UMONS
Université de Mons

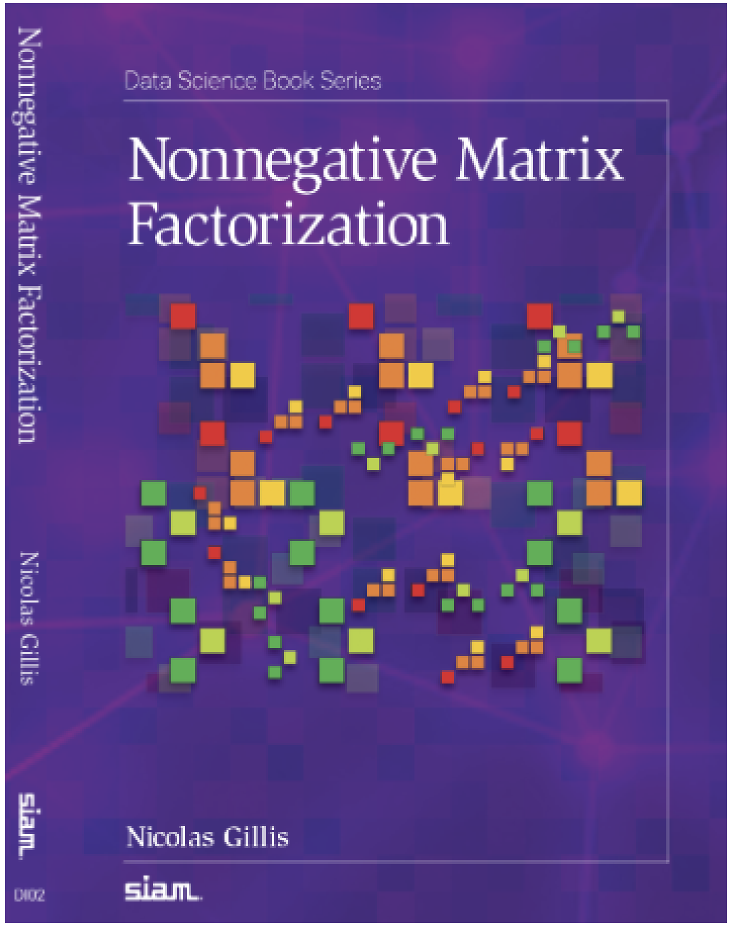
Deep Nonnegative Matrix Factorizations for Data Representation

STADIUS, ESAT, KU Leuven

February 20, 2025

Amjad Seyedi
Doctoral researcher
Department of Mathematics and Operational Research, University of Mons

Nonnegative Matrix Factorization Book



Nonnegative matrix factorization (NMF) in its modern form has become a standard tool in the analysis of high-dimensional data sets. This book provides a comprehensive and up-to-date account of the most important aspects of the NMF problem and is the first to detail its theoretical aspects, including geometric interpretation, nonnegative rank, complexity, and uniqueness. It explains why understanding these theoretical insights is key to using this computational tool effectively and meaningfully. The book also presents models, algorithms and applications of NMF. It contains 2 parts and is divided into 9 chapters:

1. Introduction

Part I - Exact factorizations

2. Exact NMF
3. Nonnegative Rank
4. Identifiability

Part II - Approximate factorizations

5. Models
6. Computational Complexity of NMF
7. Near-separable NMF
8. Iterative algorithms for NMF
9. Applications

Nicolas Gillis

Algebraic Machine Learning Team

Supervisor
Dr. Akhlaghian

Research Assist.
Amjad Seyedi

Alumni
9 Master's

Students
4 Master's &
2 PhD students

Guest
1 PhD candidate
FU-Berlin

LEARNING PROBLEMS

Robust Learning
Robust losses
Self-paced learning
Adversarial training
Distributionally robust learning

Clustering
text clustering
Graph clustering
Fair graph clustering
Directed graph clustering
Semi-supervised clustering

Matrix Completion
Link Prediction
Image Inpainting
Recommender systems

Data Fusion
Multi-view representation
Attributed graph representation
Multi-label learning/feature selection

FACTORIZATION MODELS

Robust NMFs
Weighted NMF
Self-paced NMF
Adversarial NMF
Distributionally robust NMF

