

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

E-social networks: Movies recommendations

Research Seminar, IJS, Ljubljana

Jelena Grujić

Institute of Physics
Belgrade, Serbia

05.02.2008., IJS, Ljubljana



Outline

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

- 1 Social networks**
- 2 IMDb**
- 3 Investigated properties**
- 4 Results**
- 5 Future work and conclusion**



Outline

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

1 Social networks

2 IMDb

3 Investigated properties

4 Results

5 Future work and conclusion

Social networks

Movie Recommendation

Jelena Grujić

Social networks

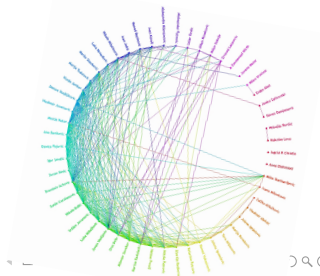
IMDb

Investigated properties

Results

Future work and conclusion

- Interactions between humans or groups
- Quantity of information
- Deeper scientific understanding
- E-social networks
- Recommendation networks
- Goal 1: Get to know social structure
- Goal 2: Use that knowledge to develop recommendation system





Outline

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

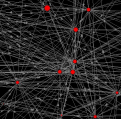
1 Social networks

2 IMDb

3 Investigated properties

4 Results

5 Future work and conclusion



The Internet Movie Database



Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

- On-line database about films, TV products, direct to video products and video games
- Launched 1990. from “Movie rating system” and “Those Eyes”
- Commercial 1995., Amazon.com 1998.
- Today: Titles 1,039,447
- Movies released theatrically: 379,871
- Users: votes, comments, message boards etc.
- Recommendations lists
- user approach

Collecting data

Movie Recommendation

Jelena Grujić


Social networks

IMDb

Investigated properties

Results

Future work and conclusion

- Name, ID, number of votes, rating, gender, IDs of recommended movies, IDs of users which left comment
- USA movies, more than 10 votes, excluded TV, straight to video and video games, around 43,000 titles
- Robot in  python™
- Power search, IDs of titles
- <http://imdb.com/title/tt0449010/>
- <http://imdb.com/title/tt0449010/recommendations>
- <http://imdb.com/title/tt0449010/usercomments?start=50>

Types of Networks

Movie Recommendation

Jelena Grujić

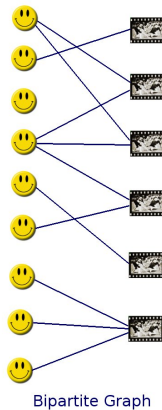
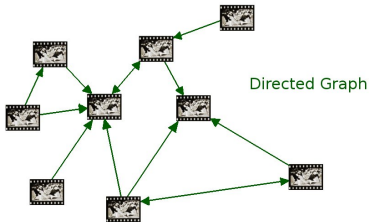
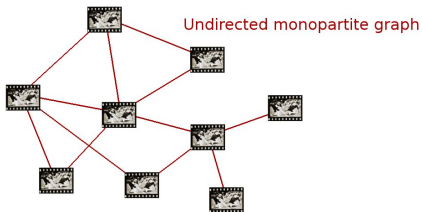
Social networks

IMDb

Investigated properties

Results

Future work and conclusion





Outline

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

1 Social networks

2 IMDb

3 Investigated properties

4 Results

5 Future work and conclusion

Investigated properties - Degree

Movie Recommendation

Jelena Grujić

Social networks

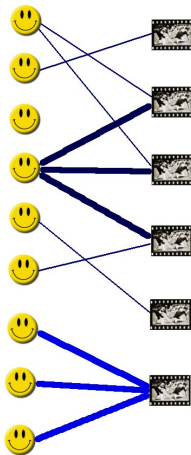
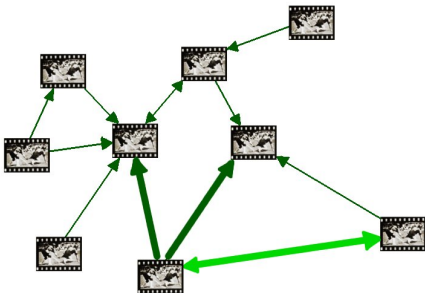
IMDb

Investigated properties

Results

Future work and conclusion

- Degree - number of neighbours
- Degree distribution
- Distinguish between types of networks
- Power law - self-organized mechanism



