

DS 210 Final Project Report

Dataset:

For my project I have used a portion of the “Social circles: Facebook” dataset found on the provided website: <https://snap.stanford.edu/data/> (Stanford Large Network Dataset Collection). These social circles are “friends list” of users on the platform. The profiles are the vertices in this case, and we are going to be observing the outer neighbors of these vertices (the profiles and their connections to outer profiles). What is a vertex neighbor? A vertex neighbor is a node (vertex) that is directly connected to a given vertex. An out neighbor is a vertex neighbor that can be reached by following an *outgoing* edge from a given vertex.

Project Idea and Implementation:

I have implemented degree distribution in my program. Degree distribution refers to the distribution of the number of edges that each vertex has in the network. The questions I have asked were, “What is the distribution of vertex degrees in my graph?” and “What if you look at the neighbors at distance 2?”. I have also used power values of vertices to question if the graph I used is a fit for the power-law distribution or not. The specific power law I have used was square law. Square law occurs where the number of vertices with a given degree is proportional to the square of that degree. This means, if the degree of a vertex is “ v ”, then the number of vertices with degree “ v ” is proportional to “ v^2 ”. When checking if a network relates to this power law or not, I took the logarithms of each vertices’ degree values at distance 1 and 2. I put the value of distance 1 to the base of the logarithm and distance 2 to the inside of the logarithm because I am looking how close each vertex’s degrees are close to the square law ideology (I specifically looked for how close is the (value of distance at 1)² is to the distance at 2 value).

Results:

What I have learned from implementing degree distribution and graphing the chosen data is that the dataset is not random. In fact, it suggests that the dataset has high level of clustering where vertices tend to be connected to other vertices with similar degrees and the distribution of connections is relatively evenly distributed among the nodes in the dataset. These connections are made under specific reasons to govern why and to which other vertices they connect. The mean power of all vertex analyzed was approximately 2.4 which suggests the previously explained result, since it is close to 2.