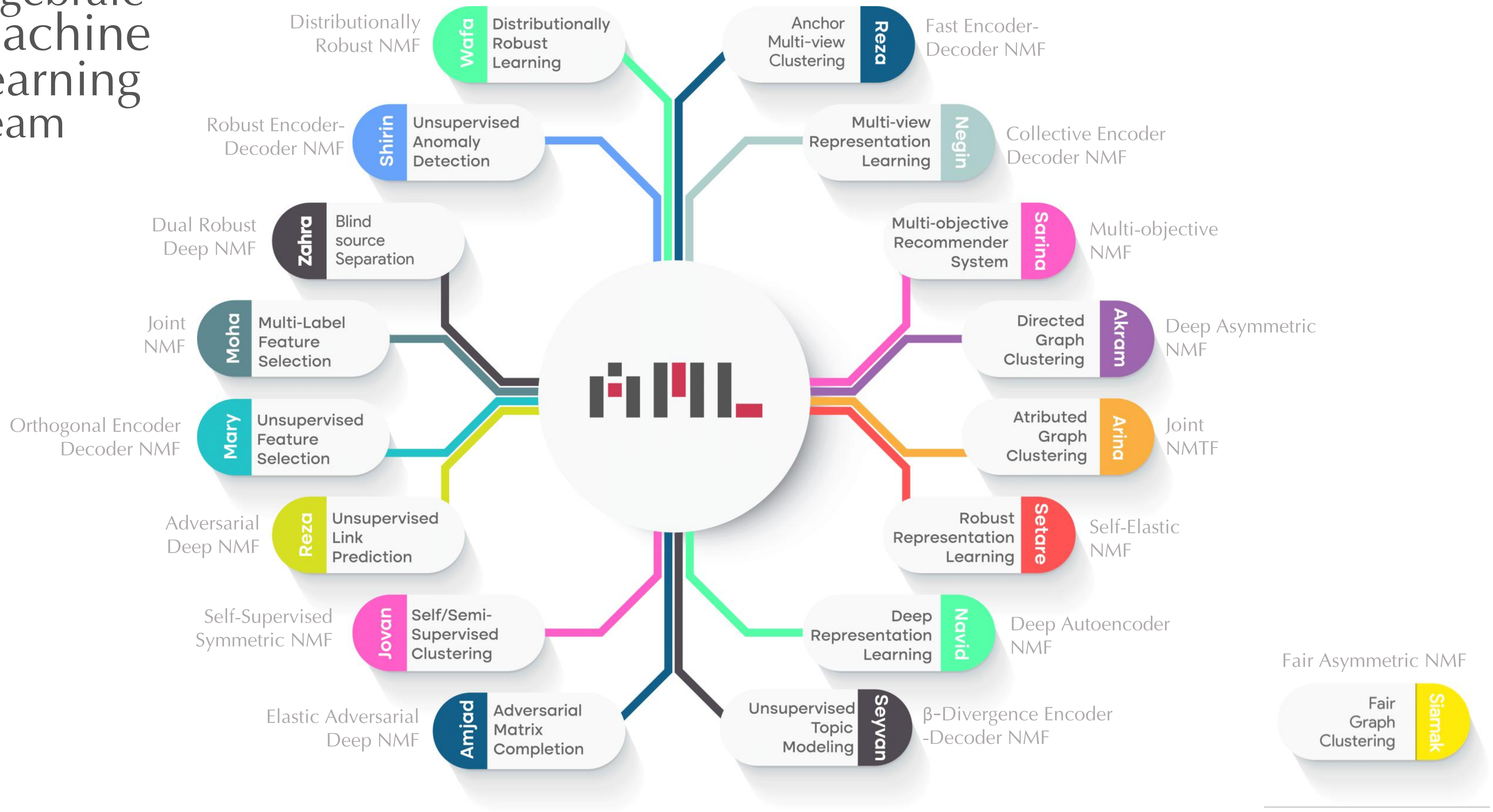
Cluster NMFs
Semi-NMF
NMF with KLD
Orthogonal NMF
constrained NMF
[A]symmetric NMF

Structured NMFs
NMTF
Deep NMF
Autoencoder NMF

Joint NMFs

Algebraic Machine Learning Team

amlteams.github.io



Contents

- Nonnegative Matrix Factorization
- Deep NMFs
  - Deep [Encoder] NMF
  - Deep Autoencoder-like NMF
- Graph-based NMFs
  - Shallow models
  - Deep models

Shallow Matrix Factorization

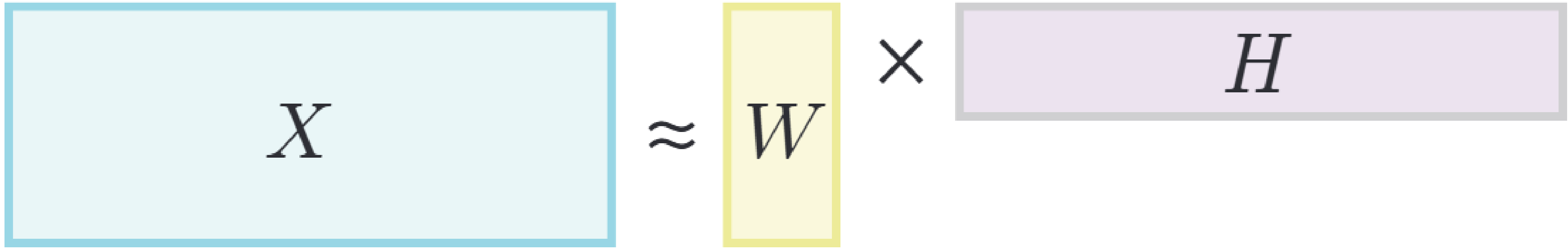
- NMF as a data representation model
- Orthogonal NMF

Basic NMF

Learning the Parts of Objects by Non-negative Matrix Factorization

Nature 1999

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}$$



$\boldsymbol{X}_{d\times n} \in \mathbb{R}^{\geq 0}$, $\boldsymbol{W}_{d\times r} \in \mathbb{R}^{\geq 0}$, $\boldsymbol{H}_{r\times n} \in \mathbb{R}^{\geq 0}$, and $r \ll \min(m, n)$.

Least Squares NMF

Learning the Parts of Objects by Non-negative Matrix Factorization

Nature 1999

$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2 = \sum_{j=1}^{d}\sum_{i=1}^{n}(\mathrm{X}_{ji} - [\boldsymbol{W}\boldsymbol{H}]_{ji})^2, \quad s.t. \quad (\boldsymbol{W},\boldsymbol{H}) \geq 0.$$

$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2 = \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{W}\boldsymbol{h}_i\|^2, \quad s.t. \quad (\boldsymbol{W},\boldsymbol{H}) \geq 0.$$

Multiplicative Update Rules

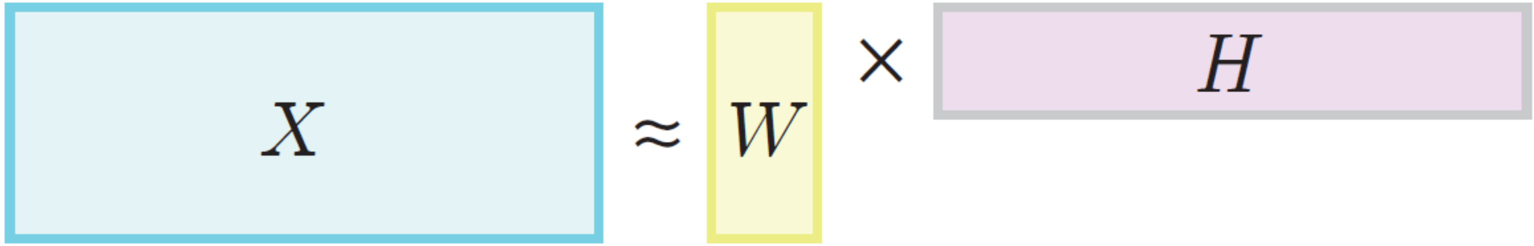
$$\boldsymbol{W} \leftarrow \boldsymbol{W} \odot \frac{\boldsymbol{X}\boldsymbol{H}^\top}{\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^\top} \qquad \boldsymbol{H} \leftarrow \boldsymbol{H} \odot \frac{\boldsymbol{W}^\top\boldsymbol{X}}{\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{H}}$$

