NTDS: Can we estimate the earnings of a movie **?**

Team 40

Roc Arandes

Romain Caristan

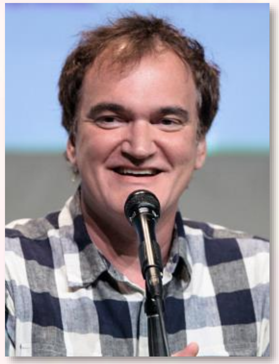
Hugo Kohli

Benoît Pasquier

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

○ Goal

" Can we ***estimate the earnings of a movie*** *by knowing its participating members? "*



○ Data acquisition and preprocessing

- TMDB 5000 Movies Dataset

- if *budget* OR *revenue* == 0 ~ **32%**
- if *budget* OR *revenue* < 1000 ~ **0.6%**

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Data acquisition and preprocessing

- *Creation of:*

$$\boxed{earnings = \frac{revenue - budget}{budget}}$$

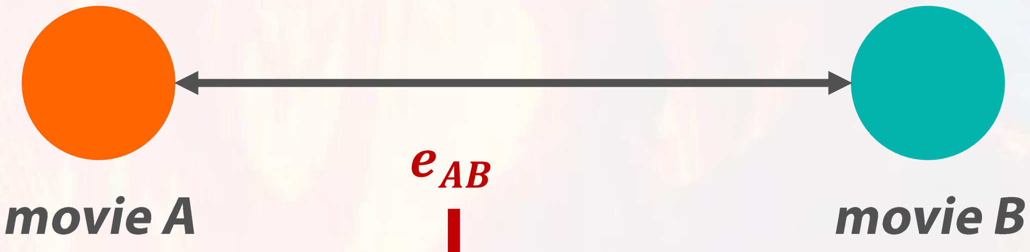
- *5 actors / 5 characters / director / producer / prod. company were kept as* ***features***

	original_title	earnings	features
0	Avatar	10.763566	[Sam Worthington, Zoe Saldana, Sigourney Weave...
1	Pirates of the Caribbean: At World's End	2.203333	[Johnny Depp, Orlando Bloom, Keira Knightley, ...
2	Spectre	2.594590	[Daniel Craig, Christoph Waltz, Léa Seydoux, R...
3	The Dark Knight Rises	3.339756	[Christian Bale, Michael Caine, Gary Oldman, A...
4	John Carter	0.092843	[Taylor Kitsch, Lynn Collins, Samantha Morton,...

○ Adding weights model

- ***weight:*** $\dfrac{\sum earnings}{n° \; of \; movies}$ …

(per feature)



$\sum weight_{common\ f.}$

*Table containing all **features** and corresponding **weights***

	feature	weight
0	Larry Mullen Jr.	0.515389
1	"Drugs" Delaney	0.041739
2	"Hickory" / The Tin Man	11.155192
3	"Hunk" / The Scarecrow	11.155192
4	"Whistling" John Shaw	-0.487703

Problem:

Outliers!

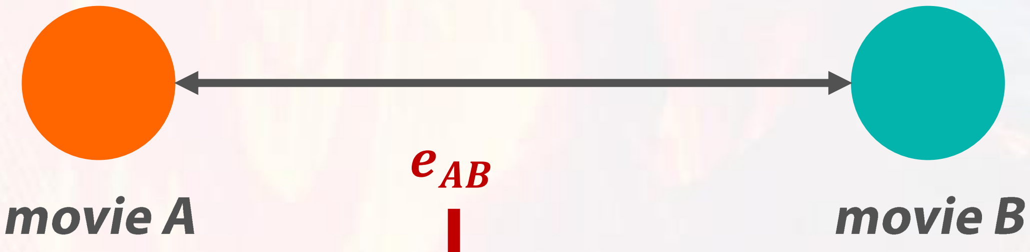
(ex: Paranormal Activity)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Adding weights model

- **weight:** $\frac{\sum earnings}{n° \; of \; movies}$ …

(per feature)



$\sum weight_{common\ f.}$

*Table containing all **features** and corresponding **weights***

	feature	weight
15145	Mark Fredrichs	12889.386667
1639	Ashley Palmer	12889.386667
15920	Micah Sloat	12889.386667
922	Amir Zbeda	12889.386667
885	Amber Armstrong	12889.386667

Problem:

Outliers!

