

History of Movie Success through GSP

Team 10

Cionca Alex, Donzier Paul, de Goumoëns Frédéric, Fluhr Hugo

DATASET

TMDb 5000 movies

Title	Budget	Release Date	Revenue	ROI	WR	Actors	Production Companies	Keywords	Genres
Avatar	240 M	10.12.2009	2'800 M	10.7	7,2 / 10	Sam Worthington Zoe Saldana Sigourney Weaver ...	Ingenious Film Partners Twentieth Century Fox...	culture clash, future, space war, space colon, ...	Action, Adventure, Fantasy, SF
Pirates of the Caribbean: At World's End	300 M	19.05.2007	1'000 M	2.3	6,9 / 10	Johnny Depp Orlando Bloom Keira Knightley ...	Walt Disney Pictures Jerry Bruck. Films ...	ocean, drug abuse, exotic island, east india, ...	Adventure, Fantasy, Action
Spectre	245 M	26.10.2015	880 M	2.6	6,3 / 10	Daniel Craig Christoph Waltz Léa Seydoux ...	Columbia Pictures Danjaq ...	spy, based on novel, secret agent, sequel, ...	Action, Adventure, Crime

Metrics

$$ROI = \frac{Revenue - Budget}{Budget}$$

$$WR = \frac{R \cdot n_{vote} + C \cdot n_{min}}{n_{vote} + n_{min}}$$

Features

Actors

Genres

Production Companies

Keywords

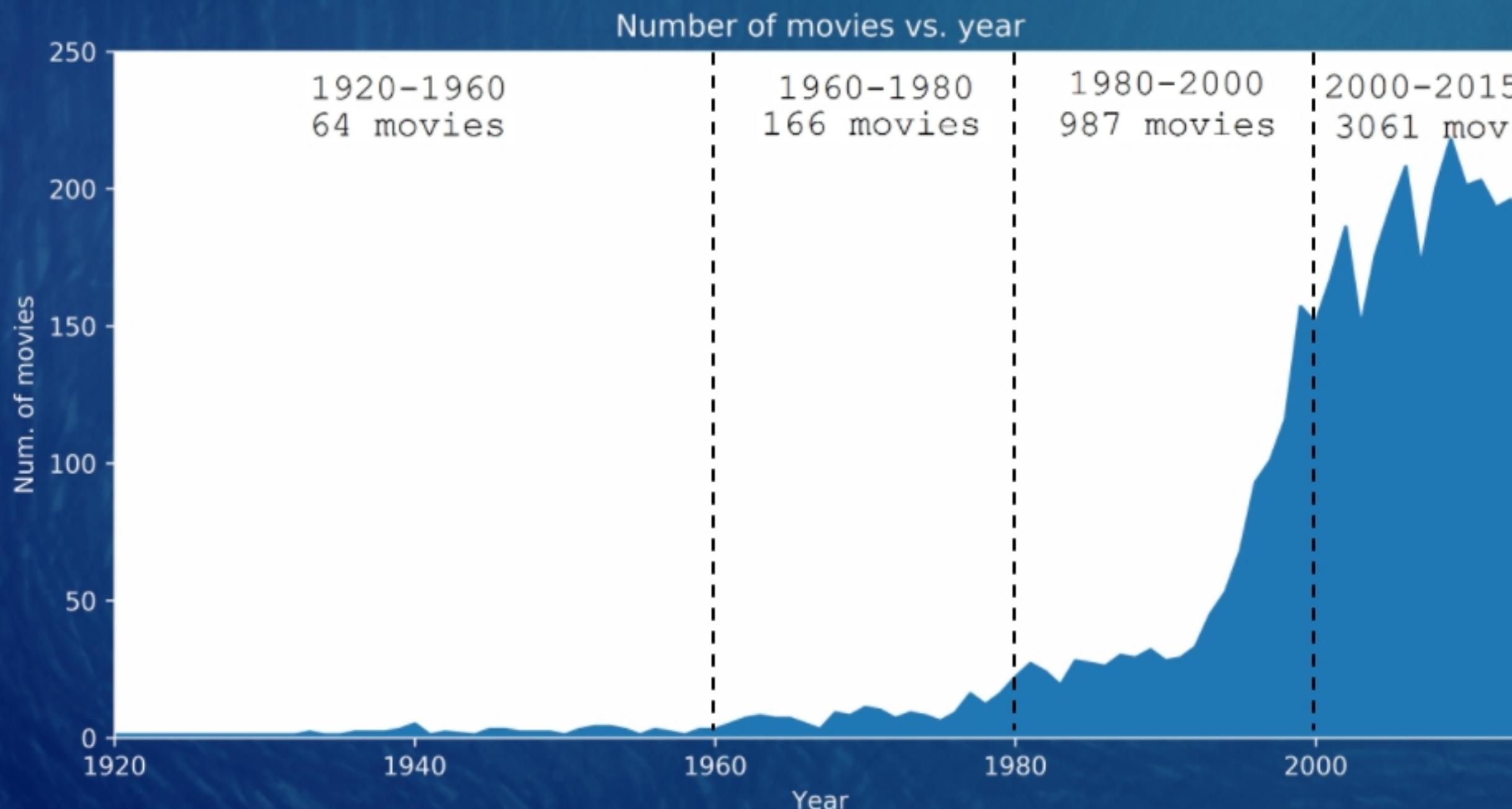
Time Periods

- | | |
|--------------------|------------------------------------|
| 1920 – 1960 | Classical Hollywood cinema |
| 1960 – 1980 | New Hollywood cinema |
| 1980 – 2000 | Boom of blockbusters/action movies |
| 2000 – 2015 | Modern (digital) cinema |