Orthogonal NMF

On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering

ICDM 2005

$$\min_{W,H} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2, \quad s.t. \quad (\mathbf{W}, \boldsymbol{H}) \geq 0, \boldsymbol{H}\boldsymbol{H}^\top = \boldsymbol{I}.$$



Deep NMFs

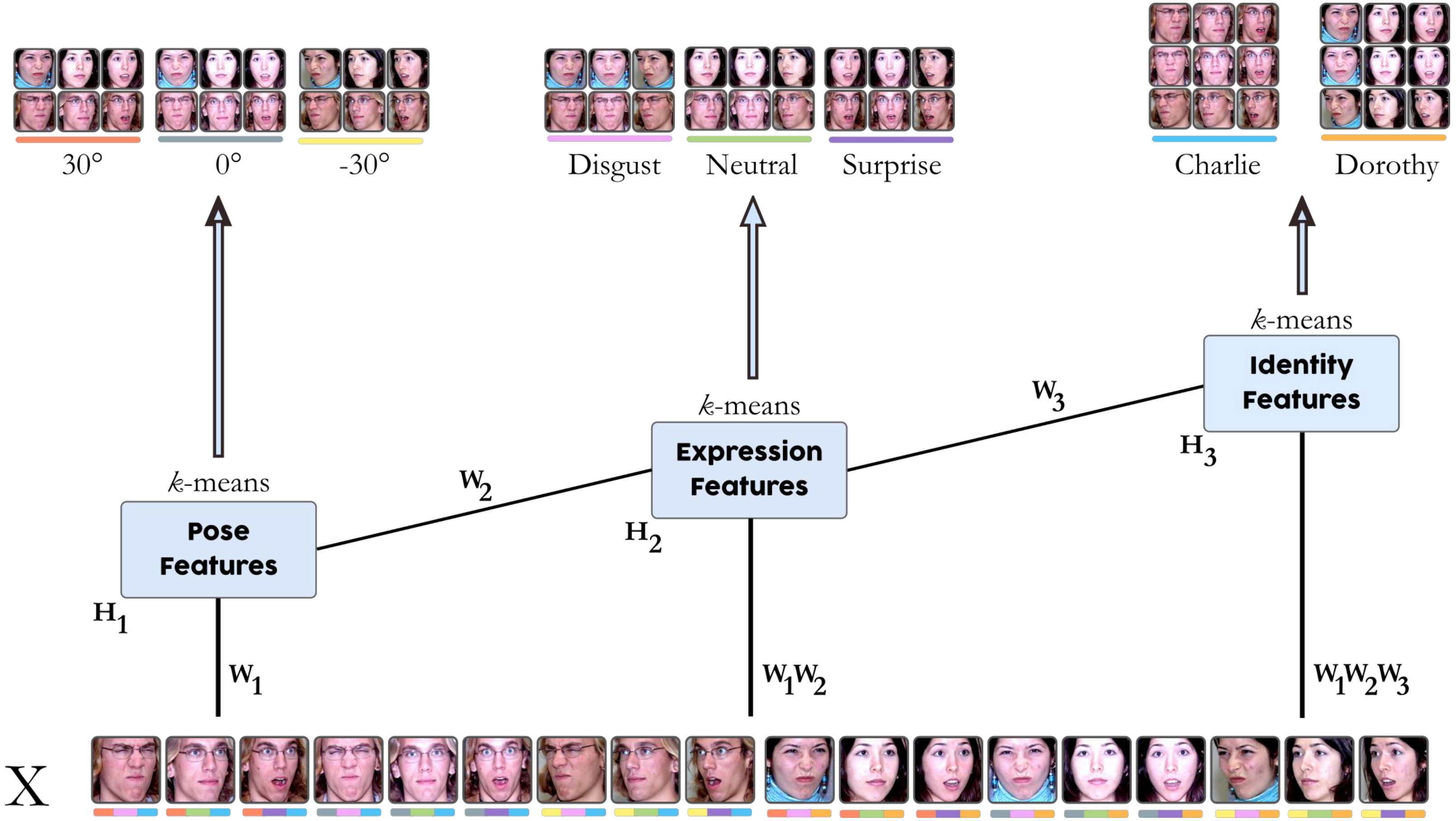
- From Multi-layer to Deep NMF
- Encoder-Decoder NMF
- Deep Autoencoder-like NMF

From Multi-layer to Deep NMF

$$
\begin{aligned}
X &\approx W_1 H_1 \\
H_1 &\approx W_2 H_2 \\
&\vdots \\
H_{p-1} &\approx W_p H_p \\
&\downarrow \\
X &\approx W_1 W_2 \ldots W_p H_p
\end{aligned}
$$

$$
\begin{aligned}
H_{p-1} &\approx W_p H_p \\
&\vdots \\
H_2 &\approx W_3 \ldots W_p H_p \\
H_1 &\approx W_2 \ldots W_p H_p
\end{aligned}
$$

A Deep Semi-NMF Model for Learning Hidden Representations
G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. Schuller
International conference on machine learning (ICML) 2014



$$\min_{\boldsymbol{W}_i, \boldsymbol{H}_p} \mathcal{L} = \|X - \boldsymbol{W}_1 \ldots \boldsymbol{W}_p \boldsymbol{H}_p\|_F^2, \text{ s.t. } \boldsymbol{H}_p \geq 0, \forall i = 1, 2, \ldots, p.$$

