Alexander Alvarez and Riley Slater

Code in D2L

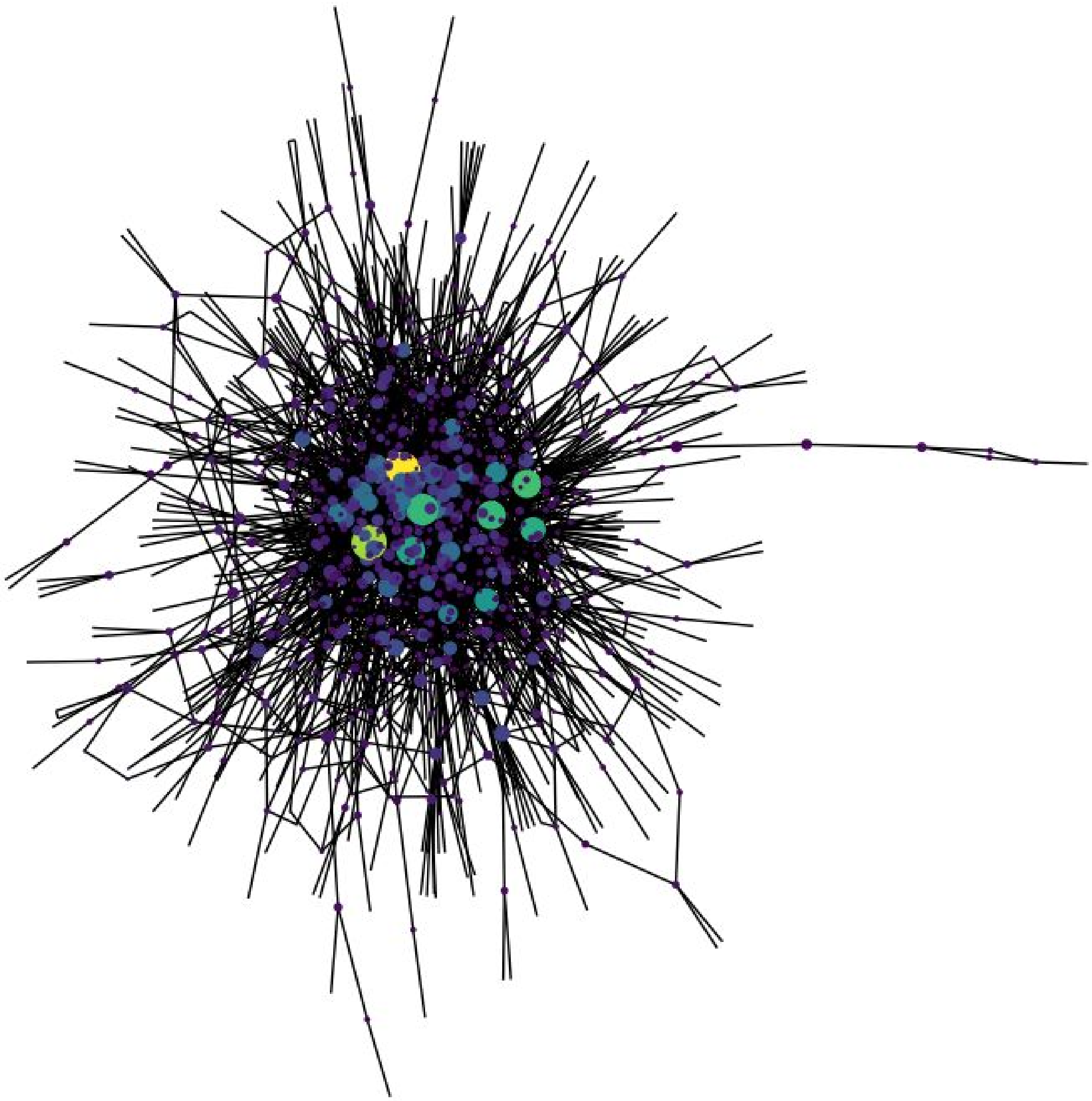
Part 1

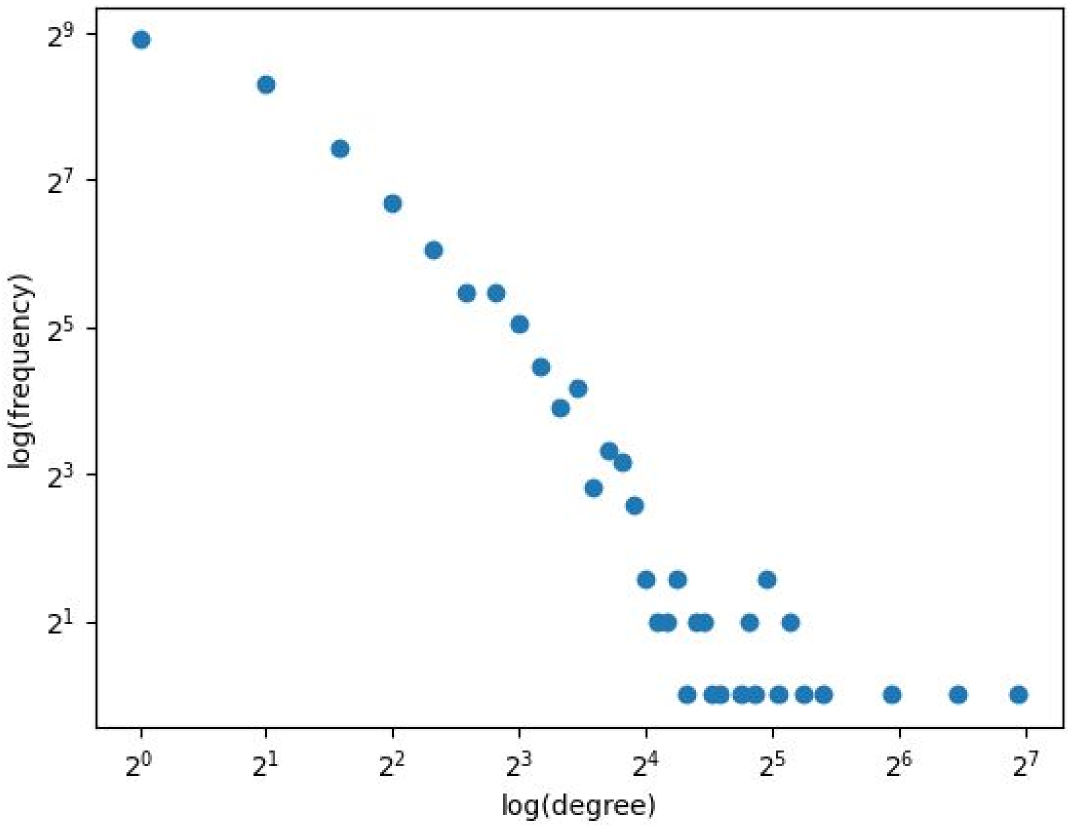
We are doing our analysis on a social network graph between friends on the Twitch streaming platform (in English). We were interested in this graph because we both personally use the Twitch app, and so the network felt very relevant to us. The graph consists of users at the nodes, with edges between users that are friends on the platform. We elected to work with the network collected from the English portion of the app, however there were still over 9,000 nodes and 33,000 edges, so preprocessing is required. We used simple random node sampling, in which we would select a node at random, and remove it (along with its edges) from the graph, repeating until we achieve the desired amount of nodes. Unfortunately, with our dataset at least, this leaves quite a few orphaned nodes after the process has concluded, and since we must pass around the graph as an edge list, these nodes are unintentionally removed once the graph is translated back into an edge list. We expect users who are more active or otherwise spend more time on the Twitch platform to have more friends on it, and from this assumption we would guess that nodes with higher centrality values are users that have the highest engagement on the platform. From this we can hypothesize that these high centrality users likely visit the site more regularly, spend more money on donations within the website, and are more active in chat rooms than users with low centrality and engagement. We believe the network's degree distribution will almost definitely exhibit the power law. While it is a social network, which usually end up exhibiting the power law, we think the fact that there is realistically no cost for adding more friends means that while many users may have 1-2 Twitch friends (that may even be friends outside of Twitch), there is nothing to stop