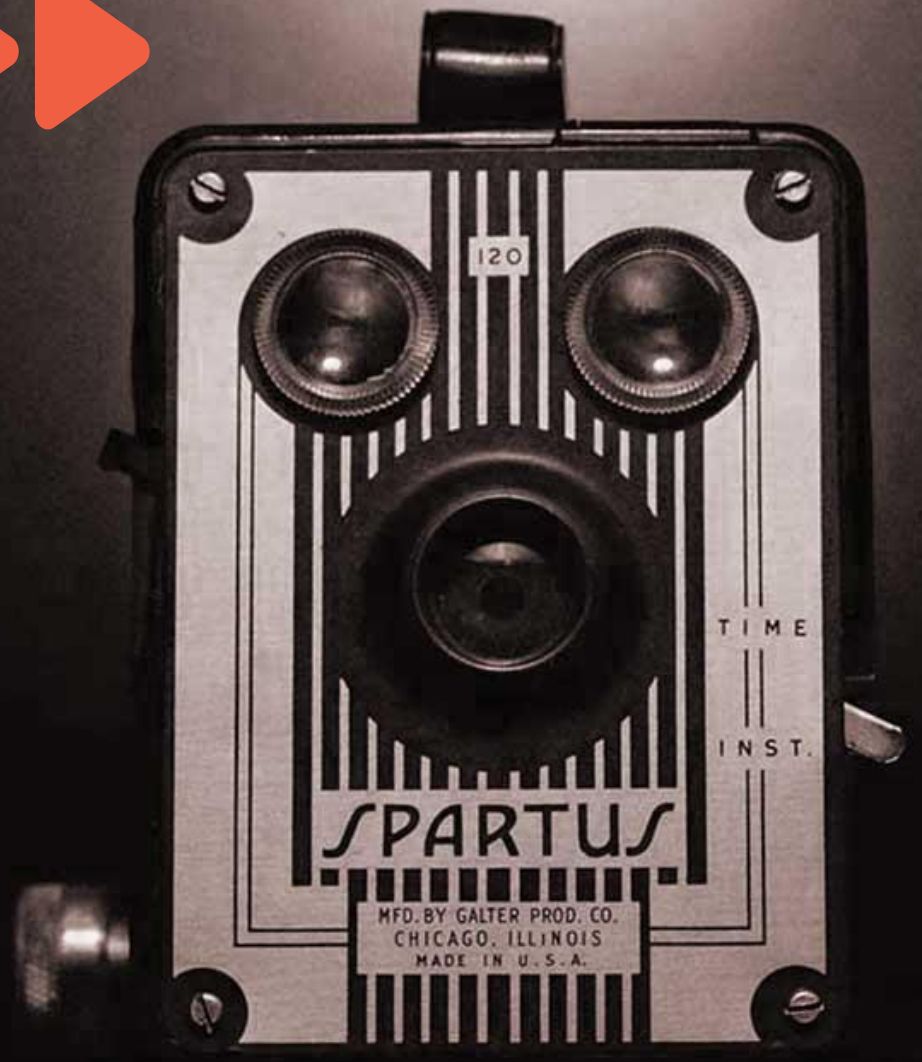
Community of the Top 50 Movies

Mehmet Ali Altunsoy 21702531

Contents



01 Project Description

What did I study in this project and what was my goal?

02 Dataset

What data did I use while doing this project and how did I get it?

03 Methodology

In this project, I examined the data I have using which methods.

04 Results

What conclusions did I reach at the end of this project?

1 Project Description

We all love to watch movies and we all have our favorite actors, directors and movies.

This project will look at the world of movies with a tabloid perspective, so to speak.

1 Project Description

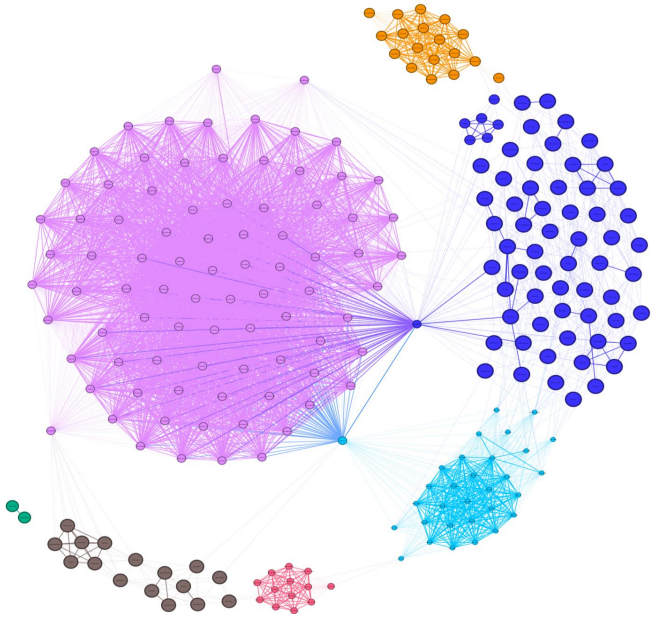
I reviewed IMDb's top 50 movies [1] with the highest ratings.

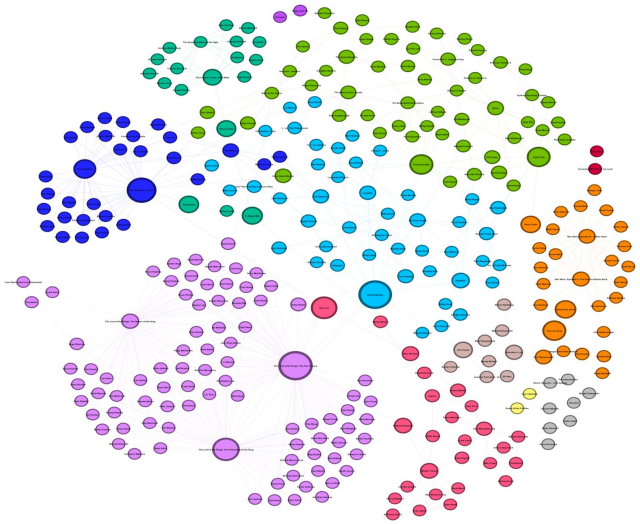
- Network of Movies
- Network of Directors
- Network of Actors

Rank & Title	IMDb Rating
1. The Shawshank Redemption (1994)	9,2
2. The Godfather (1972)	9,1
3. The Godfather: Part II (1974)	9,0
4. The Dark Knight (2008)	9,0
5. 12 Angry Men (1957)	8,9
6. Schindler's List (1993)	8,9
7. The Lord of the Rings: The Return of the King (2003)	8,9
8. Spider-Man: No Way Home (2021)	8,8
9. Pulp Fiction (1994)	8,8
10. The Good, the Bad and the Ugly (1966)	8,8

1 Project Description

By creating these networks, I cluster the similar movies, actors, and the directors to analyze their connections.





DATASET

2 Dataset

I pulled the data from IMDb with the help of IMDbPY library [2].

The data is stored as three .csv files initially, which are the following:

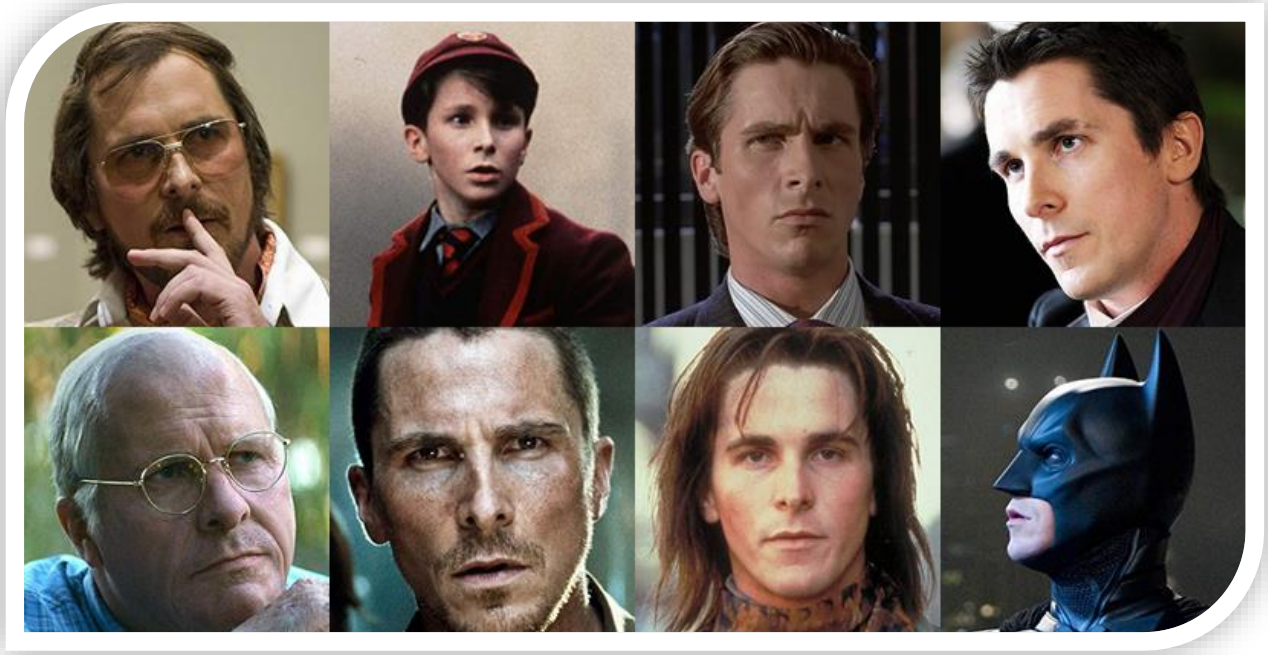
- List of top 50 movies.
- List of cast starred in the movies.
- List of directors of the movies.

2 Dataset

Movies
I created a movies to movies adjacency matrix to determine the connections of movies with each other.

The value of connections between two movies are evaluated from the number of common cast they have and the director.

Common directors weighted x5.



2 Dataset

Cast
I created a actors to actors adjacency matrix to determine the connections of actors with each other.

The value of connections between two actors are evaluated from the number of movies they starred together.

The actors who starred only one movie are eliminated from the list.

3837 -> 232

2 Dataset

Directors
I created a directors to directors adjacency matrix to determine the connections of directors with each other.

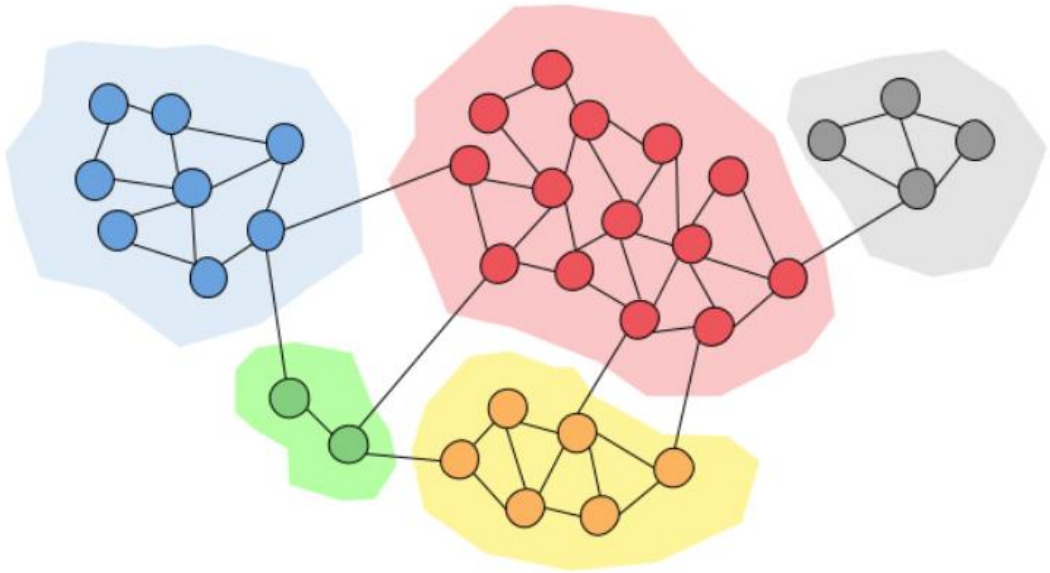
The value of connections between two directors are evaluated from the number of cast they work in common.