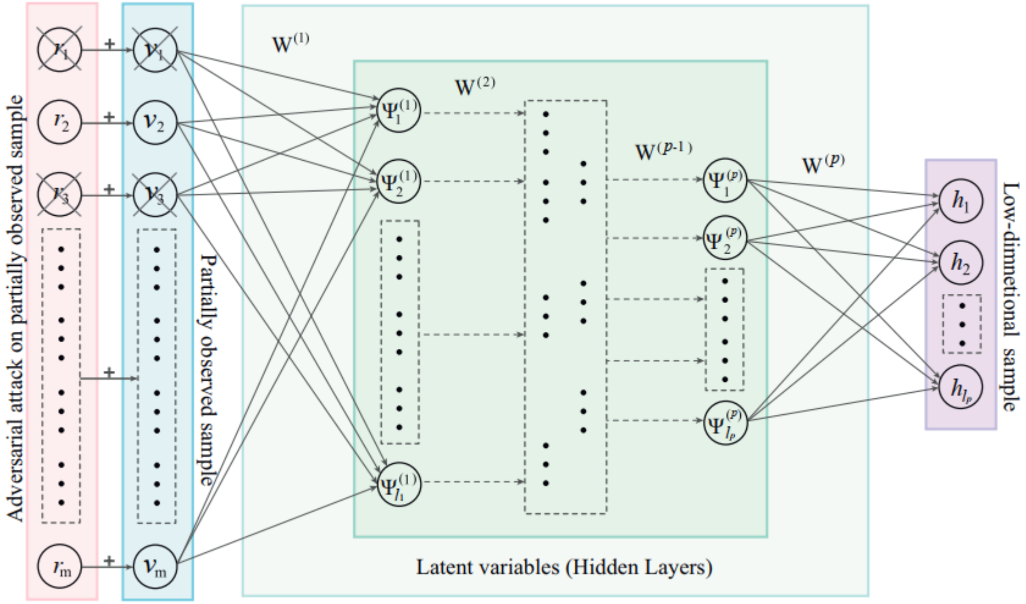
Robust Adversarial Matrix Completion

"Elastic Adversarial Deep Nonnegative Matrix Factorization for Matrix Completion"

S.A. Seyedi, F. Akhlaghian, A. Lotfi, N. Salahian, and J. Chavoshinejad

Information Sciences ,2023

$$\|\boldsymbol{V}-\boldsymbol{W}\boldsymbol{H}\|_{el}=\underbrace{\sum_i \frac{\delta\|\boldsymbol{v}^{(i)}-\boldsymbol{W}\boldsymbol{h}^{(i)}\|^2}{\delta+\|\boldsymbol{v}^{(i)}-\boldsymbol{W}\boldsymbol{h}^{(i)}\|}}_{\ell_2 \text{ pseudo-loss}}+\underbrace{\sum_i \frac{\|\boldsymbol{v}^{(i)}-\boldsymbol{W}\boldsymbol{h}^{(i)}\|^2}{\delta+\|\boldsymbol{v}^{(i)}-\boldsymbol{W}\boldsymbol{h}^{(i)}\|}}_{\ell_1 \text{ pseudo-loss}}$$



$$\max_{\boldsymbol{R}}\ \min_{\boldsymbol{W}^{(i)},\boldsymbol{H}_p} \|\boldsymbol{J}\odot(\boldsymbol{V}+\boldsymbol{R}-\boldsymbol{W}_1\ldots\boldsymbol{W}_p\boldsymbol{H}_p)\|_{el}-\lambda\|\boldsymbol{R}\|_F^2, \quad \text{s.t.} \quad \boldsymbol{V}+\boldsymbol{R}\geq 0, \boldsymbol{W}_i\geq 0, \boldsymbol{H}_p\geq 0, \forall i=1,\ldots,p.$$

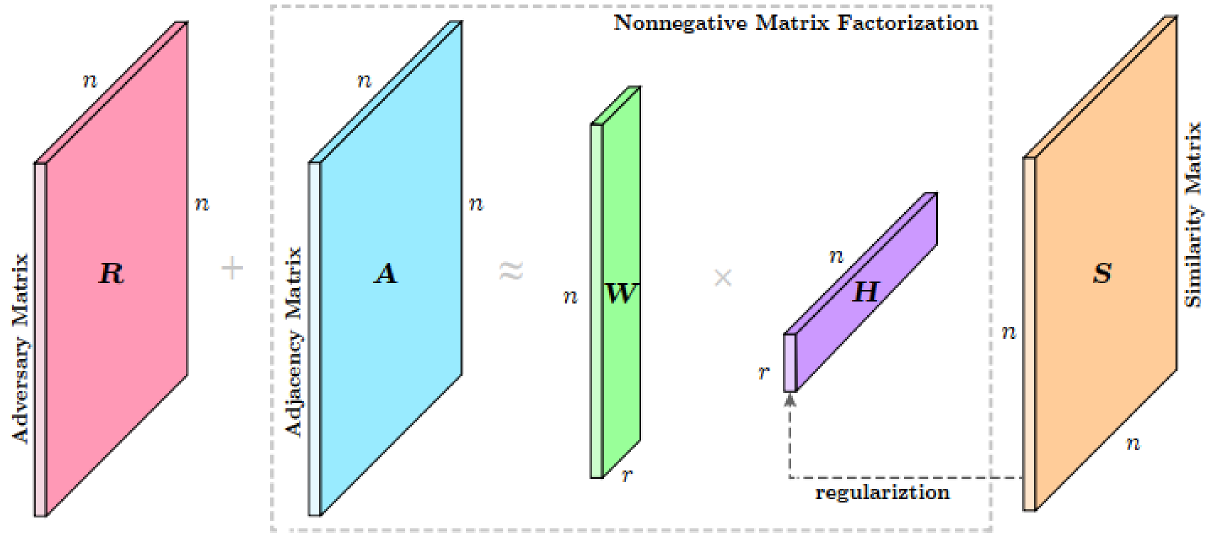
Adversarial Link Prediction

"Link Prediction by Adversarial Nonnegative Matrix Factorization"

R. Mahmoodi, **S.A. Seyedi**, F. Akhlaghian, and A. Abdolalhpouri

Knowledge-Based Systems, 2023

$$\min_{\boldsymbol{W},\boldsymbol{H}} \max_{\boldsymbol{R}} \|\boldsymbol{A}+\boldsymbol{R}-\boldsymbol{W}\boldsymbol{H}\|_F^2 - \lambda\|\boldsymbol{R}\|_F^2 + \alpha \mathrm{Tr}(\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^\top) + \beta(\|\boldsymbol{W}\|_F^2 + \|\boldsymbol{H}\|_F^2)$$



