Will this Movie be Successful?
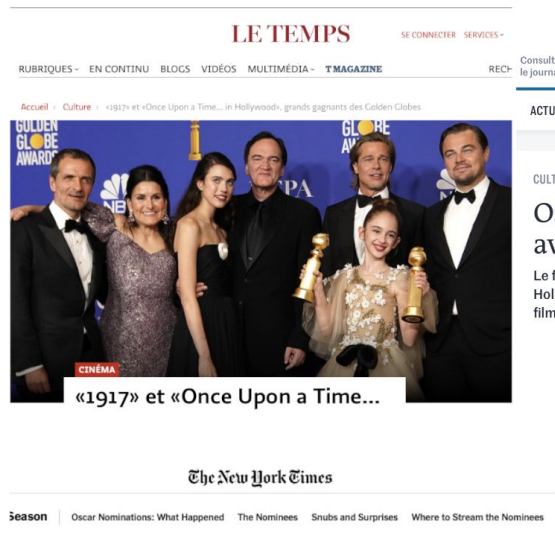
A network based prediction of Award nomination

Bizeul Alice, Carparelli Gaia, Spahr Antoine, Vinzant Hugues

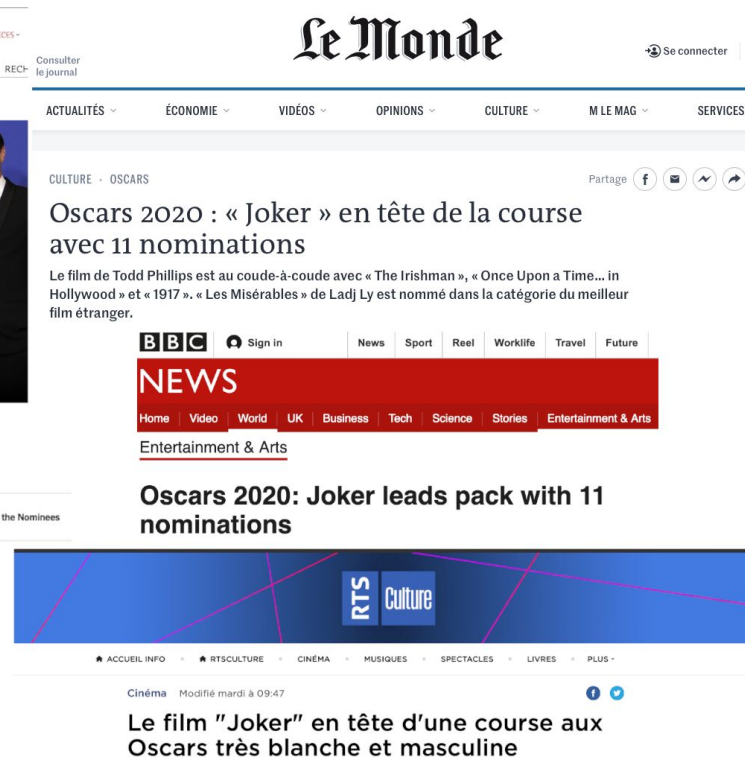EE-558 NTDS - Group 5

EPFL

Introduction

- Oscars and Golden Globes are renown awards

- In the US :
  - 29.6M watched the Oscars ceremony
  - 18.3M watched the Golden Globes ceremony
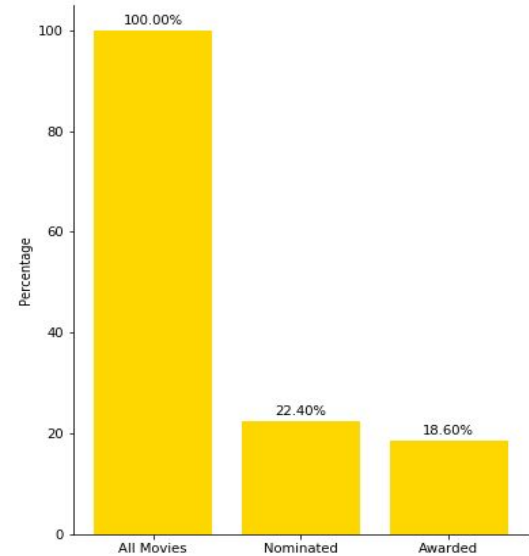
- What if we could predict them ?

