The Six Degrees of “The six degrees” by Kevin Bacon

*by Sohaib Nabeel and Dylan Taylor*

# Nodes and links

The videos on YouTube will be our nodes and two videos will be linked if one appears as a suggested video on the page of another. These links will be weighted, and directed.

# Dataset

The question of our dataset revolves around how quickly you can jump from one genre of YouTube videos to another. The idea was inspired from “The Six Degrees of Kevin Bacon” and initially asked how easily a user could get from any YouTube video to the video “The six degrees | Kevin Bacon | TedxMidwest” on YouTube. We will collect our data by creating crawlers, each of these will employ tactics to avoid the user-tracking that YouTubes suggestion algorithm does. The crawlers will store the “tags” defined by the uploader for each video it visits so that we can group the videos by genre in our analysis. They will also store the “Watch” ID associated with each video it “watches” and that of the recommended video it “watches” next, as a single entity in a csv to form a link. We will make the crawlers choose the next video depending if they have visited that video before or if the video is out of the current genre, amongst other conditions. The data gathered by each crawler will then be compiled into a final network where all nodes will be unique, and any duplicate links will be counted and used to form a weight for that one unique link in the final network.

# Expected Size of network

We will create dozens of crawlers. Each crawler will traverse over thousands of videos, resulting in thousands of nodes and links per crawler network. The “sum” of these small networks will define the size of the final network whose size will range into the tens of thousands. Once the data is obtained it will be filtered to avoid duplicate nodes and links. The weights of the links will be dependent upon the amount of duplication between crawler networks, e.g. if exactly 2 distinct crawlers both traversed from Video A to Video B, the weight of that link in the final network would be 2. This will help keep the final network sizable and provide insight into the suggestion algorithm. Since the data the crawlers will collect can increase exponentially due to the indefiniteness of YouTube’s suggestion algorithm, we will ensure the crawlers work through only a subset of the “recommended videos” (exactly how big of a subset is to be determined).

# Why We Care.

We live in a world in which everyone almost always has an opinion, or is looking to form one. We are interested in users who have yet to formalize or structure their opinions, views, or perspectives on given topics. We are curious to see how such users are led along a path of videos and which genres the YouTube suggestion algorithm tends to send them towards when it has no background data on them. That is: we want to determine which genres are easier to get away from, and which users might gravitate towards, based on videos that bubble forth from the suggestion algorithm. We are specifically interested in if these paths tend towards unfavourable genres that expose users to racial profiling, hatred, violence or any other negative topic, and if certain genres form bubbles which are difficult to escape and lead consequences like political radicalization in users. In conclusion, we aim to determine if one could start in an “innocent” genre and end up stuck in a very “negative” genre. We will analyze the rabbit holes that YouTubes suggestion algorithm leads users down, and highlight the issue of political radicalization due to “Political bubbles” which can be found on youtube.