"Enhancing link prediction through adversarial training in deep Nonnegative Matrix Factorization"

R. Mahmoodi, **S.A. Seyedi**, A. Abdolalhpouri, and F. Akhlaghian

Engineering Application in Artificial Intelligence, 2024

$$\min_{\boldsymbol{W},\boldsymbol{H}} \max_{\boldsymbol{R}\in\mathcal{R}} \|\boldsymbol{A}_S+\boldsymbol{R}-\boldsymbol{W}_1\cdots\boldsymbol{W}_l\boldsymbol{H}_l\|_F^2 - \lambda\|\boldsymbol{R}\|_{2,1} + \beta(\|\boldsymbol{W}_1\ldots\boldsymbol{W}_l\|_F^2 + \|\boldsymbol{H}_l^\top\|_F^2) \quad \text{s.t.} \quad \boldsymbol{W}_i \geq 0, \boldsymbol{H}_i \geq 0$$

Encoder-Decoder NMF

A Non-negative Symmetric Encoder-Decoder Approach for Community Detection

The Conference on Information and Knowledge Management (CIKM) 2017

Decoder

$$X \approx WH$$

Encoder-Decoder NMF

A Non-negative Symmetric Encoder-Decoder Approach for Community Detection
The Conference on Information and Knowledge Management (CIKM) 2017

Decoder

$$X \approx WH$$

Encoder

$$H \approx W^\top X$$

Encoder-Decoder NMF

A Non-negative Symmetric Encoder-Decoder Approach for Community Detection
The Conference on Information and Knowledge Management (CIKM) 2017

Decoder

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}$$

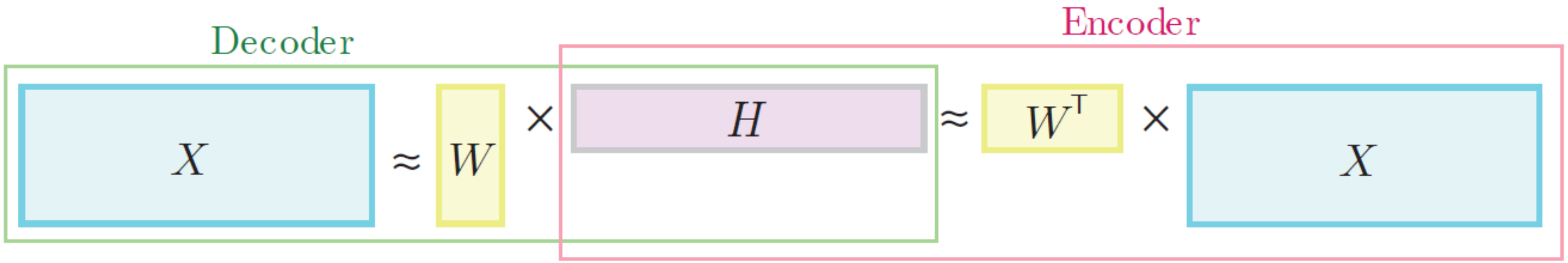
Encoder

$$\boldsymbol{H} \approx \boldsymbol{W}^\top \boldsymbol{X}$$

Projective NMF: $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{X}$

$$\boldsymbol{S}_{d\times d} = \boldsymbol{W}\boldsymbol{W}^\top$$

$$\boldsymbol{X} \approx \boldsymbol{S}\boldsymbol{X}$$



A Non-negative Symmetric Encoder-Decoder Approach for Community Detection
The Conference on Information and Knowledge Management (CIKM) 2017

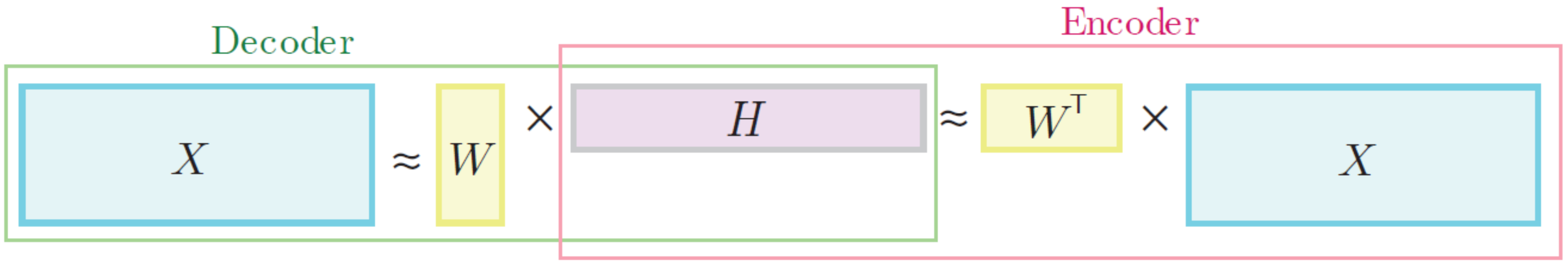
Decoder

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}$$

Encoder

$$\boldsymbol{H} \approx \boldsymbol{W}^\top \boldsymbol{X}$$

Projective NMF: $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{W}^\top \boldsymbol{X}$

