



IMDB



Network Analysis

Παντελίδου Κυριακή-Νεκταρία (AEM: 765)

Μαρινοπούλου Θεοκτίστη (AEM: 32)

Our dataset



- Kaggle dataset
- ~1000 most popular imdb movies
- 2006-2016
- Preprocessing in Python

Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced to work together to stop a fanatical warrior from taking control of the universe.	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe Saldana	2014	121	8.1	757074	333.13	76
2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a team finds a structure on a distant moon, but they soon realize they are not alone.	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fassbender, Charlize Theron	2012	124	7	485820	126.46	65
3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diagnosed 23 distinct personalities. They must try to escape before the apparent emergence of a frightful new 24th.	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richardson, Jessica Sula	2016	117	7.3	157606	138.12	62

Nodes & edges for the Gephi Platform

Id	Label
0	AamirKhan
1	SakshiTanwar
2	FatimaSanaShaikh
3	SanyaMalhotra
4	LeonardoDiCaprio

nodes.csv

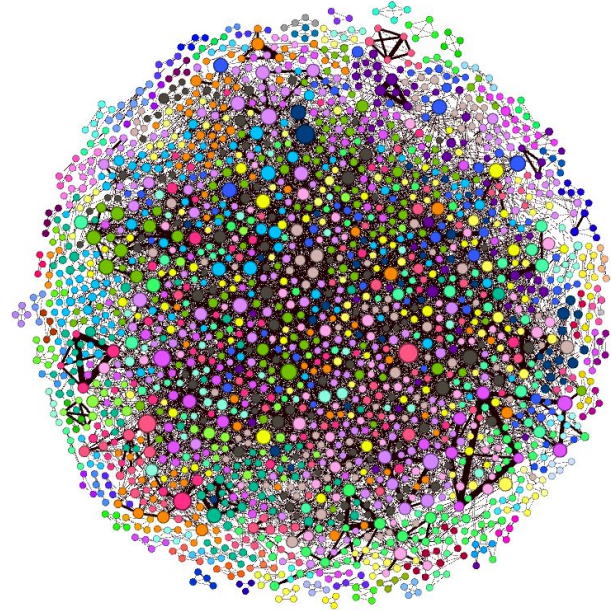
1 | Aamir Khan, Sakshi Tanwar, Fatima Sana Shaikh, Sanya Malhotra

Source	Target
0	1
0	2
0	3
1	2
1	3
2	3

edges.csv

Initial Graph

- Undirected
- 1916 nodes
- 5819 nodes
- Bigger degree \rightarrow Bigger node size
- Bigger weight \rightarrow Bolder edge
- Density: 0.003
- Diameter: 9



Q1: Who are the top influencers?

- Based on **Degree**
- High Degree = Many collaborations
- Average Degree = 5.878



2. Brad Pitt
(45)



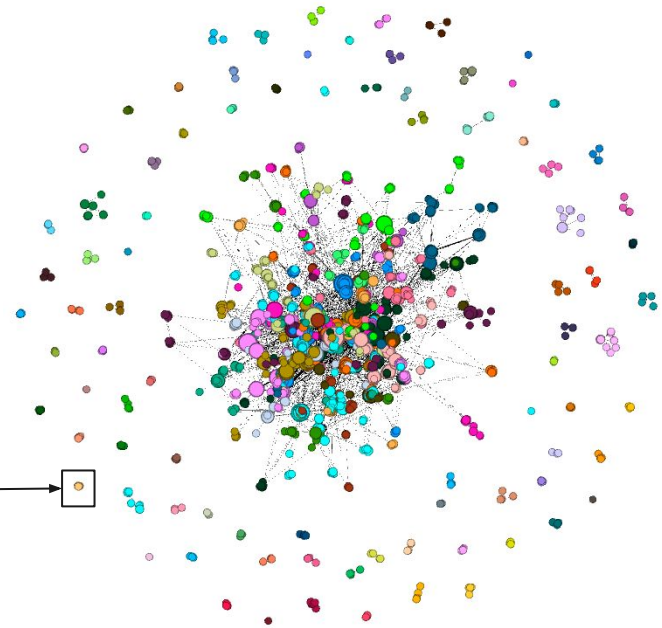
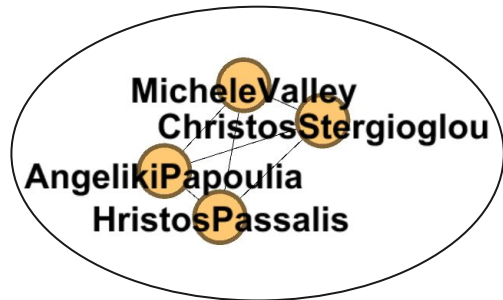
1. Hugh Jackman
(47)



3. Christian Bale
(42)

Q2: Which actors work isolated from Hollywood?

