Interdisciplinarity and Success in Science

Ahmed Alrabiah *For. Phys465-0 instructed by István Kovács*

Northwestern University

Overview

In literature usually citation networks are formed and rarely co-authorship networks are discussed. The main difference is in the first papers are the nodes and citations are the edges, while in the latter authors are the nodes and the edges are the collaborations between the authors. The authorship network may provide insights into the authorship patterns and the impact of interdisciplinary research on the success of researchers.

Tools

The tools I have used in this study are python with some of its packages specifically NetworkX that is used visualize and extract the properties of networks.

Datasets

In this study, I have used two datasets:

(1) The first dataset is provided by
"Co-Authorship Network of 402.39K authors on Google Scholar"
https://github.com/chenyang03/co-authorship-network
In which after cleaning, it contains $N = 65495$ nodes and 144475 edges. The network consists of three fields (i) computer science, (ii) biology, and (iii) sociology.

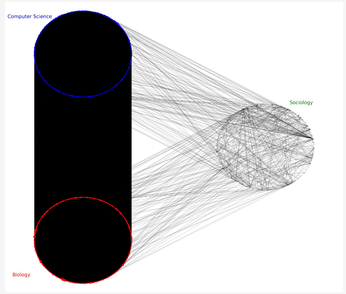
(2) The second dataset is provided by
Arxive https://www.kaggle.com/datasets/Cornell-University/arxiv
However, it is a list of papers in arxiv's different categories. The author's need to be extracted from the papers. The difficulty is in collecting the authors' fields and citations. For this I have used Scholarly package in Python to sparse through google scholars to collect the needed information. The package is a bit slow in collecting authors' information and hence in this dataset I have only a few number of authors. From 150 physicists, the total random number of nodes collected is $N = 423$ nodes with $\mathrm{L} = 1065$ edges. This is done just for the interest to see the network of physicists and with whom they collaborate mostly.

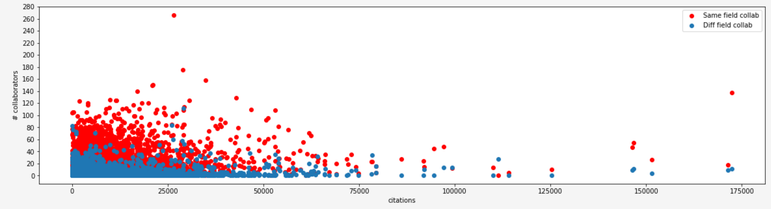
Results

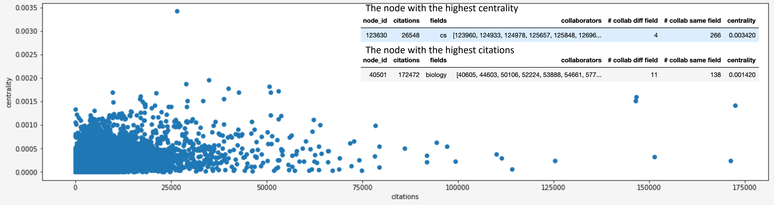
The following subsections are the results from the two datasets

(1) CS, Biology, and Sociology set

The network visualization



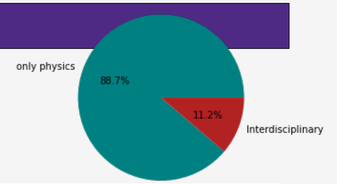


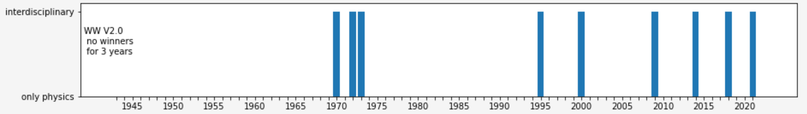


Average out-of-field coauthors	0.6939
Average in-field coauthors	4.7311
Top 100 citations average out-of-field coauthors	6.6700
Top 100 citations average in-field coauthors	28.990
lowest 100 citations average out-of-field coauthors	0.1100
lowest 100 citations average in-field coauthors	0.9000

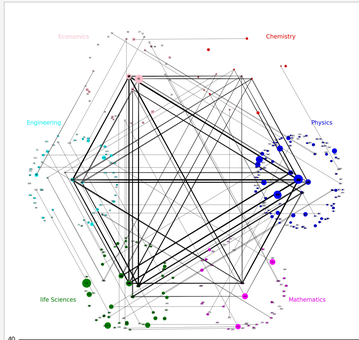
(2) Biased Physics set

Nobel prize in Physics from 1943 - 2022

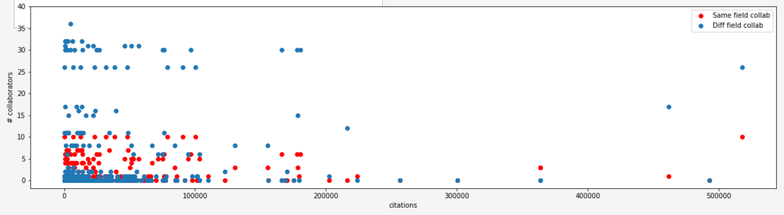




The network from the randomly selected physicists of the second dataset



Avg out-of-field coauthors	3.9385
Avg in-field coauthors	1.0969
Avg phys out-of-field coauthors	2.9533
Avg phys in-field coauthors	1.3333
Top 10 cutations avg out-of-field coauthors	4.1000
Top 10 cutations avg in-field coauthors	2.2000
lowest 10 cutations avg out-of-field coauthors	2.8000
lowest 10 cutations avg in-field coauthors	1.0000



Conclusion and Further Work

I was not able to come out with a conclusion from the given data. It was interesting how the share of the Nobel prize in Physics with non-physicists seem to increase in later years. The main thing that will be done, is to write an efficient code for data collection. We may use the data to study social biases in the scientific field. If possible, study the time evolution of scientists towards success and their collaboration.