Data

- TMDB 5000 movie Dataset from Kaggle

- Web Scrapping of Oscars and Golden Globes nominations and awards

- 22.4% of total movies at least **nominated** once

- 18.6% of total movies at least **awarded** once

Graph construction

- WHY?
  - To visualize and exploit high dimensional features

- HOW?
  - Cosine similarity matrix
  - Features used to connect each movie:
    - cast, crew, keywords and genres

  - Pruning at 0.25

- WHAT?
  - Movie characteristics:
    - revenue, budget, runtime, vote score, vote average, popularity and genres
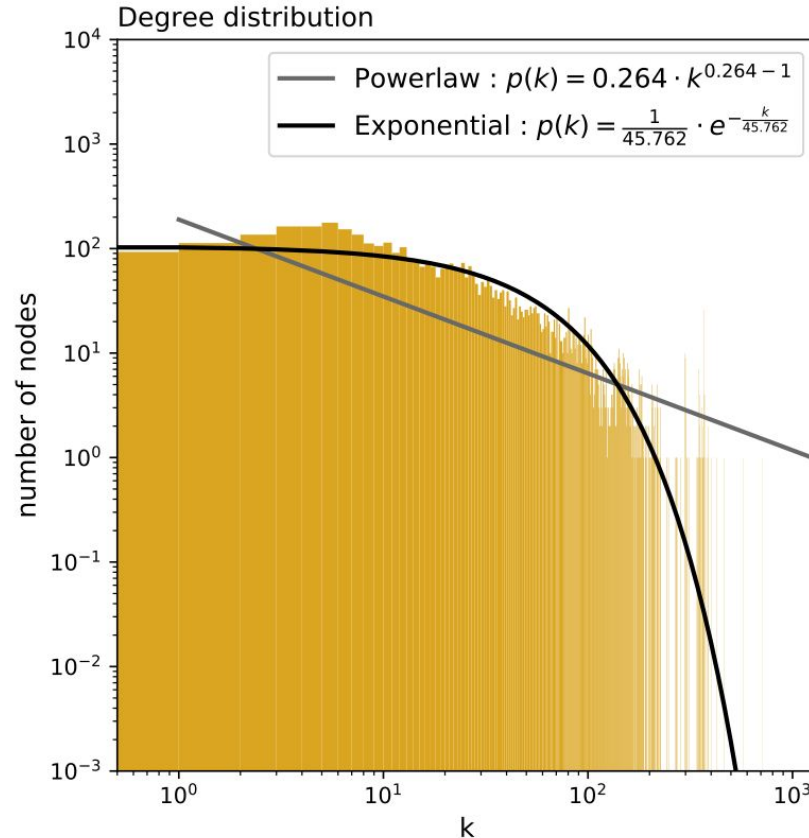
Graph Properties

	Whole Network	Giant Component
N edges	109874	109859
N nodes	4802	4690
N connected components	100	1
Giant component size	4690	4690
Avg degree	45.76	46.85
Density	0.0095	0.01
Diameter	None	11
Avg clustering coefficient	0.62	0.638

EPFL

Graph Type

Degree distribution

- Two curves fitted to the degree distribution

- Exponential law → best fit

- Exponential cutoff → fewer and smaller hubs than a scale free

- Between a scale free network in a sublinear regime and a random network [1].