(ex: Paranormal Activity)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

BRH Index

Random actor example:

	earnings
666	1.659325
1469	1.226214
2130	1.842342
2894	1.476055
3645	2.297675

⋮

	earnings
28677	-0.627225
32021	1.528698
32228	0.183054
34739	-0.437092
34951	7.355787

All earnings

	earnings
0	891.854118
1	835.000000
2	735.578720
3	441.320728
4	435.462943

Head of positive *(in %)*

Positive
BRH-Index = 20
(20 films with at least 20% earnings)

	earnings
0	-62.722540
1	-56.186133
2	-50.960690
3	-43.709183

Head of negative *(in %)*

Negative
BRH-Index = -4
(4 films with at least 4% losses)

BRH-Index = 16

BRH Index

Random actor example:

	earnings
666	1.659325
1469	1.226214
2130	1.842342
2894	1.476055
3645	2.297675

⋮

	earnings
28677	-0.627225
32021	1.528698
32228	0.183054
34739	-0.437092
34951	7.355787

All earnings

	earnings
0	891.854118
1	835.000000
2	735.578720
3	441.320728
4	435.462943

Head of positive *(in %)*

	earnings
0	-62.722540
1	-56.186133
2	-50.960690
3	-43.709183

Head of negative *(in %)*

Positive BRH-Index = 20
(20 films with at least 20% earnings)

BRH-Index = 16

Negative BRH-Index = -4
(4 films with at least 4% losses)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

BRH Index

Random actor example:

	earnings
666	1.659325
1469	1.226214
2130	1.842342
2894	1.476055
3645	2.297675

⋮

	earnings
28677	-0.627225
32021	1.528698
32228	0.183054
34739	-0.437092
34951	7.355787

All earnings

	earnings
0	891.854118
1	835.000000
2	735.578720
3	441.320728
4	435.462943

Head of positive *(in %)*

Positive BRH-Index = 20
(20 films with at least 20% earnings)

BRH-Index = 16

	earnings
0	-62.722540
1	-56.186133
2	-50.960690
3	-43.709183

Head of negative *(in %)*

Negative BRH-Index = -4
(4 films with at least 4% losses)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

BRH Index

Random actor example:

	earnings
666	1.659325
1469	1.226214
2130	1.842342
2894	1.476055
3645	2.297675

⋮

	earnings
28677	-0.627225
32021	1.528698
32228	0.183054
34739	-0.437092
34951	7.355787

All earnings

	earnings
0	891.854118
1	835.000000
2	735.578720
3	441.320728
4	435.462943

Head of positive *(in %)*

Positive BRH-Index = 20

(20 films with at least 20% earnings)

	earnings
0	-62.722540
1	-56.186133
2	-50.960690
3	-43.709183

Head of negative *(in %)*

Negative BRH-Index = -4

(4 films with at least 4% losses)

BRH-Index = 16

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

○ BRH Index

Random actor example:

	earnings
666	1.659325
1469	1.226214
2130	1.842342
2894	1.476055
3645	2.297675

⋮

	earnings
28677	-0.627225
32021	1.528698
32228	0.183054
34739	-0.437092
34951	7.355787

All earnings

	earnings
0	891.854118
1	835.000000
2	735.578720
3	441.320728
4	435.462943

Head of positive *(in %)*

Positive BRH-Index = 20
(20 films with at least 20% earnings)

	earnings
0	-62.722540
1	-56.186133
2	-50.960690
3	-43.709183

Head of negative *(in %)*

Negative BRH-Index = -4
(4 films with at least 4% losses)