METHODOLOGY

3 Methodology

I used two different clustering algorithms to cluster the data I have and examined their differences. The algorithms I use are as follows:

- Louvain Community Detection
- Girvan-Newman Clustering

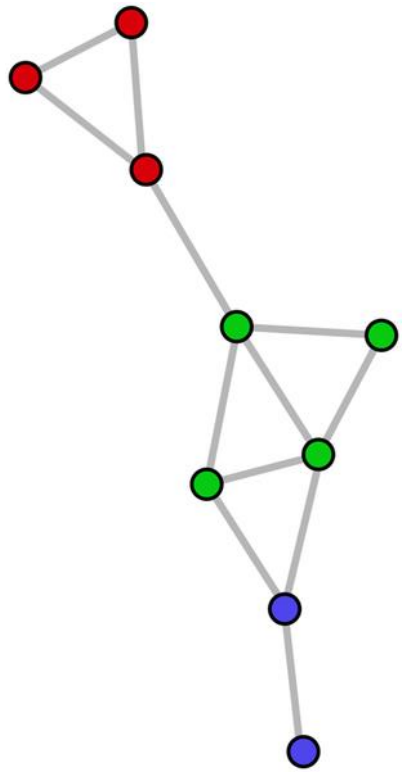


3 Methodology

3.1. Louvain Community Detection

To find the Louvain algorithm, we first need to calculate the modularity of the network.

$$M = \frac{1}{2m}\sum_{i,j}(A_{i,j} - \frac{k_i, k_j}{2m})\delta(c_i, c_j)$$



- m: sum of all edge weights
- $A_{i,j}$: edge weight between node i and j
- k_i: sum of the edge weights connected to node i
- c_i: community of node i
- δ: Kronecker delta function which $\delta(x, y) = 1\ if\ x = y, o.w.\ 0$.

After all nodes are insert in a community, community aggregation is done. [3]

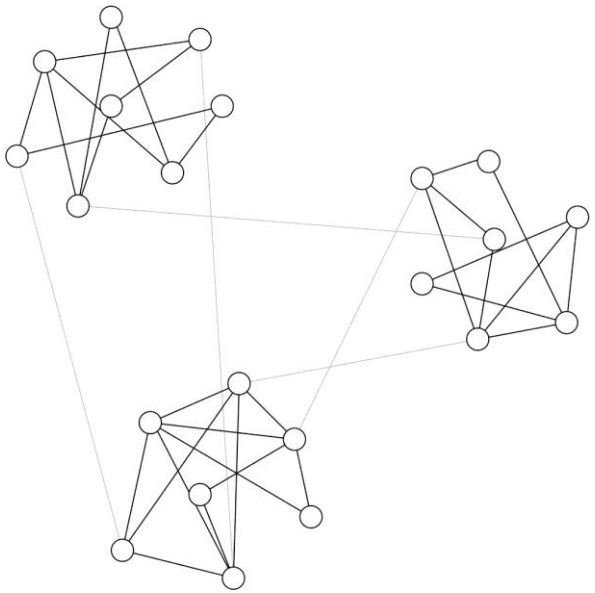
3 Methodology

3.2. Girwan-Newman Clustering

By gradually eliminating edges from the original network, the Girvan–Newman algorithm finds communities. The communities are the remaining network's connected components. [4]

The steps of the algorithm for detecting communities are as follows:

1. The network's betweenness of all existing edges is determined initially.
2. Remove the edge(s) with the maximum betweenness.
3. All edges affected by the removal have their betweenness recalculated.
4. Steps 2 and 3 are repeated until there are no more edges.



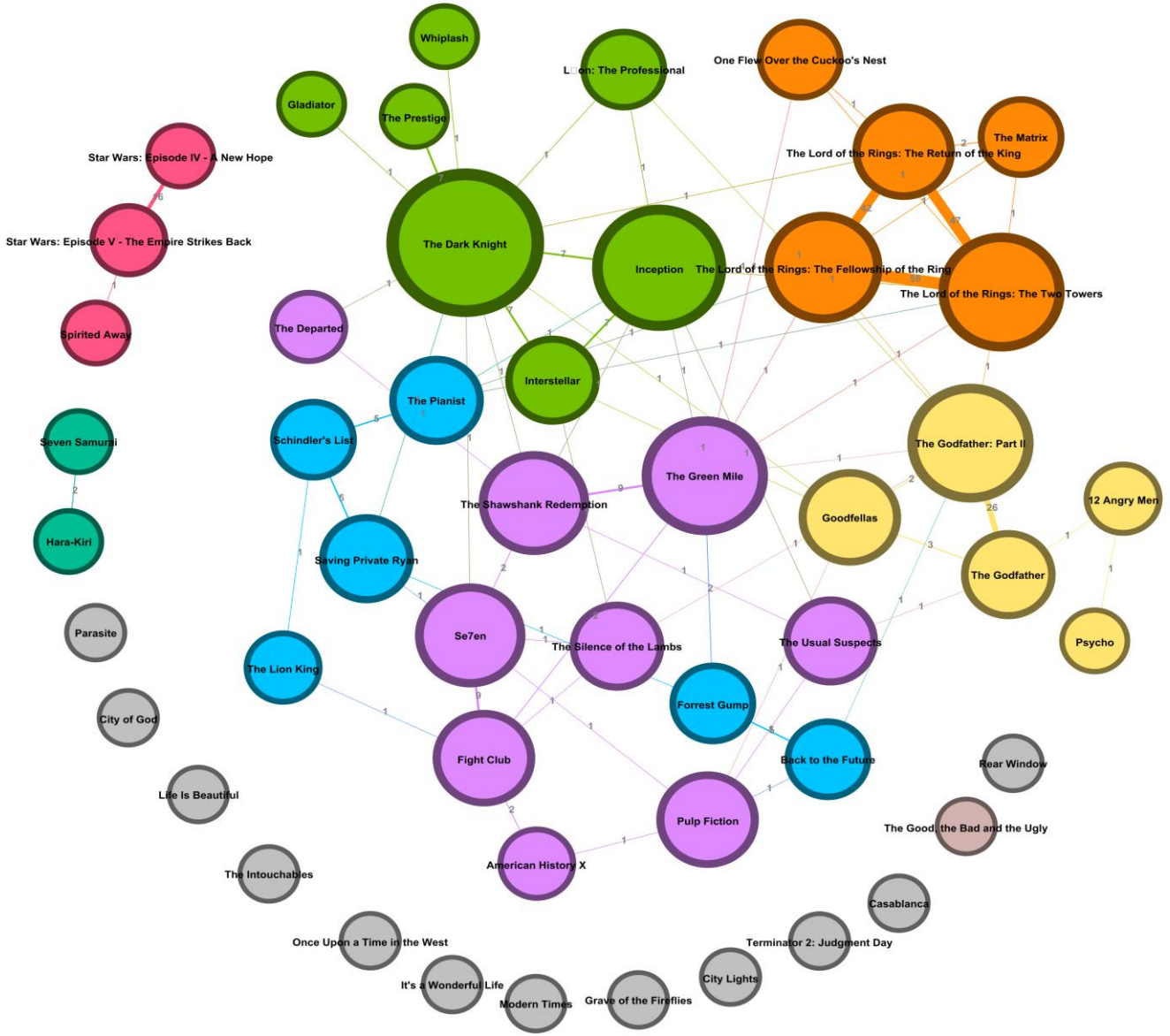

Christopher Nolan
Peter Jackson
Frank Darabont
Lana Wachowski
Martin Scorsese
David Fincher
Robert Zemeckis
Alfred Hitchcock
Quentin Tarantino
Francis Ford Coppola
Sidney Lumet

Results

4 Results

4.1. Movies (Louvain)

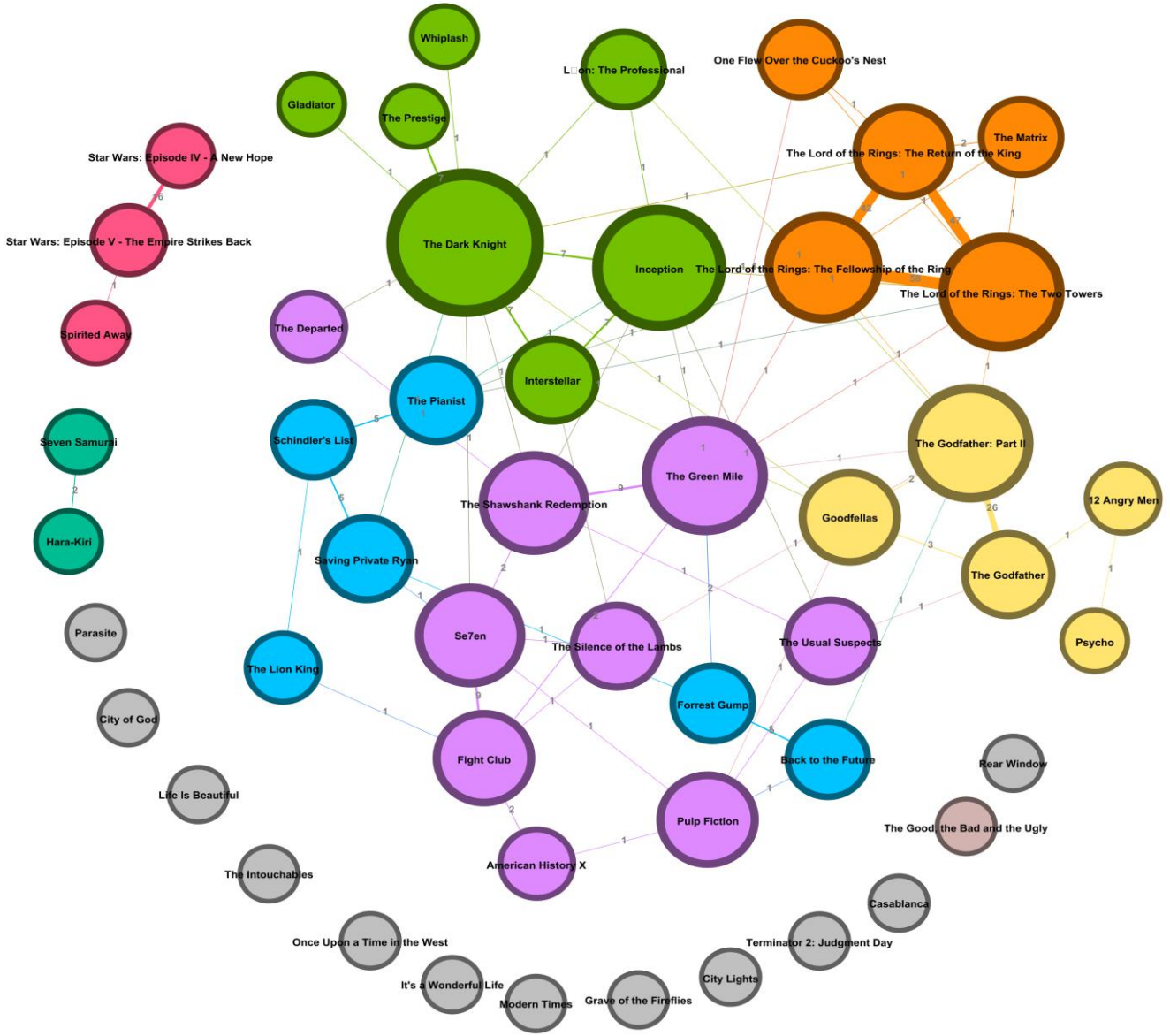
# of Communities	20
Modularity	0.597
Average Degree	2.880
Nodes	50
Edges	72
Diameter	6
Average Path Length	2.496
Density	0.059
# of Components	16