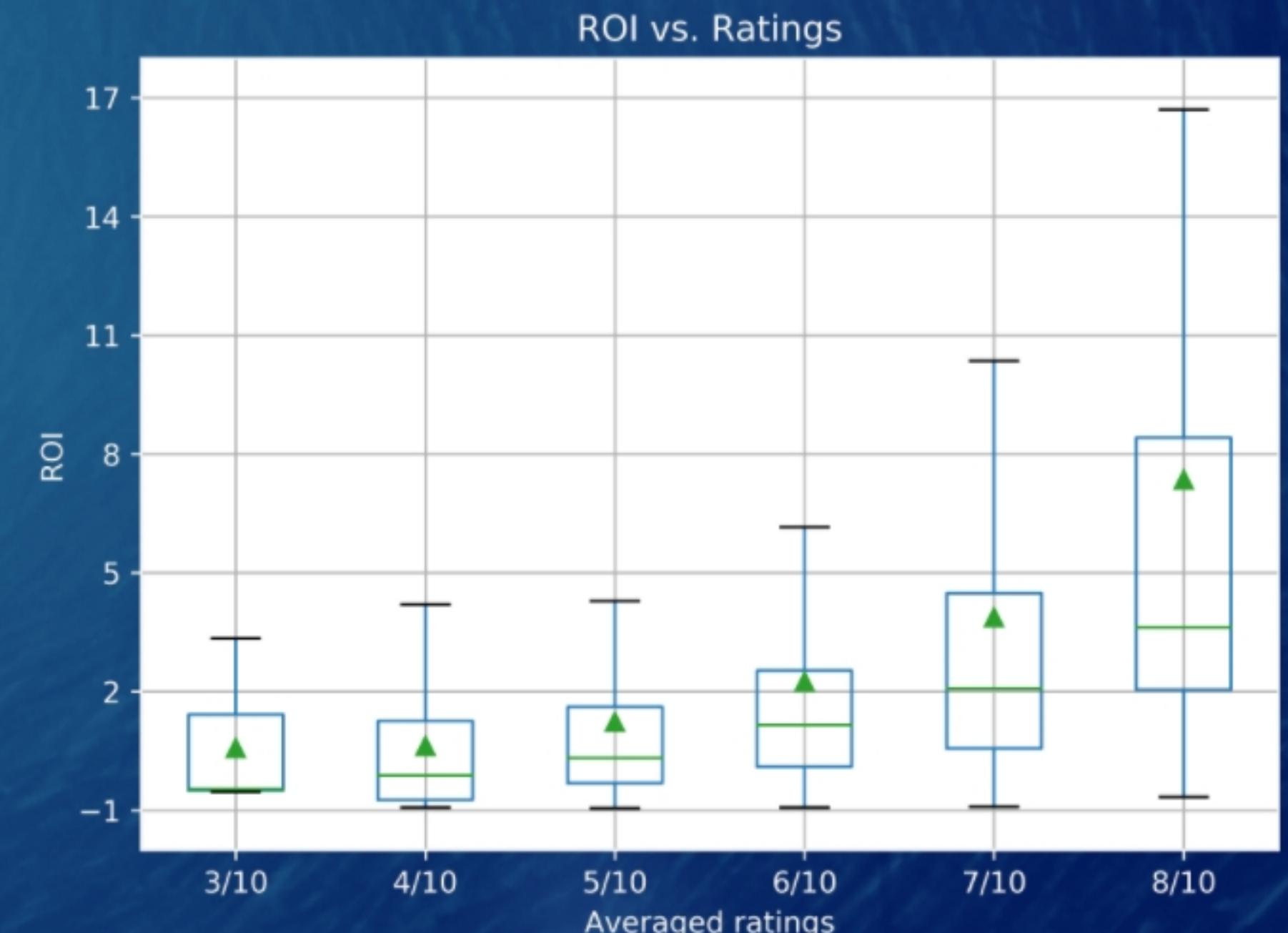
Metrics

- ▶ Different number of movies per period
- ▶ Well rated movies are more likely to give a high ROI
- ▶ Quick overview of numerical attributes links using the correlation matrix

	Min	Max
WR	2.42 (Baby Geniuses 2)	8.5 (Shawshank Redemption)
ROI	-0.95 (Shadow Conspiracy)	99.04 (Gone with the Wind)

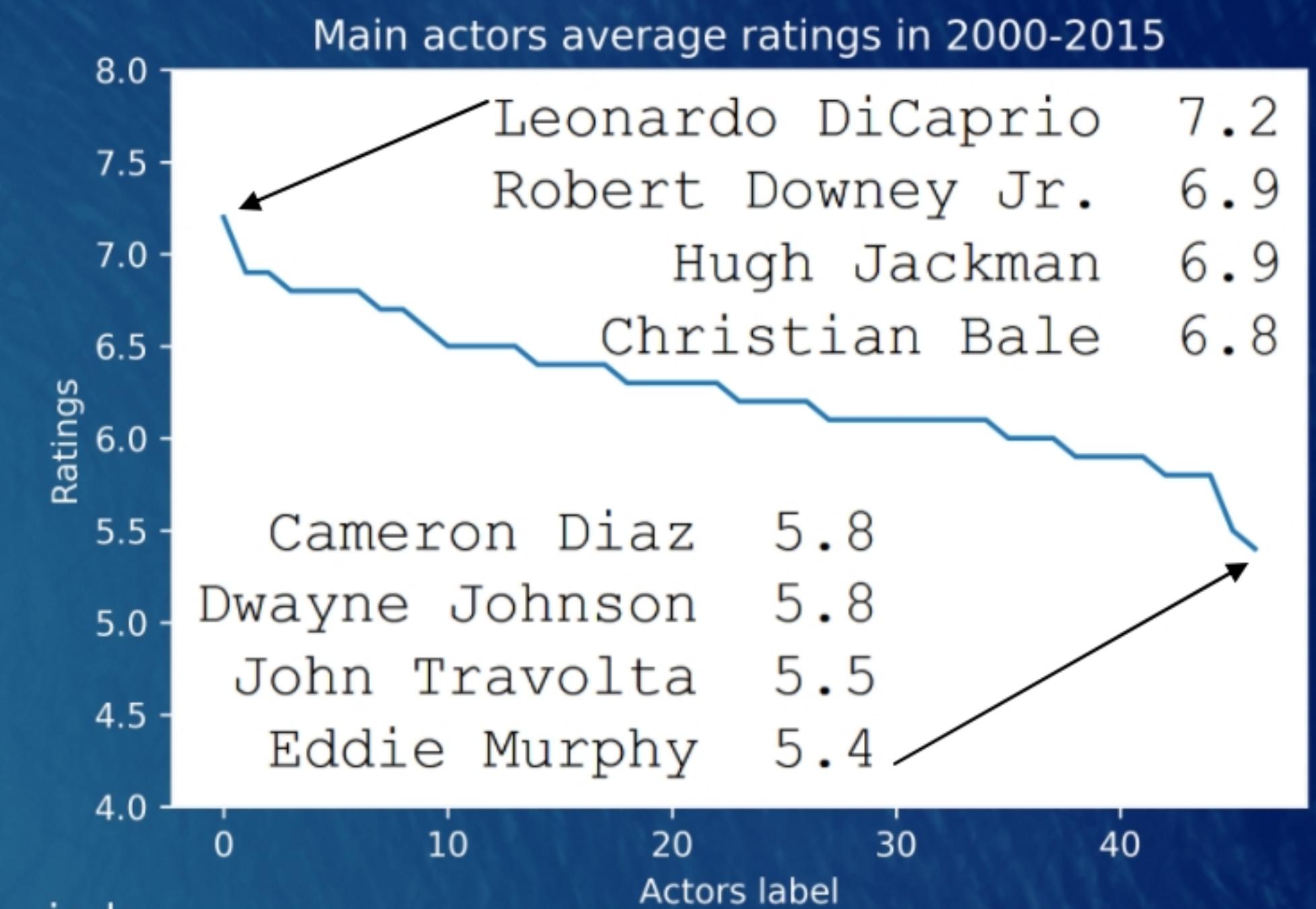


	Correlation Matrix for movies in 2000-2015					
Budget	1.00	0.73	0.35	0.57	0.08	-0.09
Revenue	0.73	1.00	0.31	0.77	0.26	0.28
Runtime	0.35	0.31	1.00	0.37	0.43	-0.03
Vote Count	0.57	0.77	0.37	1.00	0.45	0.25
Weighted Rating	0.08	0.26	0.43	0.45	1.00	0.18
ROI	-0.09	0.28	-0.03	0.25	0.18	1.00

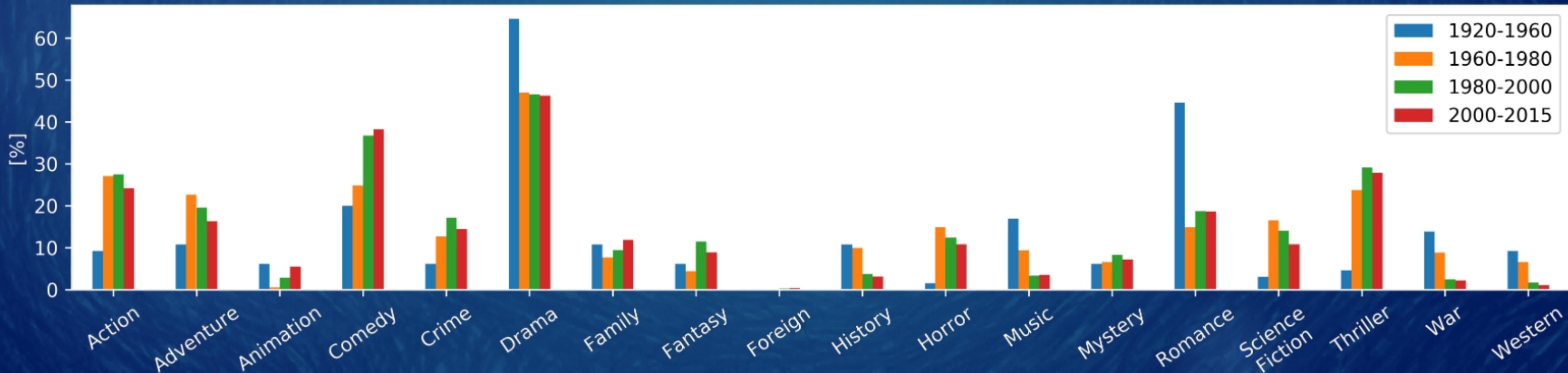


Features

- ▶ For every feature items, we looked for :
 - ▶ Occurrence
 - ▶ Link with the Weighted Rating
 - ▶ Link with the ROI
 - ▶ All this for every periods
- ▶ Slope gives a first idea of the link between features and metrics



