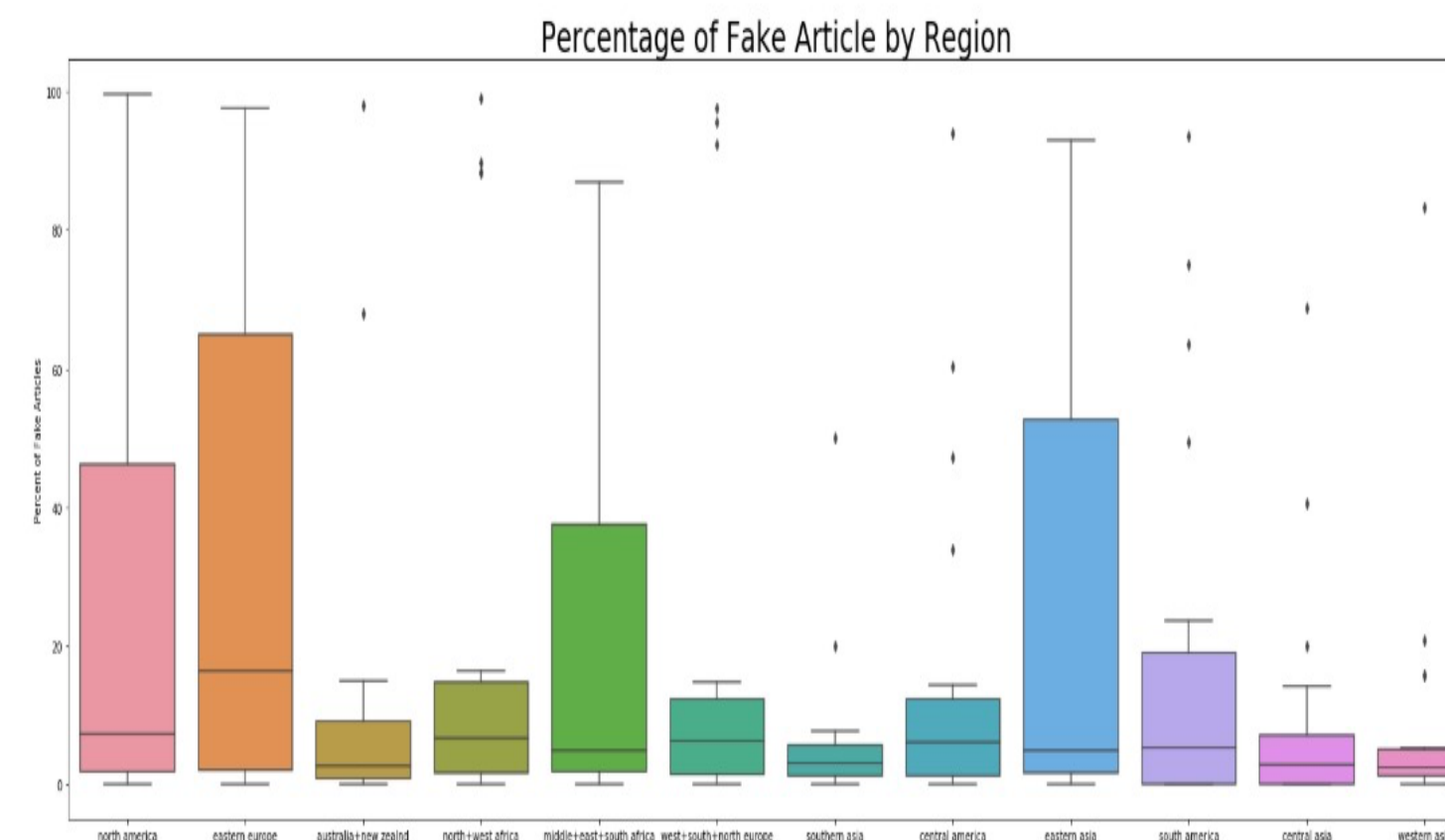


Figure 2. box plots of fake articles according to regions