4 Results

4.1. Movies (Louvain)

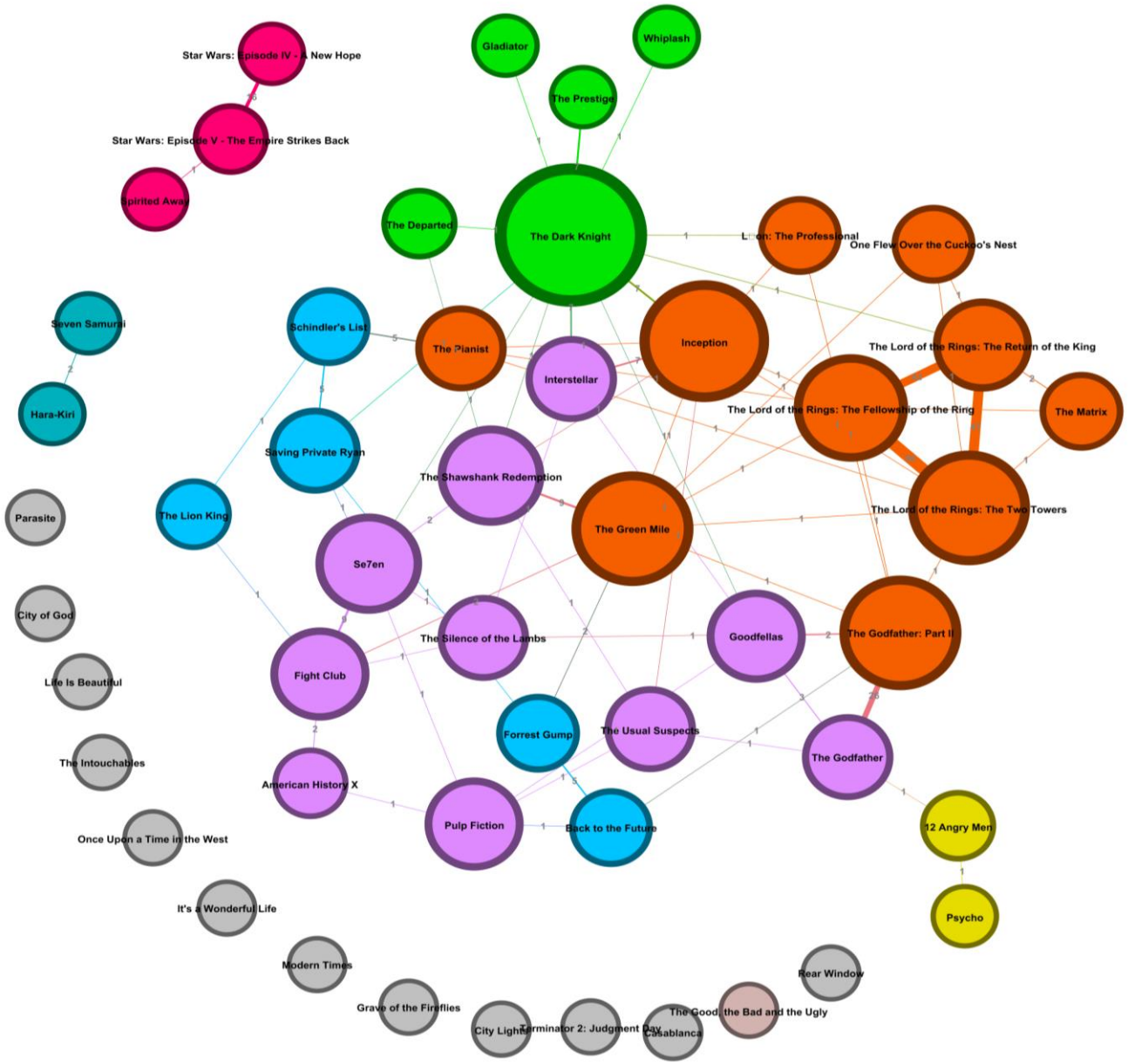
	%18	Crime Movies
	%14	Nolan Movies
	%12	Historical Movies
	%10	Mafia Movies
	%10	Lord of the Rings
	%6	Star Wars
	%4	Japanese Cinema
	%2	Indie Movies / Cult Classics



4 Results

4.1. Movies (Girvan-Newman)

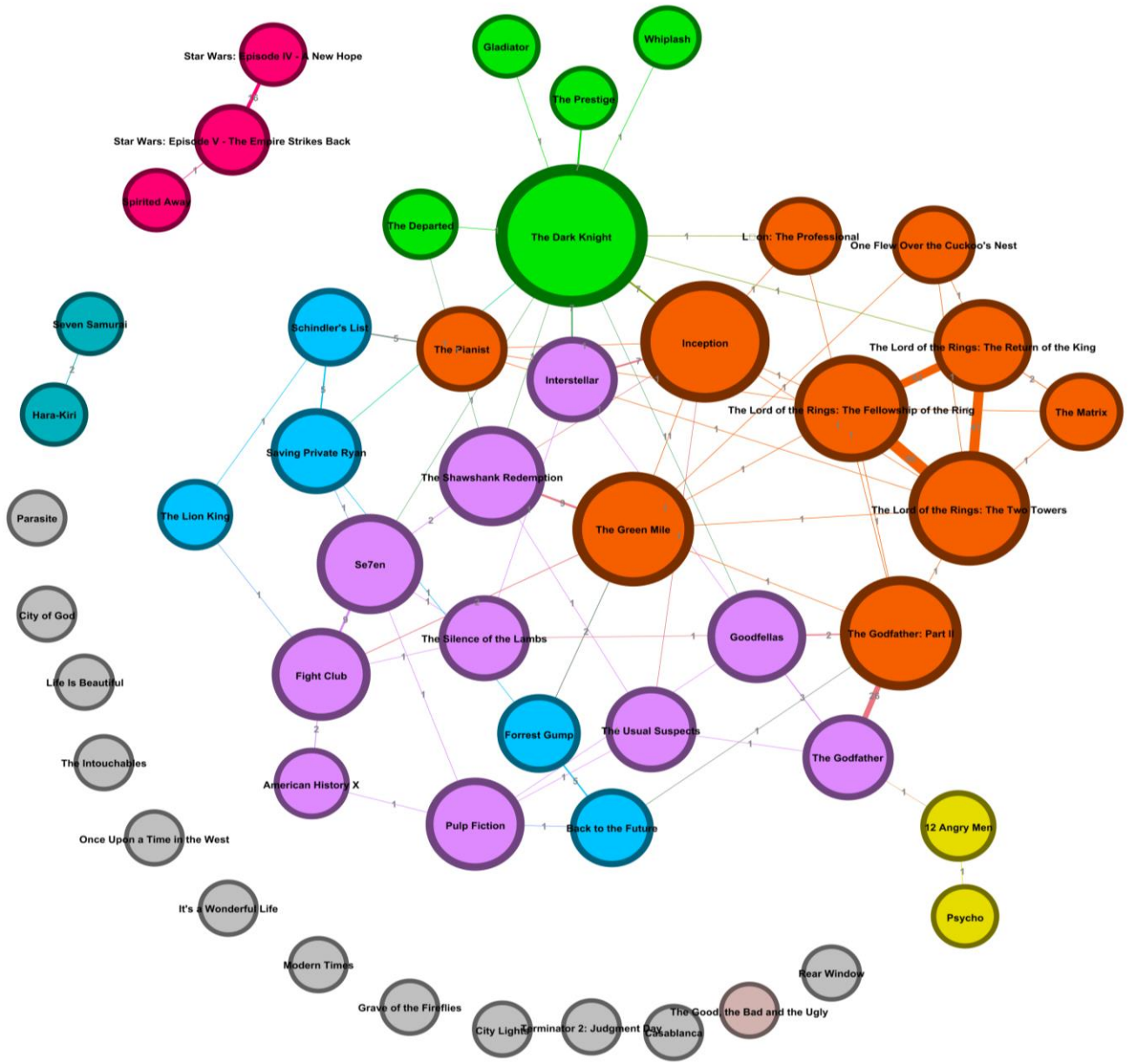
# of Communities	20
Modularity	0.38
Average Degree	2.880
Nodes	50
Edges	72
Diameter	6
Average Path Length	2.496
Density	0.059
# of Components	16



4 Results

4.1. Movies (Girvan-Newman)

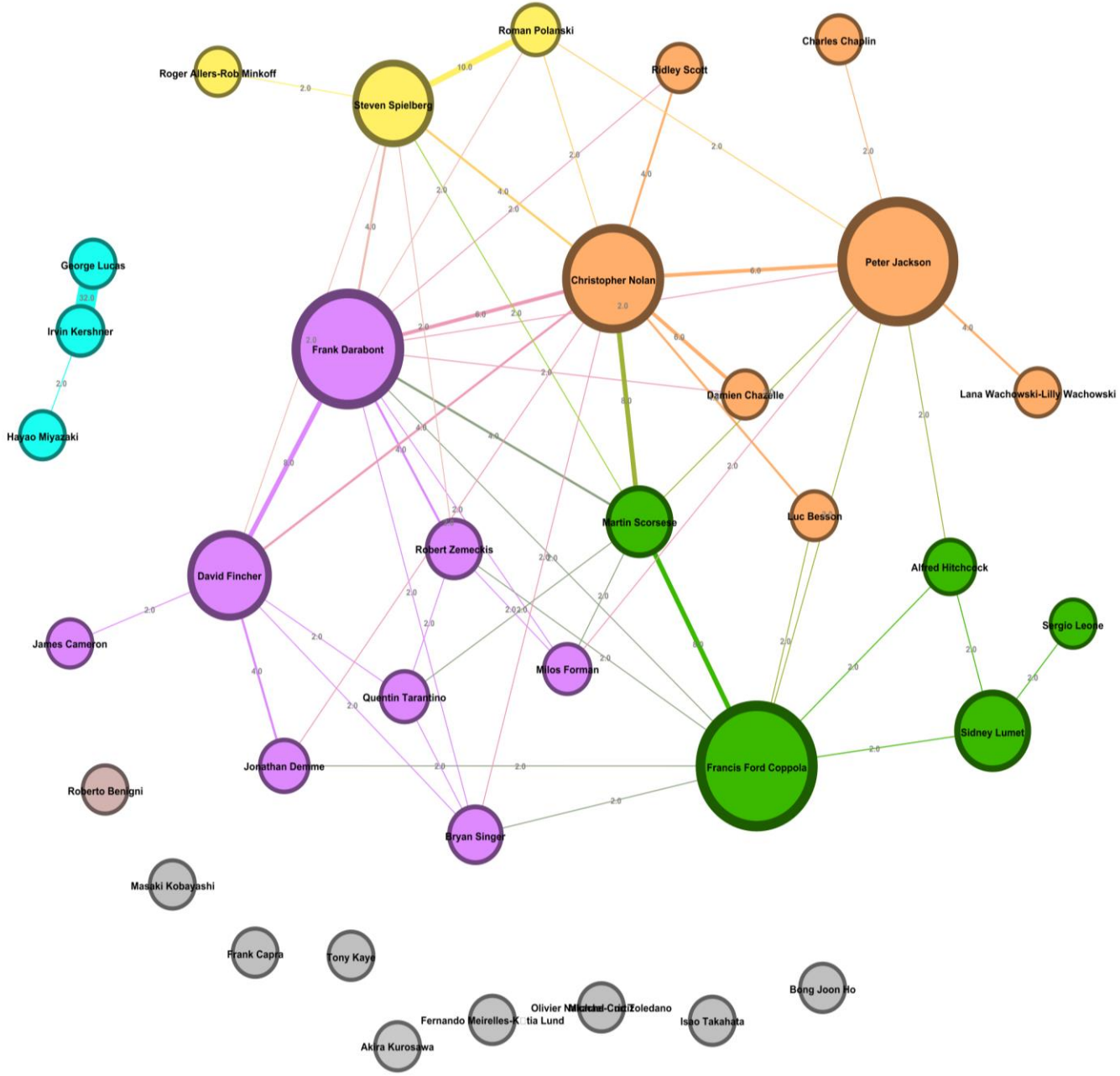
	%20	Crime Movies
	%20	Mixed
	%10	Historical Movies
	%10	Nolan Movies
	%6	Star Wars
	%4	Japanese Cinema
	%4	50’s Cinema
	%2	Indie Movies / Cult Classics



