

Ben Kahan  
DS 210  
Project Proposal  
11 November 2022

I will be using the “IMDb 5000+ Movies & Multiple Genres Dataset” sourced from Kaggle.com by Rakkesh Arv.

I had many ideas for this project, ranging from pseudo-crypto currency to computer network information. However, I decided that it would be interesting to see how much information can be extracted from relatively simple graph algorithms.

Thus, I will be researching and exploring the interconnectivity of Hollywood across a wide range of genres and years. I want to know how directors choose their leads and if those actors follow those directors. In addition, if time permits, I would love to see how the interconnectivity changes throughout time.

There are many steps that I need to accomplish:

1. Set up Rust project and VCS (< 5min)
2. Begin data structure design (1-2 hrs)
3. Input, Clean, Organize Data (1 hr)
4. Implement DS design for Data (5-10 hrs)
5. Explore different graph properties
  - a. MWST? (1-2 hrs)
  - b. Connectivity (3-4 hrs)
  - c. Density (3-4 hrs)
6. Clean Formatting (1-2 hrs)
7. Extra: Visualize Graphs over time (< 5hrs)
8. Export final project ( < 10 min)

Of course, I would love to implement more into this project and I suspect those changes will come naturally.

I would like most of the code base to be < 80 % by Thanksgiving. I understand the challenge, but I believe it could be possible. I suspect that debugging and packaging will take anywhere from 2-7 days, and I would like the project to be submittable by December 10<sup>th</sup>.