Investigated properties - Clustering

Movie Recommendation

Jelena Grujić

Social networks

IMDb

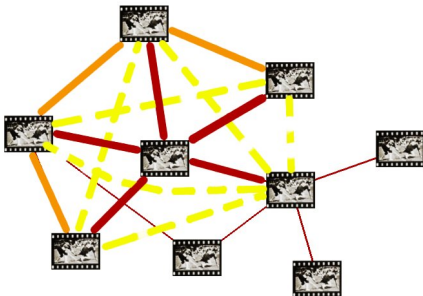
Investigated properties

Results

Future work and conclusion

- Clustering coefficient - how connected are the neighbours
- Social networks - high clustering

$$c_i = \frac{2e_i}{k_i(k_i - 1)} = \frac{2 \cdot 3}{5 \cdot 4} = 0.3$$



Investigated properties - Paths

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

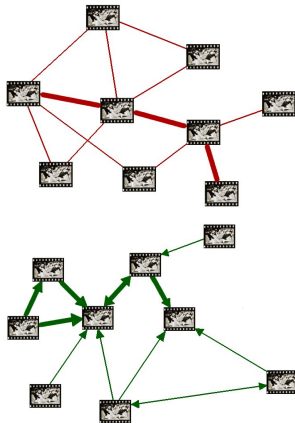
Future work and conclusion

- Communication within network
- Small world property
- Average Shortest paths

$$L = \frac{1}{N(N-1)} \sum_{i,j,i \neq j} d_{i,j}$$

- Topological Efficiency

$$E = \frac{1}{N(N-1)} \sum_{i,j,i \neq j} \frac{1}{d_{i,j}}$$



Six degrees of separation

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

- Pop culture phenomenon
- Frigyes Karinthy 1929. “Chains”
- Milgram 1967.
- Theater play 1990., movie 1993.
- Premier Magazine 1994. Oracle of Bacon
- Will&Grace, www.sixdegrees.org
- Duncan Watts 2001. e-mails
- TV seria Six Degrees, Lonely planet
- Six Degrees, ...
- Facebook experiment: Average: 6.02
Max: 14 People involved: 12,529,629





Investigated networks

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

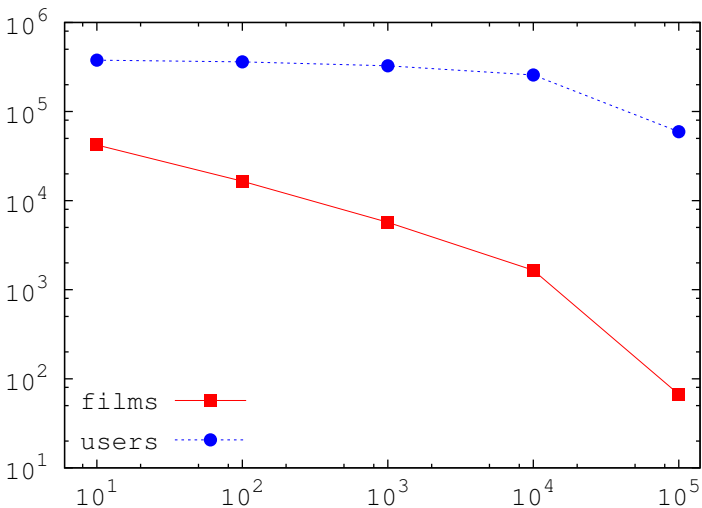
Results

Future work and conclusion

- IMDb recommendations
 - From recommendation lists on the site. User votes, genre, title, keywords, and, most importantly, user recommendations themselves
- User driven bipartite
 - Movie and user are linked if specific user left a comment on the web page of specific movie.
- User driven one-mode projection
 - Two movies will be connected if they have user which left comment for both movies, but in the recommendation list we put just ten movies with highest numbers of common users.
- Random vote-preferential
 - is generated by connecting each movie to ten other movies randomly from distribution proportional to the number of users who left comment for that movie.

