

## Final Project 2: Movies

For this project, write a Python program that creates an actor/actress network from IMDb data <https://www.imdb.com/interfaces/>, and an importance ranking of the actors and actresses. Both should be displayed in an easy-to-read figure.

- Your program should first explore and clean the data. For example, you need to drop duplicates from the original datasets.
- You can create edges, for example, by counting how often two actors/actresses appear in the same movie. You may decide on a different measure yourself.
- Use **PageRank** on the discovered network to rank the actors/actresses by importance. You may also want to consider how to include ratings of the movies in your ranking algorithm.
- Write your own graph drawing algorithm to create a visually pleasing picture of the top listed actors and actresses in your network, and come up with a nice way to display the ranking.
- Explore this network and the datasets in <https://www.imdb.com/interfaces/> and see if you can find a formula for making a successful movie. This is an open-ended project, which means that you can explore any aspect of the datasets and come up with your own interpretation.

You have until before final exam **June 11, Monday** to complete the final project. I encourage you to work together with others in the class and each team may have up to three people. You will hand in your well-commented .py files. You will also hand in a .pdf with a report that explains the problem, and explains how your program works. The report should contain pseudo-code. The report should be such that a python programmer could read it and understand how to create the program from scratch, without too much actual code. You will get extra credits if you post a demo of your program on YouTube and send me the link.

For all projects, there are many resources on the internet, and even codes that others have written. You may read these, but you are not allowed to use anyone else's code directly in yours. The final projects should be doable using mostly tools and modules from the class. Please check with me if you are not sure whether you are allowed to use a new command or module.