CS101: Discrete Mathematics
PROJECT-2

Name: Som Nainwal
Entry Number: 2023CSB1163

April 24, 2024

Question-2

Problem Statement

In this problem, we have to find missing links in the impressions network dataset stored in the CSV file.

Adjacency Matrix
If there is no connection between two nodes, a zero is inserted in the adjacency matrix at their respective positions otherwise a one is inserted . Our objective is to determine whether this zero will stay as 0 or change to 1.

Solution:-

To begin, we will gather the data from the impression network, which is stored in a CSV file. Next, we will construct a graph using `networkx`.This graph will have directed edges connecting nodes where one student was impressed by another in the experiment.

Then, we will calculate the adjacency matrix of the graph. We estimate the missing connections by using linear regression based on existing connections. If the estimated value is significant, the missing link is added to the graph. Finally, the code computes the density of the original and updated networks.

Why are we employing Least Square Method

We are using the least squares method because the problem involves solving matrix equations such as

$$x^*A=B$$

Now, to avoid the cases in which the above equation gives no solution, we are using the method of least squares.It also provides an accurate and simple approach for estimating missing values, making it suitable for practical implementation in network analysis.

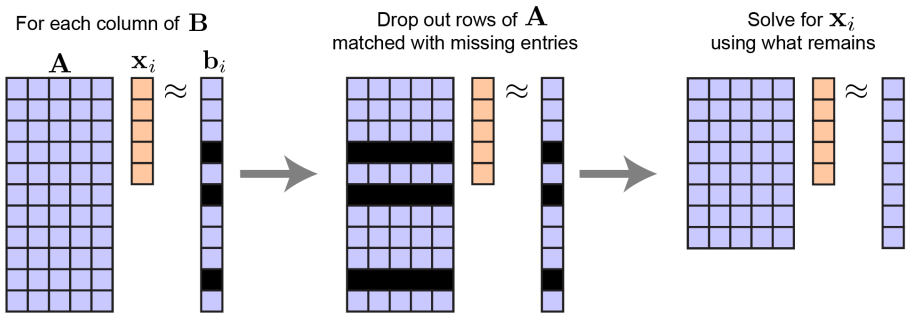

Figure 1: Least squares method of solving matrix equation

Algorithm

1. Read data from the CSV file and make a directed graph.
2. Calculate the adjacency matrix.
3. For every zero entry in the matrix that suggests a possible missing link, we will use a technique called least squares estimation(linear regression). This method takes into account the current structure of the network.
4. If the estimated value satisfies a certain threshold, we will introduce the missing link.
5. Calculate and compare the network density before and after introducing the introducing missing links.

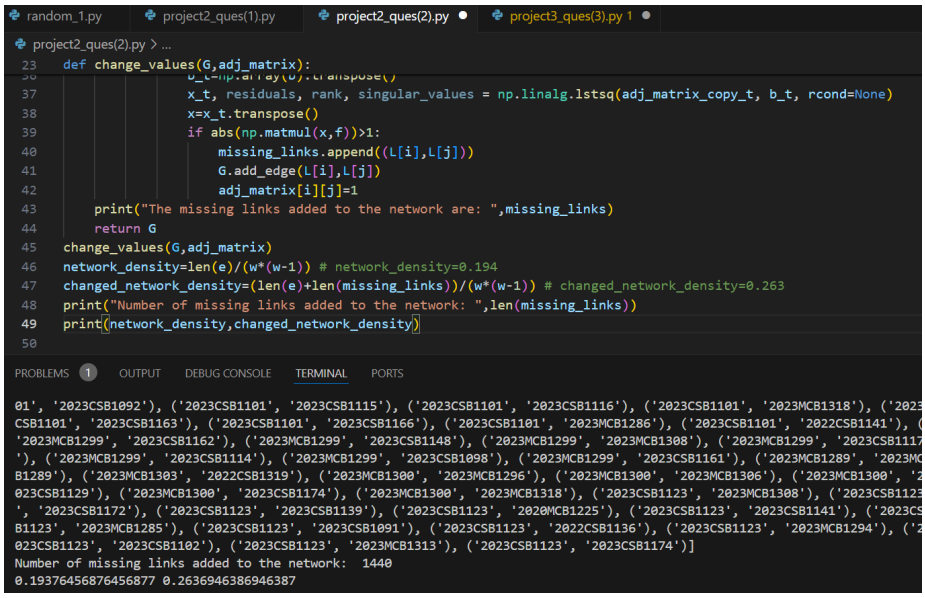

Figure 2: Snapshot of the code and the output

The number of missing links added depends on the threshold we set.
The code displays the number of missing links added, the network density before and after adding the links, and the list of missing links that were added to the network.
Number of missing links added = 1440
Initial Network Density = 0.193
Final Network Density = 0.263