

# Faculty Hiring Analysis

Vanessa Singh\*  
112657002

Chumki Acharya†  
112683478

## ABSTRACT

Faculty hiring in Computer Science depends on varied criteria across the country. Various factors like gender, the reputation of a professor, research area et cetera contribute to the hiring of faculty and this paper analyses the hiring patterns in well reputed Computer Science colleges. We aim at collecting data on Computer Science professors and making a visualization dashboard that analyses the relationships between hiring networks in top computer science schools across the United States. This dashboard will help colleges understand their shortcomings and hire a well rounded faculty that can provide quality education and can contribute towards excellent research.

**Index Terms:** Visualization—Visualization techniques—;— Visualization—Visualization design and evaluation methods

## 1 INTRODUCTION

It has been observed that the hiring trend of faculties has been changing rapidly worldwide. We are trying to figure out some of the hiring trends that has been observed in top 50 colleges of USA. Hiring trends differ mostly based on different factors as which University a candidate passed out from, their research papers, number of citations, h-index etc. Some other factors that also influence the hiring trends are demographic factors, gender of the candidate etc. Some Universities give preferences to candidates who have passed out from the same University, some universities prefer candidates who have completed their education from Universities that is higher in rank than that. Considering all the factors we have prepared a dataset that consists of information of Computer Science faculties from top 50 Universities of USA. We will use these data points to derive trends in faculty hiring across different Computer Science schools.

## 2 DATA

The analysis of faculty hiring patterns in top CS schools across the country required a consolidated dataset of faculty of these schools. We did come across some old datasets [3] which were more than five years old and only had about ten attributes. For these reasons we decided to consolidate our own data which will be more recent, have significantly more attributes and can be used for future research.

We have collected data for the top 50 Computer Science colleges (cite US News). These colleges have been divided into 5 categories having 10 colleges each. Category 1 has colleges ranking from 1 to 10, category 2 has colleges ranking from 11 to 20 and so on. The data contains information about a faculty's education as in details about their bachelors, masters, doctoral and post doctoral schools, gender information, research publication information such as total number of citations, H-index, I-10 index etc, their area of research. We have about 3500 faculty data points.

We have used different python scripts to generate google scholar links based on the faculty's name that points their google scholar page and then we use this page to scrape research information such

as total number of citations, H-index, I-10 index. We have tried to handle cases where multiple professors have the same name and ensured that we have correct data. In the case where the generated google scholar link is invalid we have manually searched the professor on google scholar and populated their data.

Populating the education background of a faculty member was a trickier problem. Every university has a different website which differently displays the professor data. Coming up with a standard way of scraping the data did not seem feasible, therefore, we manually went to every faculty member's page and tried to get the required data. Sometimes the education background was not directly available on the faculty member's webpage and we had to look into their CV/bio to get the data. This was extremely time consuming but we managed to extract the bachelors, masters, doctoral, post doctoral schools, area of research for professors teaching in the top 10 schools of US (Category 1 schools). We have also populated the state, country of these schools to analysis. To populate the gender of the professors we relied on names, google scholar profile pictures and pictures available on the faculty webpage.

Collecting accurate faculty data is extensive work therefore consolidating this data can also be a good open source project for beginners.

## 3 VISUALIZATION

The aim of this paper is to come up with visualizations that aid in the analysis of faculty data. This analysis will help us derive hiring patterns of different colleges.

### 3.1 Gender ratio per state visualization

While consolidating the data we observed a clear lack of female professors in the top 50 CS colleges. To analyse this better we have calculated the male to female professor ratio per state and used a choropleth map [6] [2] [8] to visualize it. States like Ohio, Wisconsin, Maryland have a very high male to female ratio indicating a small number of female professors. California makes for an interesting observation. Its male to female professor ratio is large even though the number of female professors is bigger than most of the other universities. This can be due to the fact that a lot of colleges under the University of California system fall in the top 50 colleges category and therefore have a very high number of total professors. States like

Colorado, Washington, New Jersey have a low male to female ratio even though these states do not house a large number of colleges that fall in the top 50 CS colleges category. This is an indicator of a diverse faculty. Having a diverse faculty is beneficial as it brings different points of view enabling a wholesome understanding of a subject and enables better research. An efficient way to correct this hiring pattern would be if different states begin to incentivize colleges to hire a diverse faculty.