[1] Network Science, by Albert-László Barabási, 2016 - Chapter 5

Network Structure

- Visualization of the network in a circle pack layout based on connected component, modularity and clustering coefficient

- Nodes colored according to number of Nominations

- Nodes sized according to number of Awards

Nominations
22
0
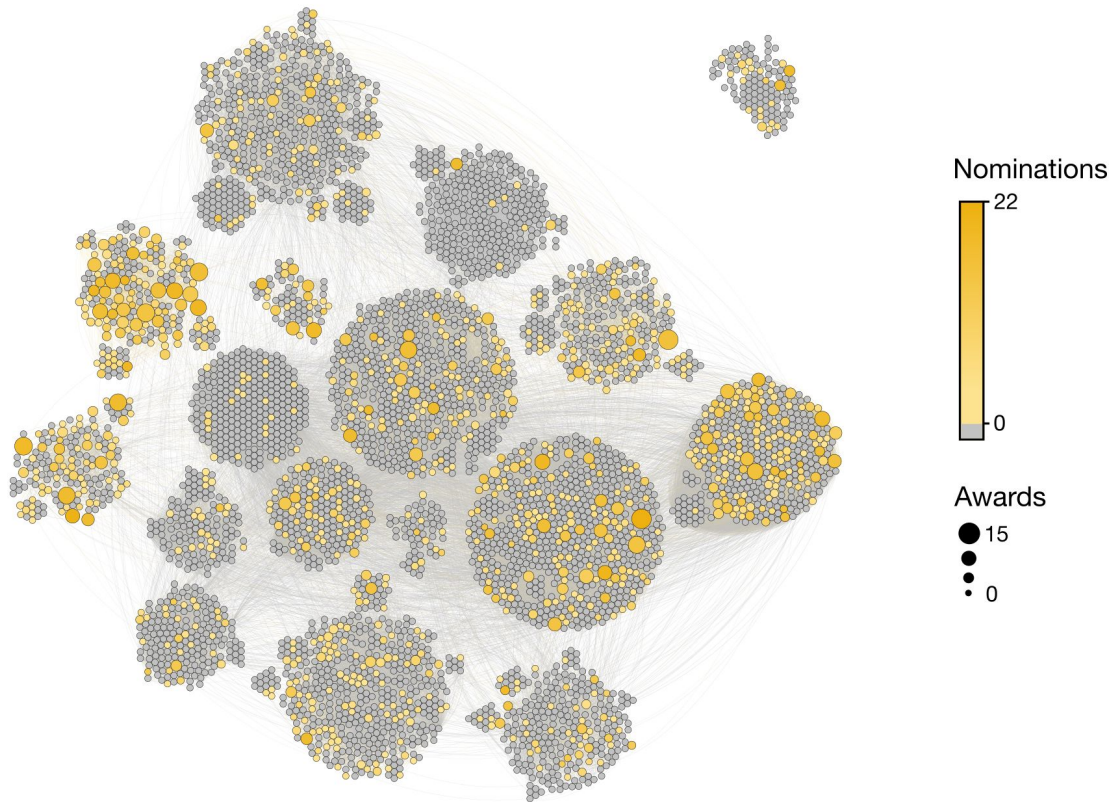
Awards
15
0

Fun Facts

- *Network* is a movie Nominated 15 times which received 4 Oscars and 4 Golden Globes. It's connected to *Julia* which is also a programming language and received awards.

- *The Untouchable* is a movie which is ironically not connected to the network.

