

MF TUTORIAL PART 4  
**BICLUSTERING**

Sibylle Hess and Michiel Hochstenbach

We have discussed the clustering of arbitrary shapes, but there is still one issue:

## THE CURSE OF DIMENSIONALITY

In high dimensions, everything  
is almost equally similar

→ **CLUSTER FEATURES  
AND OBSERVATIONS**

# BICLUSTERING:

# RECOMMENDER SYSTEMS

The image shows the Netflix homepage. At the top, there is a navigation bar with links for Home, TV Shows, Movies, Recently Added, and My List. To the right of the navigation bar are icons for search, KIDS, notifications (with a red dot), and account settings. The main banner is for the Netflix Original series "BLACK MIRROR". The title "BLACK MIRROR" is displayed in large, metallic letters against a background of shattered glass, with a prominent black smiley face in the center. Below the banner are three buttons: "Play", "My List", and "More Info". A promotional text for "Watch Season 5 Now" is present, followed by a descriptive paragraph about the show's premise. Below the main banner, there is a section titled "Popular on Netflix" featuring thumbnails for various shows and movies, including "BLACK MIRROR", "HOW TO SELL DRUGS ONLINE", "Rick and Morty", "MY NEXT GUEST", "LOVE X DEATH X ROBOTS", and "THE AMAZING SPIDER-MAN".

NETFLIX

Home TV Shows Movies Recently Added My List

NETFLIX ORIGINAL

BLACK MIRROR

Play My List More Info

Watch Season 5 Now

The sleek world of tomorrow offers opportunities beyond our wildest dreams. At a price out of our worst nightmares.

Popular on Netflix

NETFLIX BLACK MIRROR NEW EPISODES

NETFLIX HOW TO SELL DRUGS ONLINE PART 1

Rick and Morty

NETFLIX MY NEXT GUEST DAVID LETTERMAN NEW EPISODES

NETFLIX LOVE X DEATH X ROBOTS

NETFLIX THE AMAZING SPIDER-MAN

# EXCLUSIVE BICLUSTERING TRI-FACTORIZATIONS

$$\min_{Y, C, X} \|D - Y C X^\top\|^2 \quad \text{s.t. } Y \in \mathbb{1}^{m \times k}, C \in \mathbb{R}^{k \times k} \subseteq \mathcal{C}, X \in \mathbb{1}^{n \times k}$$

$Y$                      $C$                      $X^\top$

Checkerboard

$$\begin{pmatrix} \text{dark blue} \\ \text{dark blue} \\ \text{dark blue} \end{pmatrix} \begin{pmatrix} \text{dark blue} & \text{medium blue} & \text{light blue} \\ \text{medium blue} & \text{yellow} & \text{yellow} \\ \text{light blue} & \text{yellow} & \text{light blue} \end{pmatrix} \begin{pmatrix} \text{dark blue} \\ \text{dark blue} \\ \text{dark blue} \end{pmatrix} = \begin{pmatrix} \text{dark blue} & \text{medium blue} & \text{light blue} \\ \text{medium blue} & \text{yellow} & \text{yellow} \\ \text{light blue} & \text{yellow} & \text{light blue} \end{pmatrix}$$

Block-Diag.

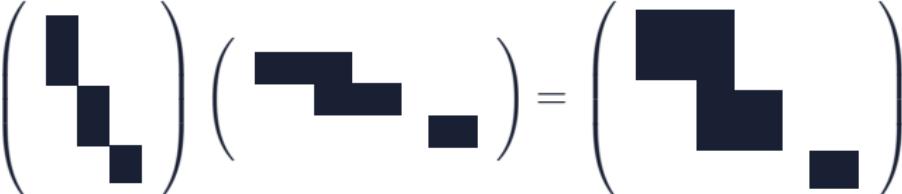
$$\begin{pmatrix} \text{dark blue} \\ \text{dark blue} \\ \text{dark blue} \end{pmatrix} \begin{pmatrix} \text{dark blue} \\ \text{yellow} \\ \text{light blue} \end{pmatrix} \begin{pmatrix} \text{dark blue} \\ \text{dark blue} \\ \text{dark blue} \end{pmatrix} = \begin{pmatrix} \text{dark blue} \\ \text{dark blue} \\ \text{dark blue} \end{pmatrix} \begin{pmatrix} \text{yellow} \\ \text{light blue} \end{pmatrix} \begin{pmatrix} \text{dark blue} \\ \text{dark blue} \\ \text{dark blue} \end{pmatrix} = \begin{pmatrix} \text{dark blue} \\ \text{dark blue} \\ \text{dark blue} \end{pmatrix}$$

## BINARY MATRIX FACTORIZATION

If the data matrix is binary, then a binary factorization is more interpretable:

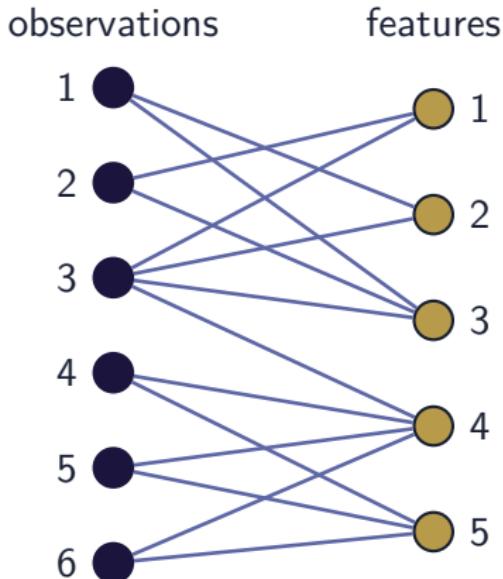
$$\min_{Y,X} \|D - YX^\top\|^2 \quad \text{s.t. } Y \in \{0,1\}^{m \times k}, X \in \{0,1\}^{n \times k}$$

Binary

$$\begin{pmatrix} Y \\ X^\top \end{pmatrix} = \begin{pmatrix} \text{Binary Matrix} & \text{Binary Matrix} \end{pmatrix}$$


The factor matrices are here not restricted to partitions, but the objective has an inbuilt **PENALIZATION** of **OVERLAP** between clusters.

# BIPARTITE GRAPH CUT INTERPRETATION OF BINARY AND BLOCK-DIAGONAL MF



Interpret  $D$  as adjacency matrix of bipartite graph, the similarity matrix is defined as

$$W = \begin{pmatrix} \mathbf{0} & D \\ D^\top & \mathbf{0} \end{pmatrix}$$

The **MINIMUM CUT** objective can then be used for the identification of biclusters.

Most cluster models assume  
**EXCLUSIVITY OF**  
**CLUSTERS** to use  
optimization tricks from  
 $k$ -means

However, especially for  
biclustering applications, the  
**EXCLUSIVITY**  
**ASSUMPTION** is often  
**INEPT.**

A movie does not only belong to one cluster (e.g. genre)

# BICLUSTERING WITH OVERLAP AND OUTLIERS

$$\min_{Y, C, X} \|D - Y C X^\top\|^2 \quad \text{s.t. } Y \in \{0, 1\}^{m \times k}, C \in \mathbb{R}^{k \times k}, X \in \{0, 1\}^{n \times k}$$

$$\text{Biclustering} \left( \begin{array}{c} Y \\ \hline \end{array} \right) \left( \begin{array}{c} C \\ \hline \end{array} \right) \left( \begin{array}{c} X^\top \\ \hline \end{array} \right) = \left( \begin{array}{cc} \text{Biclustered Matrix} & \text{Color Bar} \\ \hline \end{array} \right)$$

# BOOLEAN MATRIX FACTORIZATION

$$\min_{Y, X} \|D - Y \otimes X^\top\|^2 \quad \text{s.t. } Y \in \{0, 1\}^{m \times k}, X \in \{0, 1\}^{n \times k}$$

$$\text{Boolean} \begin{pmatrix} Y \\ \vdots \\ \text{---} \\ \text{---} \end{pmatrix} \begin{pmatrix} X^\top \\ \vdots \\ \text{---} \\ \text{---} \end{pmatrix} = \begin{pmatrix} \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \end{pmatrix}$$

The **BOOLEAN** matrix product **ALLOWS FOR OVERLAP** :

$$Y \otimes X^\top = Y_{\cdot 1} X_{.1}^\top \oplus \dots \oplus Y_{\cdot k} X_{.k}^\top, \quad \text{where } 1 \oplus 1 = 1$$

# OPTIMIZATION

## ALTERNATING MINIMIZATION

$$X_{t+1} = \arg \min_X \|D - Y_t C_t X^\top\|^2 \quad \text{s.t. } X \in \mathbb{L}^{n \times k}$$

$$C_{t+1} = \arg \min_C \|D - Y_t C X_{t+1}^\top\|^2 \quad \text{s.t. } C \in \mathbb{R}^{k \times k}$$

$$Y_{t+1} = \arg \min_Y \|D - Y C_{t+1} X_{t+1}^\top\|^2 \quad \text{s.t. } Y \in \mathbb{L}^{m \times k}$$

The **EXCLUSIVITY ASSUMPTION** makes the analytical computation of optimal  $X_{t+1}$ ,  $C_{t+1}$  and  $Y_{t+1}$  possible.

### PROS:

- ▶ fast convergence
- ▶ no hyperparameters except for rank

### CONS:

- ▶ sensitive to initialization
- ▶ only applicable to partitioning clusters

# ALTERNATING MINIMIZATION IN THE LITERATURE

## ► CHECKERBOARD BICLUSTERING

Vichi 2001, Cho et al. 2004, Wang et al. 2011

## ► DIAGONAL BICLUSTERING

Han et al. 2017; Song et al. 2020

## ► BINARY MATRIX FACTORIZATION

(one factor satisfies the exclusivity assumption)

