Aggregation Phenomena in Movie Industry

-- The secret hidden in the film industry

Team 9

Jimin Wang, Shichao Jia,
Zhuoyue Wang, Rui Chen

Contents:

Overview

Research goal & introduction

Data description & Preprocessing

Data Exploration

- Word Cloud Analysis
- Movie Genre Analysis

Data Exploitation

- Movie Graph connected by similarity of main actors
- Graph of Actors' Co-Appearance
- Movie Graph connected by similarity of main crew

Conclusion & Outlook

Overview

Investigate how various roles in the movie industry cooperate with each other and aim to discern the patterns underlying their cooperations through the approach of networks.

Build graphs of the cooperation networks mainly from three perspectives, the movies, the actors, and the crew.

Find a kind of aggregation phenomena that pervades the movie industry, and such phenomena are explored and visualized with in-depth analysis.

Data description & Preprocessing:

5000 movies (1910 ~ 2016)

Part I: 20 features of the movie (i.e. genres, vote average, production companies, revenue, etc.)

Part II: directors & actors

New features based on given data set:

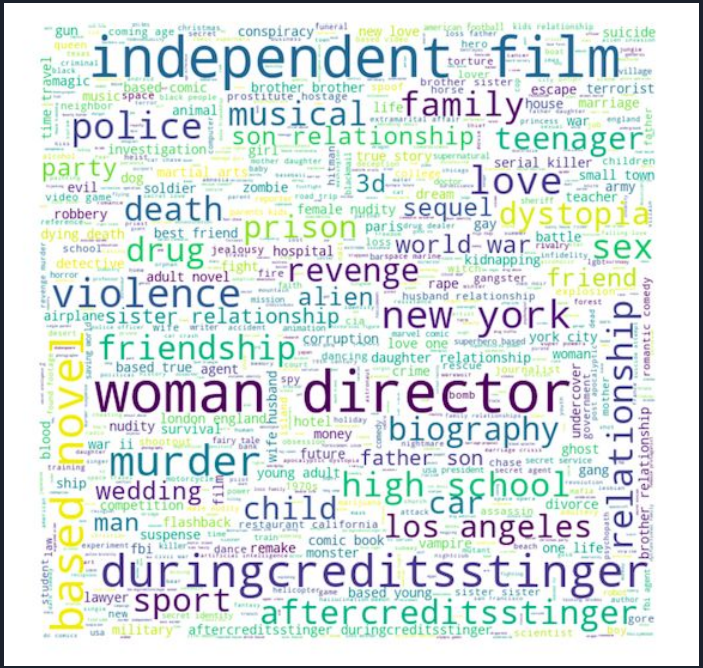
1) genre_main: The first genre of each movie.

2) genre_id: Assign an index for each genre.

3) Return of Investment (ROI): (revenue - budget) / budget

4) success: Set the success label to 1 if movie's ROI is more than 0.6, otherwise 0.

Data Exploration:
Word Cloud Analysis

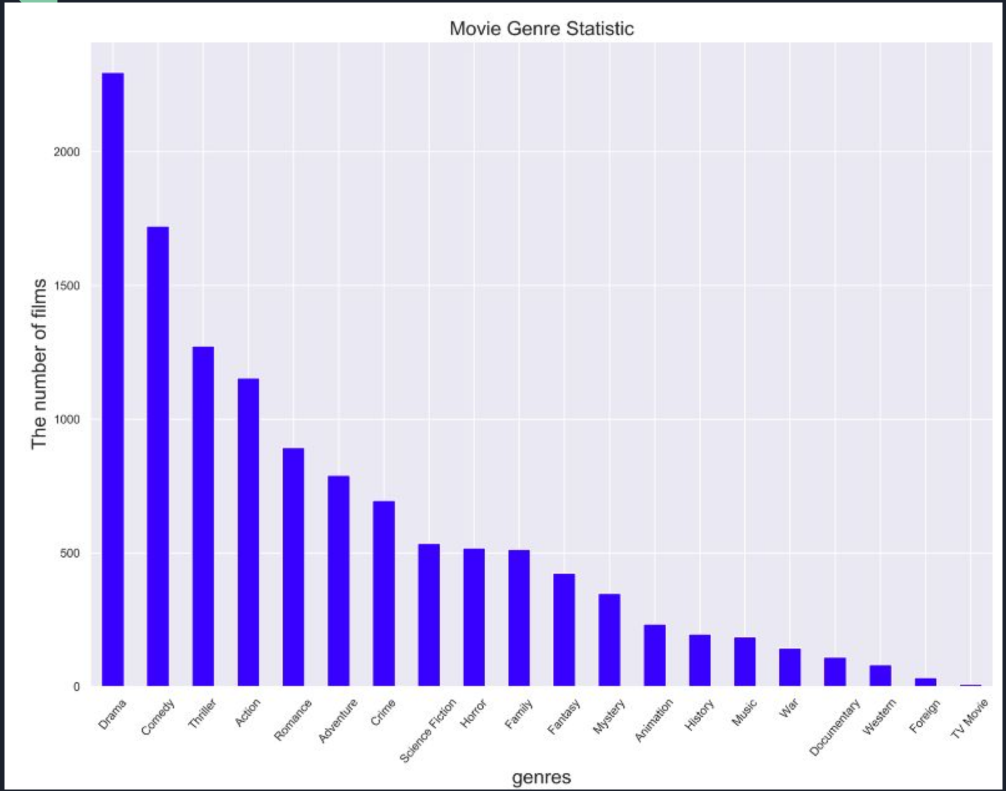


Woman director; Independent

Drama; Comedy

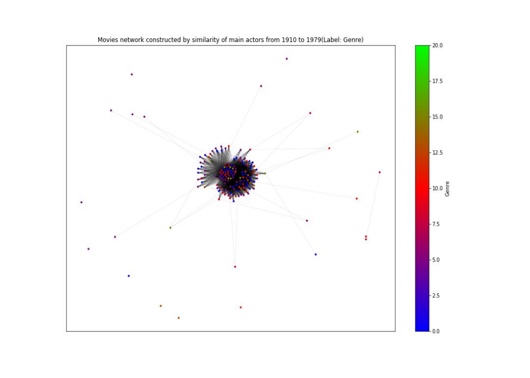
Data Exploration:

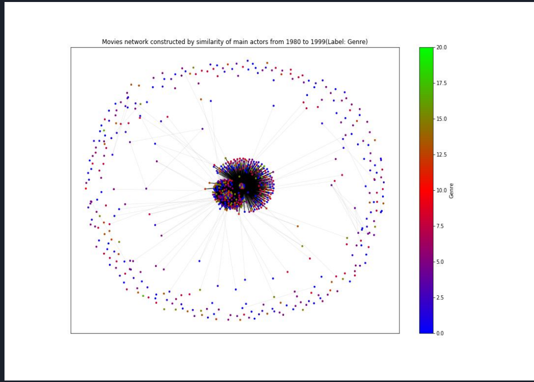
Movie Genre Analysis

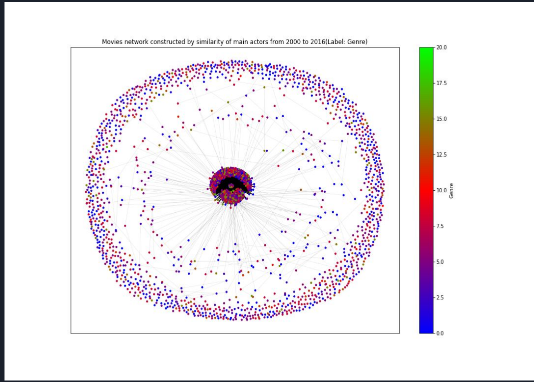


- Count the number of movies belonging to different genres from 1960 to 2017
- Drama, comedy and thriller ranked top 3
- Documentaries are perhaps not well recognized by the market
- Comedy & Adventure tend to obtain high profit

Movie graph constructed by similarity of leading actors





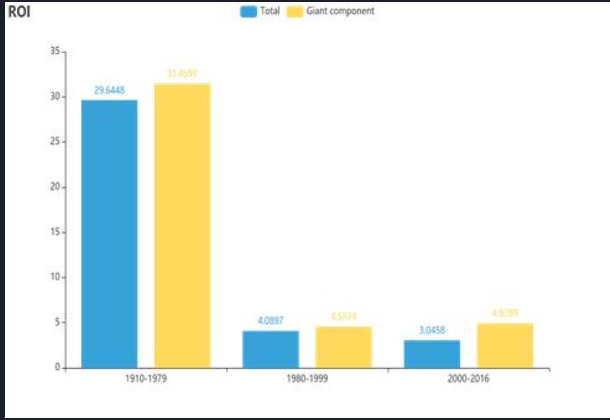


1910-1979 | 1980 - 1999 | 2000 - 2016

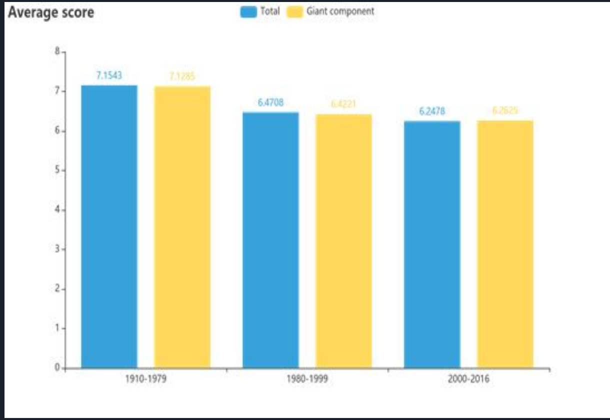
Movie graph constructed by similarity of main actors

- Built based on features denoted by the existence of actor
- Nodes: Movies
- Edges: Similarity of main actors

Aggregation effect analysis of different periods



ROI

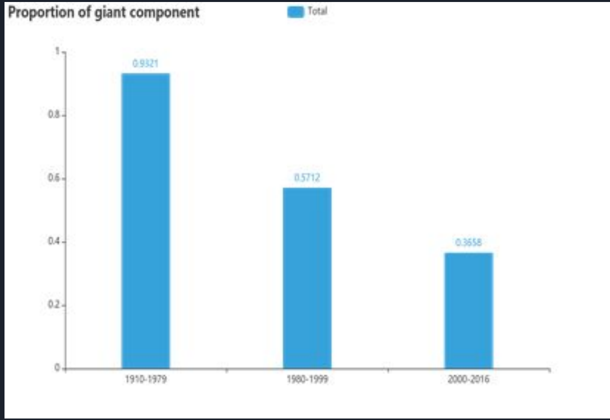


Average score

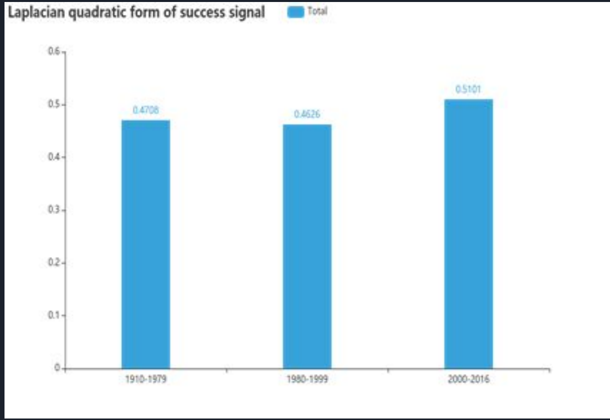
Key data Analysis of giant component(a)