Genres vs. cinema periods



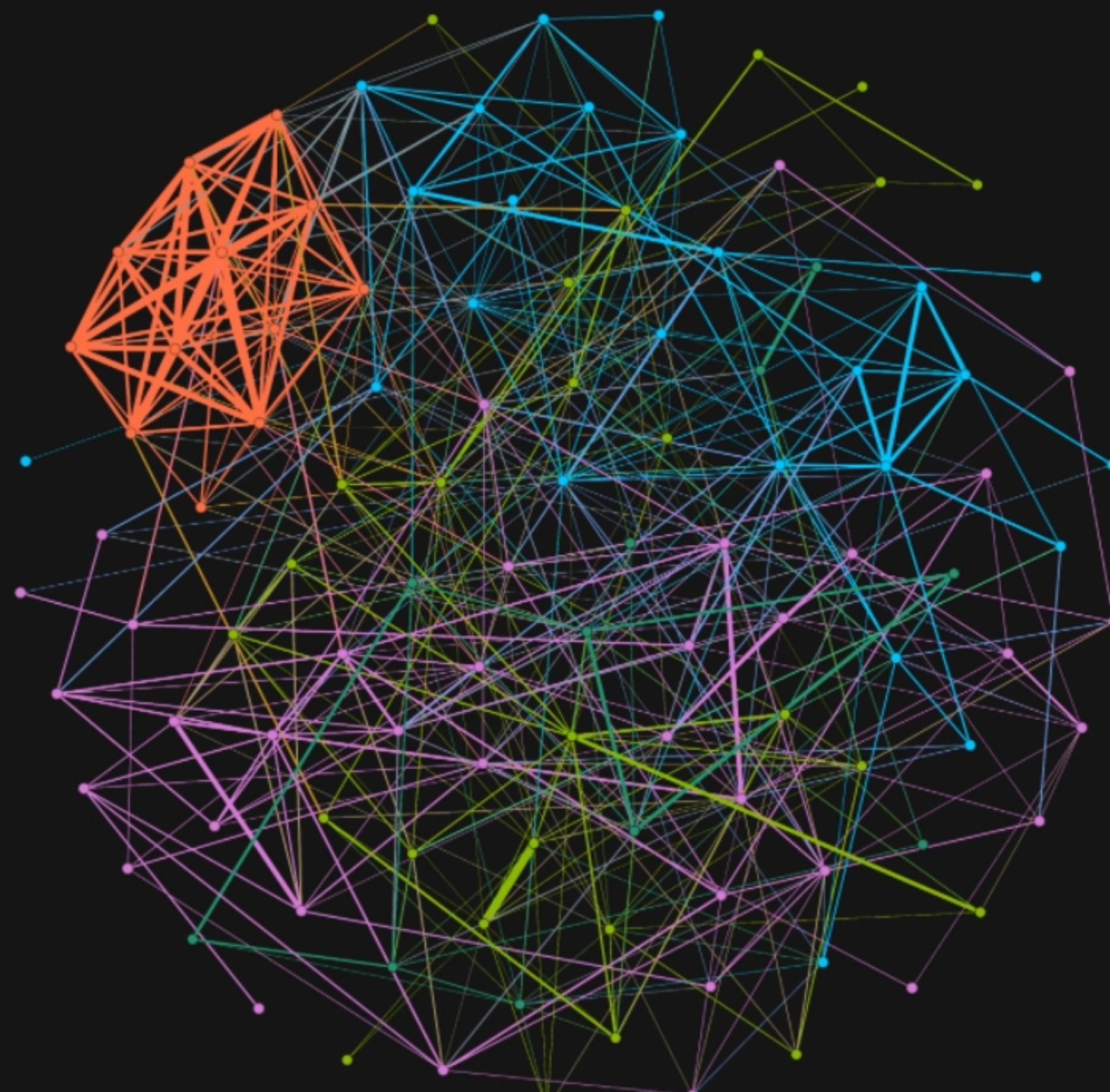
What now?

What can we learn from our data?



Building our Graphs

- ▶ Nodes = Movies
- ▶ Links = Proportion of shared items
- ▶
$$W_{i,j} = \frac{|(feature_i \cap feature_j)|}{\max_{k \in \{i,j\}} |feature_k|}$$



Keywords, 1960 - 1980

Movies 1920 - 1960

Keywords

Actors

Production Companies

Genres

1960 - 1980

Keywords

Actors

Production Companies

Genres

1980 - 2000

Keywords

Actors

Production Companies

Genres

2000 - 2015

Keywords

Actors

Production Companies

Genres

Methods

Laplacian quadratic form

Principle: Use the unnormalized Laplacian of the graphs to compute the average slope of the signal along an edge.

$$f^T L f = \sum_{(i,j) \in \mathcal{E}} W_{i,j} [f(j) - f(i)]^2$$

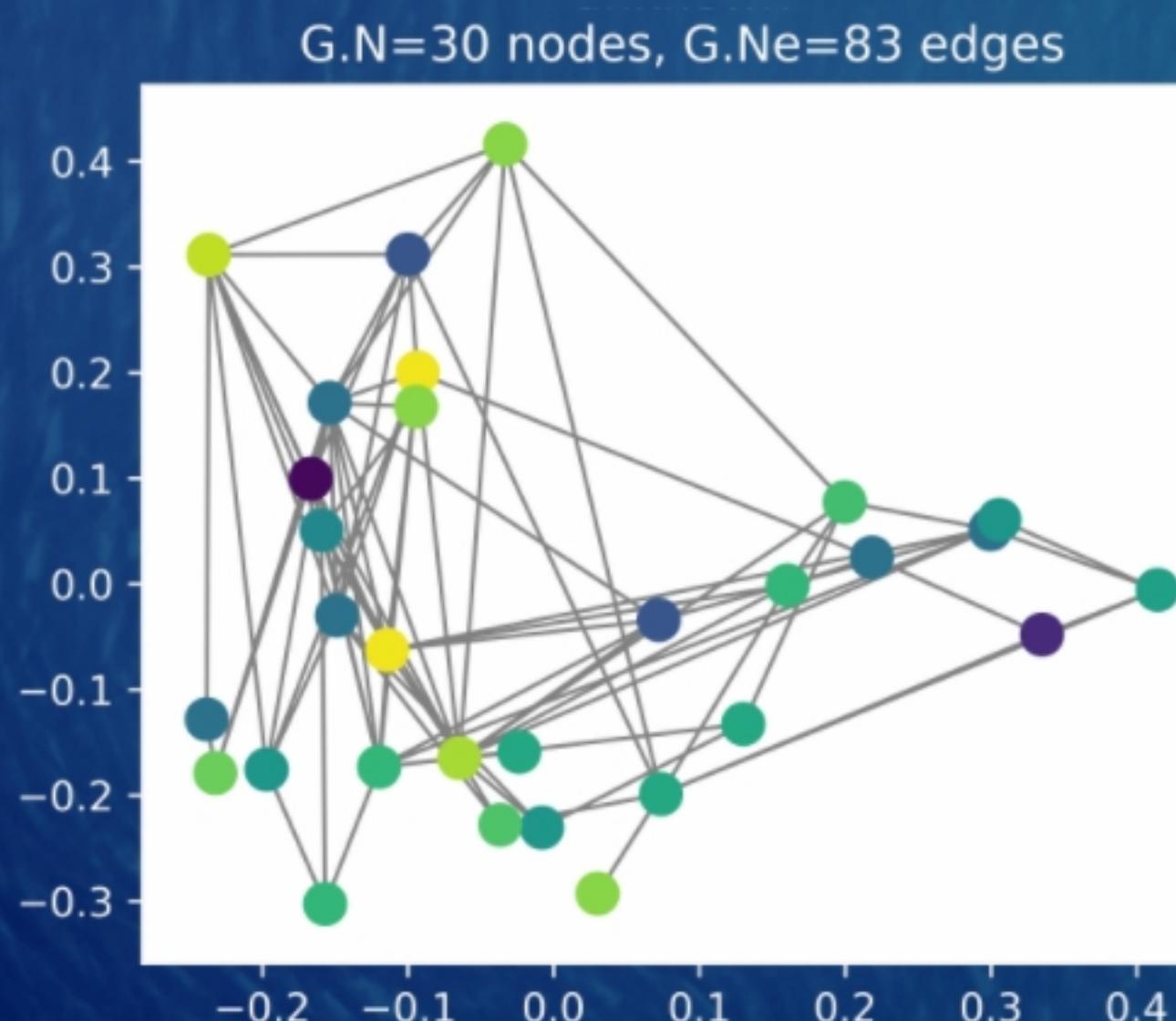
with $\begin{cases} f \text{ our signal (ROI or weighted rating)} \\ L \text{ the unnormalized Laplacian} \\ W_{i,j} \text{ the weight linking nodes } i \text{ and } j \end{cases}$