Number of votes

■ Minimal number of votes - Network of different sizes





Outline

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

1 Social networks

2 IMDb

3 Investigated properties

4 Results

5 Future work and conclusion

Results

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

Network with minimal number of votes $mV = 10$

Network	N	L	$\langle k \rangle$	C	d	E_{ff}
IMDb	41117	322203	7.83	0.19	2.6	0.024
UD-OM	42083	313784	7.46	0.25	19.2	0.008
RD-VP	42083	420830	10	0.007	5.6	0.094

- N number of nodes
- L number of links
- $\langle k \rangle$ average number of links per node
- C clustering coefficient
- d average shortest path
- E_{ff} efficiency

User driven bipartite

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

- $\langle k \rangle_m = 26.8, \langle k \rangle_u = 2.99$
- Degree distribution - power low

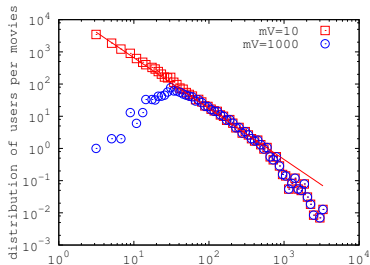
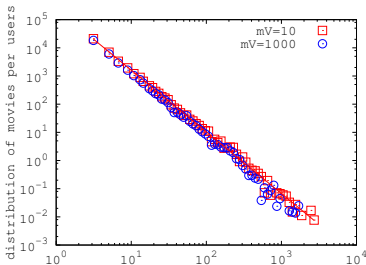


Figure: Degree distributions for bipartite networks for $mV = 10$ and $mV = 1000$. On the left number of movies per user ($\gamma = 2.16$), on the right number of users per movie ($\gamma = 1.58$).

IMDb recommendations

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

■ Godfather (1972)

The Godfather: Part II (1974) Scarface (1983) Goodfellas (1990) Beantown (2007) Once Upon a Time in America (1984) The Untouchables (1987) The Godfather: Part III (1990) Carlito's Way (1993) Léon (1994) L.A. Confidential (1997)

■ 10 things I hate about you (1999)

The Taming of the Shrew (1967) Clueless (1995) Sixteen Candles (1984) Valley Girl (1983) Slap Her... She's French (2002) Ferris Bueller's Day Off (1986) A Walk to Remember (2002) Heathers (1989) The Virgin Suicides (1999) Not Another Teen Movie (2001)