BRH-Index = 16

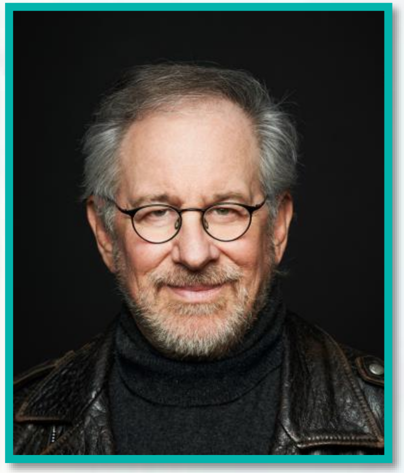
Right-shifted *for graph construction*

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

BRH Index

	feature	weight	brh_index
23505	Universal Pictures	3.901744	123.0
17856	Paramount Pictures	4.439880	114.0
23392	Twentieth Century Fox Film Corporation	3.778530	105.0
4573	Columbia Pictures	1.878315	94.0
17181	New Line Cinema	8.538247	81.0
23882	Walt Disney Pictures	2.663880	72.0
23503	United Artists	10.830685	59.0
6635	DreamWorks SKG	2.315972	47.0
7818	Fox Searchlight Pictures	4.725724	45.0
23716	Village Roadshow Pictures	1.475530	42.0
11519	Joel Silver	1.592479	41.0
21979	Steven Spielberg	8.786479	41.0
2665	Brad Pitt	2.236545	41.0
16454	Miramax Films	6.302761	40.0
23945	Warner Bros.	4.324325	39.0
4574	Columbia Pictures Corporation	3.279925	38.0
22827	Tim Bevan	2.616512	37.0
22098	Summit Entertainment	5.444916	37.0
23030	Tom Hanks	5.591047	36.0
16654	Morgan Freeman	2.868493	36.0

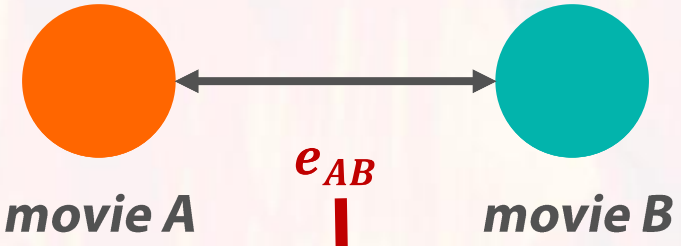
UNIVERSAL



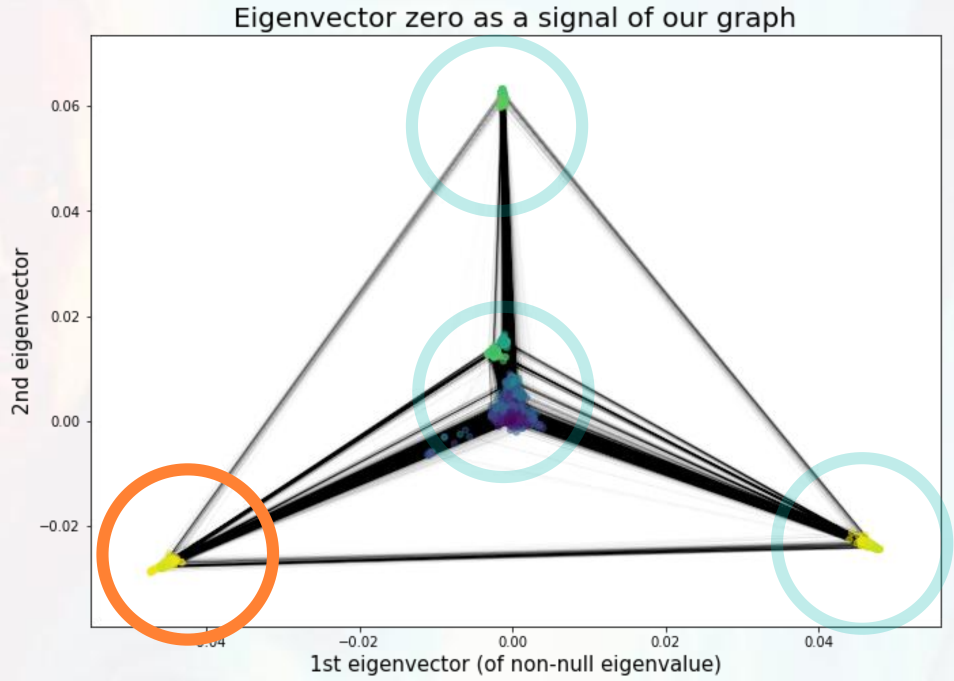
WARNER BROS. PICTURES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