Challenge: Large graphs will have large sums



Normalize by

$$\sum_{(i,j) \in \mathcal{E}} W_{i,j}$$



Example:

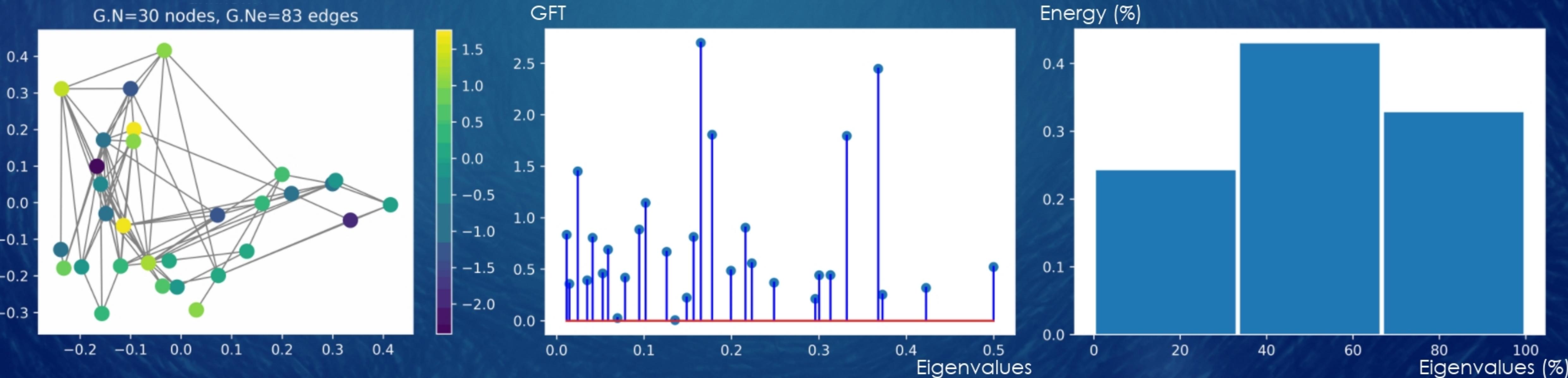
- Period: 1920-1960
- Feature: Cast
- Signal: Weighted rating

Result: 6,2386

Graph Fourier Transform

Principle: GFT of the signal on the graphs and investigate the energy distribution over the spectrum.

- ▶ Remove the continuous component
- ▶ Compute the percentage of energy located in each of the three thirds of the eigenvalues (frequencies).

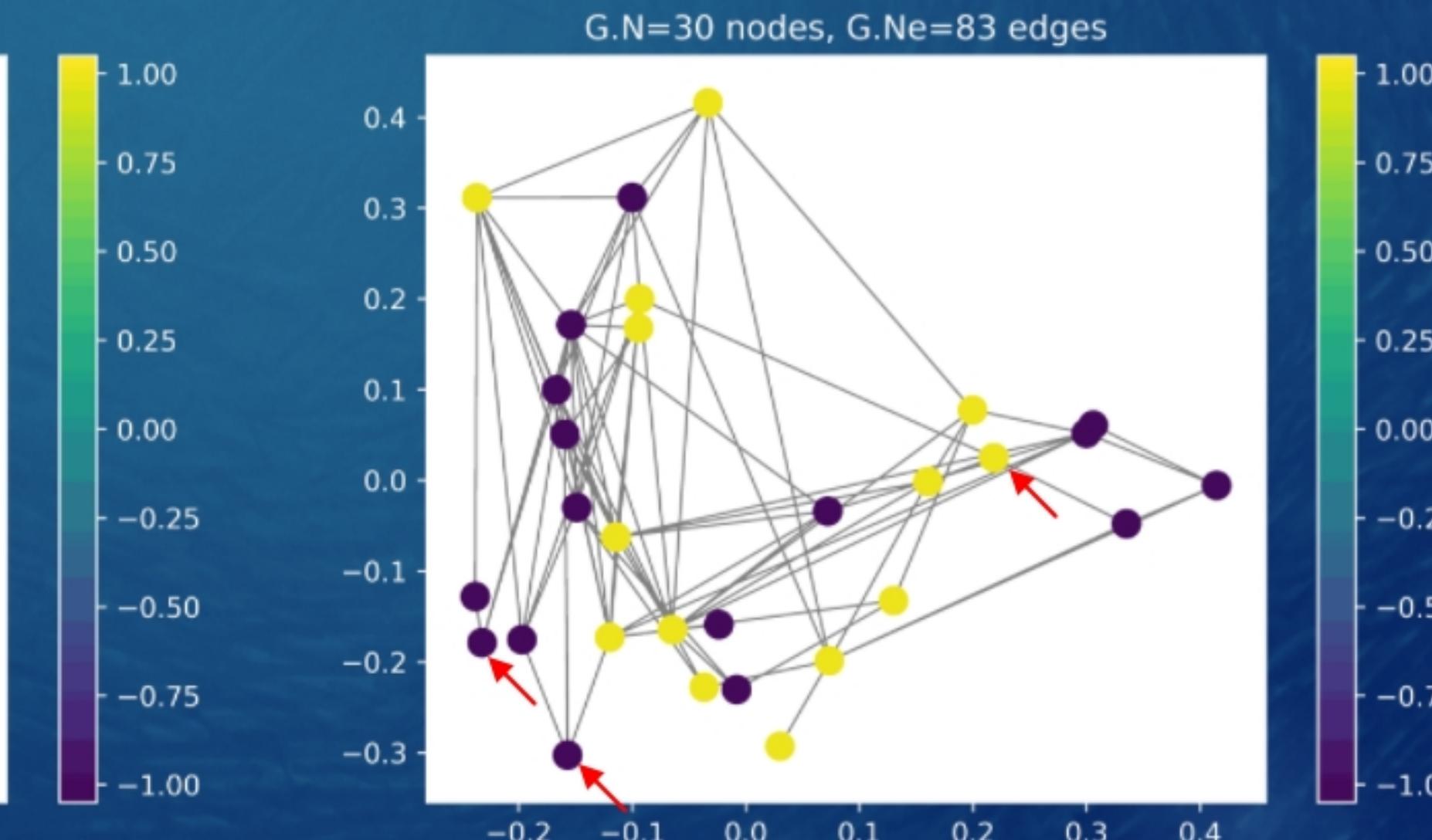
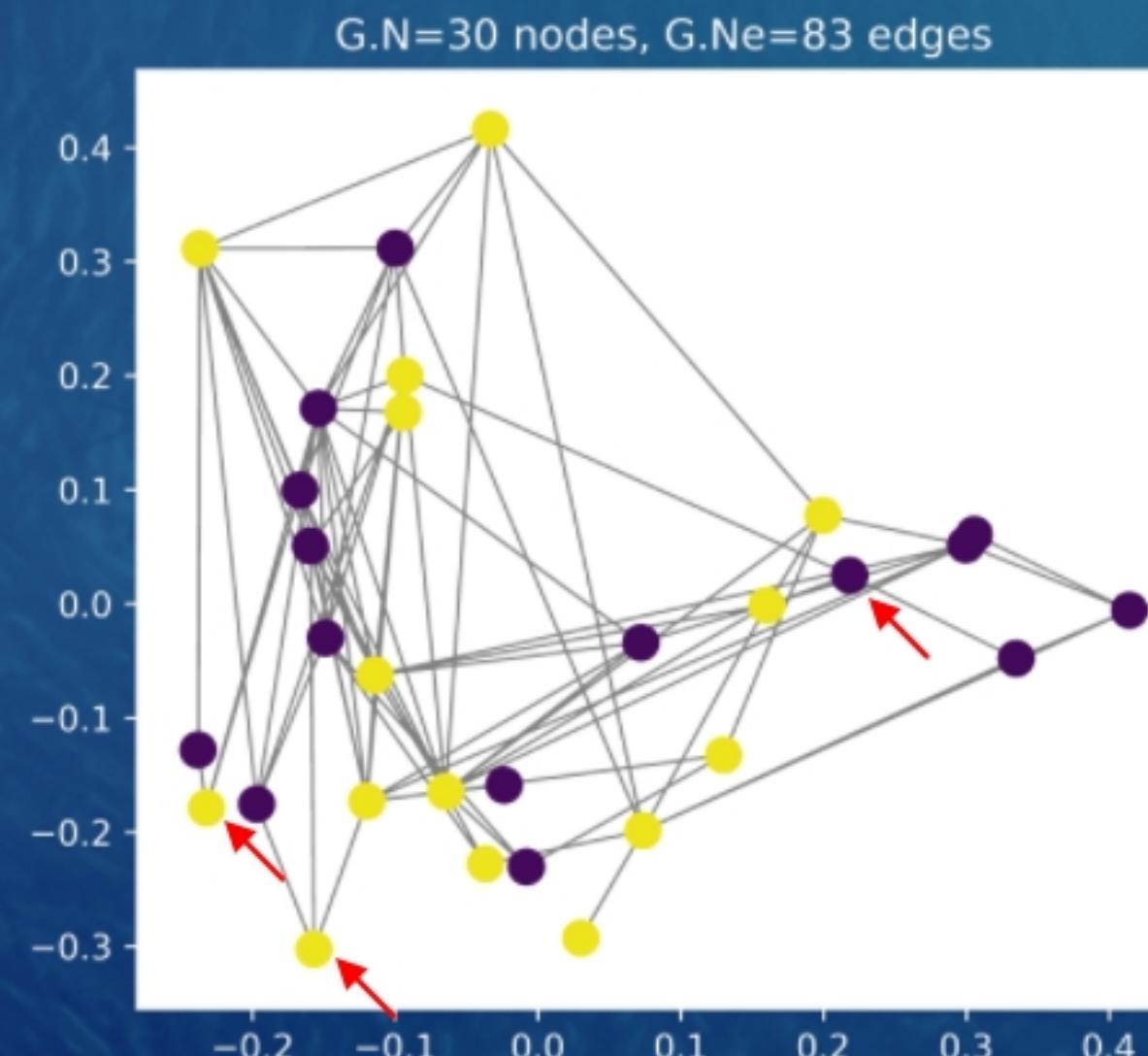