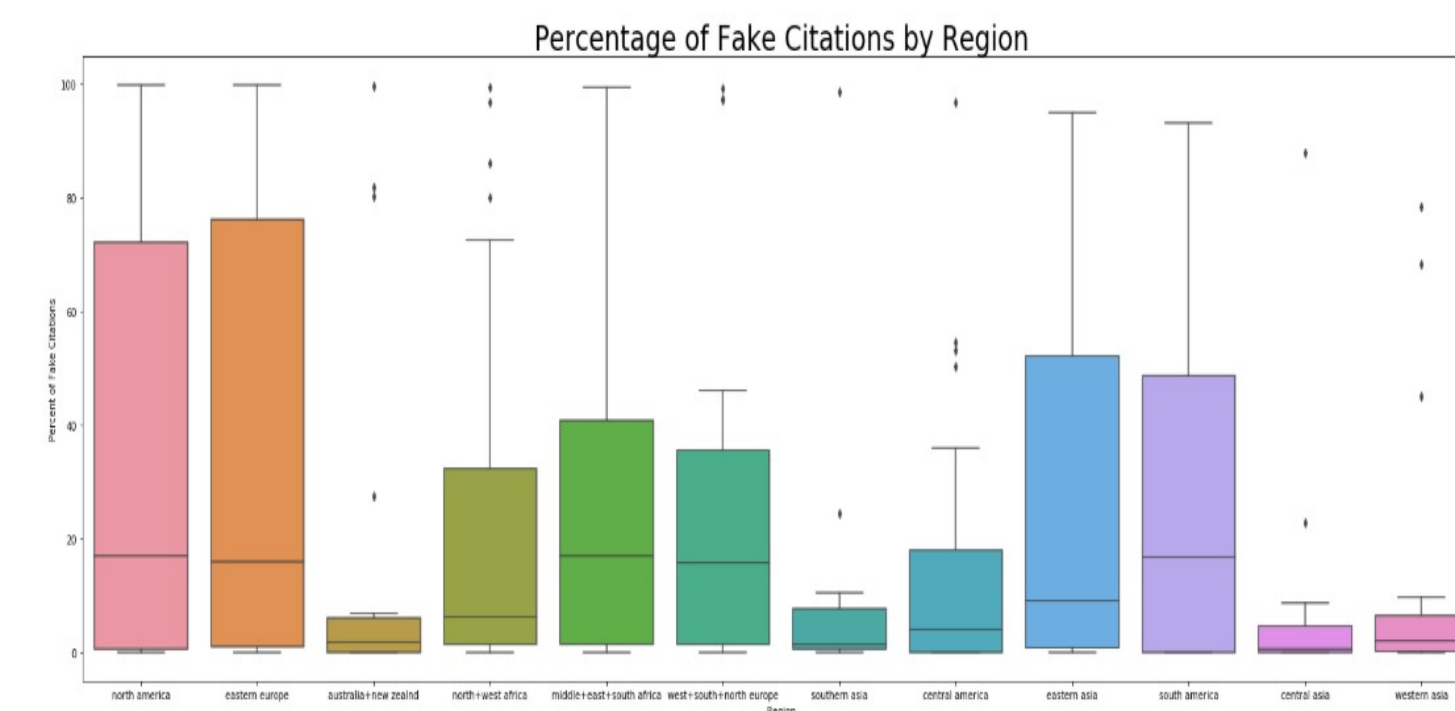


Figure 3 . box plots of fake citations according to regions