IMDb recommendations

Movie Recommendation

Jelena Grujić

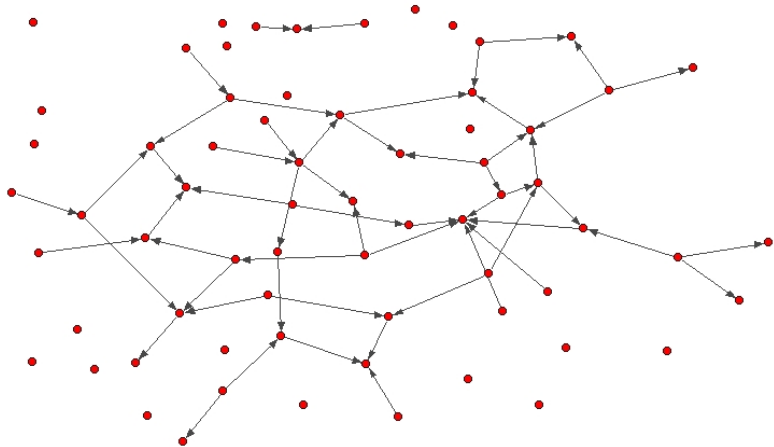
Social networks

IMDb

Investigated properties

Results

Future work and conclusion



IMDb recommendations

Movie Recommendation

Jelena Grujić

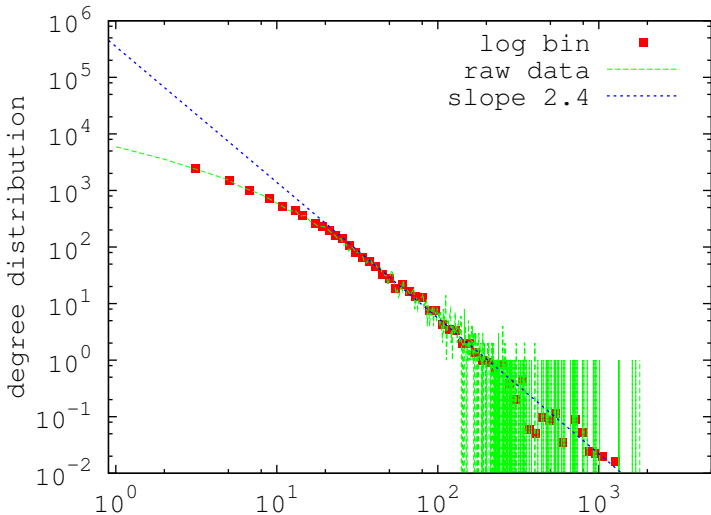
Social networks

IMDb

Investigated properties

Results

Future work and conclusion



IMDb recommendations

Movie Recommendation

Jelena Grujić

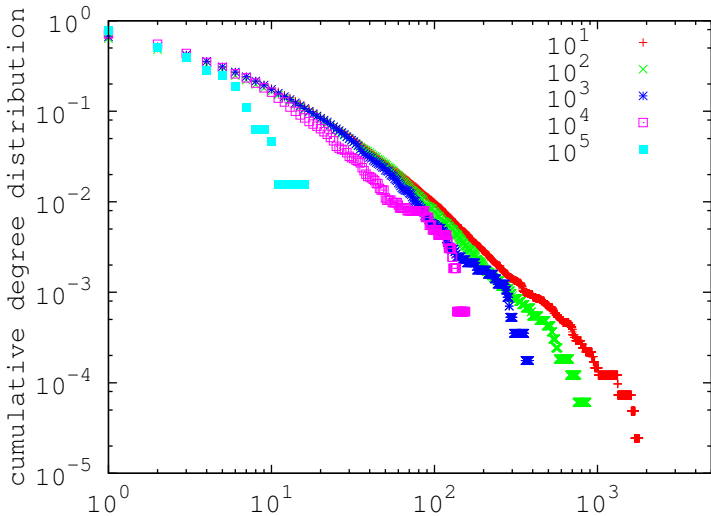
Social networks

IMDb

Investigated properties

Results

Future work and conclusion





User Driven Onemode projection

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

- Godfather (1972)

The Godfather: Part II (1974) The Lord of the Rings: The Fellowship of the Ring (2001) The Matrix (1999) Pulp Fiction (1994) Saving Private Ryan (1998) The Shawshank Redemption (1994) The Shining (1980) Star Wars (1977) Star Wars: Episode I - The Phantom Menace (1999) Titanic (1997)

- 10 Things I hate about you (1999)

American Pie (1999) The Blair Witch Project (1999) The Matrix (1999) Scream 3 (2000) She's All That (1999) The Sixth Sense (1999) Star Wars: Episode I - The Phantom Menace (1999) Star Wars: Episode II - Attack of the Clones (2002) Titanic (1997) X-Men (2000)

