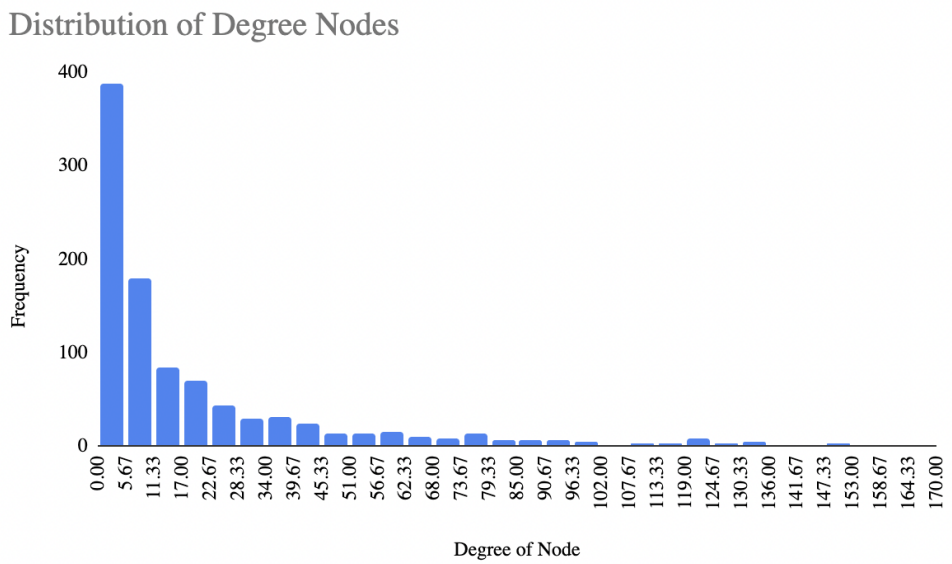
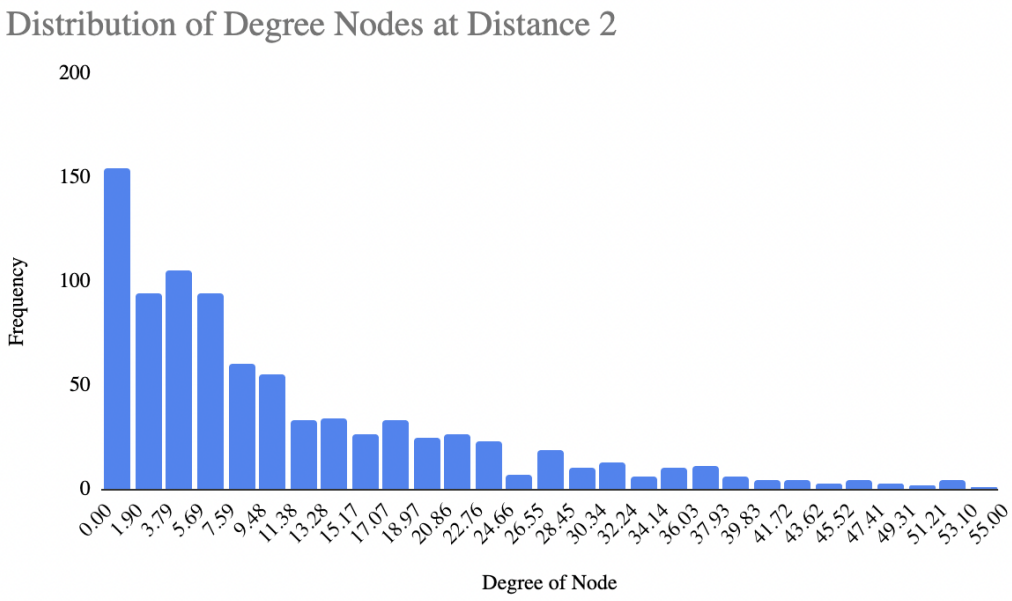
In this final project, I wanted to explore a social circle network. As someone who uses social media extensively and uses social media apps to connect with family and friends across the world, I wanted to see what types of interesting features I could find in the graph of a social network while also using the concepts (i.e the algorithms) we were learning in class. I used the Facebook social network data set found in the Stanford Large Network Dataset Collection.

My project is divided into two parts. In the first part of my project, I computed the average distance between a pair of vertices in my social network graph; I used the breadth-first-search algorithm to compute the distance between the vertexes. Essentially, I created two structs—one struct called Graph and one struct called Queue—and created an implementation on Graph where I had my breadth-first-search algorithm. Since my graph had a large number of nodes, I ran my algorithm 1000x and computed the average of those distances. Ultimately, I found the average distance between a pair of vertices to be approximately 5.243. With regard to this problem, I wanted to see if the "6 degrees of separation concept" holds up—i.e. the idea that people are 6 or less than 6 connections from another person. I realize that there are more complex approaches to figuring out the 6 degrees of separation, however, this was more of a "heuristic"/ "quick check" type of approach. No matter how many times I ran my algorithm, the distance between a pair of vertices never went above 6; it was always hovering around 5.1 to approximately 5.6. So, in my data set, I found the 6 degrees of separation concept to be true.

In part 2 of my project, I was trying to look at the idea of degree distribution in my Facebook social circle graph; I wanted to see if it resembled any type of distribution I learned in my statistics classes; however, I did not run any statistical analyses or tests. Once again, I only ran my function 1000 times to find the degree of 1000 nodes and subsequently, I averaged it out. I found the average degree of the node to be approximately 23. In my code, I also outputted the vector containing all these degrees to be able to graph the distribution. Below is the histogram distribution:



From previous reading, I've learned that social networks seem to follow a power-law probability distribution. Although I have not run any statistical tests, visually (or qualitatively) from the histogram, the general curve of the distribution seems to follow that of a power law distribution. I was also curious to find out the degree of node at distance 2 and see how the two distributions would compare to each other. Below is the distribution for degree node at distance 2:



It's interesting to note how the distribution changed when I calculated the degree of the node at distance 2. It no longer seems to "follow" or "resemble" the power-law distribution curve and it

made me wonder why. The distribution seemed to lose its dramatic curve with its harsh skewness.

Overall, this project was very insightful; I learned a lot in terms of how to clean data sets, how to decide the best methods/approaches to utilize, and the patience it takes to find the best way to approach a given problem.