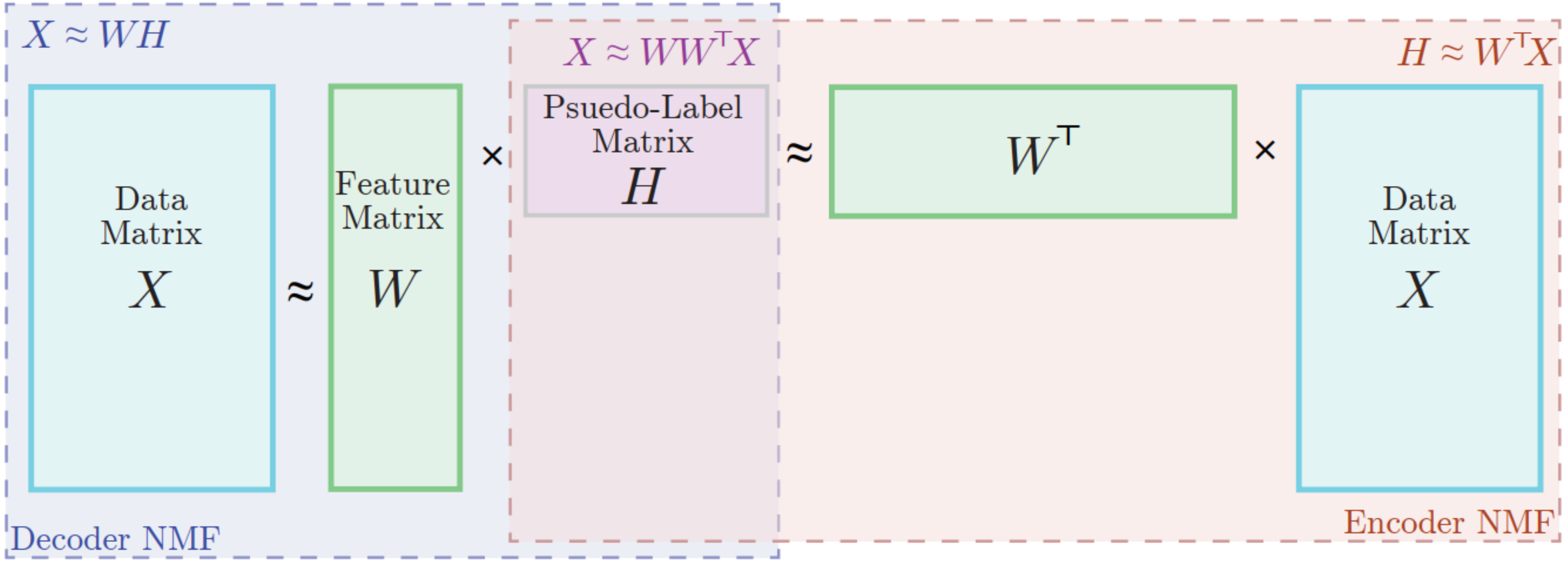


$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2 + \|\boldsymbol{H} - \boldsymbol{W}\boldsymbol{X}\|_F^2, \quad s.t. \quad (\boldsymbol{W}, \boldsymbol{H}) \geq 0.$$

"Encoder-Decoder Factorization for Unsupervised Feature Selection"

M. Mozafari, **S.A. Seyedi**, F. Akhlaghian, and A. Pirmohammaiani

Information Sciences, 2024



$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2 + \|\boldsymbol{H} - \boldsymbol{W}^\top \boldsymbol{X}\|_F^2 + \lambda \mathrm{Tr}(\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^\top) + \gamma \|\boldsymbol{W}\|_{2,1}, \quad \text{s.t.} \quad \boldsymbol{W},\boldsymbol{H} \geq 0, \boldsymbol{H}\boldsymbol{H}^\top = \boldsymbol{I},$$

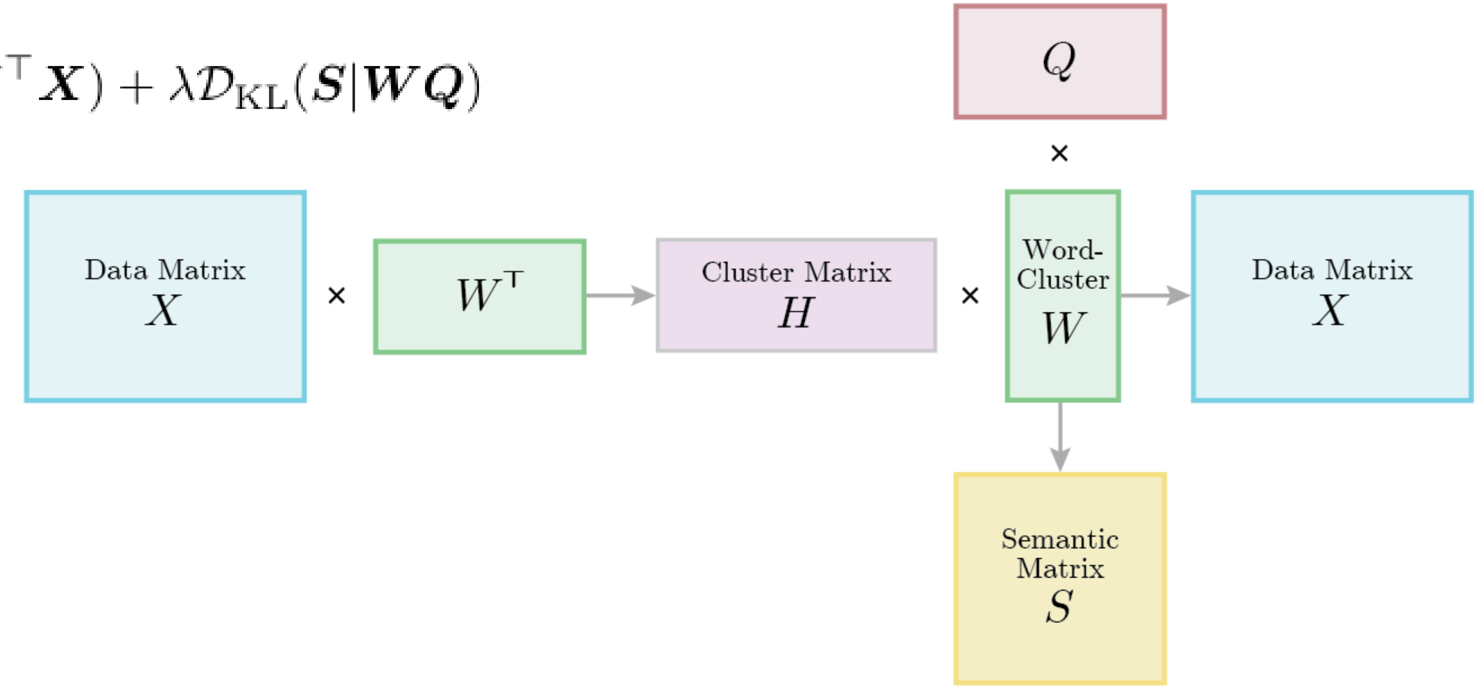
Encoder-Decoder NMF with B-Divergence

"Semantic Encoder-Decoder Nonnegative Matrix Factorization with Kullback-Leibler Divergence"

