

CS101 Project-2

April 2024

1 Introduction

In this project we were given a dataset of the experiment we performed in which we were asking each other some question and if we were impressed by other person's answer then we were writing his name in the google form. Now we have that data and we have to plot a directed graph of the data with each node representing student(via entry number of each student as identity) and direction of the edge will be from the person to the person whose name is written for e.g.- If X is impressed by the Y's answers then there will be a edge from X to Y having head towards Y.

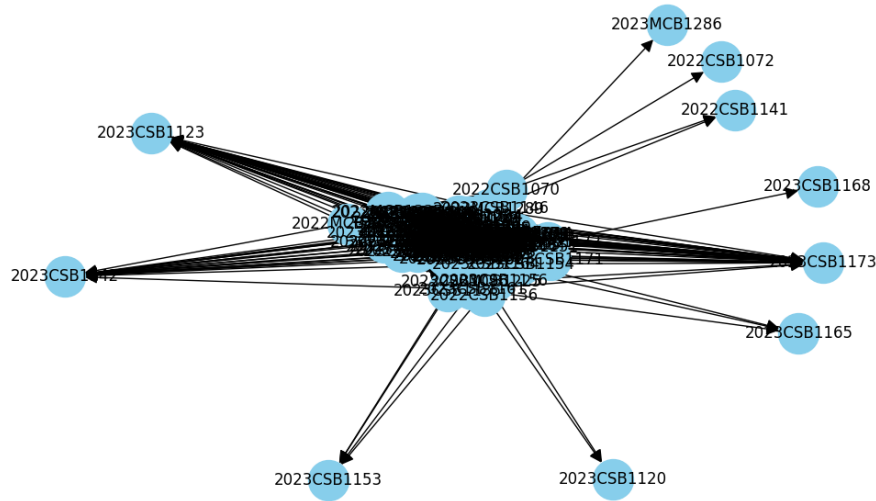


Figure 1: Graph from the given dataset

2 Random Walk With Teleportation

In the first part of the project we were asked to perform the random walk on the path with teleportation and to decide the top-leader board for which we had to perform the random walk which is explained as following:- First we have to assign zero points to each node and then randomly choosing any node and randomly choosing any number between 0 and 1 and if that number is $p=0.85$ and the chosen node has out edges then increasing the points by 1 and then choosing out of its neighbours and doing the same for a big number of times(in my case I have done it for 1000000) times and then points wise sorting them in descending order and finding the top-leader board which is entry number 2023CSB1091 , this entry number node received the maximum points in overall random walk I performed. I also verified the overall random walk result with the pagerank(inbuilt function provided by networkx library) and I was getting almost the same entry numbers but some were one or at max two place up or down as compared to the position of pagerank which is due to the sinks(no outgoing edges) explained in the videos provided by the Instructor.

3 Missing links using Matrix method

In this we are using Matrix method to find missing links in the graph. The missing links are if any person didn't met someone but if they would've met he would have been impressed by him. For this first I created a adjacency matrix of the graph which is the matrix of 0 and 1. I assigned a number to each node from (0 to 142) and then element at adjacency matrix's i th row and j th column is 1 if there is an edge from i number node to j number node or 0 if not. I iterated over the matrix and if element at i th row and j th column is 1 and element at j th row and i th column is 0 then replace 0 with -1 which means that two people met but one was not impressed then there can't be a missing link from j to i . We were using the linear combinations approach to express the row of each 0 element in terms of other rows and the finding the coefficients in the method I have described further. Now after we are gonna create one copy of the matrix. Now iterating over the matrix if the element is 0(i th row and j th column such that $i \neq j$) then deleting i th row and j th column temporarily and then remaining matrix is denoted as A and that j th column (without that 0) is denoted as B and the single column of coefficients is X and it is of the form

$$AX=B$$

Now we will be using `numpy.linalg.lstsq()` function of python to find the value of coefficients and then multiplying the coefficients with the corresponding elements of the i th row will give us the value which we will keep in the copy of the matrix and the operations will be performed on the original matrix. After that I have kept a threshold value 0.4 and if the value is below or equal to 0.4 then assigning it 0 and if the value is greater than 0.4 then assigning it 1. joining the nodes from i to j for each value 1 and then we will get the graph with missing

links. I have attached the image of the graph I have got after these operations.

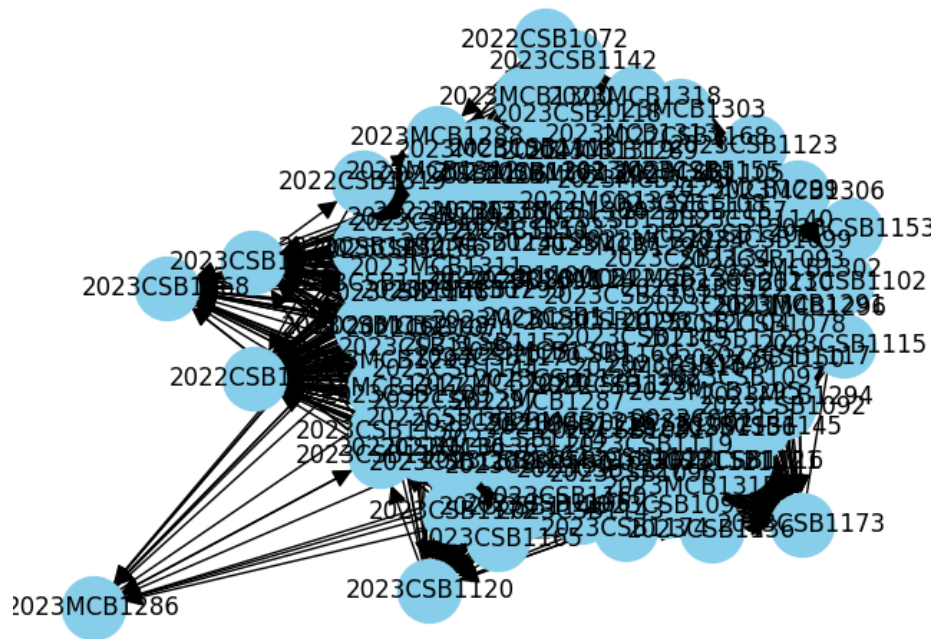


Figure 2: Graph after Matrix Method