○ Adding BRH model



$$\sum BRH\ index_{common\ f.}$$

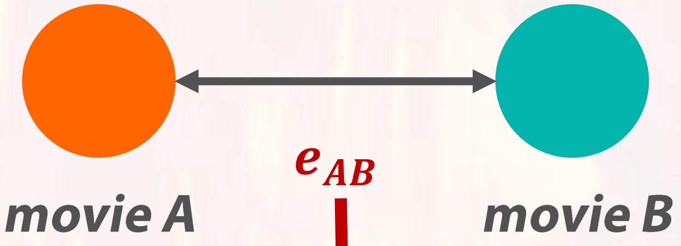


4 main clusters

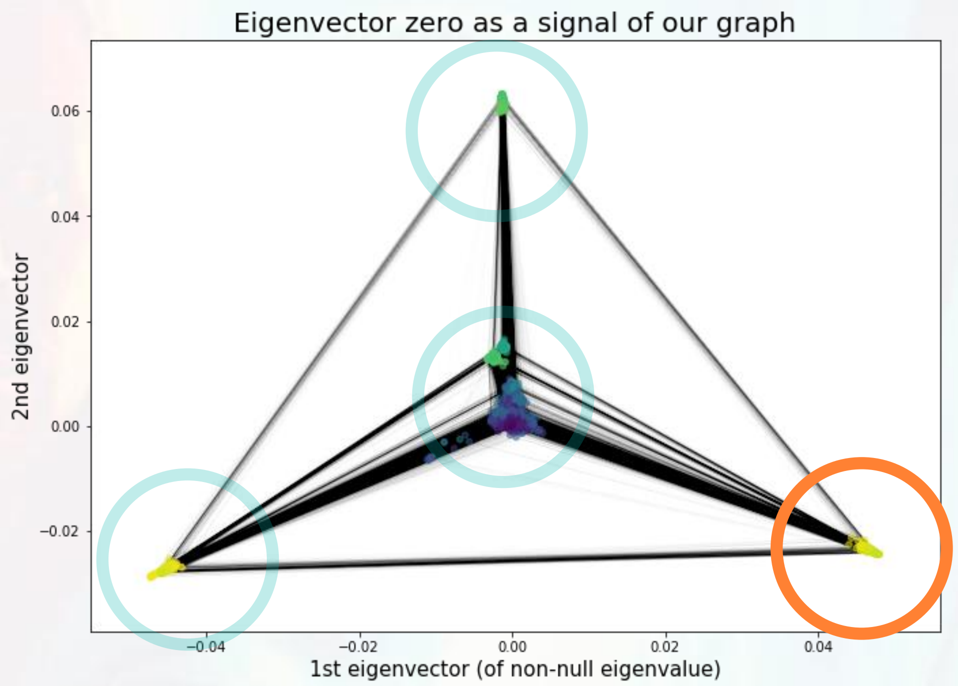
Representing production companies

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

○ Adding BRH model



$$\sum BRH\ index_{common\ f.}$$

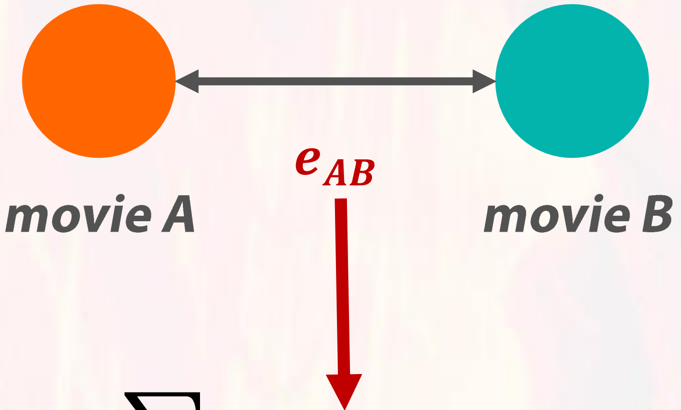


4 main clusters

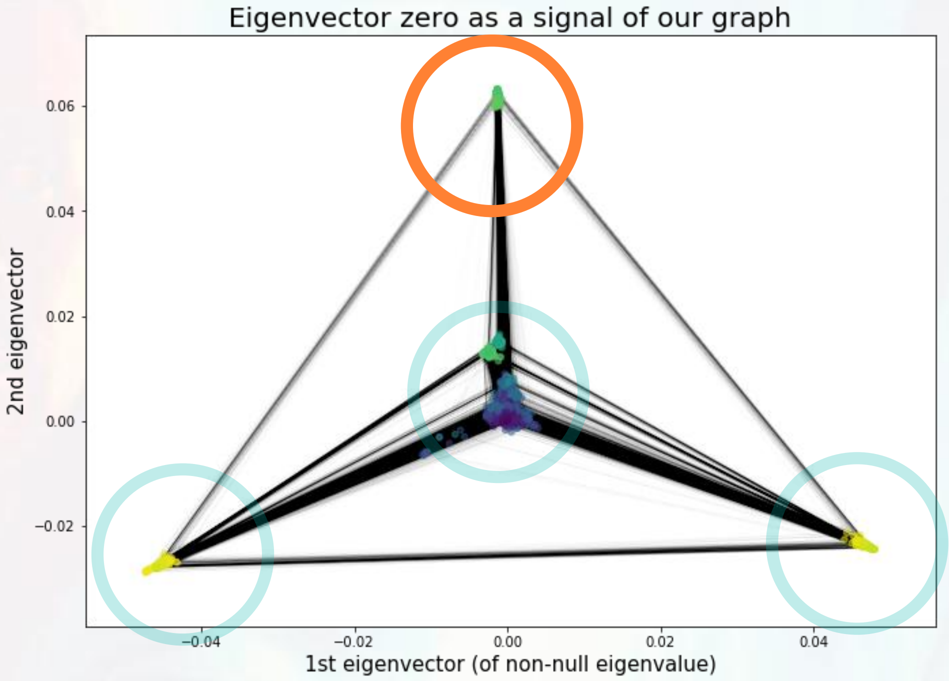