one person from mass-friending 100s or 1000s of other people anonymously. This is also our reasoning in believing that the graph will exhibit the small-world property, as most users will not be friends with each other on the network, yet if they are friends with one of these mass-friending users, they are extremely close to everyone the mass-friending user has friended.

Part 3

17) One of the biggest issues with our sampling method was that it left our graph no longer complete. After removing all of the nodes without any edges, there were still pockets of 2-3 nodes left floating around that were not connected to the rest of the graph. We removed these nodes as well, before evaluating the most highly connected nodes, which were all repeatedly at the top of our connectedness measurements. The same four nodes appeared in the top ten of every connectedness characteristic we measured, reaffirming our belief that the graph will mainly consist of regular users with friends in the single or double digits, while a very small minority would mass-friend the rest of the users and connect them all together. These mass-friending users, acting like hubs, are all ultra-connected nodes in the network.
18) The average shortest path length for this graph, from our sample, was 4.73, indicating that the network does indeed exhibit the small-world behavior. For a network with 9,000 nodes and 33,000 edges, having an average shortest path length between any 2 nodes be less than 5 edges is rather astonishing.
19) From our log-log scale graph of the network's degree distribution, it is also very obvious that this data exhibits the power-law. When both axes are scaled in logarithmic space, the points

form the linear (in log-space) line that is often associated with power-law graphs' degree distributions. We believe this is strong evidence that the network exhibits the power-law.




log(frequency)
log(degree)
2^9
2^7
2^5
2^3
2^1
2^0
2^1
2^2
2^3
2^4
2^5
2^6
2^7