**Use Python NETWORKX library (15-20 mins)**

This exercise is aimed to demonstrate this process, by requiring you to compute the different centrality measures (degree, closeness, betweenness and eigenvector centralities) over a real world dataset represented as graphs. You will try to analyze the graph data of IMDB actor network. The small version of the dataset can be found in:

<https://drive.google.com/drive/folders/1JNB782XBgMW8VWHTCQ9SLHZd1V2zYlIA?usp=sharing>

You can follow the steps from the main references <https://github.com/DataScienceUB/introduction-datascience-python-book/blob/master/ch08_Network_Analysis.ipynb>

**About the dataset**

* There are three datasets available in the folder.
* For this part, you only need two: **“imdb\_actor.names”** and **“small\_imdb\_actor.txt”**
* The dataset network of actors was collected from IMDb.
* Two actors are connected by an edge if they appeared in the same movie.
* The **“small\_imdb\_actor.txt”** file contains edge lists of the network, where each line is an edge of the form “id1< whitespace >id2”.
* The file contains a third column (to denote weight of an edge). Ignore this third column for the purpose of this assignment.
* Consider all the graphs as undirected.
* The **“imdb\_actor.names”** file contain the mapping from node id in the graph to actual names of those nodes. These will be required for semantically interpreting the results.

**Your tasks**

1. Load the dataset by using the following command:

import networkx as nx

G = nx.read\_edgelist("data/small\_imdb\_actor.txt", nodetype=int, data=(("weight", float),))

1. Print out the total number of nodes, total number of edges and average degree
2. Find the Top-5-actor’s name which have the highest centrality measures:
   1. Degree
   2. Betweenness
   3. Closeness
   4. Eigenvector
3. Discuss with you group the meaning of each centrality measure above in the context of the actor network. Which one do you think most important measure for this network?
4. Find how many communities found in the network. What does this number mean?