# Project 2 Report

Arpit

April 24, 2024

# Problem 1: Identifying the Leader in a Social Network

In this problem, we were given a unique task. We had a group of 143 students who interacted with each other randomly. If a student was impressed by another, they would write down the name of the person who impressed them. This interaction is unidirectional and the data was collected via a Google Form, which was then exported as a CSV file.

The problem posed to us was to identify the top leader in this group. To solve this, we decided to model the interactions as a graph and run a random walk algorithm with teleportation on it. The idea behind this approach is that the person who is most often "landed on" during the random walk would be the most influential or impressive person in the group.

To ensure the accuracy of our results, we ran the random walk for $10^6$ steps. This large number of steps helps to ensure that the walk has a chance to visit all nodes multiple times, thus giving a fair chance for every individual to be chosen.

# Methodology

We represented the interactions between students as a directed graph, where each student is a node and an edge from node A to node B represents that student A is impressed by student B.

We then ran a random walk with teleportation on this graph. At each step, the walker has a chance to either follow one of the outgoing edges (representing being impressed by someone) or "teleport" to a random node (if that node is not impressed by anyone).

# Results and Discussion

The result came out to be AADIT MAHAJAN , entry no. 2023CSB1091.On running the program multiple times, we found that the result was consistent and the same person was identified as the leader each time. This indicates that the random walk algorithm is working as expected and is able to identify the most influential person in the group.

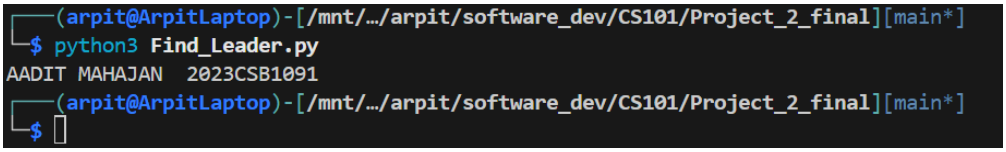


Figure 1: Output in terminal

# Problem 2: Recommending missing links

In the second part of this project, we were tasked with recommending missing links in the social network. This problem is essentially about predicting future interactions between students based on the existing interactions.

## Methodology

To solve this problem we used linear regression. We first created an adjacency matrix of the graph, where a 1 in the matrix represents that the student in the corresponding row is impressed by the student in the corresponding column. Then predicted the values which were initially 0 in the matrix.

There are several methods to solve this but we will use the basic idea of linear regression to predict the missing links. Since there is no disliking values in the impression netework, we have only 2 possibilities, either the student is impressed by the other student or we do not know what would have happend when they intracted. So, we there is a possibility that the graph would have been completed if everyone intracted with everyone. So, we will just complete the missing link in given data and only complete 30 imperssions of each student(if s/he has less than 30 impressions and did participate in the survey).

To do this, we used the method of matrix. First we created adjacency matrix of the graph and for each row, we completed that one. We represented the row in terms of linear combination of other rows ( by looking only those columns which have value 1 in the row ). Asuming that the row is linar combination of others, the complete row can be written as linear combination of other rows ( for incomplete values too). Then , sorting the new values and putting heighest being one ( representing being impressed) until there are 30 impressions or no more impressions possbile. This completes the matrix after doing the algorithm for all rows.

## Results and Discussion

The output of the program is a csv file created in same directory named "impressions.csv". This file comtains the completed survey data with 30 impressions of each student.

# Problem 3: Choosing problem ourselves

## Problem Statement:

As the survey was conducted on people who knew each other before the experiment, the experiment was not completely random. So from the given data, predict the top friends of each student in their general life.

## Solution

### Idea

Since no feature of the student is given ( like their hobbies, address, origin, gender, nature ), we will use their property of impressing others and being impressed by others to predict their friends. This can be done simply using very basic intutions and ideas of vector calculus and linear algebra without any regrous mathematical proof.

### Methodology

We will define two features: impressing vector and being impressed vector. So, to calculate the impressing ability of a node,

We have to make a vector with dimension equal to number of students. The value of impressing someone will be more for those who are directly impressed by him/her and less for chains. ( If a person is directly being imperssing by him/her and also got impressed friend, s/he is even more impressed as common intrests are even more). so start by giving the node 500 points. Then divide those points equally to all the people who found him/her impressive. Do this agian and again ( 3 iterations). [ With very high probability the whole graph will be completed in 3 iterations]. Now , for each dimention, the value in the defined impressing vector will be sum of all points ever received by the dimension's corresponding student.

Similarly we can find the being impressed ability of a node. ( further the adjacency matrices are just transpose of each other, so we can use the same code for this too).

Now, to compare two students, we have to calculate the norm or the sum of square of difference of the two vectors. The smaller the norm, the more the two students are friends.( both vector are normalized with respect to each other so both have equal weightage.)

### Results and Discussion

The output of the program is a csv file created in same directory named `Best_Friends.csv`. This file contains the top 8 friends of each student in the given data. The data is in general satisfying to the real life data as the student who are prediceted to be friends in the data are mostly friends in real life too.