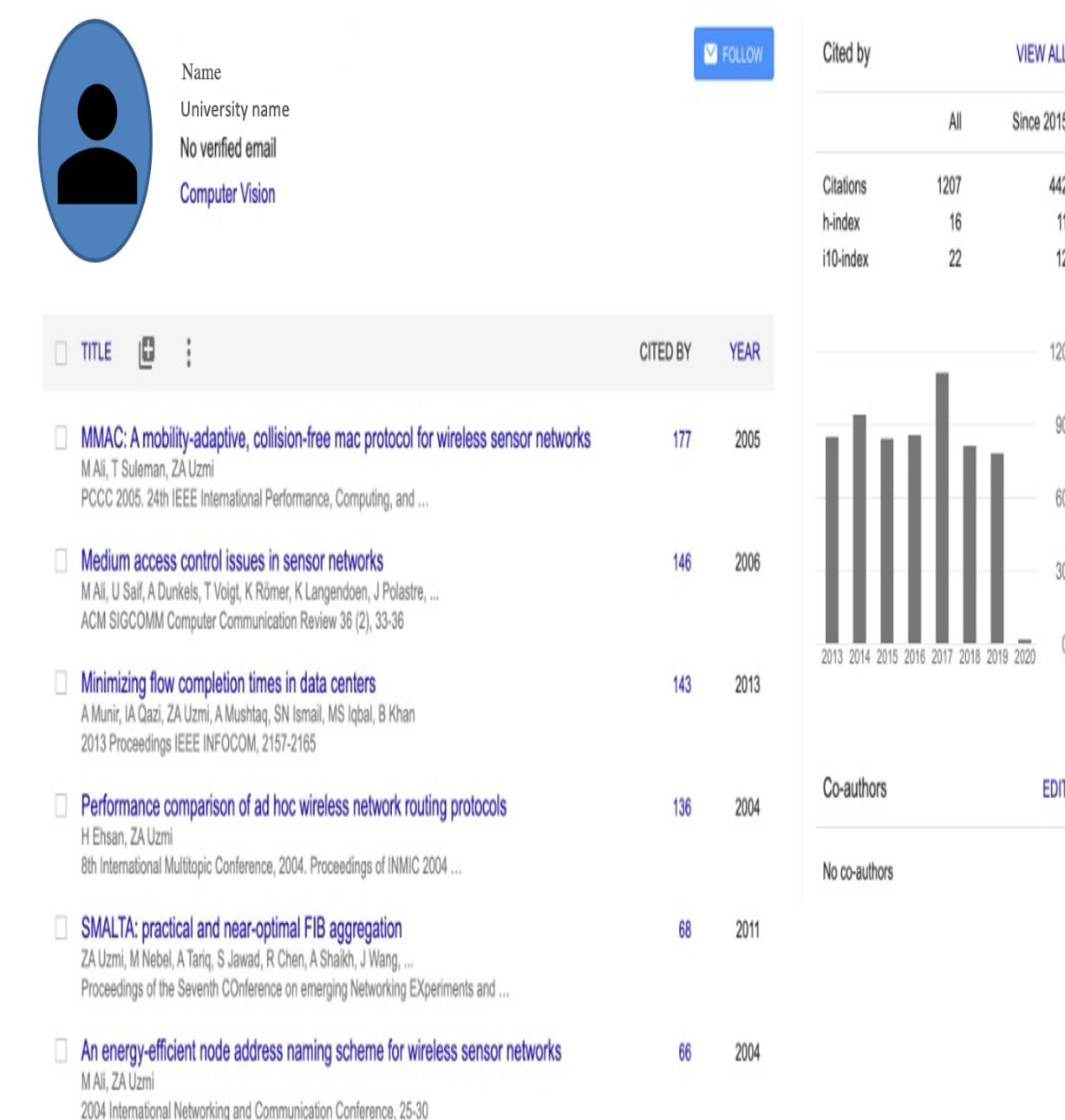


Figure 1. Example of a Google Scholar profile.

