Phillip Kim

DS210 Final Project

For this project, I ended up choosing a dataset from Facebook (https://snap.stanford.edu/data/gemsec-Facebook.html). The datasets represent blue verified Facebook page networks of different categories. Out of the categories, I chose the athlete-specific page. There are 13,866 nodes, representing the pages and edges are mutual likes among them. After retrieving the CSV file, I converted the data into a text file where I then was able to read it into VSCode. My goal for this project was to figure out the degree distribution of this dataset. My original hypothesis was that the distribution would follow the power-law distribution, as Facebook is a social network with undirected edges.

For my analysis, I created an undirected graph and with the graph, I performed a Breadth First Search to find the distances between each node and to see how connected each node is to each other. With the distances, I calculated the average distance for each vertex to see how far it is on average to travel to another node. I also found the degrees of each vertex. After collecting each degree, I compiled all the degrees into an empty vector, which would be used to plot the distribution of degrees for each vertex.

After producing the graph of the distribution of degrees, as seen in Figure 1, there does not seem to be a power-law distribution. There seems to be no specific degree distribution from the plot. Although my hypothesis was incorrect, I did find that the average distance for each vertex is 3, which means that the average distance between two vertices is 3 in the shortest path connecting them. This also means that each person, who is represented by a vertex, is pretty

closely connected to each other as they do not have to travel far to reach another person.

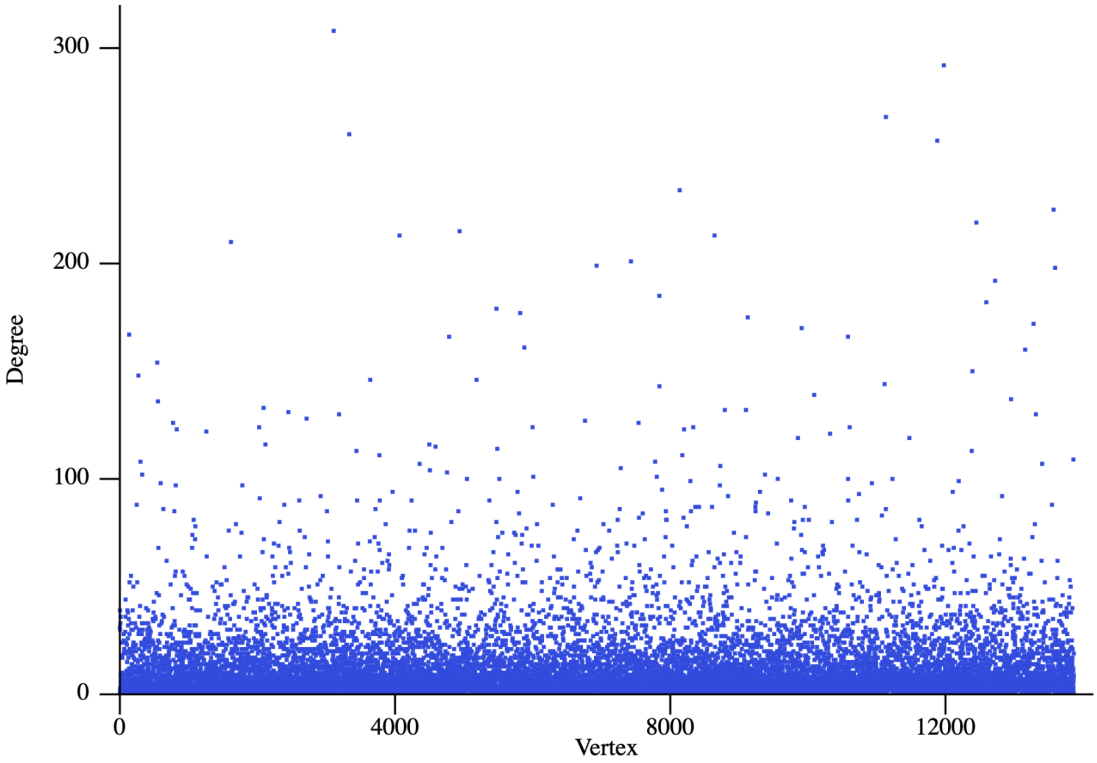

Figure 1. Graph of Degree Distribution