CS101 PROJECT-2 Q2

Name - Nachiket Avachat
Entry number - 2023CSB1106

We are provided with a **Social Impression Network**. We should plot the graph and study its various properties.

We are expected to plot a directed graph. To create the graph, firstly, I manually made changes in the file since all the entries were not in the same format (some were in the form of email ID while some were in the form name+entry no.). Each entry was converted in the form of an entry number. Doing so, it became easier to plot the graph. I have attached that file in the project submissions. If A is impressed by B, we will add a directed edge from A to B. In this way, a graph is plotted where nodes are the entry numbers of students while edges represent the impressions.

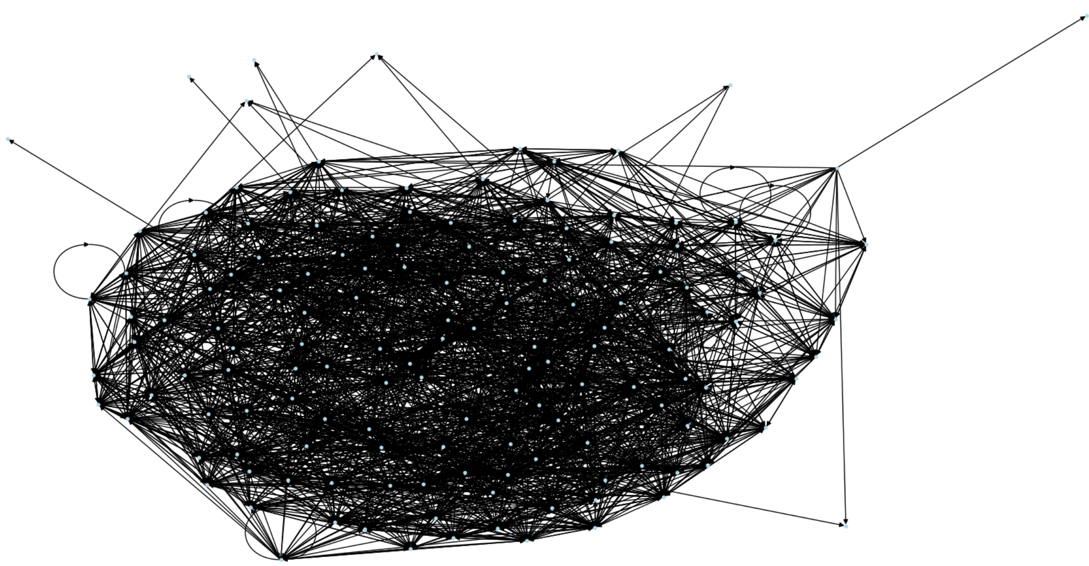


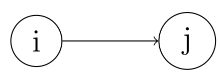
Figure 1: PLOTTED GRAPH

QUESTION II -

Finding Missing Links

Problem Statement : Recommend missing links using the matrix method explained in the class.

Logic : In our social impression network, people had limited options to fill. Also it is obvious that everybody wasn't able to meet each person. So, we have to determine if person A met person B would he like him or not. To solve this problem, I made the adjacency matrix of a graph whose a_{ij}'th entry is 1 if i points to j; otherwise, it is zero.



$$a_{ij} = 1$$

We will select a zero from the matrix and see if it becomes 1 or not.

As discussed in the class, to determine this for a_{ij} th cell, we will try to represent row i as the linear combination of the remaining n-1 rows. If we get the coefficients we will multiply them with the respective column j to get the expected value of our a_{ij} th cell.
Let's take an example where the shape of our adjacency matrix is (5,5).

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & \boxed{0} & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Let's select cell a_{32}, the entry of which is zero. Now we have to represent R_3 as the linear combination of remaining rows.
$\therefore \; R_3 = c_1R_1 + c_2R_2 + c_4R_4 + c_5R_5$

While finding the coefficients we cannot consider the 2nd column since we have considered the value of a_{32} as unknown. Thus, we will eliminate the 2nd column. Store the remaining row 3 in a matrix (say B). Therefore, $B = R_3$ except a_{32}. Thus the

shape of B is (1,4).
We will store the matrix except R_3 and C_2 in another variable (say A).
$\therefore A = R_1 + R_2 + R_4 + R_5$ (except column 2 entries of each row). Thus, the shape of A is (4,4). Let the matrix of coefficients be $X = [c_1 c_2 c_4 c_5]$ To find X we have to solve, $XA = B$ i.e. $X = BA^{-1}$. The problem here is that in most cases, we find A to be singular. Thus, A^{-1} is undefined.

To solve this, we use the least square function of numpy. Since there can be more than one solution, the function gives us the best fit for a set of data. It minimizes the Euclidean 2-norm $||b - a||$. If there are multiple minimizing solutions, the one with the smallest 2-norm $||x||$ is returned.

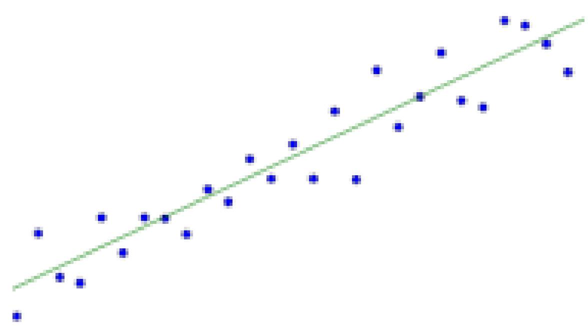


Figure 2: Visual representation of least square method

Once we obtain X, we will multiply it with the 2nd column to find a_{32}. Now, the value of a_{32} may not be necessarily 0 or 1.
Thus, to determine whether to put a_{32} is 1 or 0, some threshold needs to be set. By the method of trial and error, I determined this as if $|a_{32}|$ comes to be greater than 0.8, change a_{32} to 1 else, keep it zero.
Similarly, we will conduct these steps on all entries of our adjacency matrix and observe the changes. In this way, we can find the missing links on our graph.