Transductive Learning

Principle: Reconstruct incomplete signal using Gradient Minimization.
The reconstruction error is expected to be lower for smoother signals.

- ▶ Binarization done using the signal's mediane.
- ▶ 80% of signal used for reconstruction.

Reconstruction error :
$$\frac{\|signal - reconstruction\|_2}{\|signal\|_2}$$

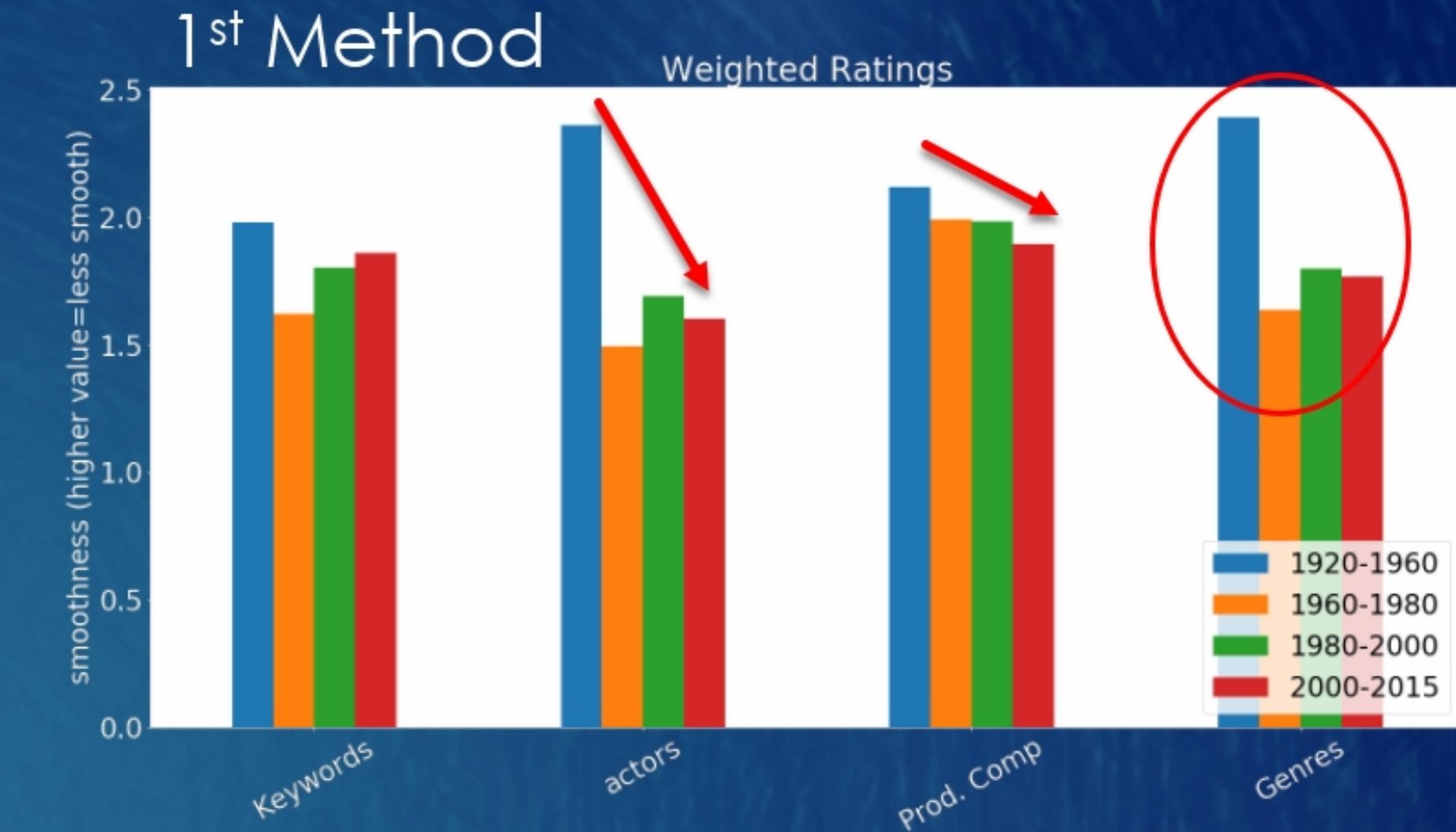


Results

Results:

Weighted movie rating

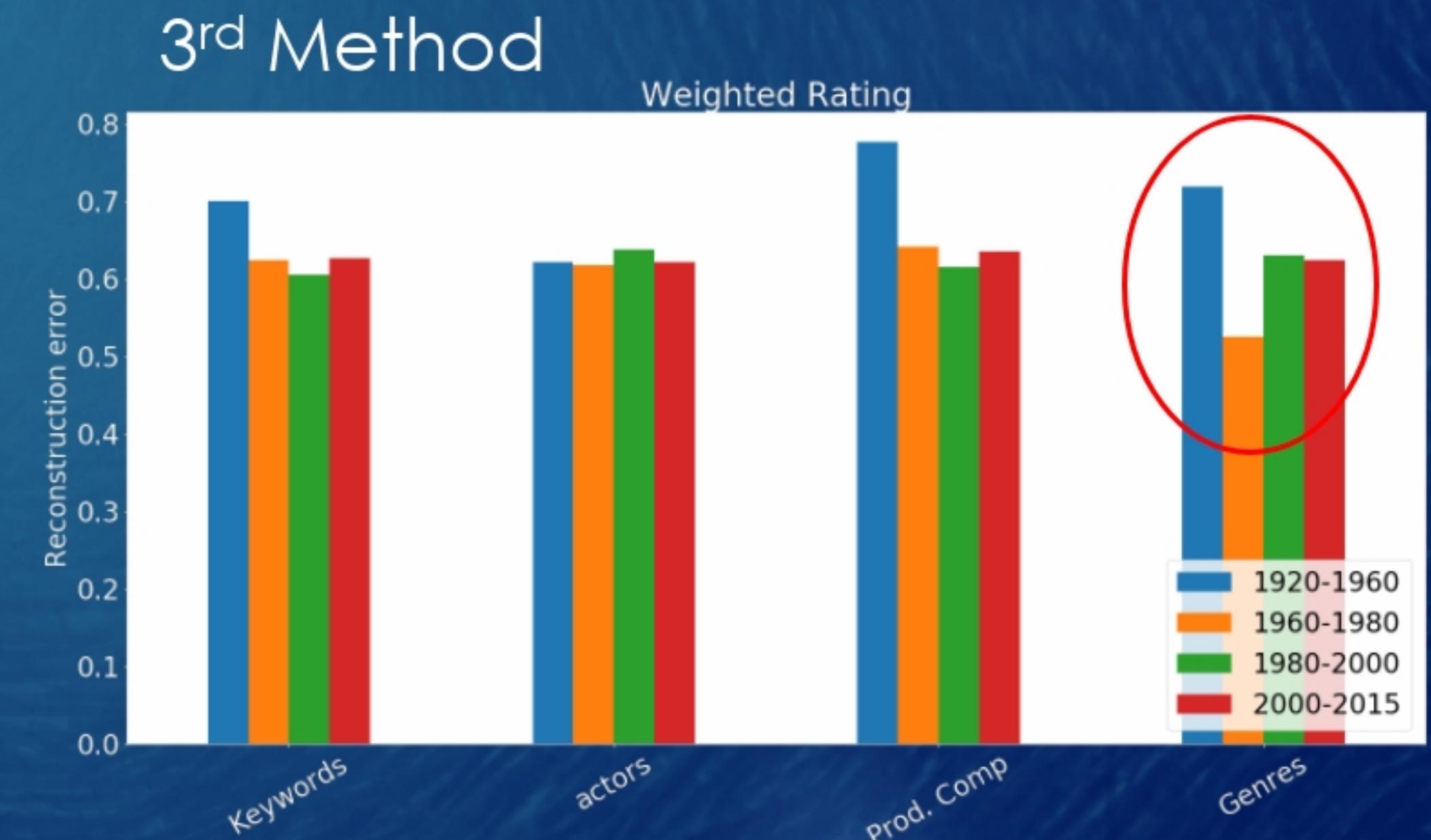
- ▶ 1st method:
 - ▶ Actors & Prod.Comp: Increase
 - ▶ Genres: 1960-1980
- ▶ 2nd method:
 - ▶ Actors & Prod.Comp: Increase
 - ▶ Genres: 1960-1980
- ▶ 3rd method:



Results:

Weighted movie rating

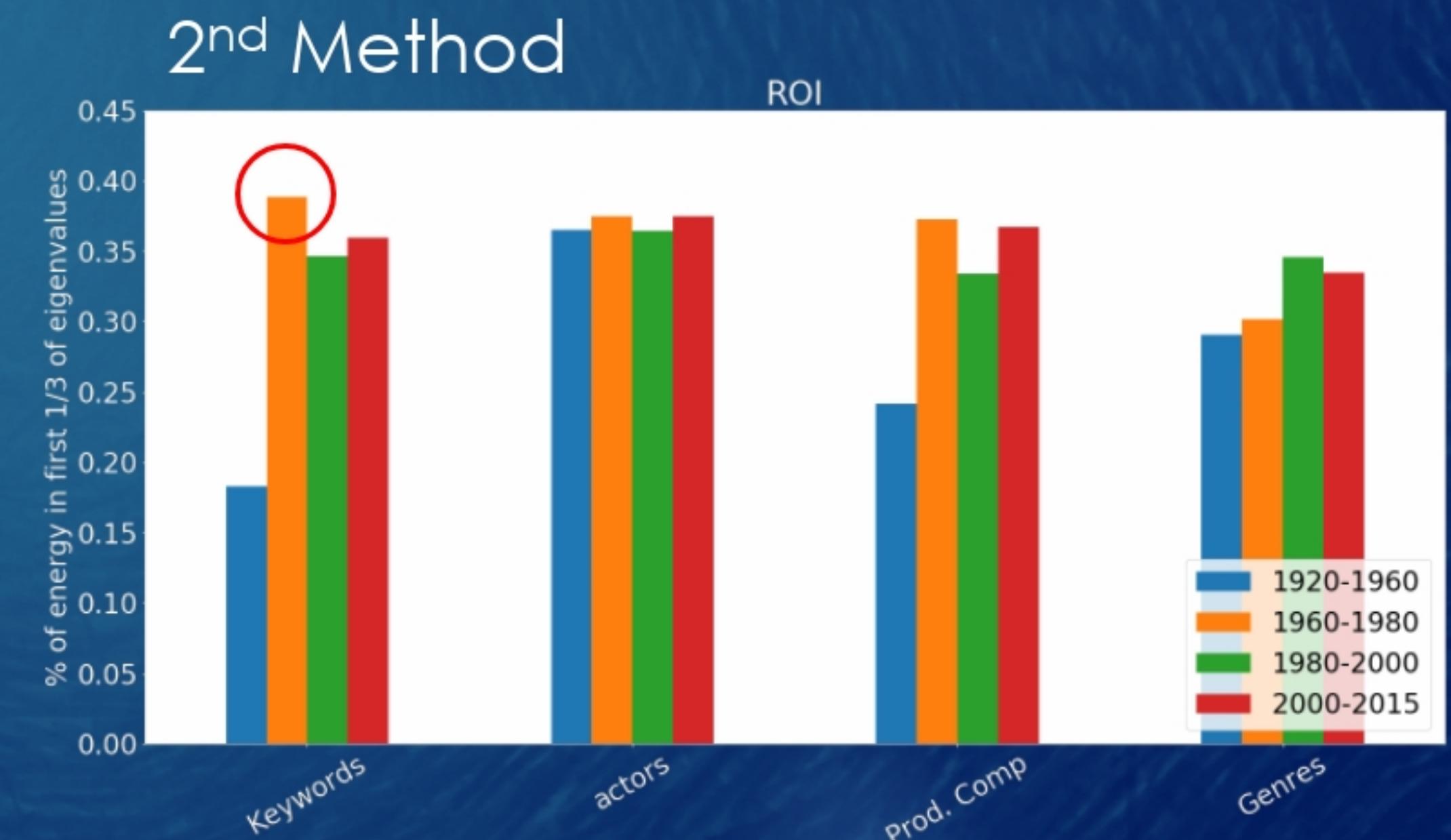
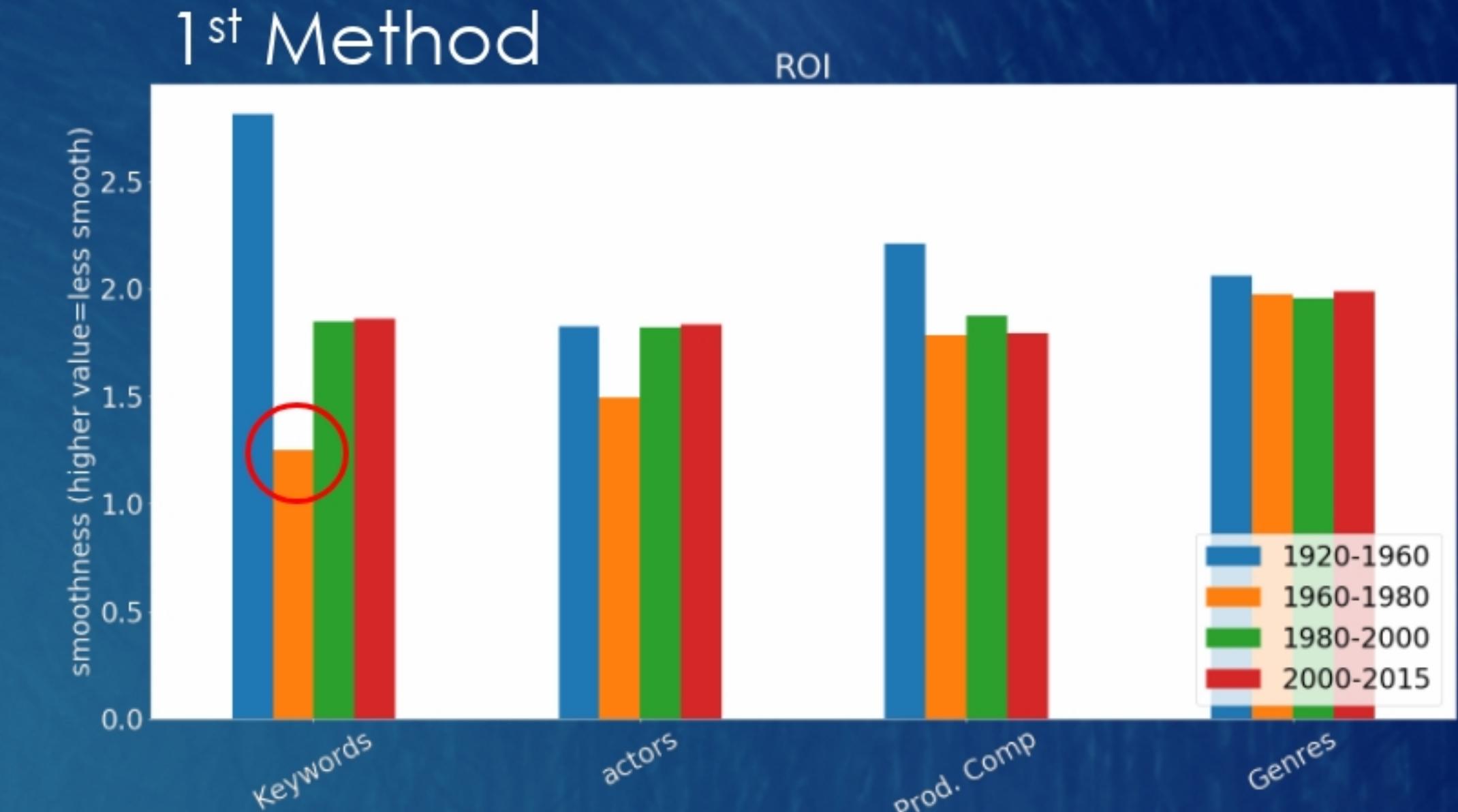
- ▶ 1st method:
 - ▶ Actors & Prod.Comp: Increase
 - ▶ Genres: 1960-1980
- ▶ 2nd method:
 - ▶ Actors & Prod.Comp: Increase
 - ▶ Genres: 1960-1980
- ▶ 3rd method:
 - ▶ Genres: 1960-1980