User Driven Onemode projection

Movie Recommendation

Jelena Grujić

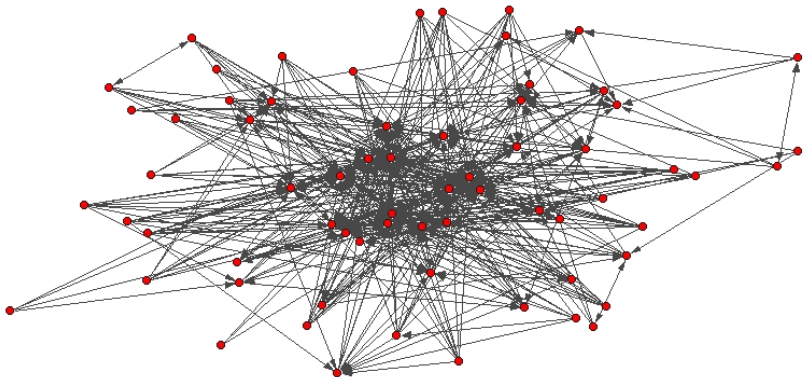
Social networks

IMDb

Investigated properties

Results

Future work and conclusion



User Driven Onemode projection

Movie Recommendation

Jelena Grujić

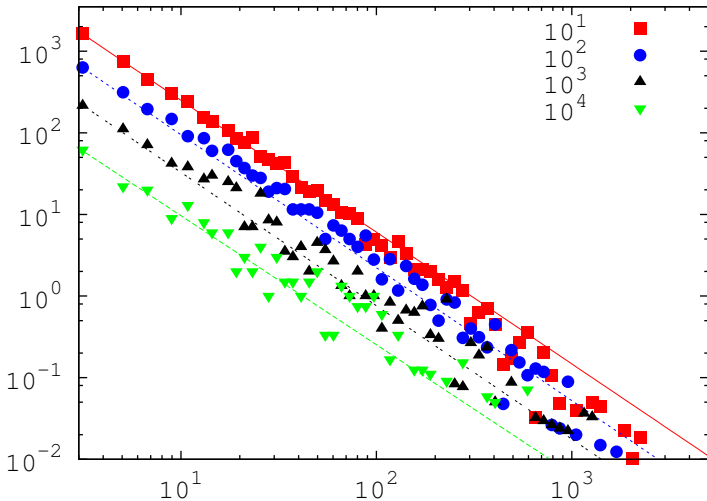
Social networks

IMDb

Investigated properties

Results

Future work and conclusion



Random vote-preferential

Movie Recommendation

Jelena Grujić

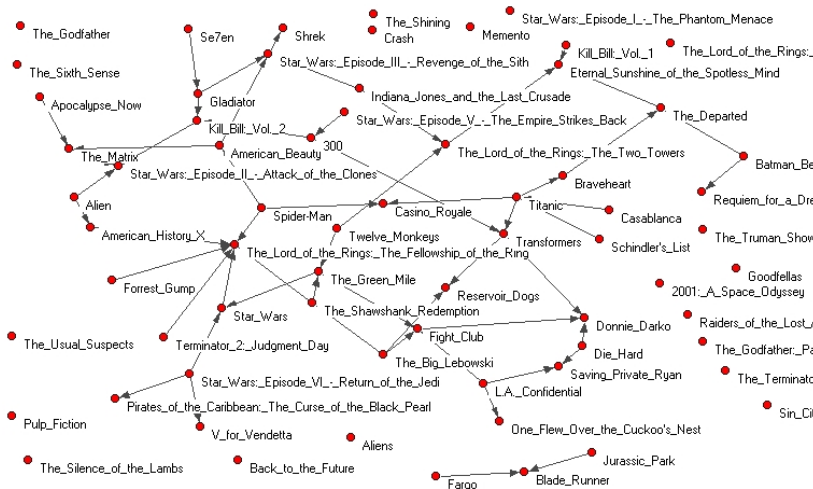
Social networks

IMDb

Investigated properties

Results

Future work and conclusion



Comparison - degree

Movie Recommendation

Jelena Grujić

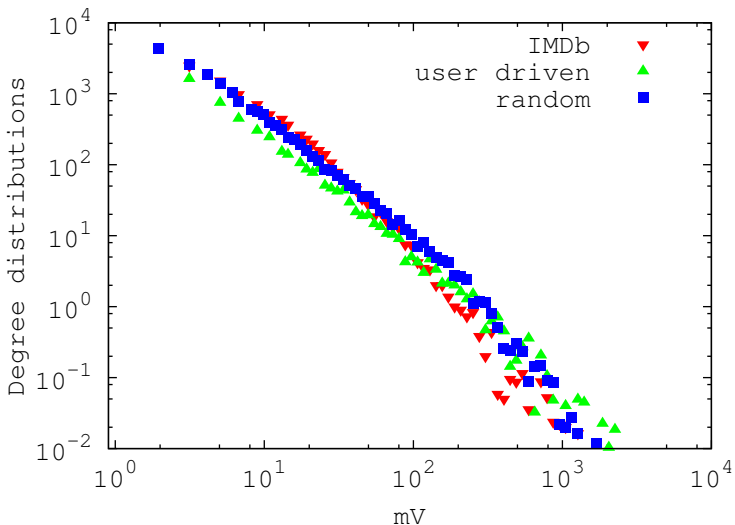
Social networks

IMDb

Investigated properties

Results

Future work and conclusion



Comparison - clustering

Movie Recommendation

Jelena Grujić

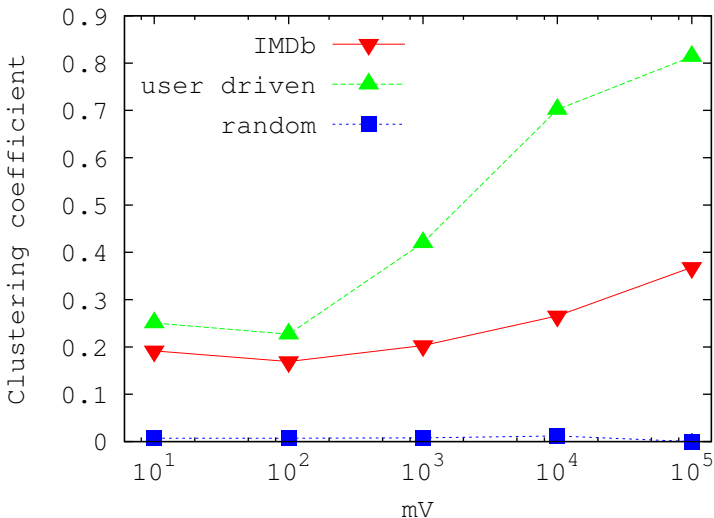
Social networks

IMDb

Investigated properties

Results

Future work and conclusion



Comparison - paths

Movie Recommendation

Jelena Grujić

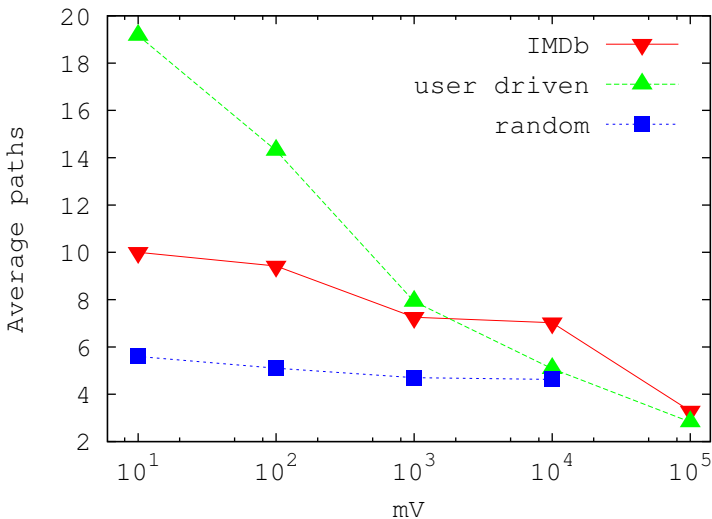
Social networks

IMDb

Investigated properties

Results

Future work and conclusion



Comparison - efficiency