Representing production companies

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

○ Adding BRH model



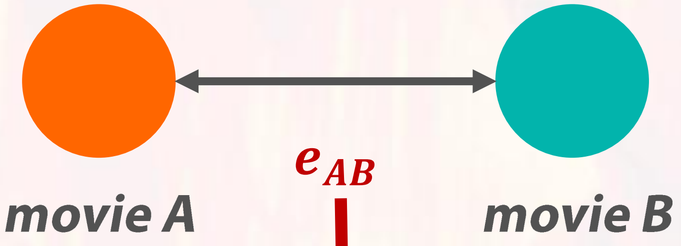
$$\sum BRH\ index_{common\ f.}$$



4 main clusters
Representing
production companies

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

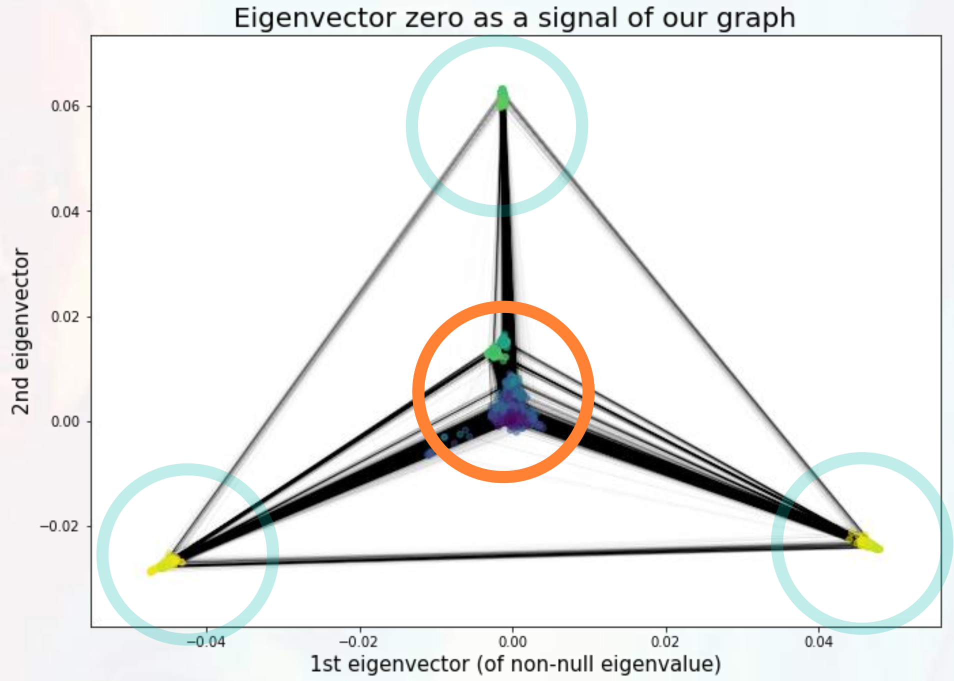
Adding BRH model



$$\sum BRH\ index_{common\ f.}$$

4 main clusters