Results:

Return on Investment

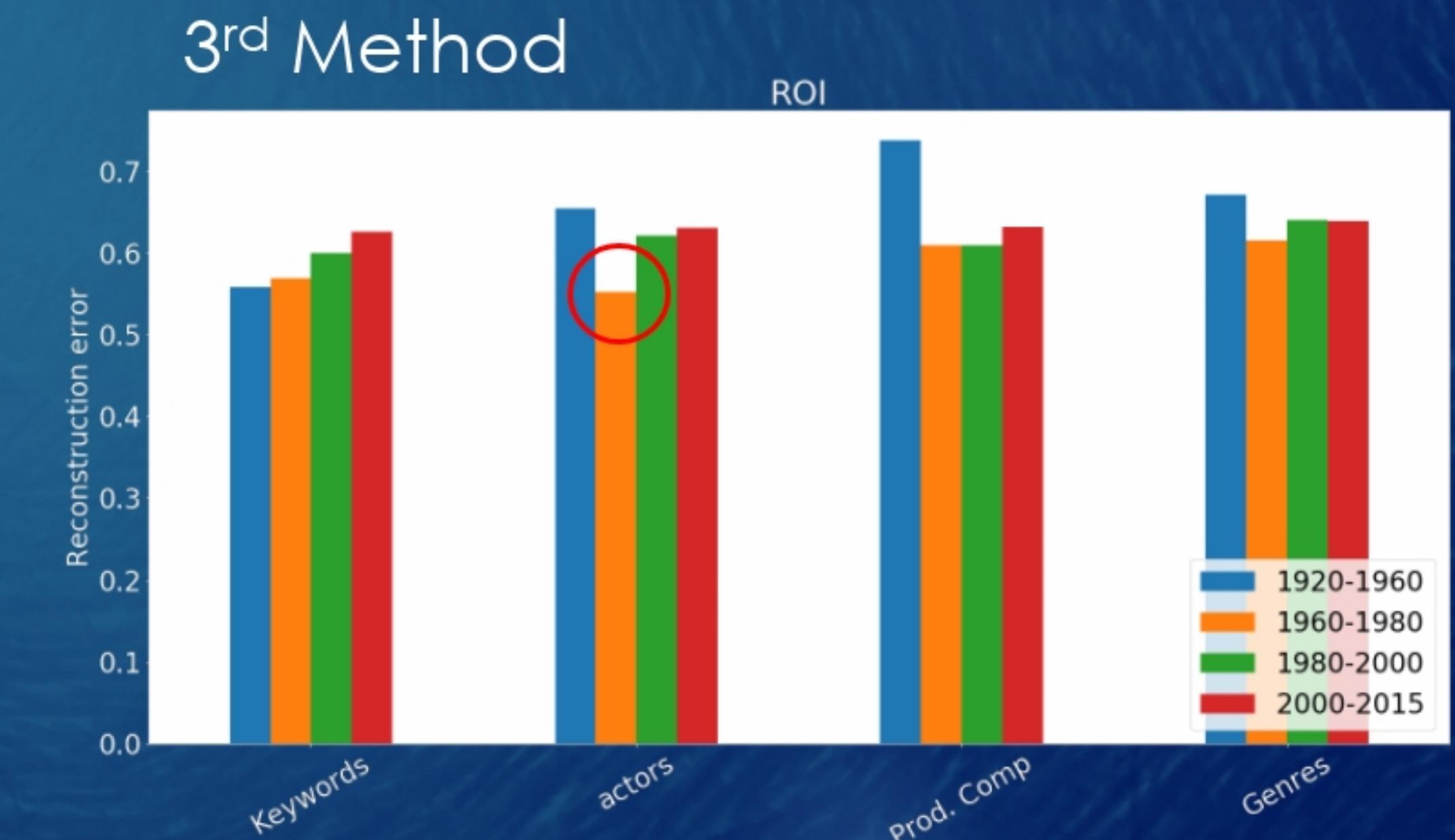
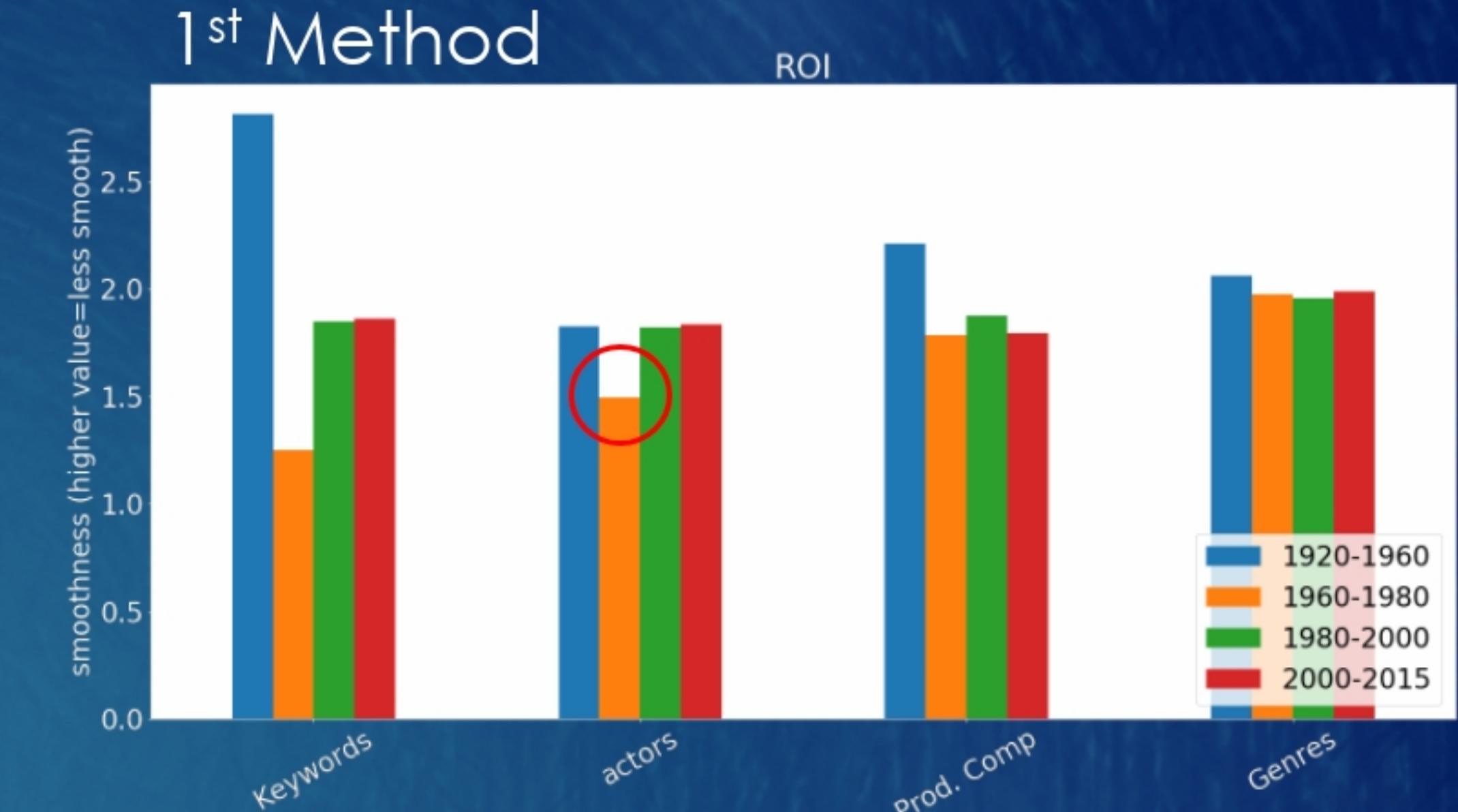
- ▶ 1st method:
 - ▶ Keywords & actors: 1960-1980
- ▶ 2nd method:
 - ▶ Keyword: 1960-1980
- ▶ 3rd method:



Results:

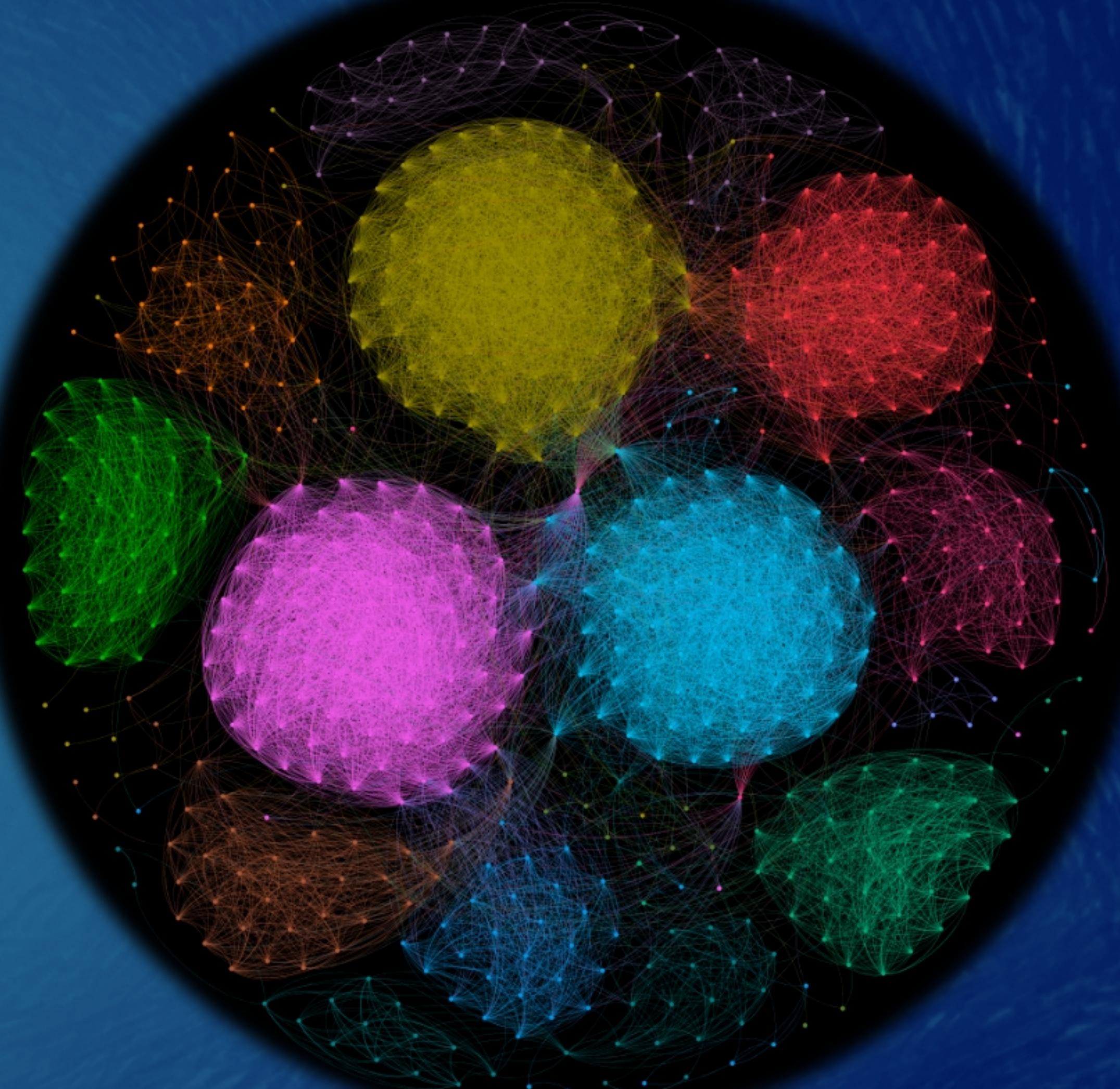
Return on Investment

- ▶ 1st method:
 - ▶ Keywords & actors: 1960-1980
- ▶ 2nd method:
 - ▶ Keyword: 1960-1980
- ▶ 3rd method:
 - ▶ Actors: 1960-1980



CONCLUSION

- ▶ Challenges:
 - ▶ Data filtering
 - ▶ Signal normalization
 - ▶ Results comparison
- ▶ What next?
 - ▶ Combine graphs
 - ▶ Optimize smoothness
 - ▶ Predict success evolution through time



1980 – 2000, Production companies