- Based on **Closeness Centrality**
- more central node = closer to all other nodes
- communities found outside = cliques
- not american actors seems to be more isolated from the center



Q3: Which actors are important for the communities' cohesion?

- Based on **Betweenness Centrality**
- How often an actor appears on shortest paths between other actors
- Average Path length: 4.171



2. Hugh Jackman
(0.025)



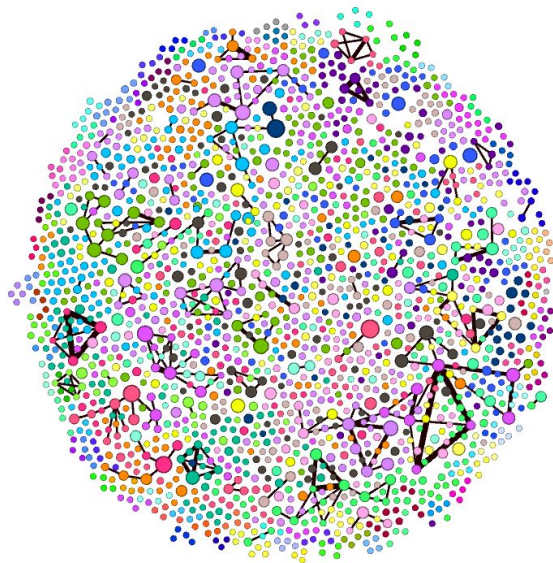
1. Mark Wahlberg
(0.026)



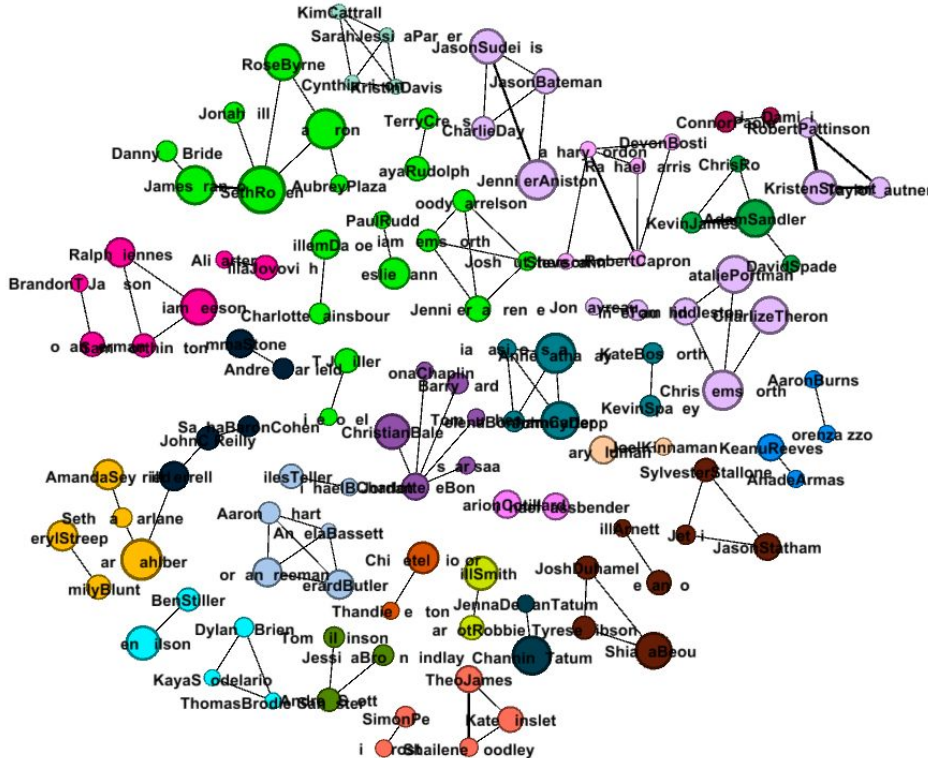
3. Christian Bale
(0.023)

Q4: How successful are actors' collaborations?