### 3.2 Gender ratio per college category visualization

On analysing the gender ratio in every college category through a pie chart [4], we observed that colleges having a lower rank have more female professors whereas the top 20 colleges have a high ratio indicating poor gender diversity. Some mid level colleges (between the ranks of 31 to 40) have a poor gender ratio. This problem can be solved by accounting for faculty diversity while calculating college

\*e-mail: vassingh@cs.stonybrook.edu

†e-mail: cacharya@cs.stonybrook.edu

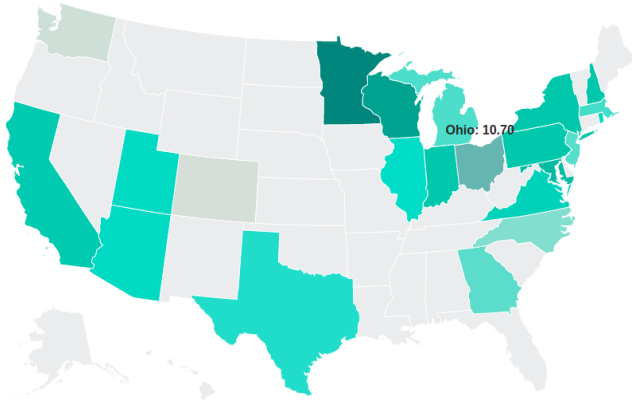


Figure 1: Gender ratio in top 50 CS colleges per state

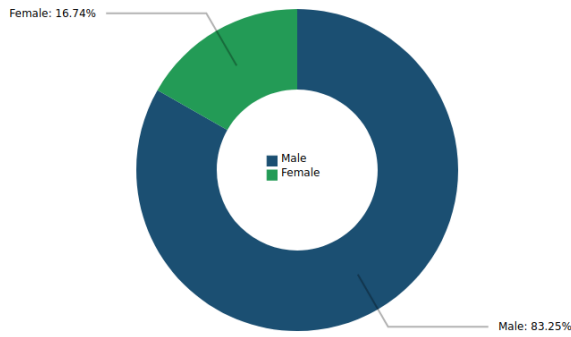


Figure 2: Percentage of male to female professors in a Category 1 college

ranks. It will act as an incentive for lower rank colleges to hire a diverse faculty and for higher rank colleges to give opportunity to upcoming professors. It will make sure that well reputed colleges have a more well rounded faculty and opportunities are given to all races and genders.

### 3.3 Comparison of H-index, Citations in each college category visualization

To analyse the merit of professors working in a college we are analysing the number of citations to their name and their H Index on Google Scholar. This helps in establishing the reputation of the professors working at a college. We have grouped the colleges in to five categories based on rank and then taken the median of citations and H-Index of the professors that fall in that category. For citations, we observe that higher rank colleges have higher median citations and lower rank colleges have lower median citations. The same trend is observed for H-Index. This trend makes perfect sense as professors at well ranked colleges would have done better research. and there would be less professors having less than ten citations. H-Index is a good indicator of a professor's work as it can be used measuring both quality and quantity of the research. It helps in estimating both productivity and impact of the research. I10-Index indicates the number of papers with more than and equal to ten citations and is only used by Google Scholar as a metric, therefore, we have not analysed it. Therefore, professors who want to move to higher ranking college should ensure high citations and h-index as it a factor in hiring.

Table 1: Male to Female Ratio per State

State	Male	Female	Ratio
Colorado	54	18	3.0
Pennsylvania	232	35	6.628571428571429
New Hampshire	27	4	6.75
Rhode Island	20	4	5.0
Washington	59	19	3.10526 31578947367
Ohio	107	10	10.7
Wisconsin	100	11	9.090909090909092
Maryland	99	13	7.615384615384615
New Jersey	84	20	4.2
Massachusetts	356	83	4.289156626506024
North Carolina	144	37	3.891891891891892
California	588	92	6.391304347826087
Indiana	60	9	6.666666666666667
Illinois	256	54	4.7407407407407405
Texas	132	30	4.4
Michigan	17	4	4.25
Arizona	79	16	4.9375
Georgia	25	6	4.166666666666667
Virginia	98	17	5.764705882352941
Minnesota	64	6	10.666666666666666
New York	205	30	6.833333333333333
Utah	45	9	5.0

Table 2: Male to Female Ratio per College Category

Category	Male	Female	Ratio
1 (College Ranks 1 to 10)	726	146	4.972602739726027
2 (College Ranks 11 to 20)	522	85	6.141176470588236
3 (College Ranks 21 to 30)	367	88	4.170454545454546
4 (College Ranks 31 to 40)	775	104	7.451923076923077
5 (College Ranks 41 to 50)	461	104	4.4326923076923075

