Eugene Reedy

DS 210 Project

Title: Fraudulent Transactions

I. Introduction

I find this project particularly intriguing due to my academic background in economics with a specialization in data science. My passion lies in uncovering intricate connections within financial transactions, especially given the sheer volume of such transactions and the occasional occurrence of fraudulent activities. The prospect of leveraging graph theory and vertices to predict and prevent fraud greatly excites me.

At its core, the problem I aim to address revolves around predicting fraudulent transactions within the realm of financial services. In today's fast-paced world, an incessant stream of financial transactions flows among individuals and corporations. The challenge lies in discerning patterns and anomalies amidst this data deluge to identify fraudulent activities. This endeavor strikes me as a captivating pursuit, as it holds the potential to safeguard financial institutions and their clients from the perils of fraudulent transactions.

II. Methodology

- I decided to focus on a few interesting aspects in my project. First, I looked into degree distributions, which is basically about understanding how many connections each point in my graph has. It's like seeing how popular or connected each point is in a simple way.
- Next, I explored the number of neighbors that are two steps away from each point in the graph. It's like checking out the friends of your friends, trying to see how connected the whole network really is.
- Lastly, I played around with power-law distributions. It's a way to figure out if there are any hidden patterns in the data. It's kind of like trying to spot some regularity in how things are connected in the network.

- All of these things combined helped me get a better picture of the network's structure and how everything is connected. It's like putting together the pieces of a puzzle to understand how the whole thing works.

III. Data Collection (if applicable

- Made 4 modules
    - main.rs: primary module
    - graph.rs: module that worked with creating graphs from the csv file
    - io.rs: module that read the csv file
    - Test.rs:module that ran the tests for the distributions

IV. Results

- After a magnitude of testing and running the code, I was able to let it run but was not able to read the csv file and get the output of the degree distribution graph

V. Discussion

Although I was not able to get the result I wanted, I know that it would be good to see that there were a small number of nodes with high degrees(many transactions) or low degrees(few transactions). However, another thing that had to be considered is anomalies or outliers, which include nodes with degrees significantly higher or lower than expected. High degree anomalies could be an indication of entities engaged in excessive transactions activities, potentially for fraudulent purposes.

VI. Conclusion

- Degree distributions can reveal nodes (entities) with exceptionally high degrees. In a financial transaction network, these high-degree nodes may represent financial institutions, accounts, or individuals that are central to many transactions.
- Connection: High-degree nodes, also known as hubs, can be targets for fraud detection. Some hubs may be involved in money laundering, fraudulent transactions, or other illegal activities.

- Degree distributions can reveal nodes (entities) with exceptionally high degrees. In a financial transaction network, these high-degree nodes may represent financial institutions, accounts, or individuals that are central to many transactions.
- Connection: High-degree nodes, also known as hubs, can be targets for fraud detection. Some hubs may be involved in money laundering, fraudulent transactions, or other illegal activities.
-

VII. References

https://www.kaggle.com/datasets/isabbaggin/transaction-fraudulent-financial-syntheticdata