4 Results

4.2. Directors (Louvain)

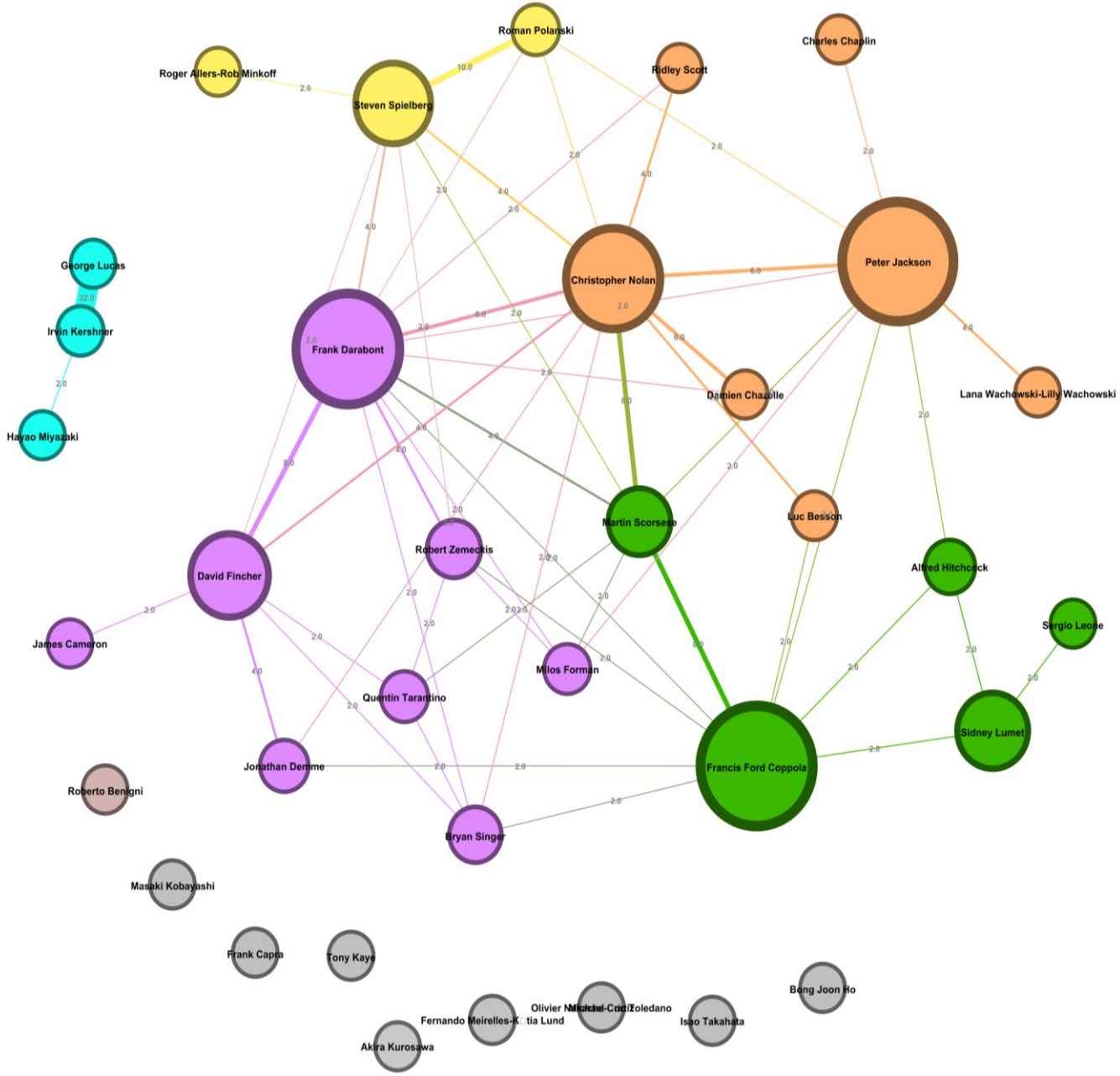
# of Communities	14
Modularity	0.597
Average Degree	3
Nodes	36
Edges	54
Diameter	5
Average Path Length	2.20
Density	0.086
# of Components	11



4 Results

4.2. Directors (Louvain)

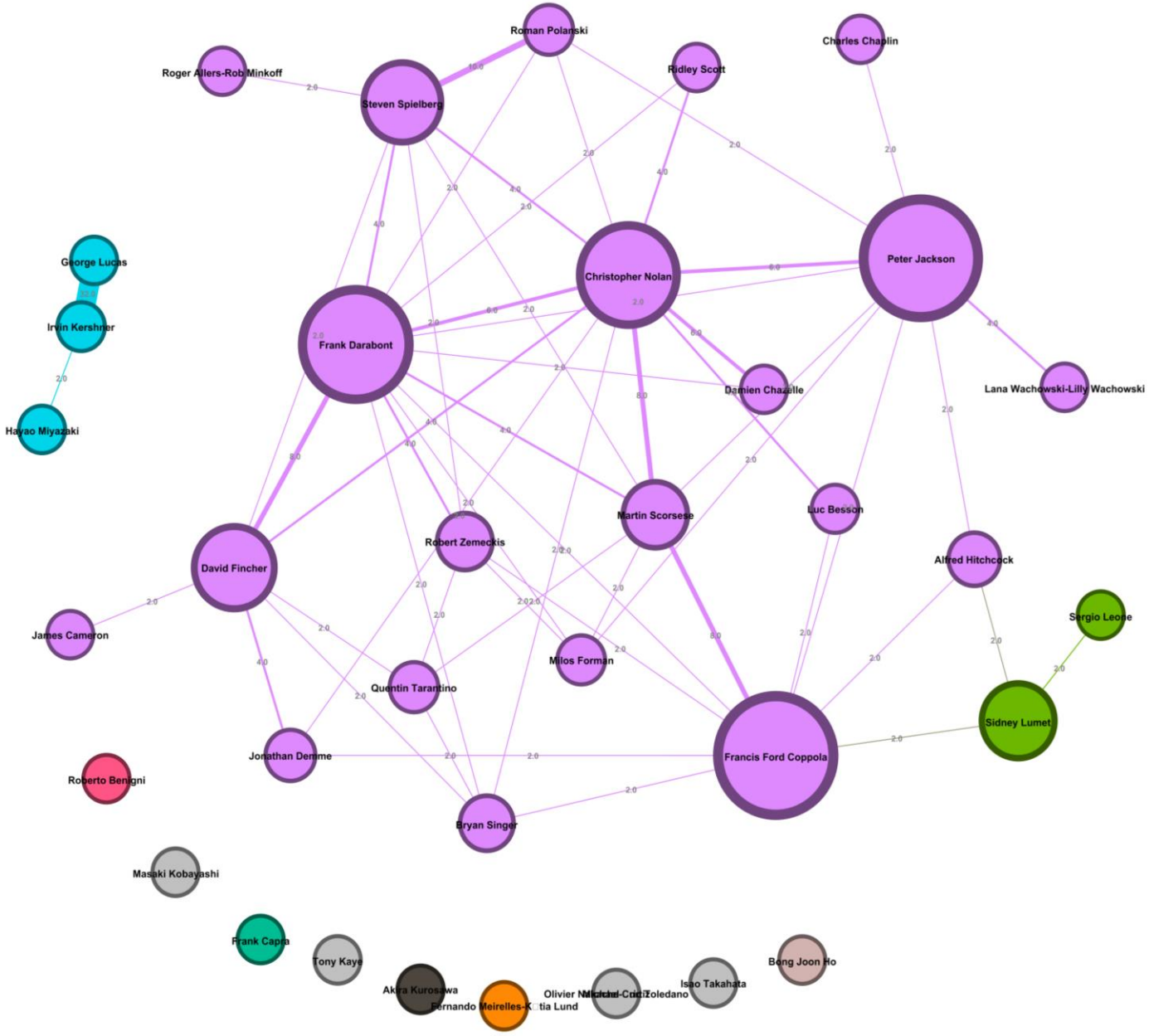
	%22	Mixed
	%19	Sci-Fi / Fantasy Directors
	%14	Mafia / Crime Directors
	%8	Spielberg - Polanski
	%8	Star Wars Directors
	%2	Foreign Directors



4 Results

4.2. Directors (Girvan-Newman)

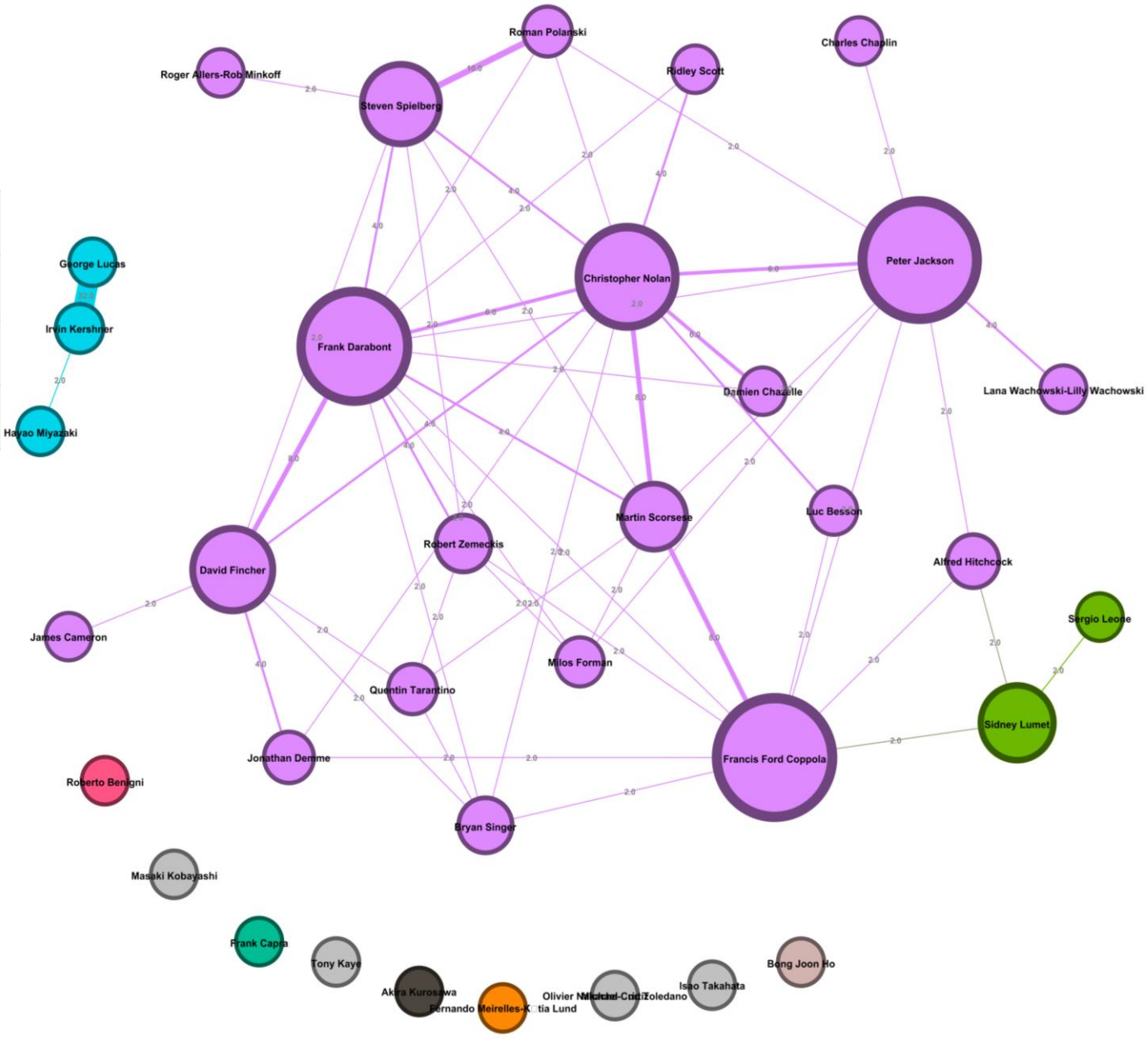
# of Communities	13
Modularity	0.10
Average Degree	3
Nodes	36
Edges	54
Diameter	5
Average Path Length	2.20
Density	0.086
# of Components	12



4 Results

4.2. Directors (Girvan-Newman)

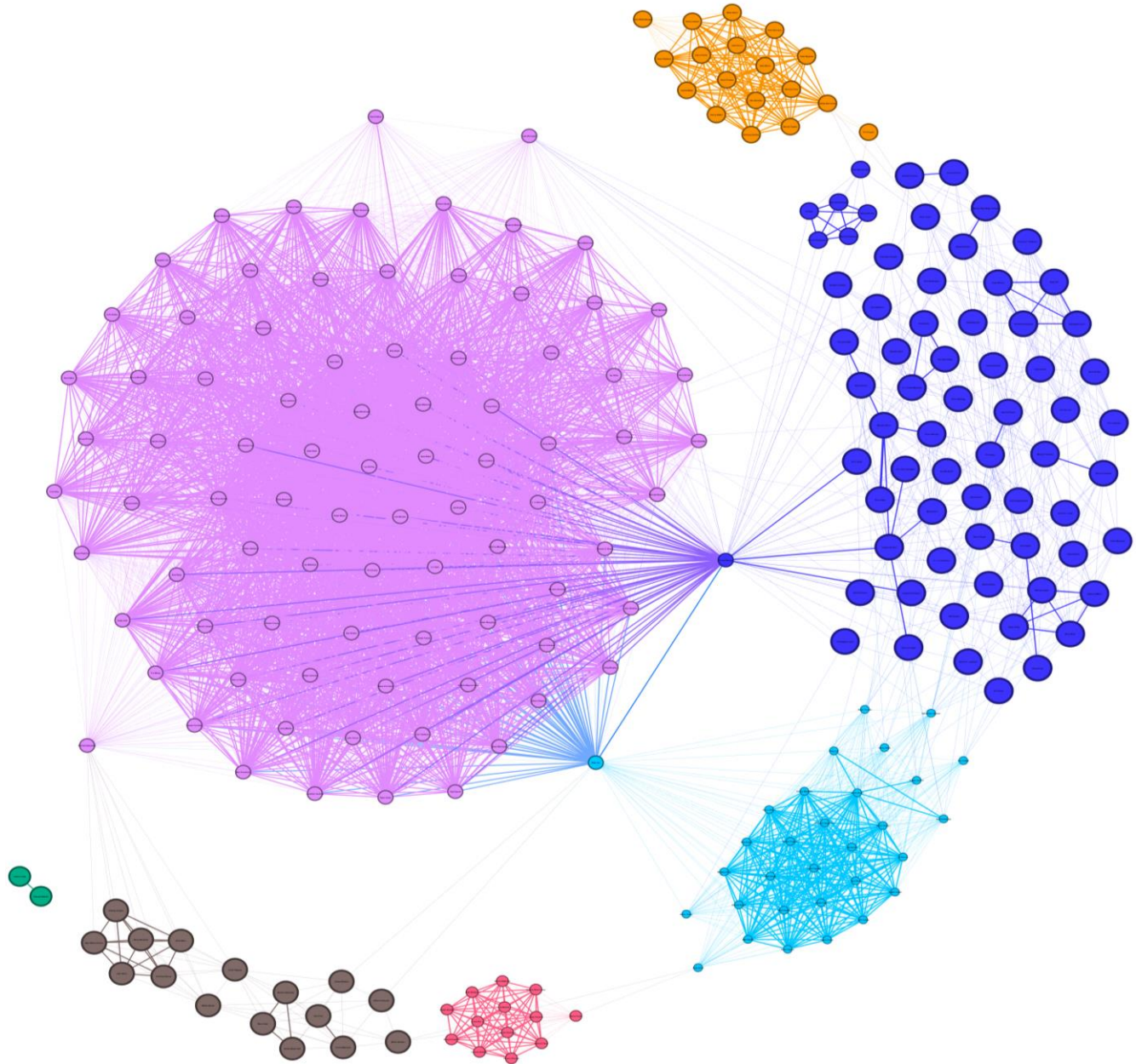
	%58	Mixed
	%8	Star Wars Directors
	%5	50’s Directors
	%2	Foreign Directors



4 Results

4.3. Cast (Louvain)

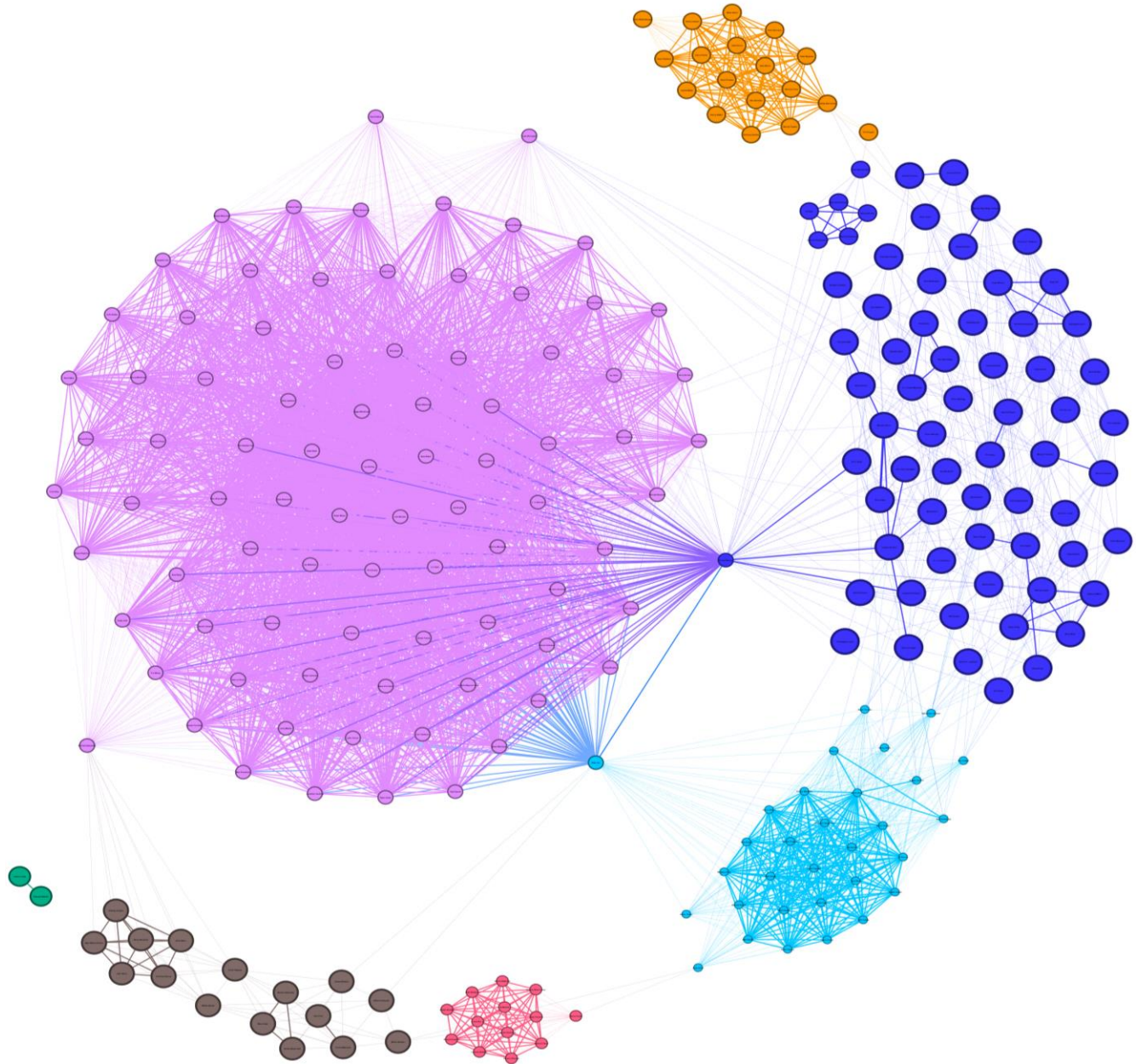
# of Communities	7
Modularity	0.341
Average Degree	41.957
Nodes	232
Edges	4867
Diameter	8
Average Path Length	2.768
Density	0.182
# of Components	2



4 Results

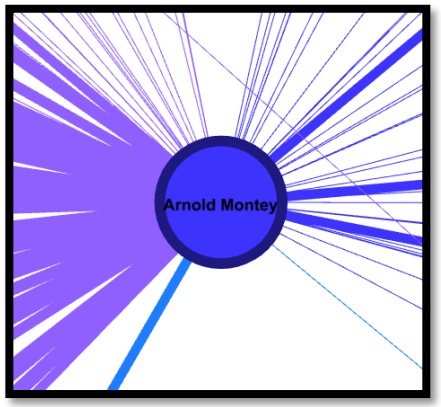
4.3. Cast (Louvain)

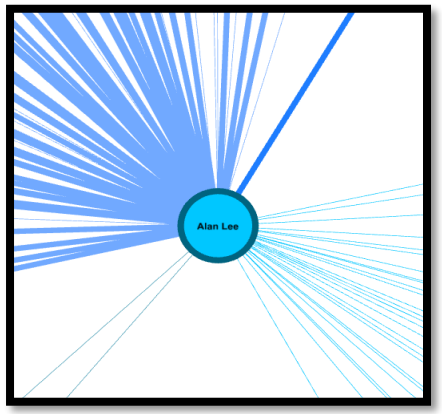
	%37	Cast of the Lord of the Rings Series
	%29	Mixed
	%13	Cast of the Godfather Series
	%8	Cast of the Star Wars Series
	%7	Chaplin and Hitchcock Cast
	%6	Cast of mid 20th century
	%1	Japanese

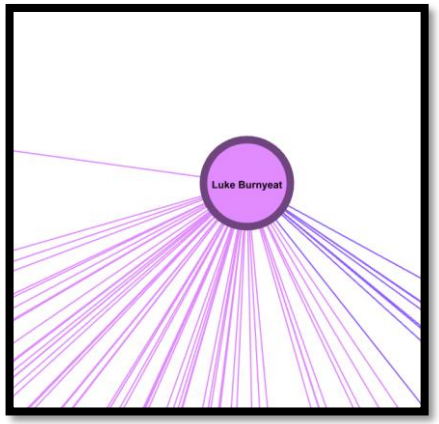


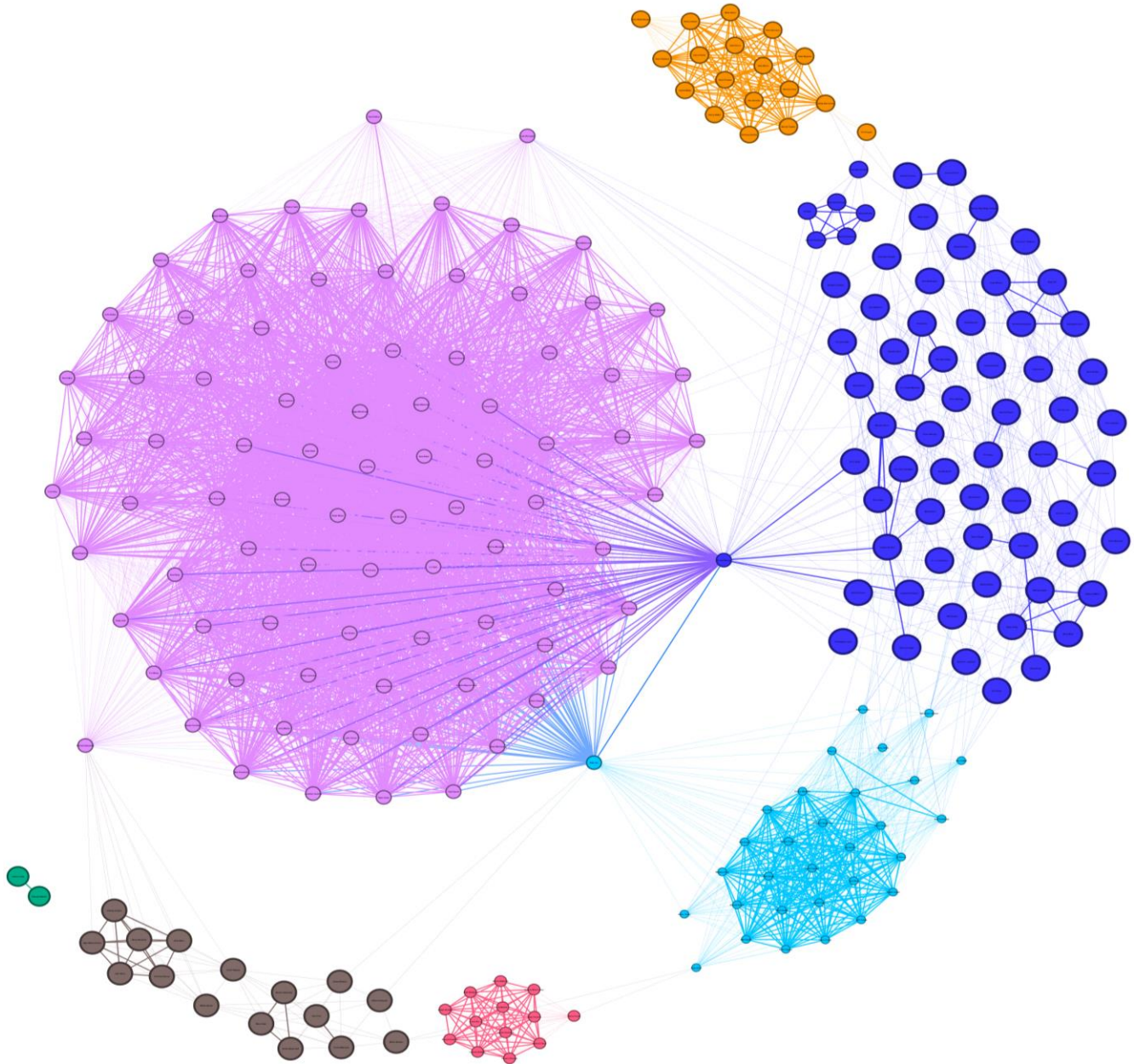
4 Results

4.3. Cast (Louvain)









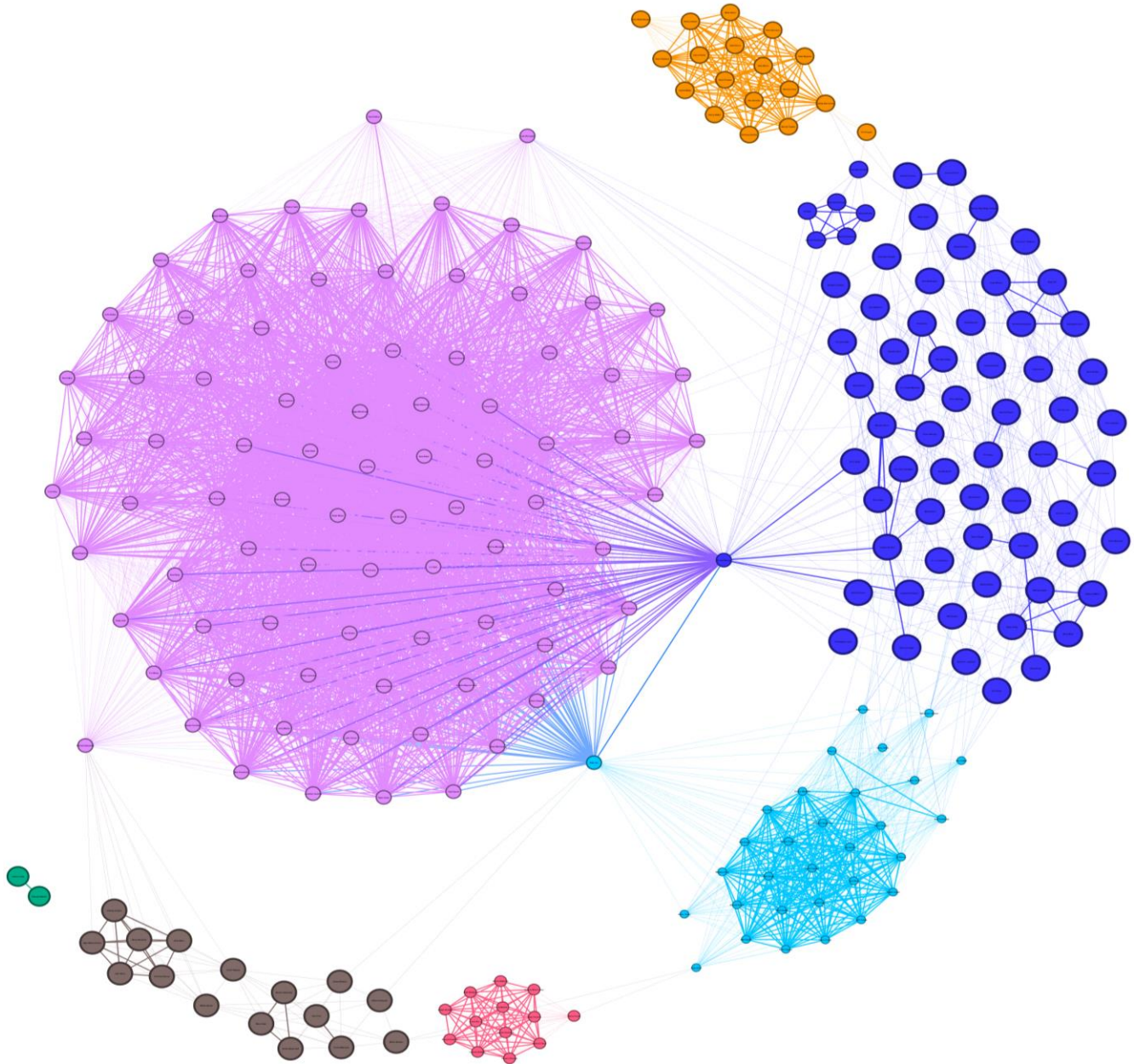
4 Results

4.3. Cast (Louvain)

Arnold Montey

Alan Lee

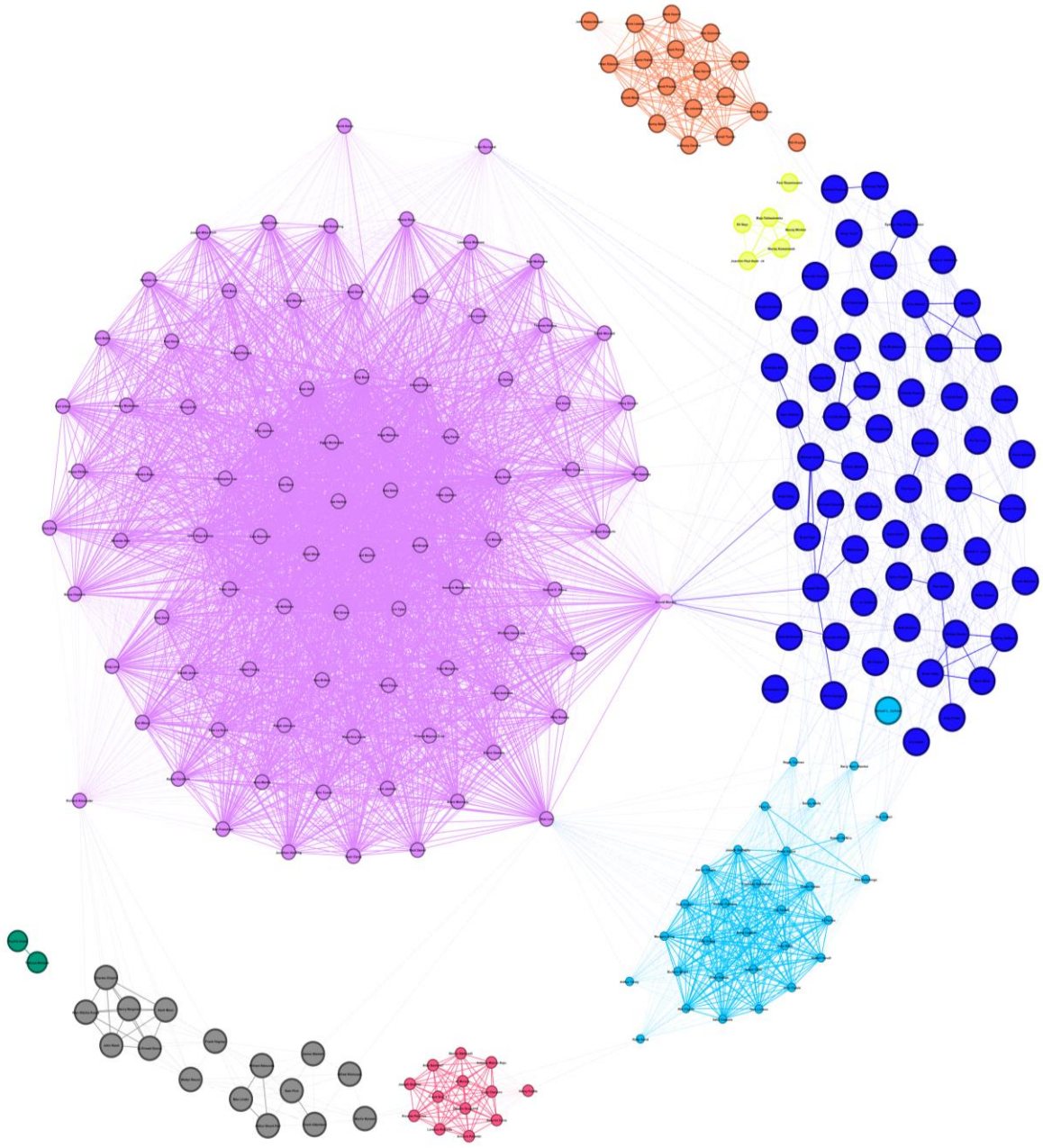
Luke Burnyeat



4 Results

4.3. Cast (Girvan-Newman)

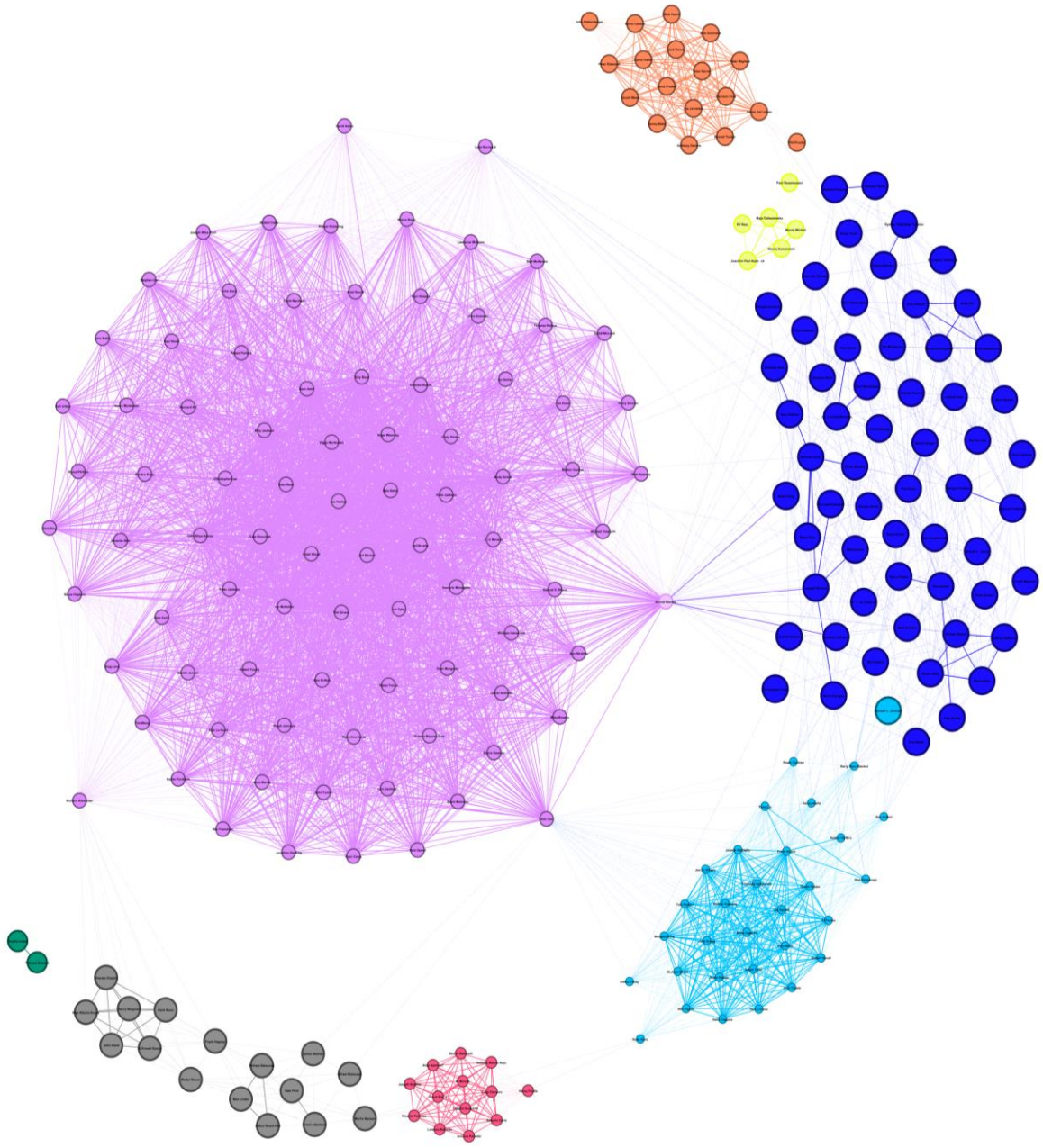
# of Communities	**8**
Modularity	0.374
Average Degree	41.957
Nodes	232
Edges	4867
Diameter	8
Average Path Length	2.768
Density	0.182
# of Components	2



4 Results

4.3. Cast (Girvan-Newman)

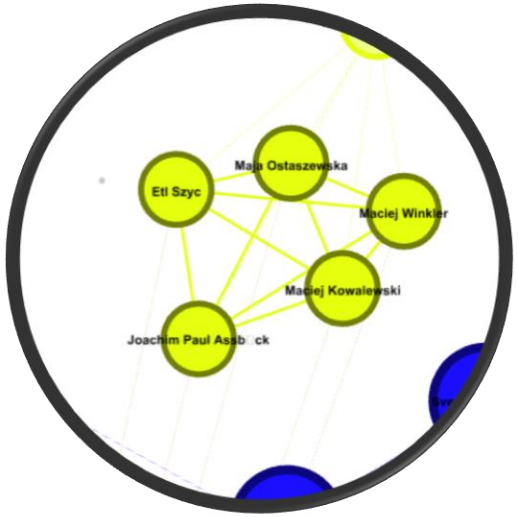
	%37	Cast of the Lord of the Rings Series
	%25	Mixed
	%13	Cast of the Godfather Series
	%8	Cast of the Star Wars Series
	%7	Chaplin and Hitchcock Cast
	%6	Cast of mid 20th century
	%3	Schindler’s List & Pianist
	%2	Japanese

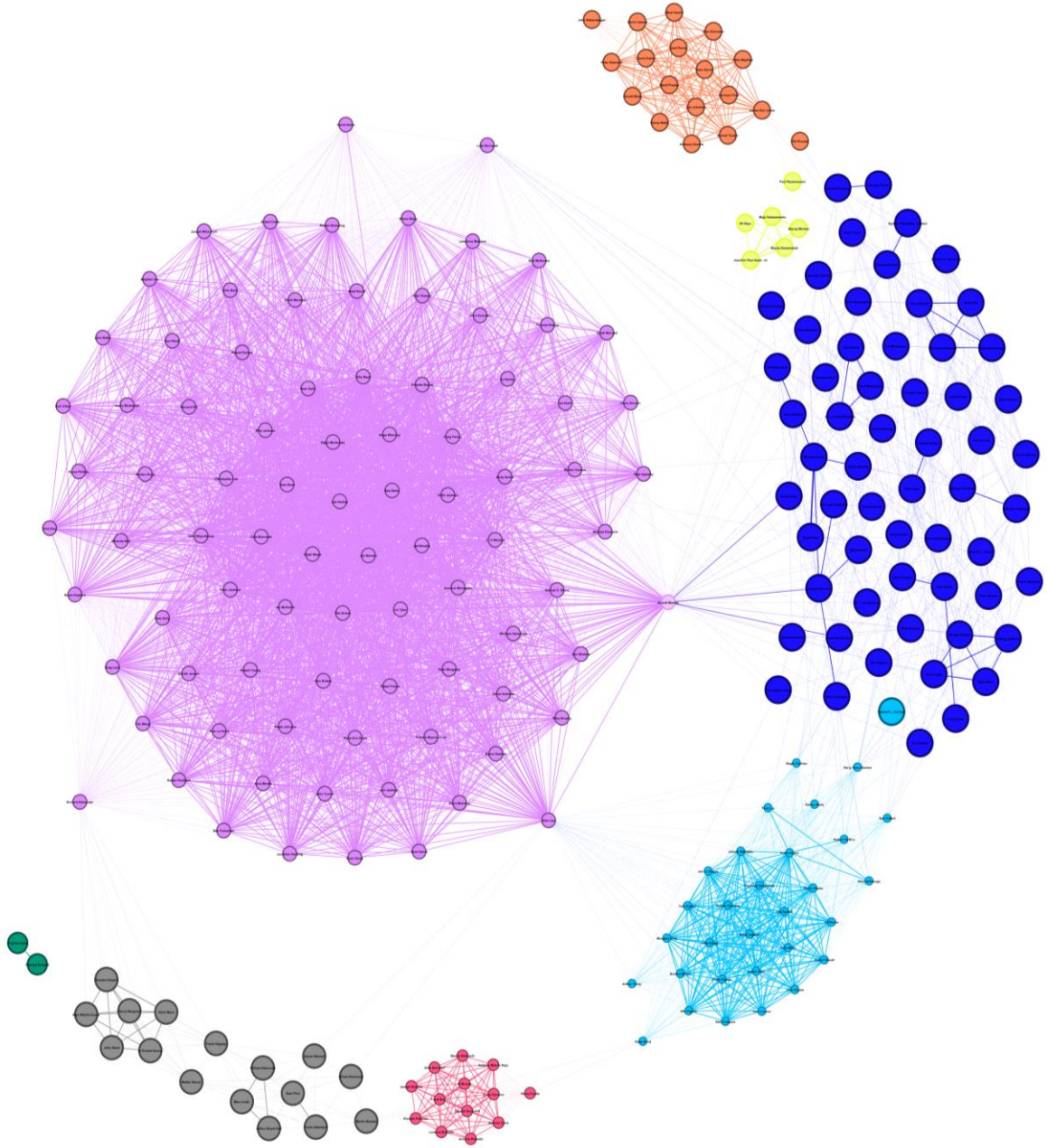


4 Results

4.3. Cast (Girvan-Newman)

Differences from Louvain



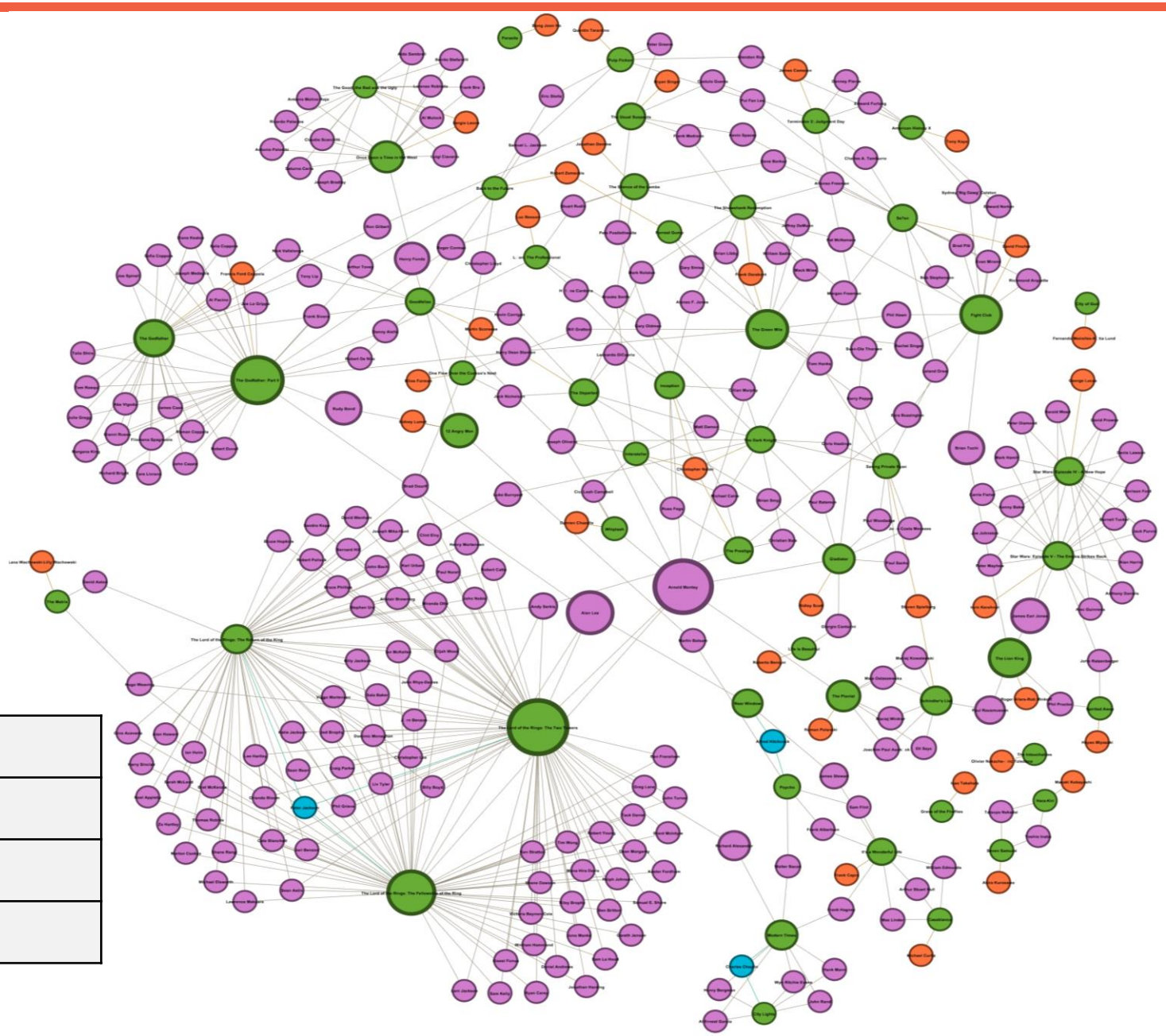


4 Results

4.4. General (Genre Graph)

Average Degree	3.511
Nodes	315
Edges	553
Diameter	17
Average Path Length	5.768
Density	0.011
# of Components	6

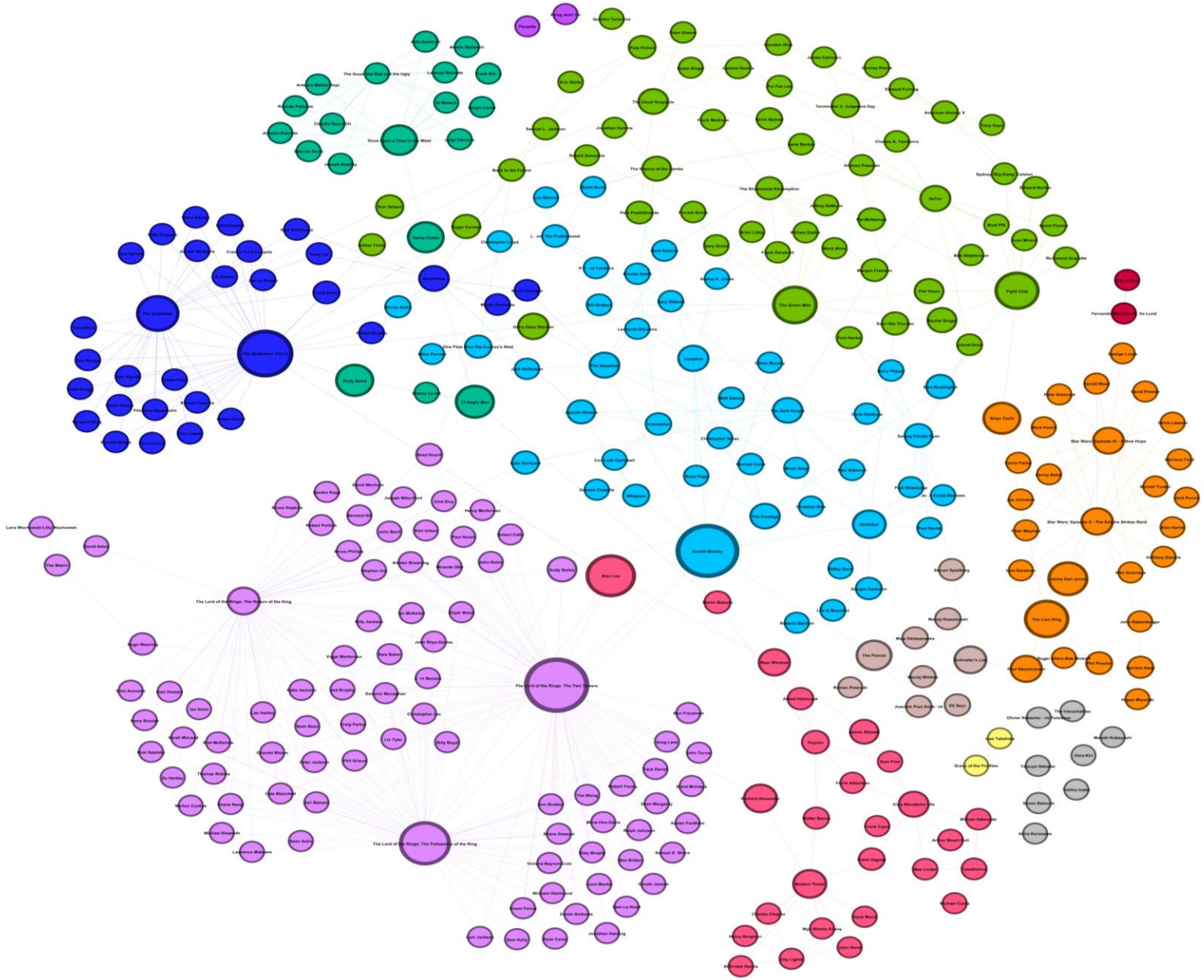
	%73	Actors
	%16	Movies
	%11	Directors
	%1	Actor and Director



4 Results

4.4. General (Louvain)

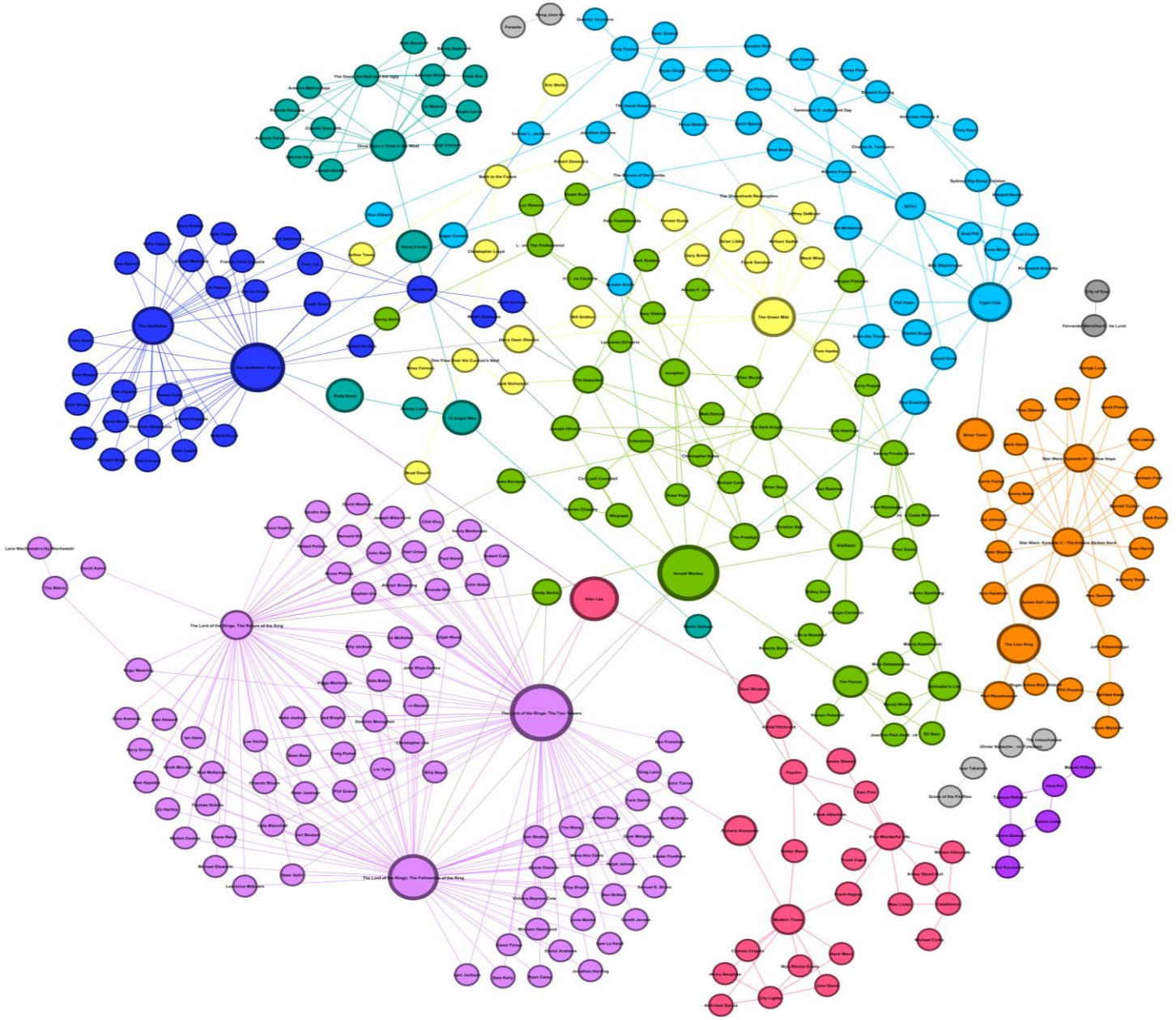
# of Communities	13
Modularity	0.734
Average Degree	3.511
Nodes	315
Edges	553
Diameter	17
Average Path Length	5.768
Density	0.011
# of Components	6



4 Results

4.4. General (Louvain)

	%27	Lord of the Rings
	%16	Drama
	%12	Crime
	%10	Godfather
	%9	Star Wars
	%8	60’s
	%7	Cult Drama
	%6	Western
	%1	Japanese
	%1	Foreign

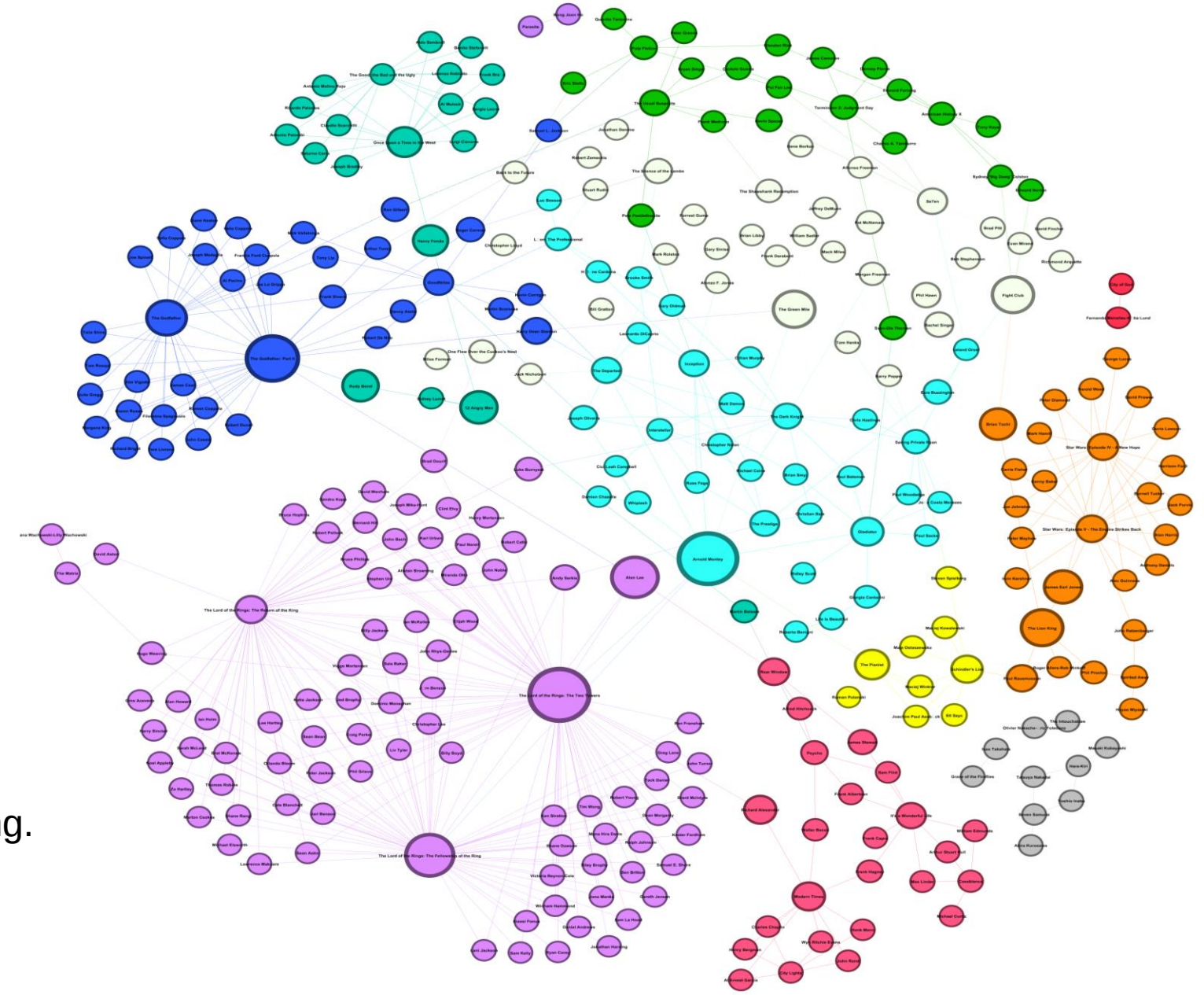


4 Results

4.4. General (Girvan-Newman)

# of Communities	14
Modularity	0.735
Average Degree	3.511
Nodes	315
Edges	553
Diameter	17
Average Path Length	5.768
Density	0.011
# of Components	6

Almost same with the Louvain clustering.



5 References

[1] “Top 250 Movies,” Imdb.com. [Online]. Available: https://www.imdb.com/chart/top/. [Accessed: 06-Dec-2021].

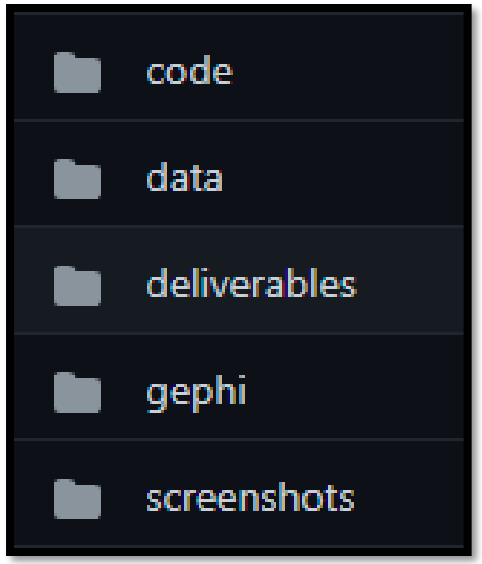
[2] “IMDbPY,” Github.io. [Online]. Available: https://imdbpy.github.io/. [Accessed: 25-Oct2021].

[3] L. Rita, “Louvain Algorithm,” Towards Data Science, 09- Apr-2020. [Online]. Available: https://towardsdatascience.com/louvain-algorithm93fde589f58c. [Accessed: 06-Dec-2021].

[4] B. Kong, L. Zhou, and W. Liu, “Improved Modularity Based on Girvan-Newman Modularity,” in 2012 Second International Conference on Intelligent System Design and Engineering Application, 2012.

GitHub

- https://github.com/malialtunsoy/network-of-movies



THANKS

FOR WATCHING