# 19ZO02-Social and Economic Network Analysis

Project Report

**Topic: Credit Card Fraud Analysis**

**Team Members:**

19Z203 ANUHARSHINI B

19Z206 CHARANYA AG

19Z239 SAMSITHA BANU S

19Z248 SUBASHININ B

# BACHELOR OF ENGINEERING

**Branch: COMPUTER SCIENCE AND ENGINEERING**

Of Anna University



PSG College of Technology Coimbatore – 641004

# Problem Statement:

Objective is to analyze and detect Fraudulent transactions among Legitimate ones. For this we make use of dataset from Kaggle for representing transactions as a graph with

* + - Customers and merchants serve as Nodes
    - Transactions serve as Edges.
    - The Edges are weighed according to the amount of the transaction
    - label of edge as "Fraudulent."

# Dataset Description:

* The dataset is comprised of simulated card transactions. It contains both genuine and fake exchanges of clients with a pool of Vendors.
* This dataset was extracted from the Kaggle database.
* Dataset Attributes:
  + - index - Row ID
    - cc\_ num - Credit Card Number of the customer
    - merchant - Merchant Name
    - amt-Transaction Amount (in USD)
    - is\_ fraud -Target Variable. (0=genuine, 1=fraudulent)
* Dataset: <https://www.kaggle.com/kartik2112/fraud-detection>

1. **Tools used:**

* **Python:** We have used the Python Language for the coding part because of its User-friendly Data Structures.
* **NetworkX**: The Python programming language's Network X package is used to create, modify, and research the structure, dynamics, and functionalities of complex graph networks.
* **Google Colab**: Google Colab is particularly well suited to machine learning, data analysis, and education since it enables anyone to develop and run arbitrary Python code through the internet. The cloud-based, open-source Jupyter Notebook, which provides free computing resources, is essentially another name for Google Colab. Python code may be written and run through a browser using Google Colab.

**4.Challenges Faced:**

* There was some error in the code, so we were unable to visualize the graph Initially.
* Even though our code was debugged and ran, the expected output in terms of degree and centrality were all 0.
* Since we were new to Network X, it was tiring to understand and visualize the graphs.

# 5.Contribution:

|  |  |
| --- | --- |
| **Name (Roll number)** | **Contribution** |
| ANU HARSHINI B(19Z203) | Documentation and Statistical analysis |
| CHARANYA A G (19Z206) | Collecting dataset |
| SAMSITHA BANU S (19Z239) | Graph visualization and coding |
| SUBASHINI B (19Z248) | Graph analysis and coding |

**Annexure – I:**

https://github.com/clouding-land/SENA-PROJECT

**Annexure- II:**



Figure 1: Loading the Libraries and describing the dataset information

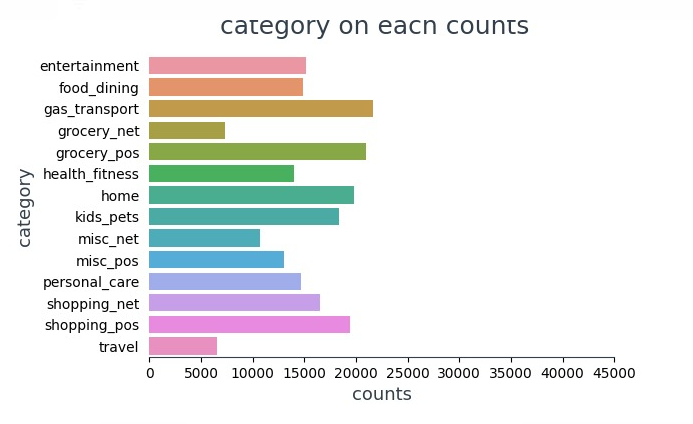


Figure 2: Analysis of Transactions according to the Feature called Category.

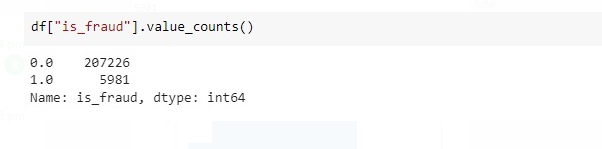


Figure 3: Total count of Fraudulent and legit nodes

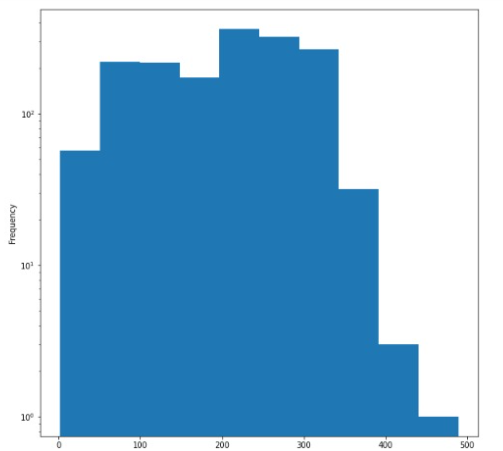


Figure 4: Histograms of node degrees using inbuilt function nx.degree()