Representing production companies

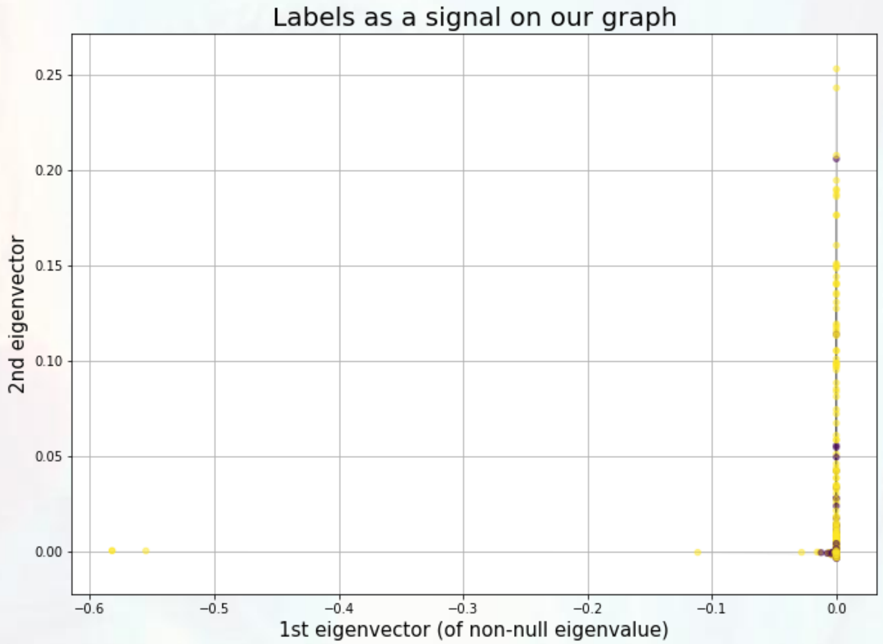


other companies

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Averaging BRH model

$$\sum \frac{BRH\ index_{common\ f}}{N_{common\ f}}$$



+1 : *earnings* >= **100%**

-1 : *earnings* < **100%**

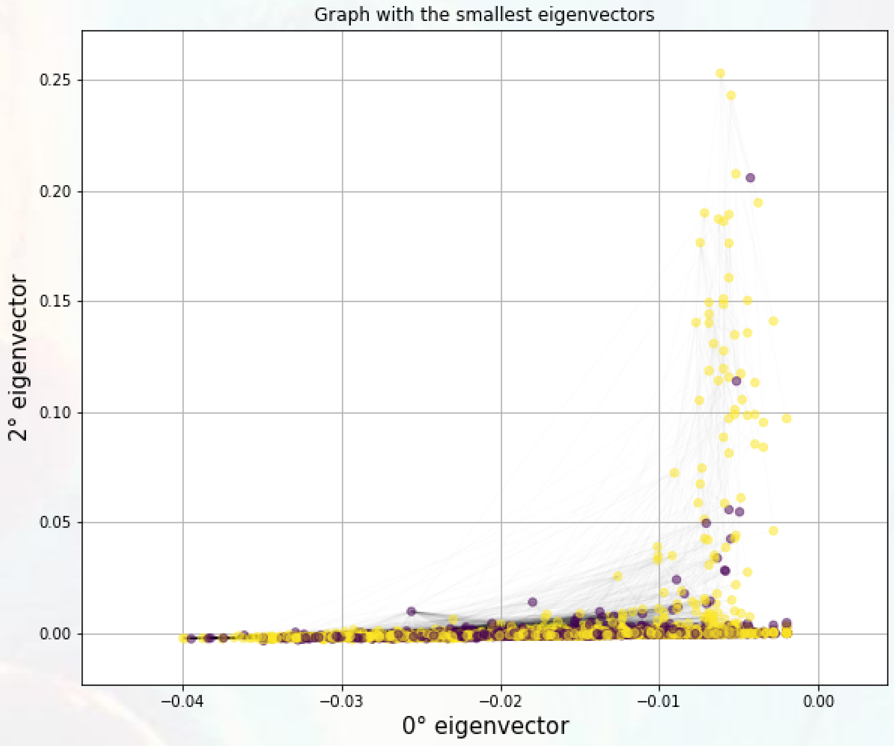
- **Next step :**
  Display labels on most representative eigenvectors

