

PROJECT-2

FOUNDATION OF COMPUTING

OVERALL STATUS:

For this project, we examined the project's problem statement and determined the project's requirements. First, we looked at Google Co-lab, as though we were going to write the code in Co-lab. We were both new to python, so we learned the basics by watching videos and taking foundations of computing seminars. For learning purposes, we also used the Python documentation.

Next we have started working on the importing of the data of the Data set of the given files into the project notebook of the google co-lab. And mounted to the google drive and given the file location for the data to be read.

In the next step we have taken the file and selected the the city given that is saint Charles, and taken the necessary data from the file business.json file and extracted the necessary data from the .json file. Then extracted the necessary data from all the given files of the review.json and user.json and connected the nodes of the users. The data we got from the preprocessing we have connected to the edges for the user id. All the preprocessed data we got have connected by using the networks module. After taking that model into consideration we have created the layer degree distribution graph where the shows the number of nodes and degree in the

graph. After taking this into consideration we have created the histogram for the top 5 business and for the top 5 users.

FILE DESCRIPTIONS:

The given data set are in the .json file and we have loaded the file and read it with the json.loads(line) which uses the to read the file line to line. We have used the matplotlib, plotly.express, json, geopandas, random, networks, os, statistics, collections, csv, sys, pandas, network.algorithms.flow some of the files for the connecting of the nodes and edges for the nx-graph and the network of the modules. We have used for the matplotlib for the building of an histogram.

We have the .json file we have read the file and load and extracted the necessary data of the given city.

DIVISION OF THE LABOUR:

We have equally divided the work for the team members and the solving and the analysis part for the project 2. We have shared the ideas and analysis of the project-2 by the both and making a note what are required for the project-2. Every day we have working for the two-hours on the project and making the project moving forward. We have spent approximately 60- hours on the project and we have discussed about the problems encountered on the working of the project.

PROBLEMS ENCOUNTERED AND HANDLING:

1.ANALYSIS RELATED:

we have the json files and the loaded in to the drive and copied the path of the file and loaded it into the google co-lab and using this co-lab we have taken the required city for the pre-processing of the data. The given three files are in the json and we have pre-processed all the files. Firstly we have taken the business.json file for the required city and got the values of the city. We have created an empty list and mapped the values of the to the business id and eliminated the duplicate values in the business id. From the business id we have appended the values in to the user id.

2.LOGICAL RELATED:

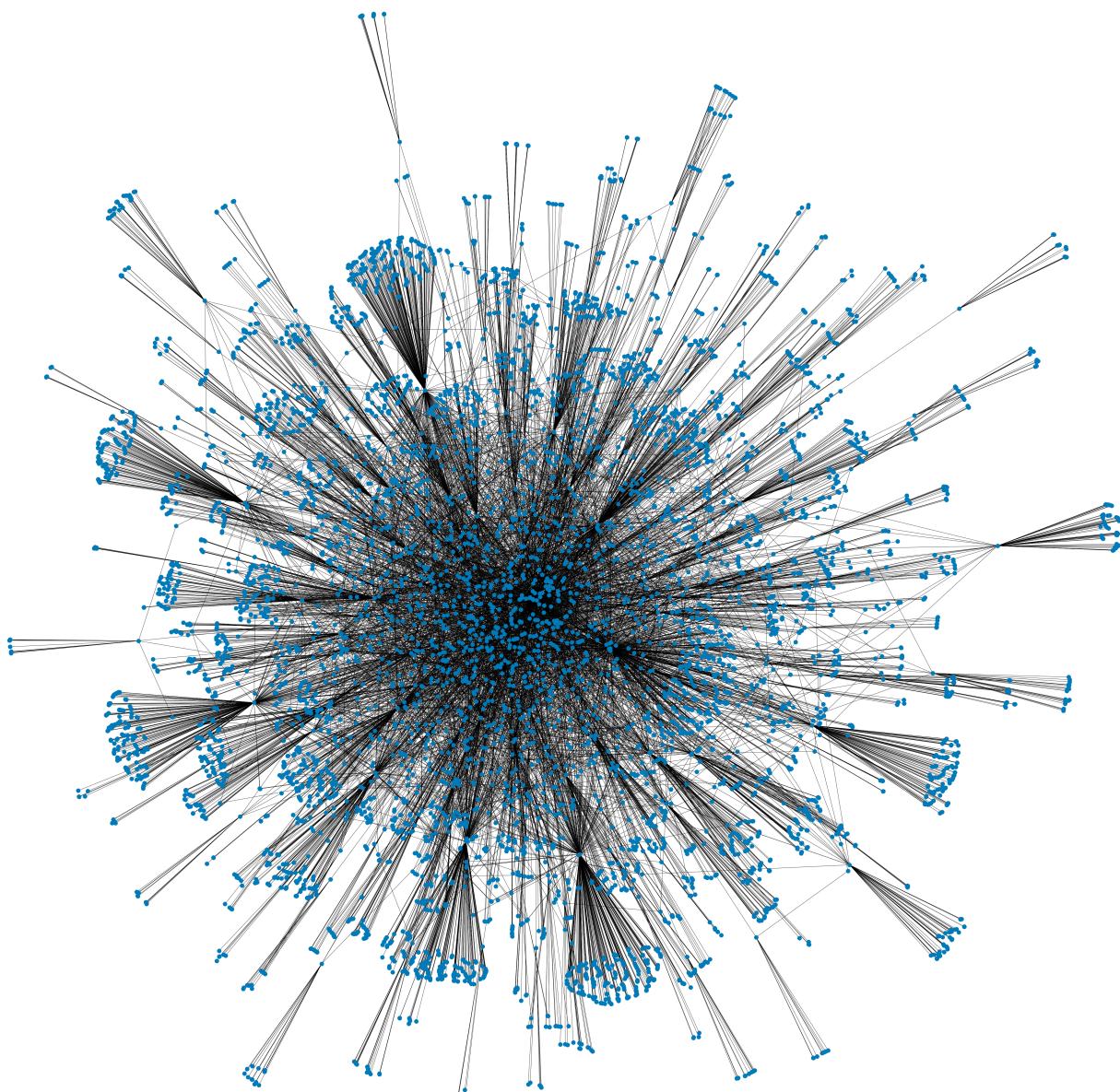
For the average of the values in the we have taken the stars values as an average and taking these values for the getting of the average stars and append to the business id from there we get the values of the user id and taking the values from the review file and append it to the business id and making these values from the user id. From the business id we have taken the values from all the reviews and append it to the connected to the nodes.

