



CS 429
Dynamic & Social Network Analysis
Fall 2021

Project
Community of the Top 50 Movies

Milestone 2: Revised Short Proposal

Mehmet Ali Altunsoy, 21702531

1. Description of the Project:

I will analyze the connections of the top 50 movies according to IMDb [1], actors performed in these movies and directors of these movies. This project will try to achieve the following:

- Cluster similar movies in terms of their cast and director.
- Analyze the actors performed in the top 50 movies and create their network.
- Analyze the directors who shot the top 50 movies and create their network.

2. Dataset

The following data will be used in the project:

- Top 50 movie from the IMDb's top 250 movie list [2] as in Appendix A.
- The cast of these top 50 movies. Movie-by-actor matrix will be created. The cast of "*The Shawshank Redemption*" is pulled as in Appendix B.
- The directors of these top 50 movies. Movie-by-director matrix will be created.

3. Methods

In order to cluster the similar movies I will analyze the common actors performed in the movies and common directors shot the movies, then create a movie-by-movie adjacency matrix according to these analyzes which will show the connections and similarity of the movies. I will use the following clustering algorithms and compare their outcomes:

- **Highly Connected Subgraphs:** I will find the movies that have common directors or cast and run the HCS algorithm [3] to separate the network graph to different clusters. The result will be represented with a basic graph.
- **Louvain community detection:** I will valuate the similarity of two movies by looking their directors and cast and use the Louvain Method [4]. Movies with more common cast will be valuate more similar (If two movies have same director it will increase the similarity more than one common actor/actress. Hence, the weight of the common director is more than common actors/actresses). The result will be represented with color based graph that different clusters have different colors.

In order to create actor and director networks I will create actor-by-actor and director-by-director adjacency matrices according to their contributions to the movies in top 50. The results will be represented with a graph.

4. Expected Challenges

- Collecting the data autonomous may require some time to work efficient.
- The IMDbPY library may have a time limit for batch requests. (I will look cron jobs or loops to solve it.)
- The amount of data could be too small or too big for creating a meaningful network. In order to overcome this problem more or less movies could be covered from the top 250-movie list.

References

- [1] "Top Rated Movies," Imdb.com. [Online]. Available: <https://www.imdb.com/chart/top/>. [Accessed: 25-Oct-2021].
- [2] "IMDbPY," Github.io. [Online]. Available: <https://imdbpy.github.io/>. [Accessed: 25-Oct-2021].
- [3] F. Hüffner, C. Komusiewicz, and M. Sorge, "Finding highly connected subgraphs," in Lecture Notes in Computer Science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 254–265.
- [4] L. Rita, "Louvain Algorithm," Towards Data Science, 09-Apr-2020. [Online]. Available: <https://towardsdatascience.com/louvain-algorithm-93fde589f58c>. [Accessed: 22-Nov-2021].

Appendices

Appendix A

- 1 The Shawshank Redemption
- 2 The Godfather
- 3 The Godfather: Part II
- 4 The Dark Knight
- 5 12 Angry Men
- 6 Schindler's List
- 7 The Lord of the Rings: The Return of the King
- 8 Pulp Fiction
- 9 The Good, the Bad and the Ugly
- 10 The Lord of the Rings: The Fellowship of the Ring
- 11 Fight Club
- 12 Forrest Gump
- 13 Inception
- 14 The Lord of the Rings: The Two Towers
- 15 Star Wars: Episode V - The Empire Strikes Back
- 16 The Matrix
- 17 Goodfellas
- 18 One Flew Over the Cuckoo's Nest
- 19 Seven Samurai
- 20 Se7en
- 21 The Silence of the Lambs
- 22 City of God
- 23 Life Is Beautiful
- 24 It's a Wonderful Life
- 25 Star Wars: Episode IV - A New Hope
- 26 Saving Private Ryan
- 27 Interstellar
- 28 Spirited Away
- 29 The Green Mile
- 30 Parasite
- 31 Léon: The Professional
- 32 Hara-Kiri
- 33 The Pianist
- 34 The Usual Suspects
- 35 Terminator 2: Judgment Day
- 36 Back to the Future
- 37 Psycho
- 38 The Lion King
- 39 Modern Times
- 40 American History X
- 41 Grave of the Fireflies
- 42 City Lights
- 43 Whiplash
- 44 Gladiator
- 45 The Departed
- 46 The Intouchables
- 47 The Prestige
- 48 Casablanca
- 49 Once Upon a Time in the West
- 50 Rear Window

Appendix B

Tim Robbins	
Morgan Freeman	
Bob Gunton	
William Sadler	
Clancy Brown	
Gil Bellows	
Mark Rolston	
James Whitmore	
Jeffrey DeMunn	
Larry Brandenburg	
Neil Giuntoli	
Brian Libby	
David Proval	
Joseph Ragno	
Jude Ciccolella	
Paul McCrane	
Renee Blaine	
Scott Mann	
John Horton	
Gordon Greene	
Alfonso Freeman	
Vincent Foster	
John E. Summers	
Frank Medrano	
Mack Miles	
Alan R. Kessler	
Morgan Lund	
Cornell Wallace	
Gary Lee Davis	
Neil Summers	
Ned Bellamy	
Joe Pecoraro	
Harold E. Cope Jr.	
Brian Delate	
Don McManus	
Donald Zinn	
Dorothy Silver	
Robert Haley	
	Dana Snyder
	John D. Craig
	Ken Magee
	Eugene C. DePasquale
	Bill Bolender
	Ron Newell
	John R. Woodward
	Chuck Brauchler
	Dion Anderson
	Claire Slemmer
	James Kisicki
	Rohn Thomas
	Charlie Kearns
	Rob Reider
	Brian Brophy
	Paul Kennedy
	James Babson
	Dennis Baker
	Fred Culbertson
	Richard Doone
	Shane Grove
	Rita Hayworth
	David Hecht
	Alonzo F. Jones
	Gary Jones
	Sergio Kato
	Michael Lightsey
	George Macready
	Christopher Page
	Neil Riddaway
	Brad Spencer
	Jodiviah Stepp
	Mark A. Strain