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

○ Averaging BRH model

$$\sum \frac{BRH\ index_{common\ f}}{N_{common\ f}}$$

- **Again, not separable**
  But nodes are more spread out



+1 : *earnings* >= **100%**

-1 : *earnings* < **100%**

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

○ Averaging BRH model: interpolation

Initial goal :

Earning estimation → Positive or negative earning

Compromise:

- Earning x1
- Earning x4
- Earning x2
- Earning x…

- Negative earning
- Positive earning

- From a small subset… …**interpolate** the rest of the graph

○ Averaging BRH model: interpolation

Initial goal :

Earning estimation → Positive or negative earning

Compromise:

- Earning x1
- Earning x4
- Earning x2
- Earning x…

- Negative earning
- Positive earning

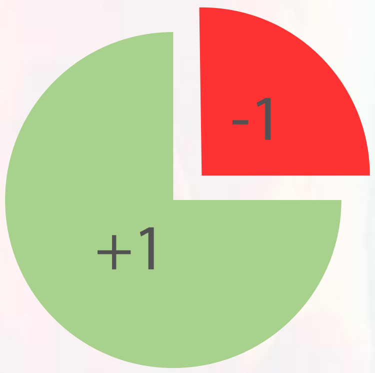
- From a small subset…

…**interpolate** the rest of the graph

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

○ Averaging BRH model: interpolation

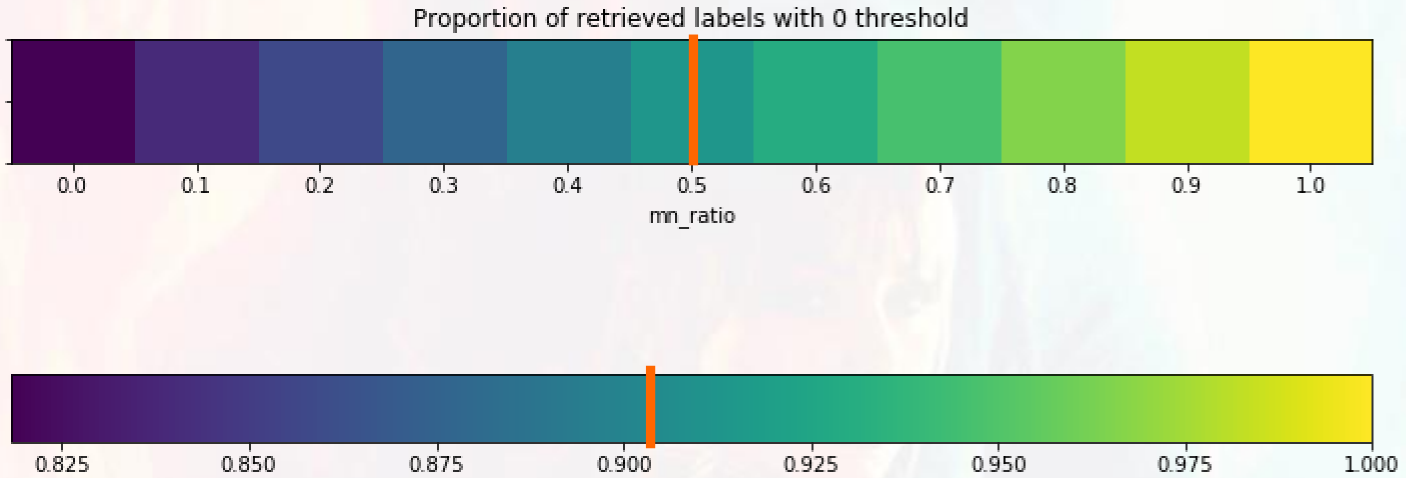
Binary labels :



Loss-making movies

Profitable movies

- **~90%** of correct interpolation with **50% of labels**

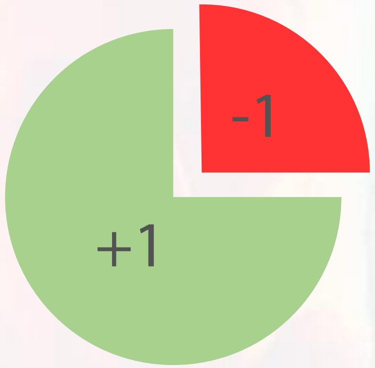


1 Introduction
2 Data
3 First attempt
4 BRH models
5 Conclusion

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Averaging BRH model: interpolation