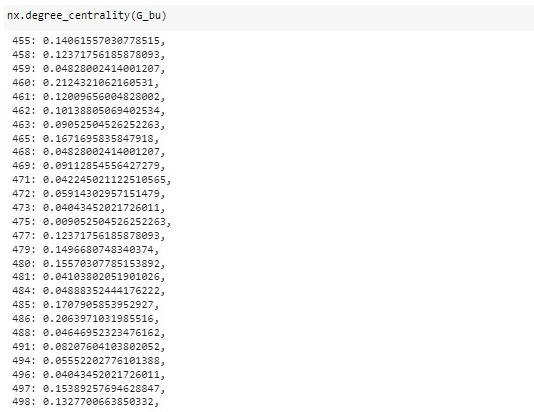


Figure 5: Finding Closeness and Degree centrality using nx.degree\_Centrality and nx.Closeness\_Centrality

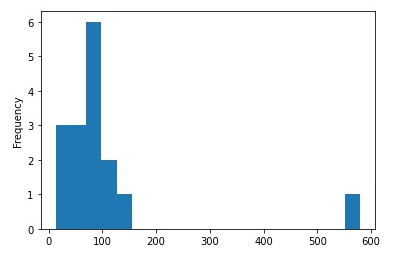
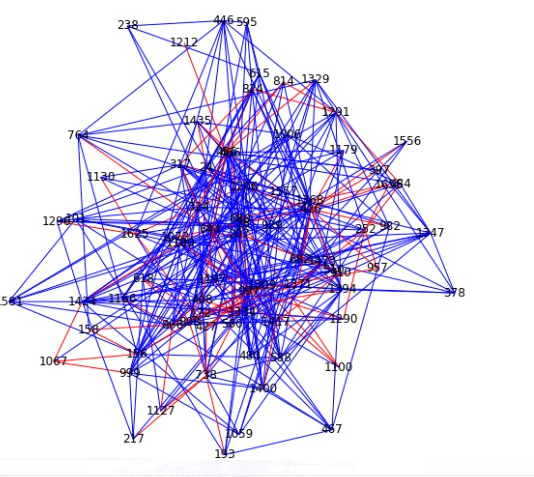
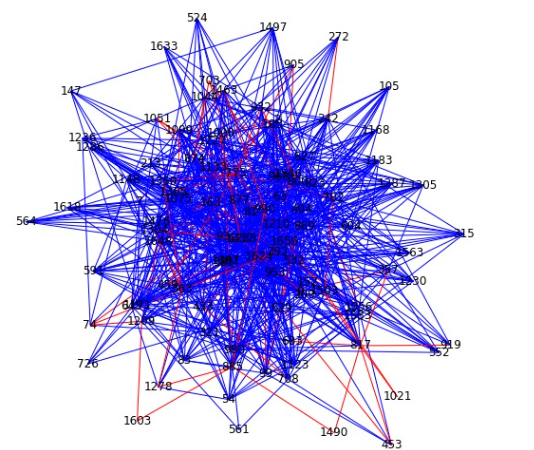


Figure 6: Histogram of communities present in the dataset including both fraudulent and legit Transactions.





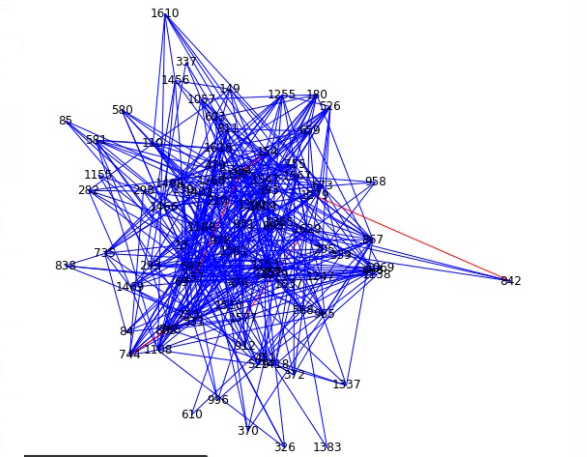


Figure 7: The above three pictures represent the communities of both fraud and legit transaction based on the community .

# References:

1. Kaggle Dataset — https://www.kaggle.com/kartik2112/fraud-detection
2. NetworkX —https://networkx.org/documentation/stable/reference/algorithms/index.html
3. Pandas- <https://pandas.pydata.org/>
4. https://towardsdatascience.com/social-network-analysis-community-detection-2b1
5. <https://arxiv.org/abs/1706.01953>
6. <https://www.altexsoft.com/blog/credit-card-fraud-detection/>
7. <https://ieeexplore.ieee.org/document/9835751>
8. <https://dl.acm.org/doi/10.1145/3368756.3369082>
9. <https://youtu.be/NCgjcHLFNDg>
10. <https://youtu.be/MZGuz-o7Fl0>