- most connected actors
- Gephi Filter → Edge Weight
- 2-4 weight
- 404 connected actors (202 collaborations)
- Split network
 - Low rated (≤ 7)
 - High rated (> 7)



Collaborations in low rated movies



- 103 collaborations
 - ◆ 50% of initial network
- graph cliques = sequels
 - ◆ 42% sequels
- sequels' genre
 - ◆ adventure
 - ◆ comedy
 - ◆ action
 - ◆ sci-fi
 - ◆ romance
 - ◆ drama

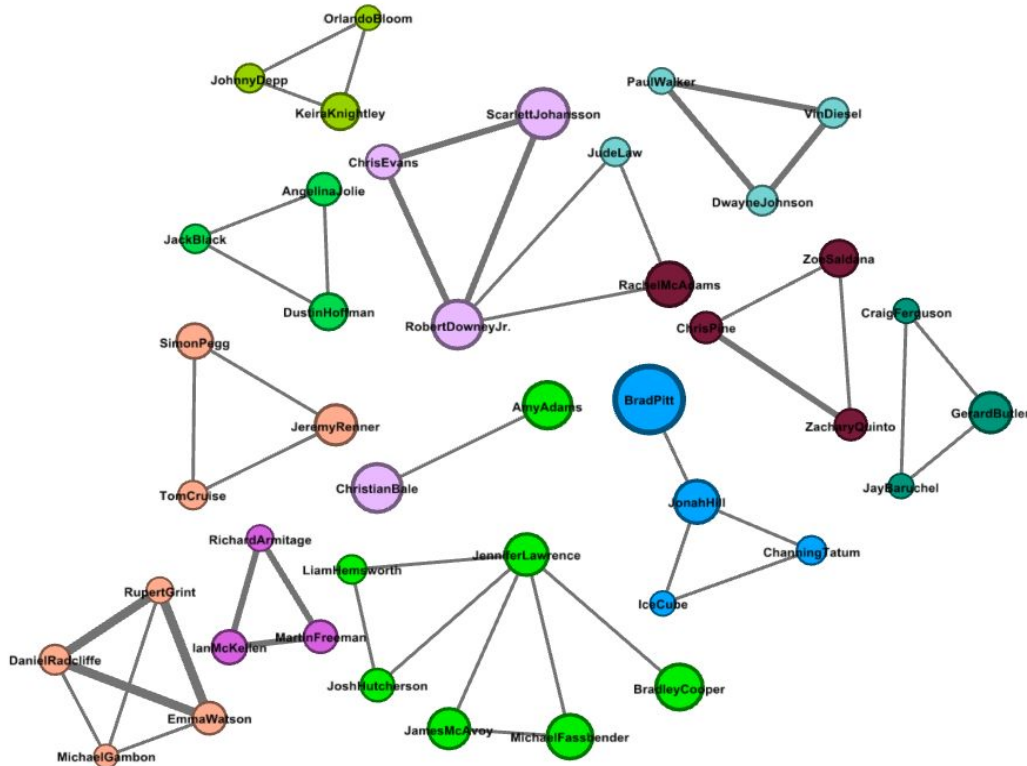
Bad collaboration pairs

Michael Fassbender & Marion Cotillard

Jennifer Aniston & Jason Sudeikis

Natalie Portman & Chris Hemsworth

Collaborations in high rated movies



- 60 collaborations
 - ◆ 30% of initial network
 - ◆ (20% lost because of split)
- graph cliques = sequels
 - ◆ 65% sequels
- sequels' genre:
 - ◆ adventure
 - ◆ fantasy

