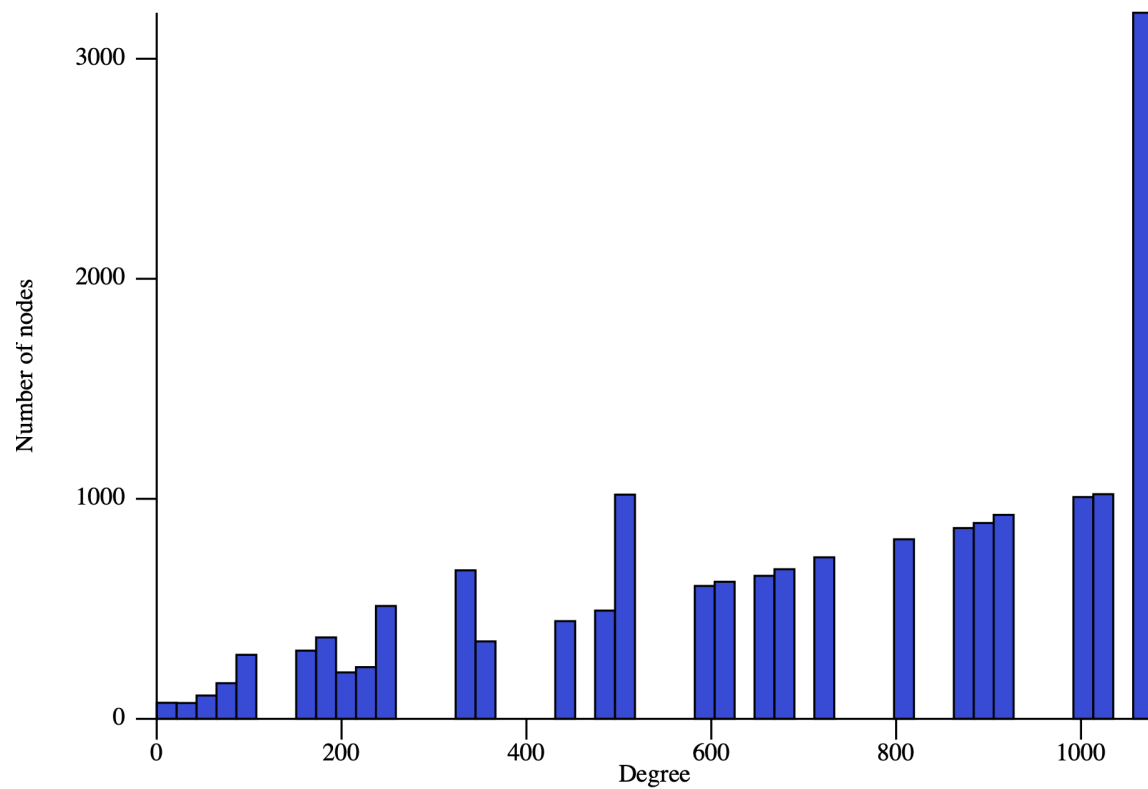Final Report

I chose a dataset about the video game FIFA and its players. The dataset shows all the players that were in the game in 2020 and all the attributes that correlate with a player's overall rating. The original dataset had 18,278 players, but I cleaned the dataset just in case there were any duplicate players and ended up with 17,354 players. My goal was to study the relationship between players based on their overall and form a degree distribution of this dataset. My hypothesis was that the degree distribution would follow a normal distribution instead of a power-law distribution due to there being less high rated and low rated overall players, than average rated players. I also wanted to take a look at the relationship between overall and the market value of players rated a certain overall. Market value represents the amount of money a player is estimated to be worth based on various factors such as their age, skills, performance, and market demand. The value is calculated by FIFA based on their own algorithms and is not necessarily the actual amount of money a player would cost in the transfer market. I believe that a higher rated player will be worth more than a lower rated player.

To start the analysis, the program reads player data from a CSV file and uses the data to construct a graph with players as nodes, and edges are added between nodes if their overall ratings are the same as another player. After creating the graph, the degree distribution was calculated, which shows the frequency of nodes with a given degree. The results are then printed, along with the percentage of total nodes for each degree. To get a visualization of the degree distribution, a histogram was created using plotlib. As shown in the histrogram, the distribution was not a normal distribution or a power law distribtuion. It ended up not having a specific degree distribution. Therefore, my hypothesis about the distribution was incorrect, however we can see a small trend that the higher the degree there is the more nodes there are connected with

it. This could correlate with there being more average rated players than high and low rated players , so there are more nodes connected with it. On the other hand, the relationship between overall and market value seemed to be a positive one. The results show that a higher overall correlates with a larger average market value, which was expected.

Sources

Dataset: https://www.kaggle.com/datasets/stefanoleone992/fifa-20-complete-player-dataset

https://serde.rs/impl-deserialize.html