### 3.4 Comparison of H-index, Citations in each college visualization

We have calculated the median citations and median h-index in each college to analyse the reputation of the professors that work there. This helps in analysing the rank of the college with respect to the research work being done there. Parallel coordinate [9] graph has been used to visualize this better. A school having high number of citations should also have a high h-index as it indicates higher quality of research. Professors doing new and impactful research should be given an opportunity at higher ranked colleges. The opportunities given out by a university to upcoming professors should be weighed in while assigning a rank to a university. The faculty at high ranking colleges like Yale, University of Washington is observed to have a significantly high number of citations with a high H-index. Some of these universities like UC Berkely and UCLA have faculty which have a large number of citations and an excellent median h-index. This can be the reason why a high number of beginning level professors happen to have completed their doctoral studies from UC Berkeley. Colleges like Princeton and Stanford have a high rank, a high number of citations but a low H-Index indicating that they should give opportunities to professors having a better profile and a better H-Index.

### 3.5 US/Non-US Professors Visualization

We have made a visualization that analyses the place where different professors have completed their education level from. This is an attempt to understand the diversity of faculty in each school. We observe that most of the faculty working in CS schools have completed their Bachelors degree in the US but the number of professors holding an out of country Bachelors is also high. For Masters degree,

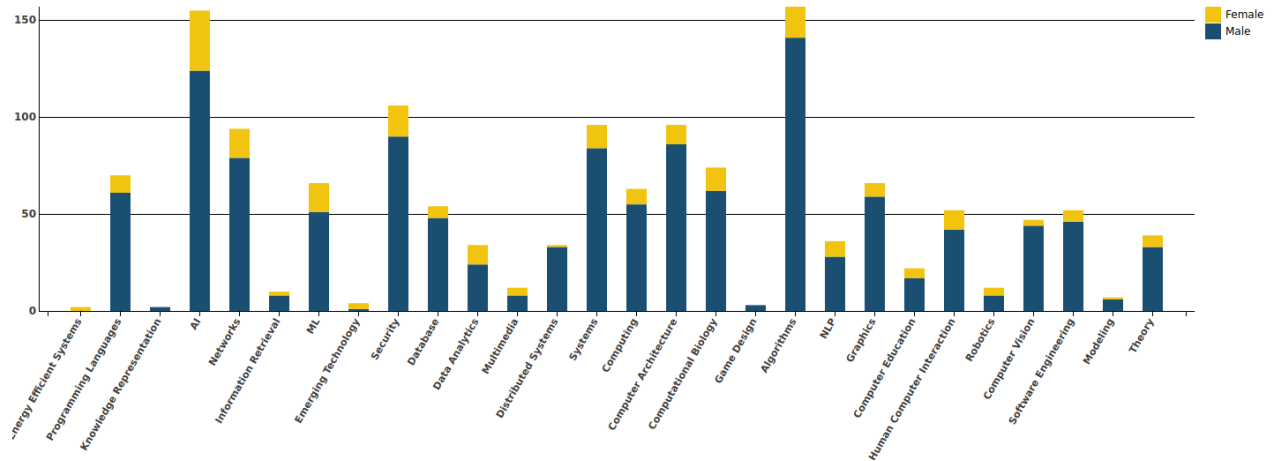


Figure 3: Count of male and female professors in a CS research area

Table 3: Top 4 countries where faculty has completed their education from

BS	MS	Phd
India	China	Canada
China	Canada	Israel
Canada	Israel	Switzerland
Israel	India	France

we observe a different trend, most of the professors have a Masters degree from the US itself and only a handful of them have an out of country Masters. A similar trend is observed for Phd - a large number of professors pursue their doctoral studies in the US. Only a small number of professors come to the US to pursue their academic career after completing their doctoral studies outside the US.

### 3.6 Popular countries for out of country professors visualization

On analysing professors who have completed one of their degree levels outside the US we observe that some countries emerge as favourites. We have complete professor data for top 10 colleges and on analysing that through a horizontal bar chart [5] [1] we find that India, China, Israel, Canada emerge out to be favourites for Bachelors and Masters degrees. This means a lot of professors from these countries who come to the US to pursue a career in academia have completed their their bachelors and masters in these countries. Favourite places completing doctoral studies are Canada, Israel, Switzerland etc. On analysing the hiring pattern of each college, some colleges have a higher number of professors who have completed graduation levels outside the US indicating that these colleges welcome professors from all over the world. Massachusetts Institute of Technology, University of Illinois - Urbana Champaign have a large number of professors that have completed their degrees out of the US. But this can also be a consequence of these colleges having a large number of faculty. More data is required to yield a clear analysis result and then analysis can also be performed on which college likes to hire professors from which country.