Most succesful pairs

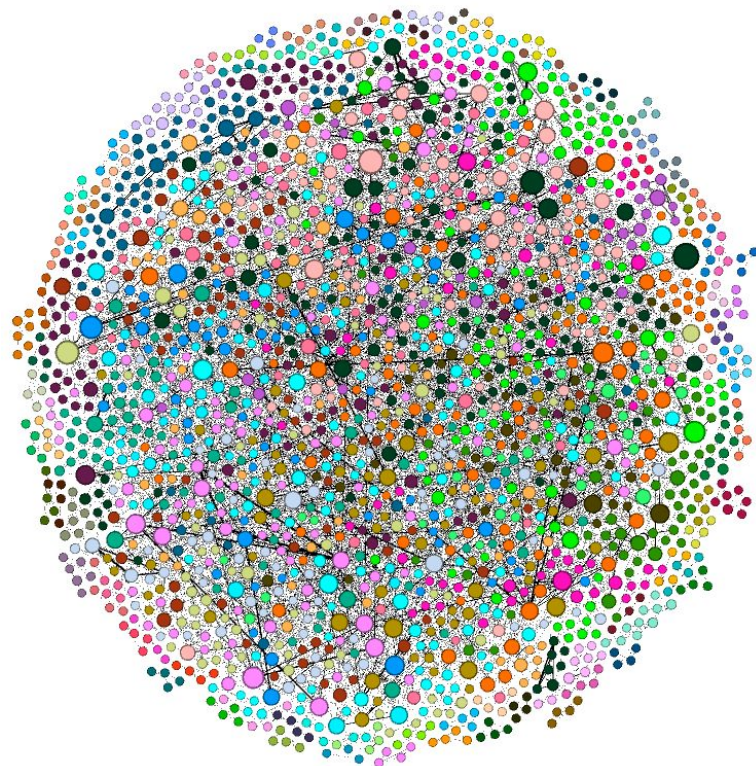
Jennifer Lawrence & Bradley Cooper

Amy Adams & Christian Bale

Brad Pitt & Jonah Hill

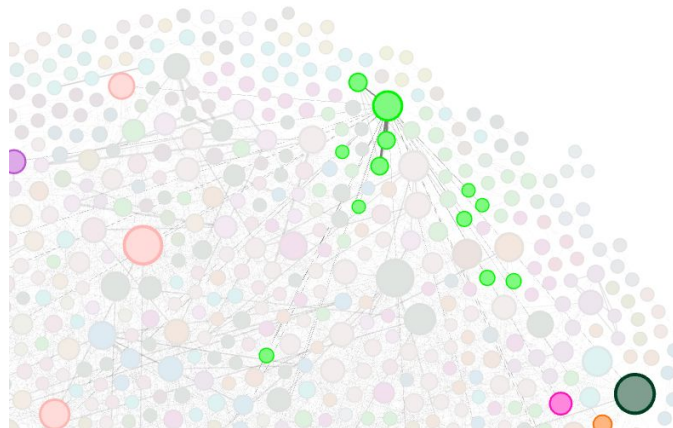
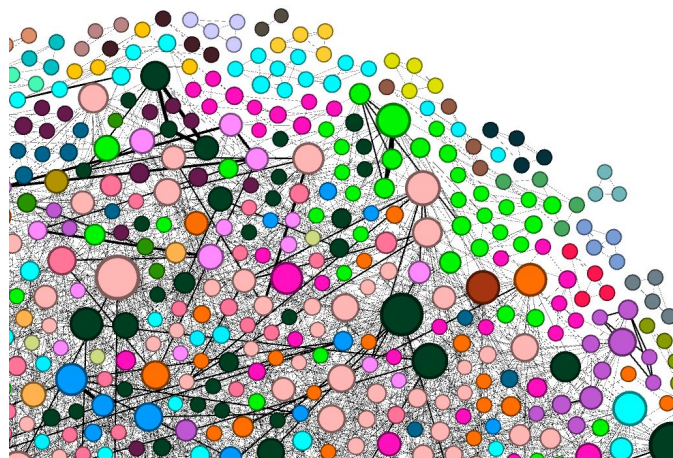
Q5: Which collaborations would we suggest?

- Based on communities
- Modularity (projection: 1.0)
- Layout
 - OpenOrd
 - Noverlap
- 172 communities
- 22 bigger clusters
- A lot of small cliques



Recommendations

- Recommend new collaborations between actors of same community
- Export nodes table
 - Node
 - Modularity
 - Degree
- Sort by degree (>10)
 - 285 actors
- Python code
- 1965 new pairs



Feature Collaborations



Chris Hemsworth

+



Charlize Theron



Brad Pitt

+



Gerard Butler



Jake Gyllenhaal

+



Matthew McConaughey

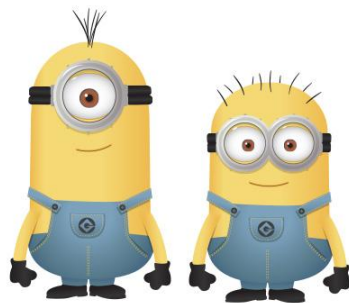


Ryan Reynolds

+



Jennifer Aniston



Sandy

Theo