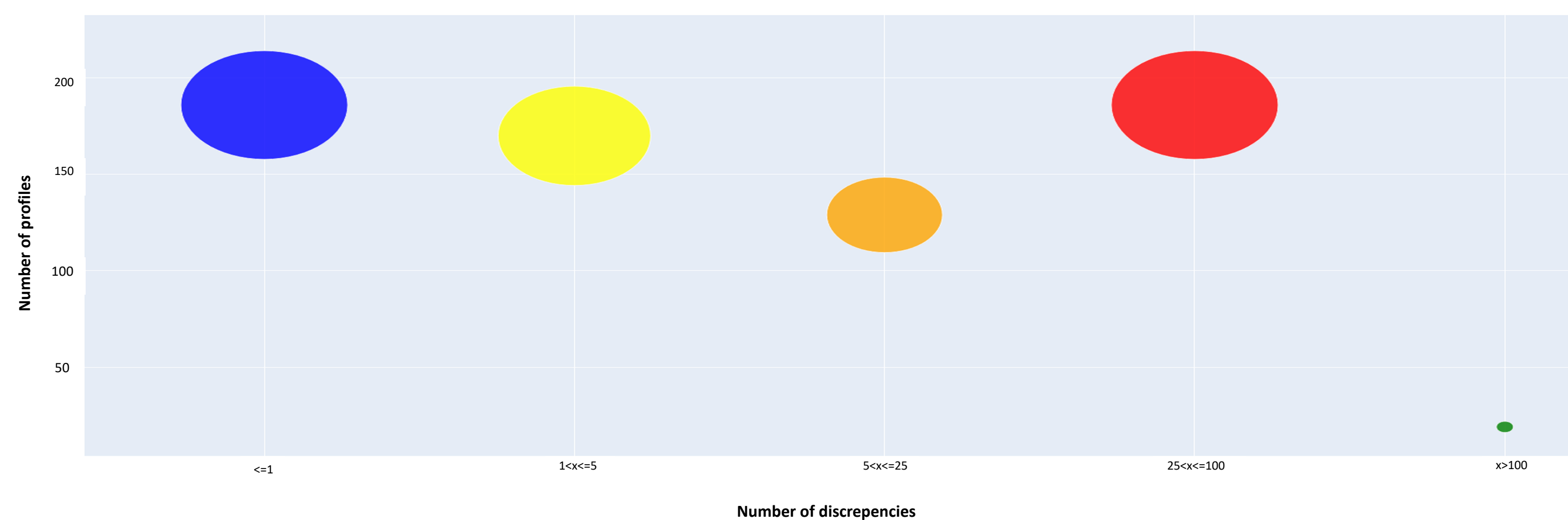


Figure 4 . bubble chart of number of discrepancies and number of profiles

## Results

- Figure 4 shows that out of the 736 profiles scraped 690 profiles contained at least one article which did not belong to the author of the profile.
- 19 profiles had more than 100 discrepancies.
- Out of the 690 the highest number of discrepancy (false count) in a profile was 377 articles from the total of 500 articles.
- The very high False Citation Counts of the top few authors was mainly due to the first few highly cited articles posted on their profile which do not list them as one of the authors for the articles.
- Figure 2 shows that other than Eastern Europe the mean percentage of fake articles is less than 10% in each region.
- Regions of Eastern Europe, Eastern Asia, North America, and South Africa show very huge 3<sup>rd</sup> quartiles going up to 65 percent in case of Eastern Europe.
- In figure 3 North America, South Africa, East Asia and South America rose by a great margin. this shows that the number of fake articles in profiles belonging to these regions are lesser but the citations in those articles are very high.
- Asian and Australian regions have the lowest means and quartiles.
- While creating a profile google suggested articles by matching the authors name with the list of authors in the articles. And asked for no further verification to prove the articles belonged to the author.
- Any email address from the university the author is affiliated with can be used to verify the profile even if it belongs to a student.

## References

- Google Scholar as a new source for citation analysis <<https://www.int-res.com/articles/esep2008/8/e008p061.pdf>>
- Advisor Reviews—Standard Review Google Scholar <[https://www.researchgate.net/publication/263572400\\_Google\\_Scholar](https://www.researchgate.net/publication/263572400_Google_Scholar)>
- Using Google Scholar to track the scholarly output of research groups <<https://link.springer.com/content/pdf/10.1007/s40037-019-0515-4.pdf>>
- Google Scholar: An Internet-Based Program for Finding and Tracking Citations <<https://archives.ourheritagejournal.com/index.php/oh/article/view/4273>>
- Google Scholar revisited <[https://www.researchgate.net/profile/Peter\\_Jacso/publication/220207410\\_Google\\_Scholar\\_revisited/links/0deec528451d9f2884000000.pdf](https://www.researchgate.net/profile/Peter_Jacso/publication/220207410_Google_Scholar_revisited/links/0deec528451d9f2884000000.pdf)>
- Use of Google Scholar public profiles in orthopedics: Rate of growth and changing international patterns <<https://journals.sagepub.com/doi/pdf/10.1177/2309499017690322>>