### 3.7 Gender ratio in each research area visualization

Analysing gender ratios helps us in identifying the lack of diversity in particular research fields. To visualize this we have used stacked bar plots [7] [10]. Not a large number of faculty is interested in fields like Game Design, Multimedia, Robotics, Modeling. Research

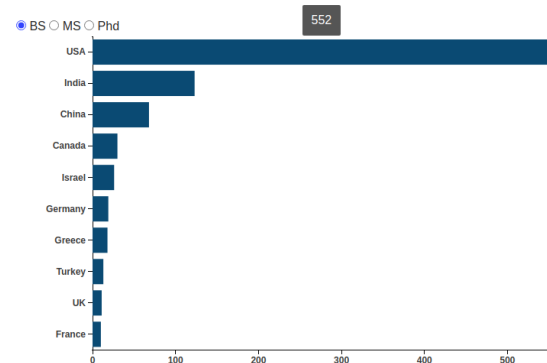


Figure 4: Popular country for pursuing Bachelors degree for US professors

Table 4: Research areas with the lowest female to male professor ratio

Research Area
Game Design
Knowledge Representation
Distributed Systems
Computer Vision
Algorithms

in these fields should be actively encouraged by giving grants and creating opportunities at the college or state level. Moreover, the gender ratio in these fields is poor. Game Design, Computer Vision have a small number of female professors. A diverse set of professors is interested in pursuing research in AI as it has a good female to male ratio. On the other hand, a lot of professors are interested in researching in Algorithms but only a handful of them are female. Colleges should encourage active hiring in fields where there is a clear lack of professors to prevent research in that field from dying out and increase the the number of women in computer science.

## 4 RESULTS

After analysing the given data we have observed a few key points that have been summarised below.

- There is a clear lack of faculty diversity in some states. States should work on promoting diverse faculty hiring.
- Diverse faculty hiring should be weighed in while calculating college ranks.
- Professors should work on doing quality research which will reflect in getting more citations and a higher h-index. These factors count when getting hired in a high ranking college.
- Colleges should also look into hiring well rounded senior professors from all over the world. Only a small number of professors who complete their doctoral studies abroad work in a US University.
- Research in less popular fields like Game Design, Emerging Technologies should be encouraged. Universities should open up faculty positions for professors who have specialized in these subjects, otherwise academic research in these fields will gradually die out.
- Even in popular fields like AI, Algorithms the female to male professor ratio is low indicating that only a handful of women pursue a career in academia. Universities and states should actively ensure that a good number of female professors are hired for a research area and more opportunities are given out.

## 5 CONCLUSION

Consolidating faculty hiring data is an extensive task and requires a lot of time. Scripts for this purpose cannot be automated completely and manual intervention is required. Therefore, making it an open source project would be beneficial for accurately analysing the hiring trends. Analysing faculty hiring can be beneficial for both colleges and people wanting to pursue a career in academia.

## ACKNOWLEDGMENTS

The authors wish to thank Ayush Kumar and Klaus Mueller for guiding us through this project.

## REFERENCES

- [1] <http://bl.ocks.org/curran/8ac62d283a7aa6e95de8>.
- [2] <http://bl.ocks.org/lwhitaker3/e8090246a20d9515789b>.
- [3] [http://cs.brown.edu/people/apapouts/faculty\\_dataset.html](http://cs.brown.edu/people/apapouts/faculty_dataset.html).
- [4] <https://bl.ocks.org/antonorlov/6b42d8676943cc933f48a43a7c7e5b6c>.
- [5] <https://bl.ocks.org/hrecht/f84012ee860cb4da66331f18d588eee3>.
- [6] <https://bl.ocks.org/wboykinm/dbbe50d1023f90d4e241712395c27fb3>.
- [7] <https://gist.github.com/michellechandra/0b2ce4923dc9b5809922>.
- [8] [https://www.d3-graph-gallery.com/graph/barplot\\_button\\_data\\_simple.html](https://www.d3-graph-gallery.com/graph/barplot_button_data_simple.html).
- [9] [https://www.d3-graph-gallery.com/graph/parallel\\_custom.html](https://www.d3-graph-gallery.com/graph/parallel_custom.html).
- [10] <http://www.d3noob.org/2013/01/adding-tooltips-to-d3js-graph.html>.