- Guarantee of film investment success
- No improvement in public praise

Aggregation effect analysis of different periods



Proportion of giant component



The Laplacian quadratic form of giant component by success label

Key data Analysis of giant component(b)

- This aggregation effect is attenuated
- From 1980 to 1999 this aggregation is more effective to determine investment success

Graph of Actors’ Co-Appearance

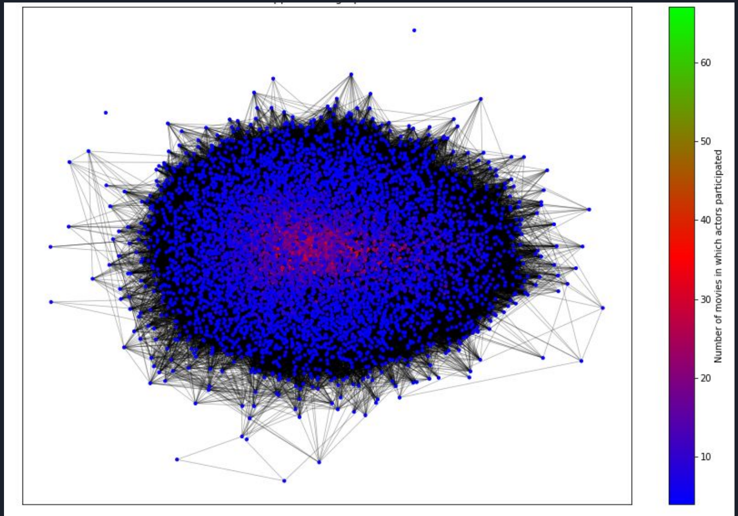
Nodes

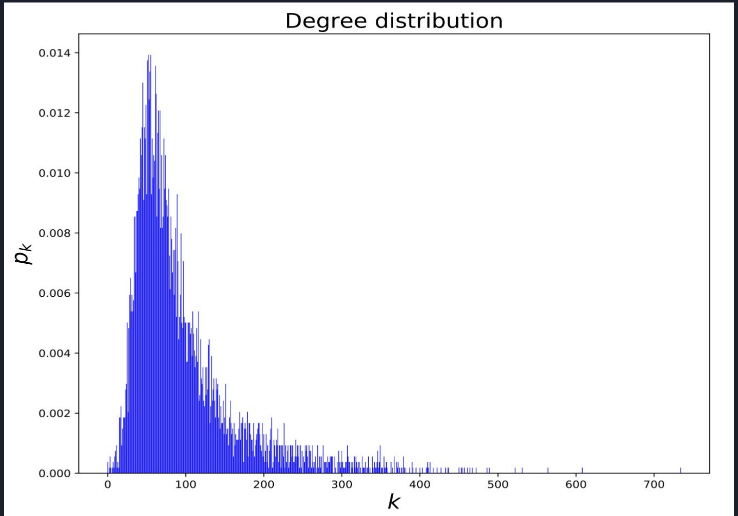
- Actors
- Only those who participated in more than 3 movies are kept
- 54,198 nodes ⇒ 5,383 nodes

Edges

- Co-appearance between actors
- Two nodes are linked only if these two actors co-appeared in <u>at least *w* movies</u>

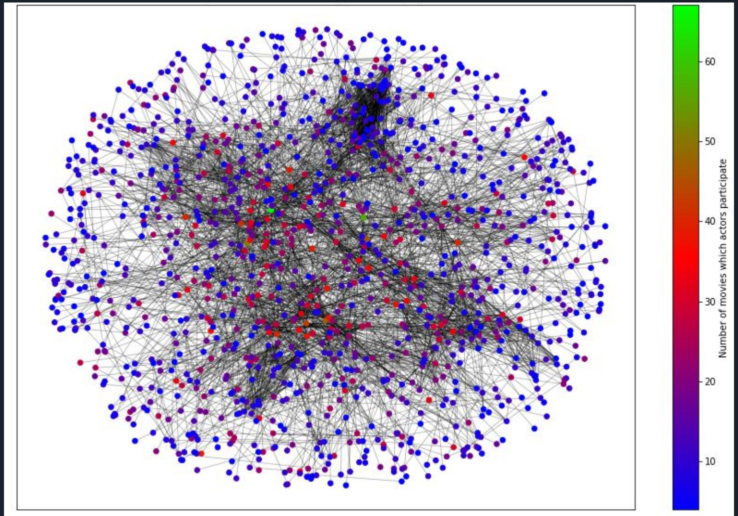
Graph of Actors' Co-Appearance (w = 1)

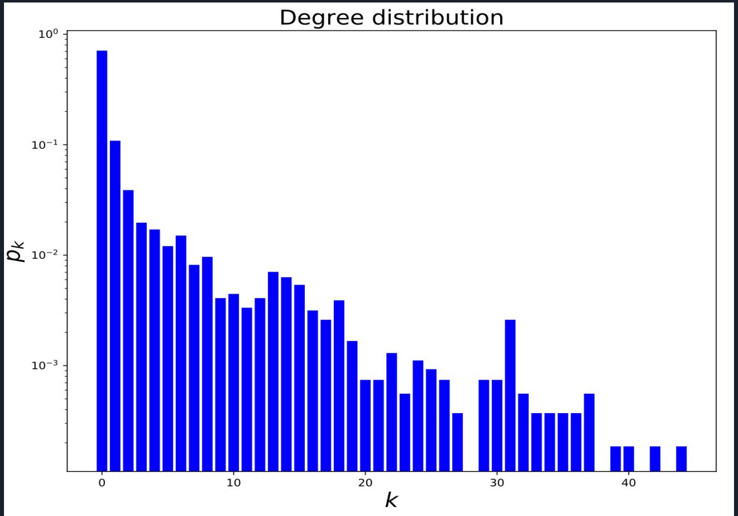




Graph of actors' co-appearance when $w = 1$ and the degree distribution

Graph of Actors’ Co-Appearance (w = 3)