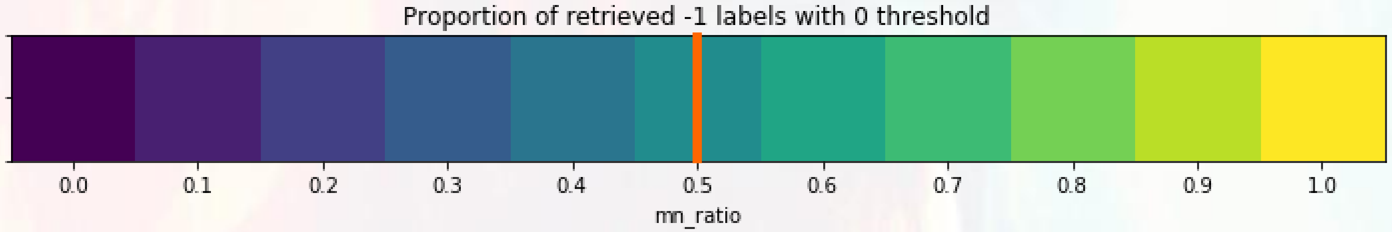
Binary labels :

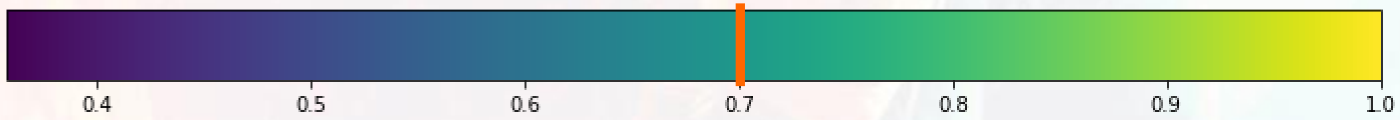


Loss-making movies

Profitable movies

- **~70%** of negative labels are **correctly** found





ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Can we estimate the earnings of a movie?

- **BRH index :**

  Promising way to describe ***economical consistency***

- **Edge calculation :**

  Further investigation required (Jensen – Shanon Divergence, etc.)

- **Interpolation :**

  Successfully separate ***profitable VS loss-making*** movies

- **Earnings estimation :**

  Many other features are linked to the financial success of a movie

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Q&A

Team 40

Roc Arandes

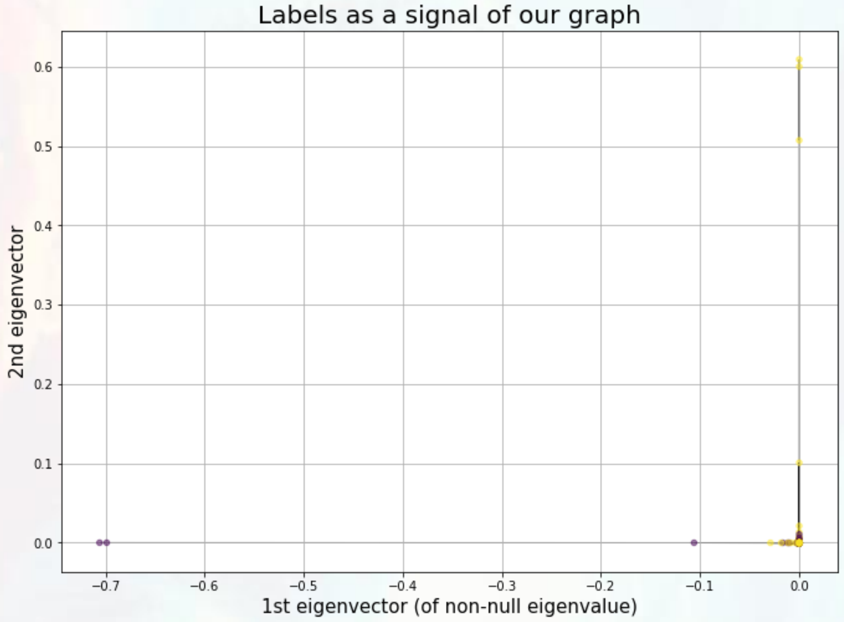
Romain Caristan

Hugo Kohli

Benoît Pasquier

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

○ Adding BRH model, no companies



movie A $\longleftrightarrow$ ***movie B***

e_{AB} → $\sum BRH\ index_{common\ f.}$

Problem:

Graph not separable

Not possible to get earnings from the graph

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE