Mini Project 3 Part 1 *Kevin Koljonen*

The question:

I want to find out how connected/similar english speaking twitch streamers are and if there are separated clusters of streamers. I would also like to check how popular these clusters are in terms of views.

Importance:

Does not have much value for society at large but to satisfy my personal curiosity. Might be informative to the twitch streamers themselves or just general twitch users.

Approach:

Using google colab to check the network density and the average path length for all possible node pairs. I will attempt to answer the question using NetworkX and by following the examples made by Sepinoud in her lectures. Using NetworkX I will visualize the network of the dataset. This shows both the average path length and density of groups. The average path length can just be checked by adding all the path lengths together and dividing it by the amount of edges. The visualization of the paths will also show how similar certain “communities” are by looking at how close knit the nodes are. Network density can also be calculated using the networkX density function. Using the datasets features and displaying each node with the views they’ve accumulated I should be able to see which streamers are more popular. The network could be sorted by each feature too to see even more how similar twitch streamers are. Like how long they’ve been streaming for, if they’re partnered or not etc

Expected outcome:

My expectation for this project is that a majority of twitch streamers will be largely similar to one another. Although there may be outliers I expect the majority to have a dense network. I expect there to be many separate dense clusters of nodes. Some clusters will be more popular than others and many of them will probably be less popular rather than more popular.

Dataset used:

I’m using a Twitch Social Networks dataset which for the english dataset includes over 7,000 nodes and over 35,000 edges.

https://snap.stanford.edu/data/twitch-social-networks.html