Koyotürk and Grama 2003, Li 2005

# BIPARTITE GRAPH CUTS

Minimize the graph cut of

$$W = \begin{pmatrix} \mathbf{0} & D \\ D^\top & \mathbf{0} \end{pmatrix}$$

## PROS:

- ▶ Fast, TSVD of  $D$  is sufficient to get the eigendecomposition of  $W$

## CONS:

- ▶ Sensitive to similarity measure and noise

## LITERATURE

- ▶ **BLOCK-DIAG. MF** Zha et al. 2001, Dhillon 2001  
Learn suitable graph representation: Nie et al. 2017
- ▶ **BINARY MF** Neumann 2018

# NONNEGATIVE (SOFT)-ORTHOGONAL RELAXATION

$$\min_{\substack{Y \in \mathbb{R}_+^{m \times k}, \\ C \in \mathbb{R}_+^{k \times k}, \\ X \in \mathbb{R}_+^{n \times k}}} \|D - YCX^\top\|^2, \text{ s.t. } Y^\top Y = I, X^\top X = I.$$

Orthogonal NMF is NP-hard, in practice only soft-orthogonality of the matrices is achieved

Asteris et al. 2015

## PROS:

- ▶ NMF optimization methods can be applied
- ▶ Soft orthogonality allows for overlap and identification of outliers

## CONS:

- ▶ Discretization is necessary  
→ quality of binary end-result unclear

$$\min_{\substack{Y \in \mathbb{R}^{m \times k}, \\ C \in \mathbb{R}_+^{k \times k}, \\ X \in \mathbb{R}^{n \times k}}} \|D - YCX^\top\|^2 + \phi(X) + \phi(Y)$$

$\phi$  penalizes nonbinary elements in the factor matrices  $X$  and  $Y$ .

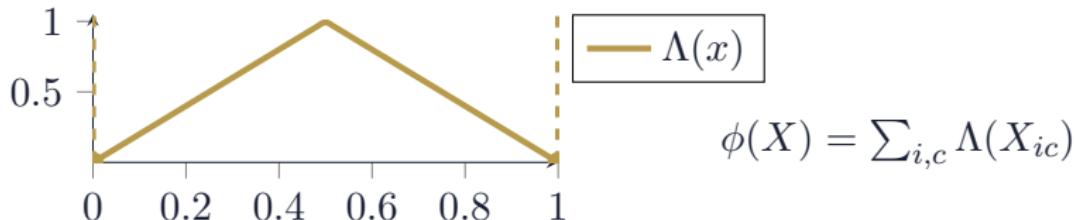
### PROS:

- ▶ NMF optimization methods can be applied
- ▶ No exclusivity, models overlap and outliers
- ▶ Results are very close to or equal to binary matrices

### CONS:

- ▶ Discretization might be still necessary
- ▶ Scaling of matrices during optimization is an issue

# PROX-OPERATOR FOR NONBINARY PENALIZATION



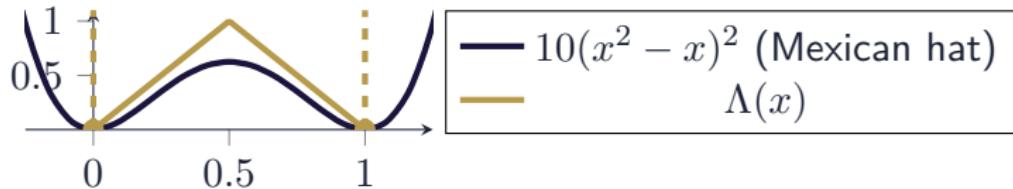
The Prox-Operator for  $\phi$  is given by

$$(\text{prox}_{\alpha\phi}(X))_{ic} = \text{prox}_{\alpha\Lambda}(X_{ic}) \text{ with}$$

$$\text{prox}_{\alpha\Lambda}(x) = \begin{cases} \max\{0, x - \alpha\} & x \leq 0.5 \\ \min\{1, x + \alpha\} & x > 0.5. \end{cases}$$

# NONBINARY LITERATURE

## PENALIZATION IN THE



### ► CHECKERBOARD BICLUSTERING

Stochastic PALM with  $\Lambda$ -Penalization: Hess et al. 2021

### ► BINARY MATRIX FACTORIZATION

Multiplicative Updates with Mexican hat penalization

Zhang et al. 2007/2010/2013

### ► BOOLEAN MATRIX FACTORIZATION

PALM with  $\Lambda$ -Penalization:

Hess et al. 2017

# CONCLUSIONS

## DISCUSSION

- ▶ Many of the one-sided clustering tricks are taken over for biclustering, but having two binary factor matrices poses additional challenges
- ▶ Theory for discretization steps is needed
- ▶ Different penalization functions could be tried (inspiration from regularized regression)
- ▶ Good biclustering optimizations are also relevant for tensor factorizations **Balažević et al. 2019**
- ▶ Modeling overlap and outliers in biclustering is still an active field of research

## SOME REFERENCES PART IV

- |                                     |   |
|-------------------------------------|---|
| <b>Miettinen &amp; Neumann 2020</b> | Survey Boolean MF   |
| <b>Hess et al. 2021</b>             | Penalized optimization of biclustering with overlap and outliers  |
| <b>Del Buono &amp; Pio. 2015</b>    | Analysis of nonnegative matrix tri-factorization for biclustering |