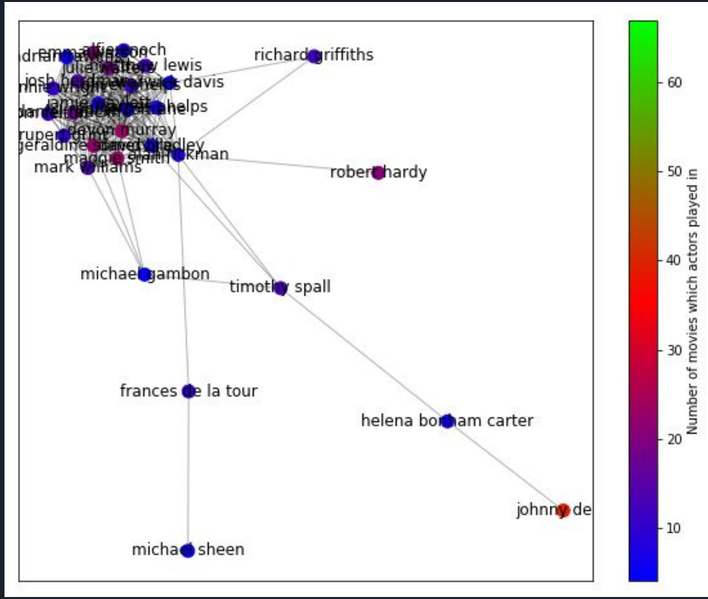




Largest component of actors’ co-appearance
and the degree distribution of the entire graph when w = 3

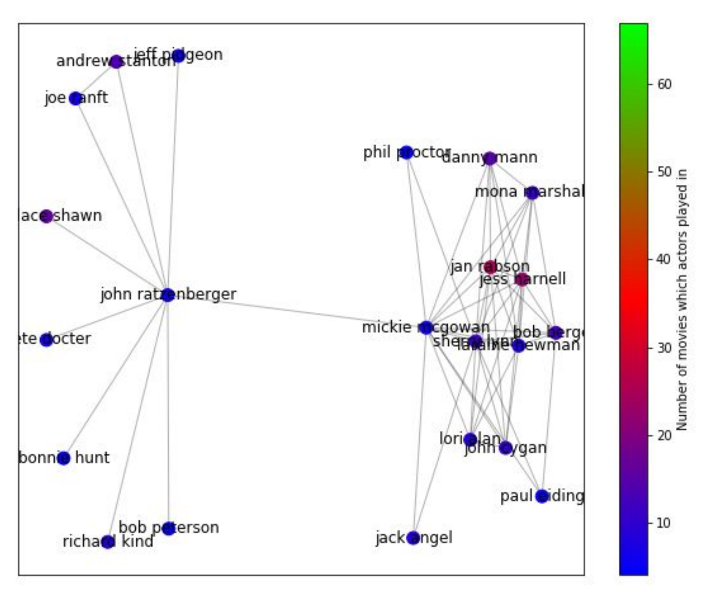
Search of Frequent Co-operations (w = 5)

Harry Potter



2nd largest component

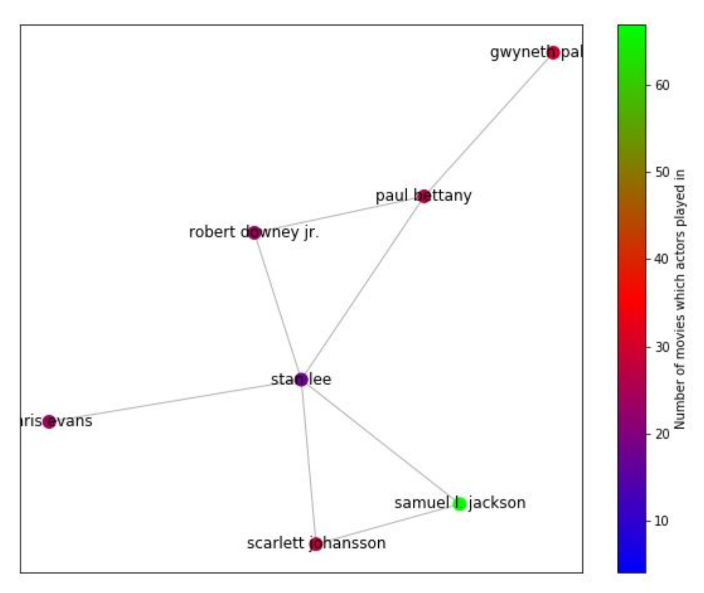
Pixar



4th largest component

Search of Frequent Co-operations (*w* = 5)

Marvel Universe



8th largest component

Findings of Frequent Co-operations

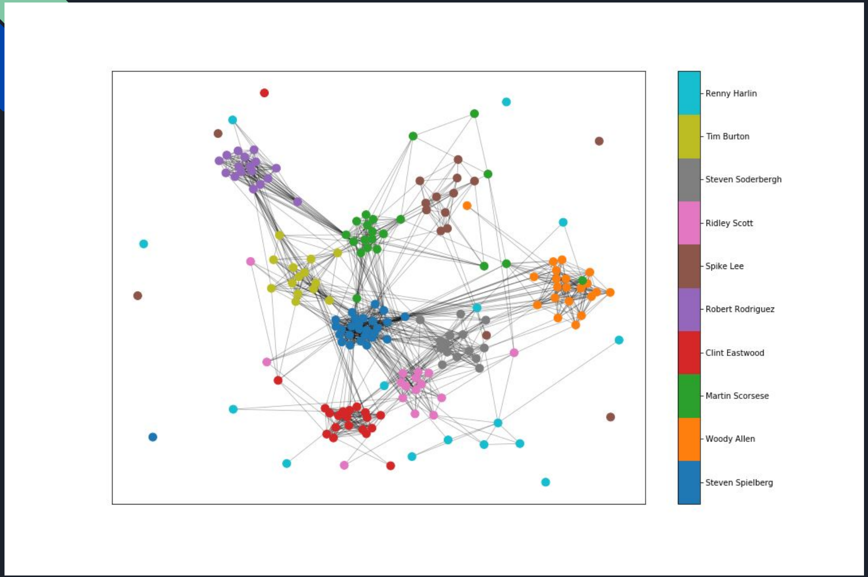
Movies

- Series movies (Harry Potter, Lord of Rings)
- Media franchise movies (Marvel Universe)
- Animation movies

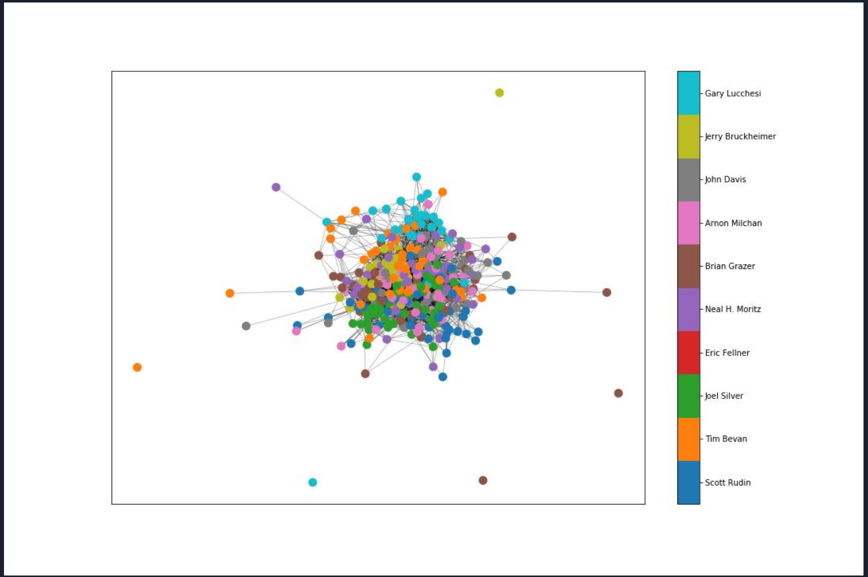
Dominant Genres

- Comedy & adventure
- 2 most lucrative genres

Movie graph constructed by main crews



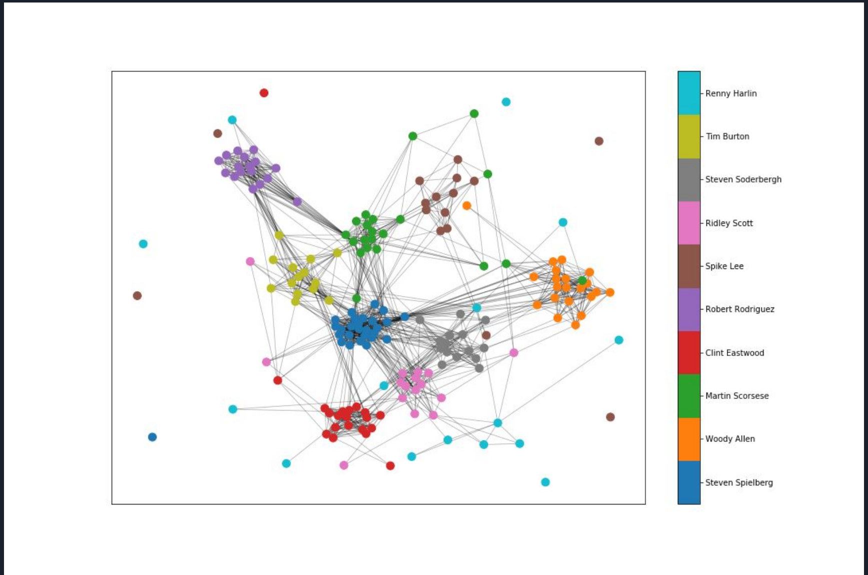
top directors' movies



top producers' movies

- Built based on features denoted by the existence of top crews
- Nodes: Movies
- Edges: Similarity of main crews

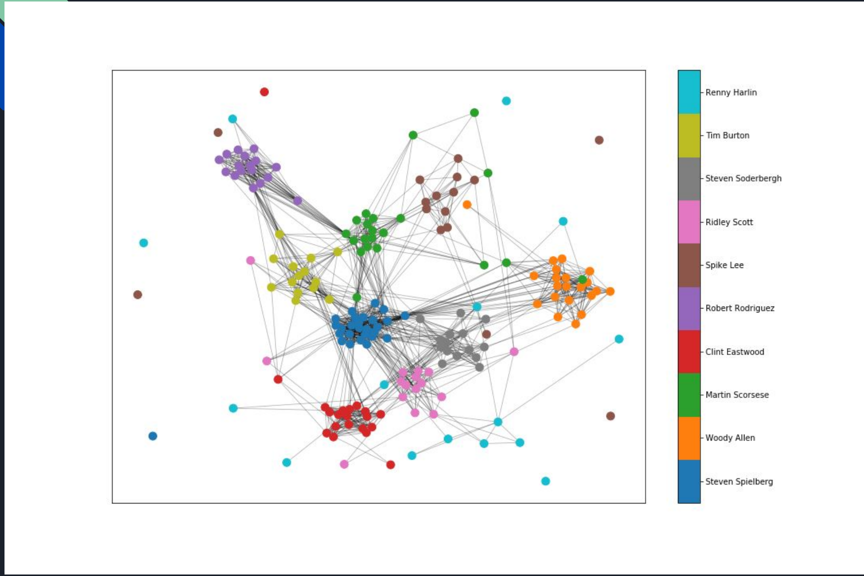
Movie graph constructed by main crews



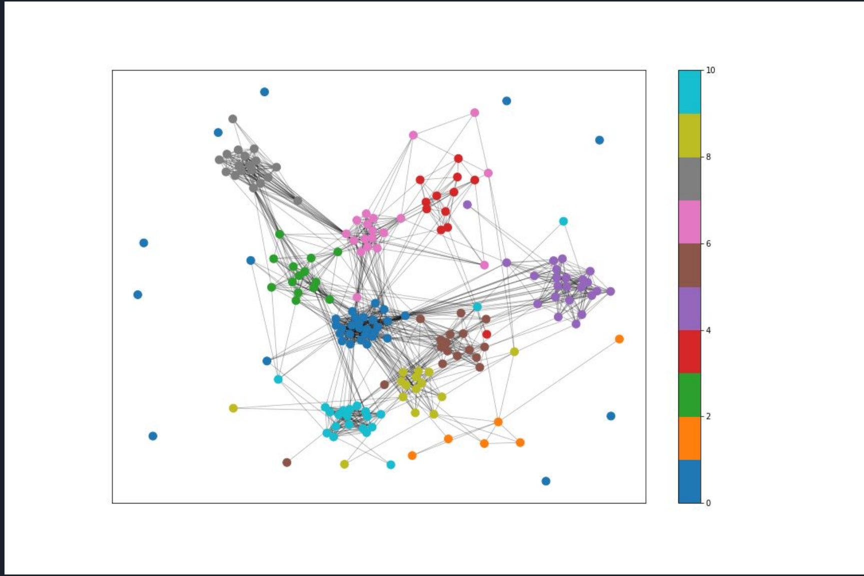
top directors' movies

- Obvious clustering

Movie graph constructed by main crews



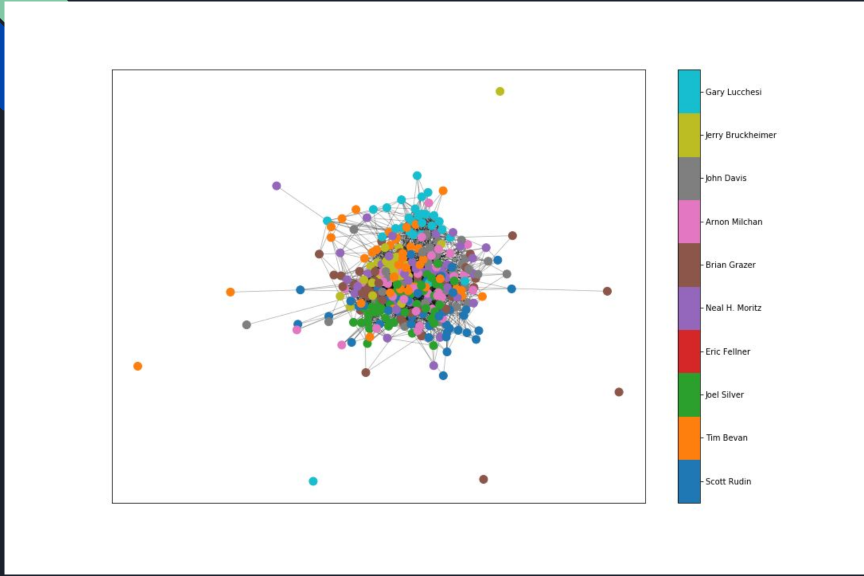
top directors' movies



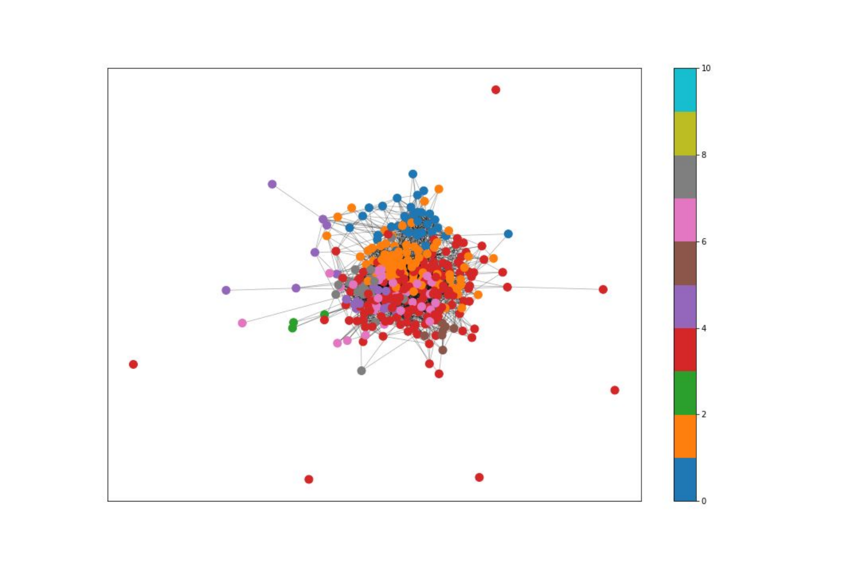
predicted by spectral clustering

- Obvious clustering
- Nice performance of Spectral custering with 90.1% of F1 score

Movie graph constructed by main crews



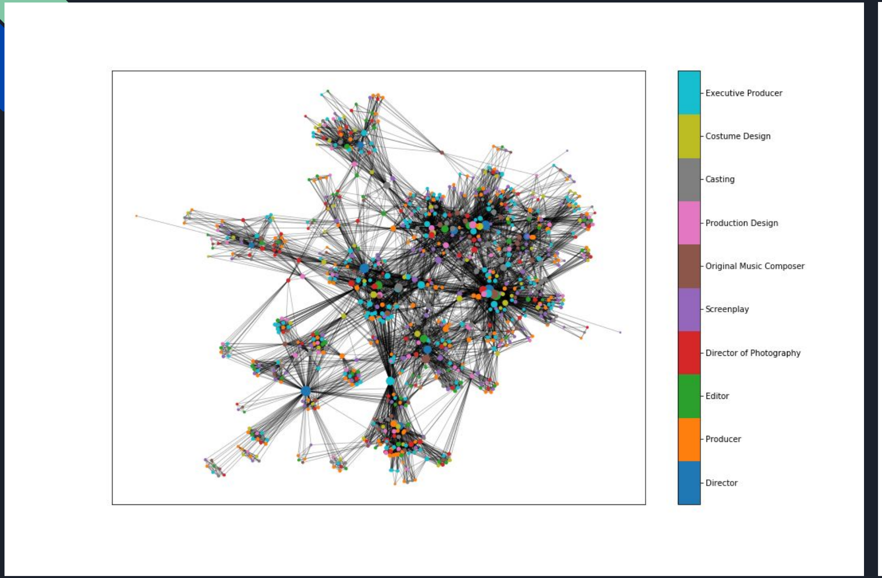
top producers' movies



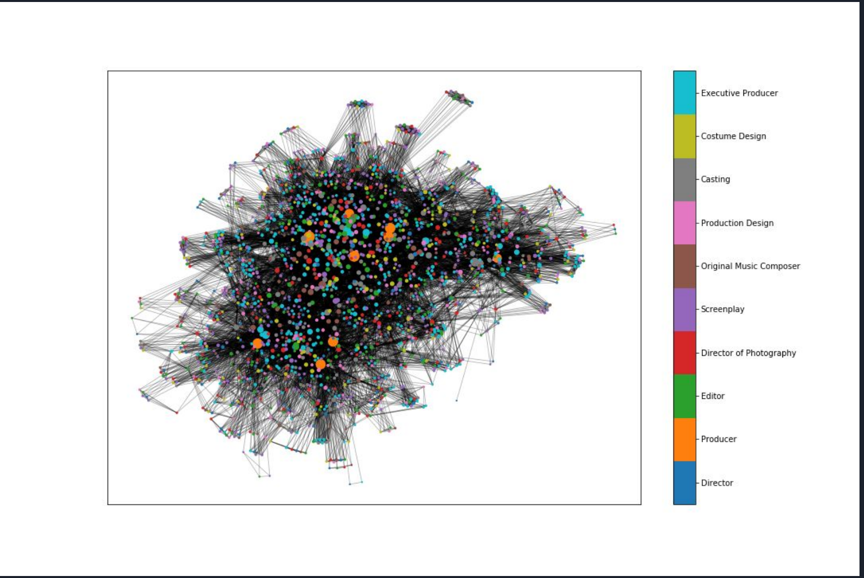
predicted by spectral clustering

- No obvious clustering
- Spectral clustering with 40.9% of F1 score

Movie graph constructed by main crews



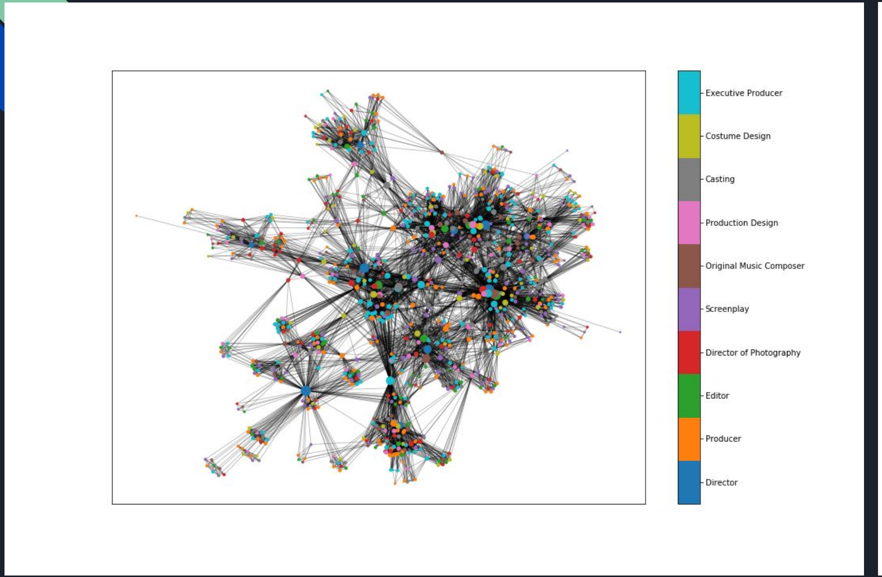
top directors' crews



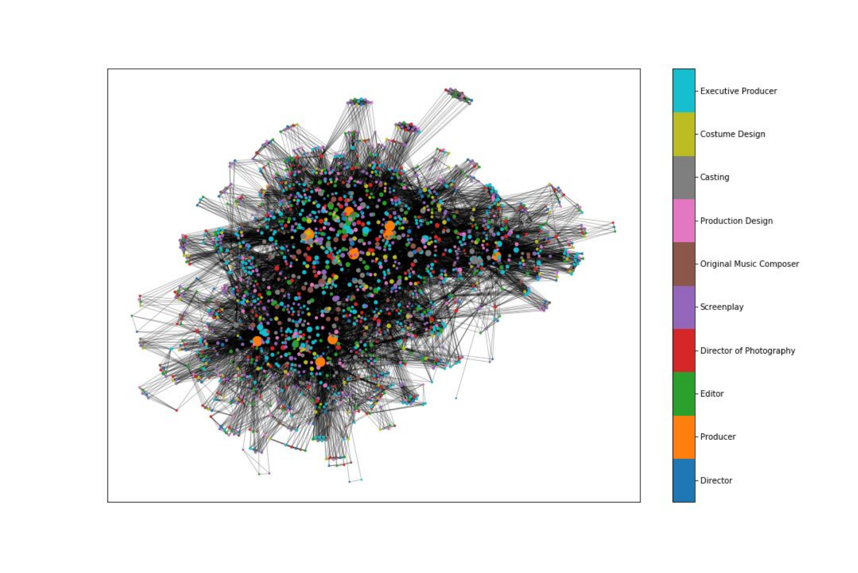
top producers' crews

- Built based on features denoted by the existence of top crews
- Nodes: crews appeared in selected movies
- Edges: Co-appear times in same movie production

Movie graph constructed by main crews



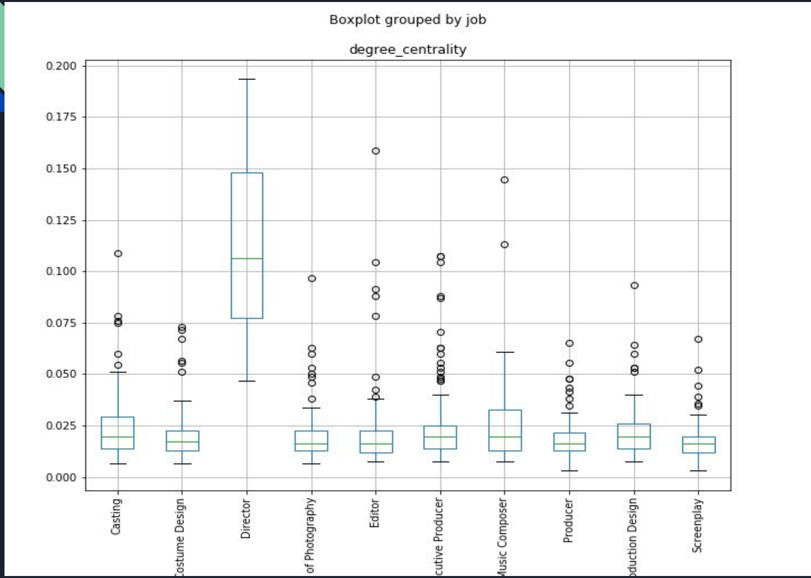
top directors' crews



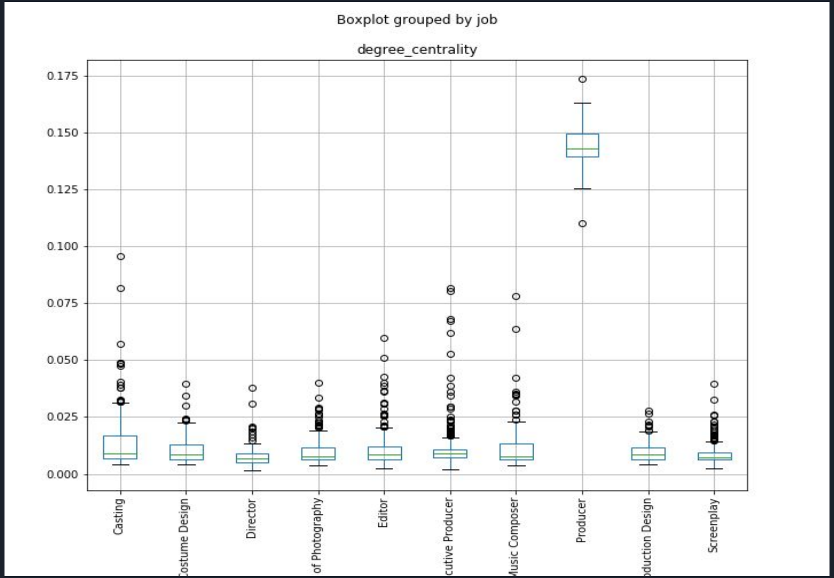
top producers' crews

- Clustering is more obvious in director's graph

Movie graph constructed by main crews



centrality degree of top directors' crews



centrality degree of top producer' crews

- Compare to the centrality degree of top producers, top director's centrality is closer to other crews.

Conclusion & Outlook

- Three networks to disclose the aggregation phenomena
- Movie networks based on similar actors: the movies in the aggregation are more profitable
- Actors' co-appearance networks: actors tend to frequently cooperate in series movies and media franchise that belong to lucrative genres
- Crew networks: compared to productive producers, productive directors prefer a similar crew to work with
- Unbiased selection of movie subsets is required to assess the dominance of directors and producers

Thank you for your attention