- There is a movie called *AI Artificial Intelligence* which got nominated twice for the Oscars.
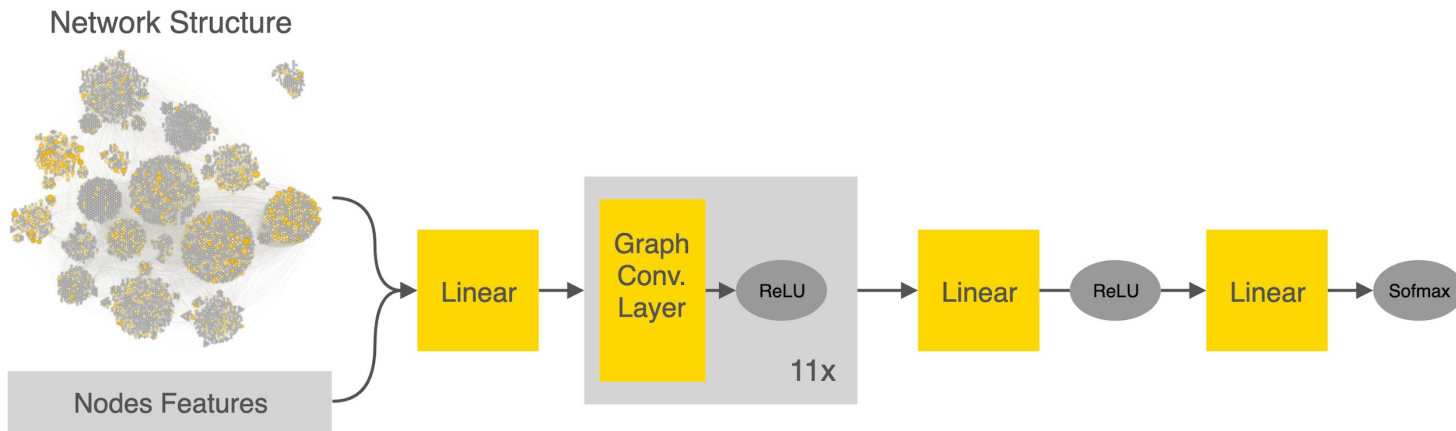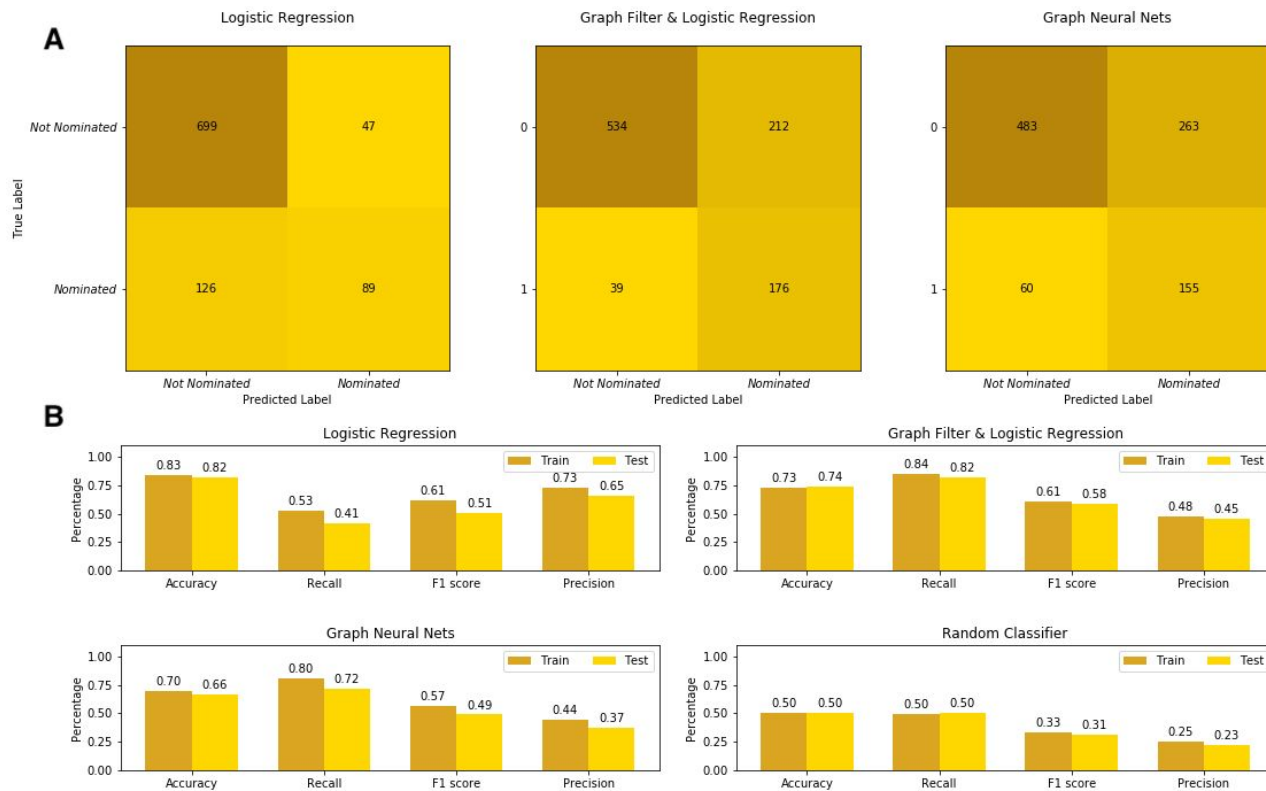
EPFL
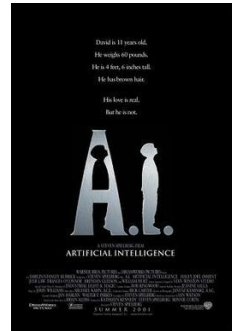
Prediction Models

Could one predict if a new movie will be nominated at the next Award season ?

- Logistic Regression
- Graph Polynomial Filter & Logistic Regression
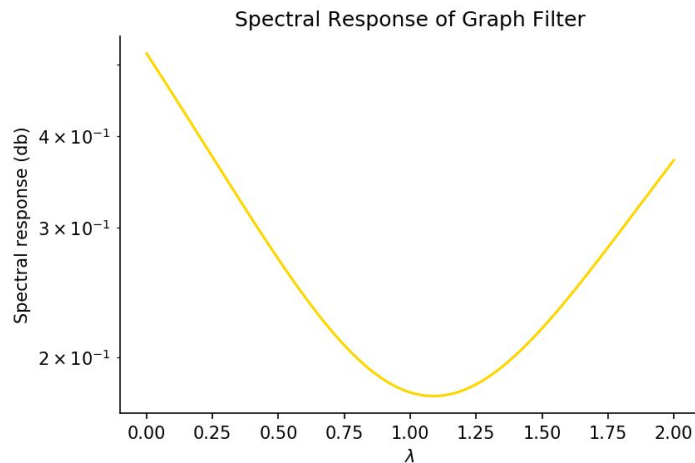- Graph Convolutional Neural Network
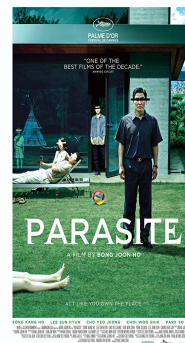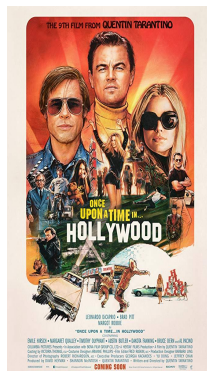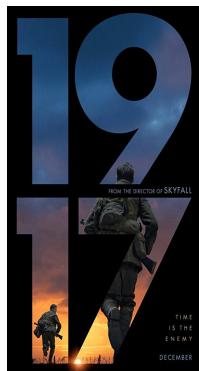
Prediction Performances

Discussion

- Combining Laplacian **Polynomial Graph Filter** with **Logistic Regression** gives highest F1-score and lowest overfit (Train F1 - 61% , Test F1 - 58%)

Our model predicts the success of a film based on the criterias of American Film Awards institutions with a 82% of true positive rate

Spectral Response of Graph Filter

LR : Logistic regression   GF_LR : Graph Laplacian + Logistic Regression   GCN : Graph Convolutional Network

EPFL

Limitations & Further Improvements

Limitations...

- Features sparsity
- Signal-Nomination correlation
- Class imbalance

... & Further Improvements

- Get more movies features
- Different graph construction
- Rebalance the dataset

Conclusion

- Our study question was partially answered
- Yet many information could be gathered from the network structure