Movie Recommendation

Jelena Grujić

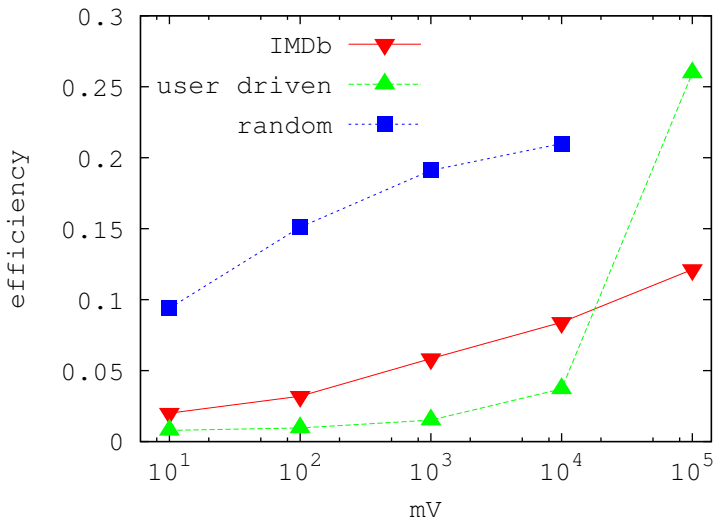
Social networks

IMDb

Investigated properties

Results

Future work and conclusion





Outline

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

1 Social networks

2 IMDb

3 Investigated properties

4 Results

5 Future work and conclusion

Future works

Movie Recommendation

Jelena Grujić

Social networks

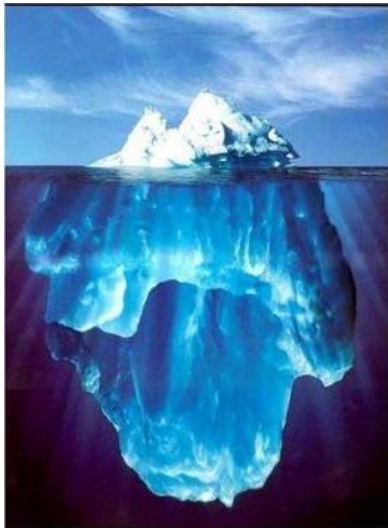
IMDb

Investigated properties

Results

Future work and conclusion

- Weighted one mode projection
- Users modularity
- Model for bipartite network





Conclusion

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion

- Two type of data considered: IMDb and collaborative filtering
- All network: small world properties and high clustering
- Small-world: sparse, short distances and high clustering coefficient. Good navigation properties
- IMDb: non-universal
- Collaborative filtering: universal power law
- Electronic fingertips

Acknowledgments

Movie Recommendation

Jelena Grujić

Social networks

IMDb

Investigated properties

Results

Future work and conclusion



- Author thanks Bosiljka Tadić for numerous useful comments and suggestions.