3.LOGICAL RELATED:

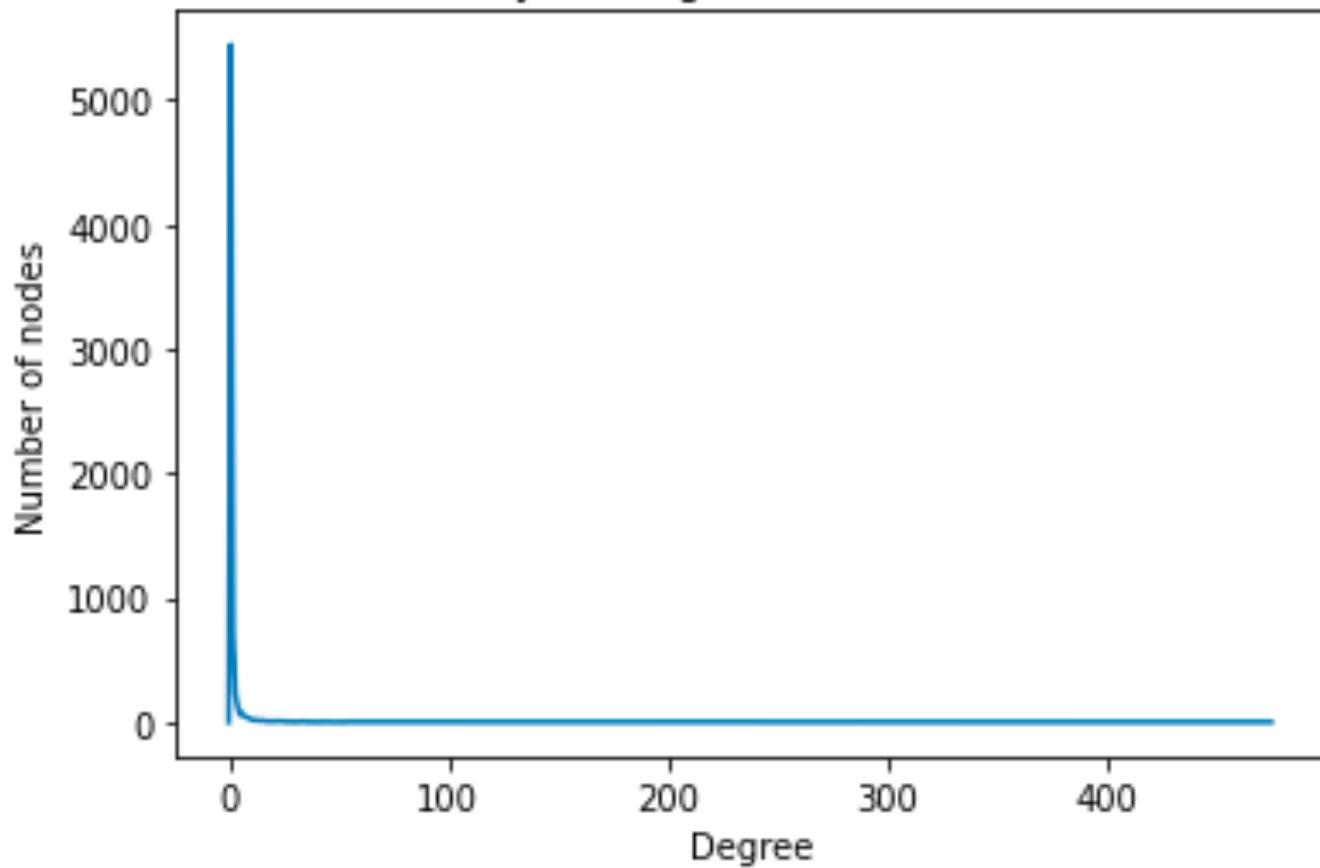
For the nodes we have connected the edges with all the nodes and got connected to with the each other using the networks module and connected to the edges with the nodes. We have created the the nodes for all the edges by the connecting all the nodes. By the using of the network module we got the connected all the edges. Bu using the networks module we have connected all the node with the used Id and mapped the values to the user id and making the nodes.

PRE-PROCESSING:

Output:



layer0 Degree Distribution



Number of nodes in the layer 0 = 7052

Number of edges in the layer 0 = 10117

All the nodes and the edges are connected to each other. The above graph shows the layer0 degree distribution of the nodes and the degree.

ANALYSIS-1

TEAM: 28(for reference)

VENKATA GOPI VALLURI

BAVINENI BHEEMESWARA RAO