S. Soleymanbaigi, **S.A. Seyedi**,,F. Akhlaghian, and F. Daneshfar

[International Journal of Machine Learning and Cybernetics – First Revision]

$$\min_{\boldsymbol{W},\boldsymbol{H},\boldsymbol{Q}} \mathcal{D}_{\mathrm{KL}}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) + \mathcal{D}_{\mathrm{KL}}(\boldsymbol{H}|\boldsymbol{W}^\top\boldsymbol{X}) + \lambda\mathcal{D}_{\mathrm{KL}}(\boldsymbol{S}|\boldsymbol{W}\boldsymbol{Q})$$



"Data Clustering by Encoder-Decoder Nonnegative Matrix Factorization with β-Divergence"

S. Soleymanbaigi, **S.A. Seyedi**,,F. Akhlaghian, and F. Daneshfar

[Pattern Recognition – First Revision]

$$\mathcal{L}(\boldsymbol{W},\boldsymbol{H}) = D_\beta(\boldsymbol{X},\boldsymbol{W}\boldsymbol{H}) + \lambda D_\beta(\boldsymbol{H},\boldsymbol{W}^\top\boldsymbol{X}) + \gamma\mathrm{Tr}(\boldsymbol{H}\boldsymbol{L}\boldsymbol{H}^\top)$$

Decoder

$$X \approx W_1 H_1$$

$$H_1 \approx W_2 H_2$$

$$\vdots$$

$$H_{p-1} \approx W_p H_p$$

$$\downarrow$$

$$X \approx W_1 W_2 \ldots W_p H_p$$

Encoder

$$H_1 \approx W_1^\top X$$

$$H_2 \approx W_2^\top H_1$$

$$\vdots$$

$$H_p \approx W_p^\top H_{p-1}$$

$$\downarrow$$

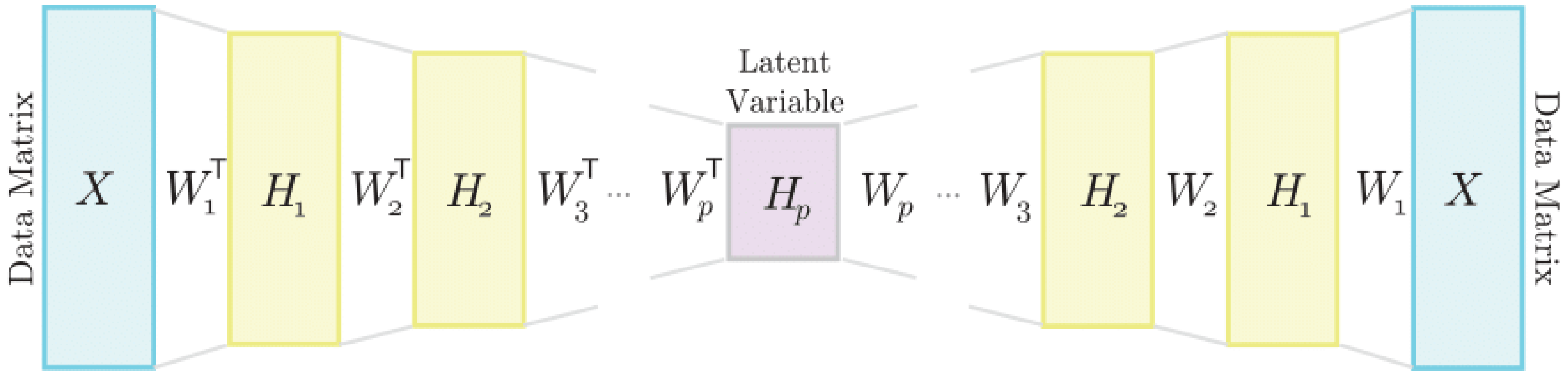
$$H_p \approx W_p^\top \ldots W_2^\top W_1^\top X,$$

Deep Encoder-Decoder NMF

Deep Autoencoder-like Nonnegative Matrix Factorization for Community Detection
The Conference on Information and Knowledge Management (CIKM) 2018

$$\min_{\boldsymbol{W}_l, \boldsymbol{H}_l} \|\boldsymbol{X} - \boldsymbol{W}_1 \boldsymbol{W}_2 \ldots \boldsymbol{W}_p \boldsymbol{H}_p\|_F^2 + \|\boldsymbol{H}_p - \boldsymbol{W}_p^\top \ldots \boldsymbol{W}_2^\top \boldsymbol{W}_1^\top \boldsymbol{X}\|_F^2$$

$$s.t. \quad \{\boldsymbol{W}_l, \boldsymbol{H}_l\} \geq 0, \forall l = 1, 2, \ldots, p.$$



Extended models:

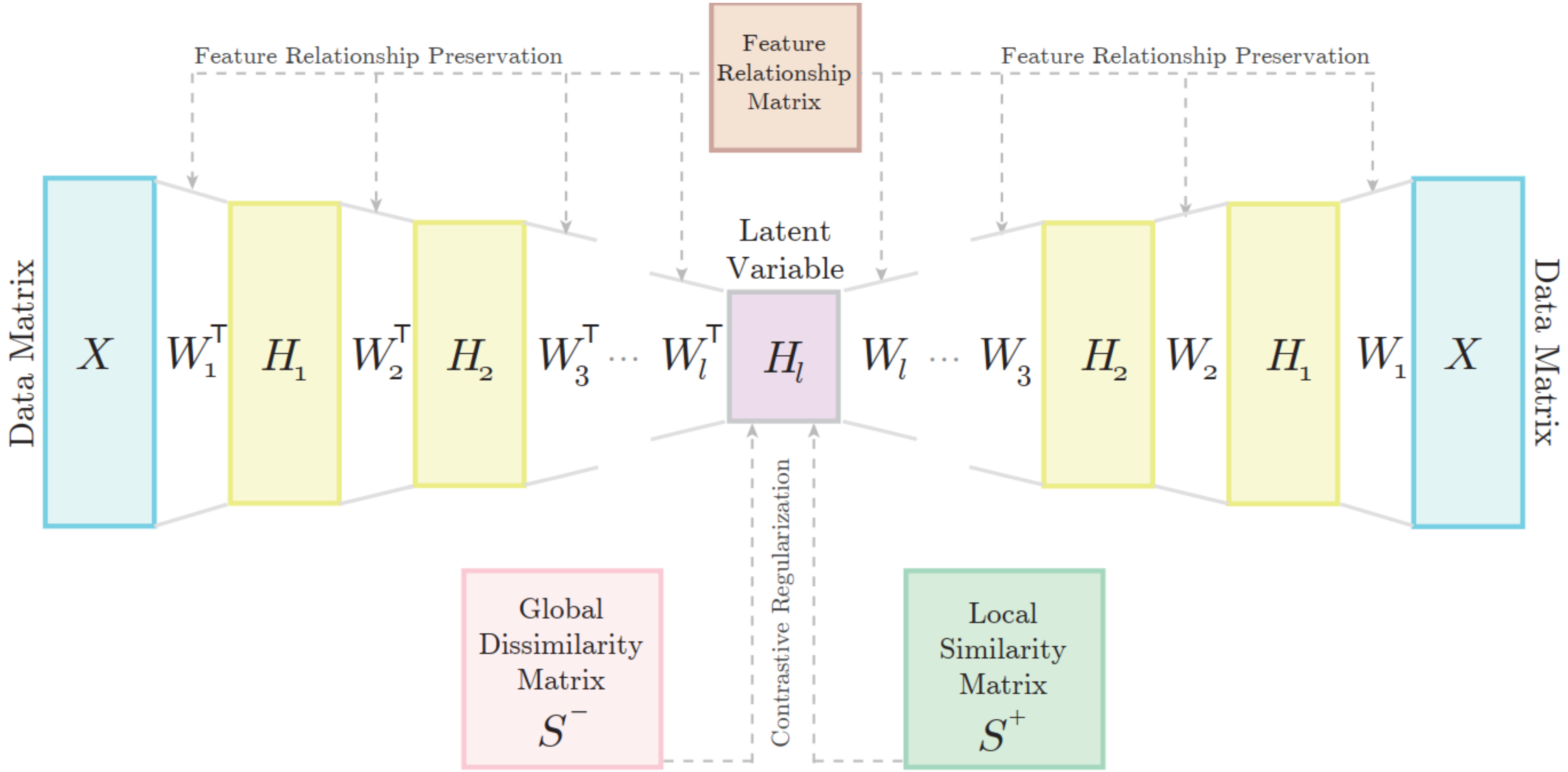
- Structural Deep Autoencoder-like Nonnegative Matrix Factorization for Community Detection, Applied Soft Computing 2020
- Community detection in networks through a deep robust auto-encoder nonnegative matrix factorization, Engineering Applications of Artificial Intelligence 2023

Data Representation / Clustering

"Deep Autoencoder-Like NMF with Contrastive Regularization and Feature Relationship Preservation"

N. Salahian, F. Akhlaghian, **S.A. Seyedi**, and J. Chavoshinejad

Expert Systems with Applications, 2023

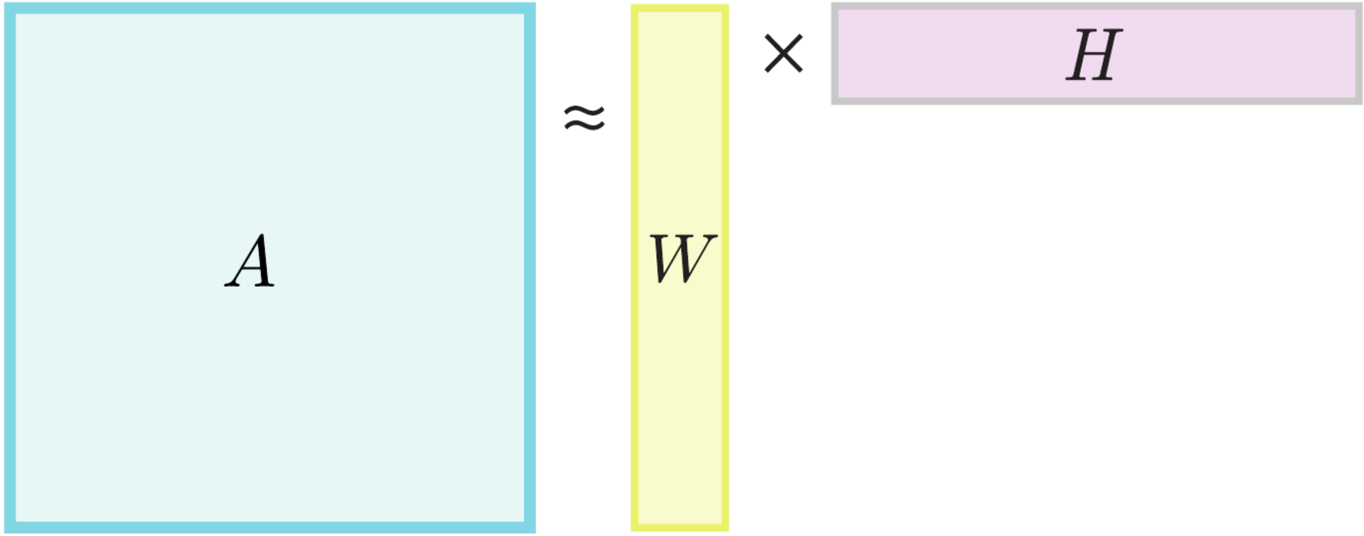


$$\min_{W_i, H_l} \mathcal{L} = \|X - W_1 \ldots W_l H_l\|_F^2 + \|H_l - W_l^\top \ldots W_1^\top X\|_F^2 + \lambda_1 \|S^- \odot H_l H_l^\top\|_1 + \lambda_2 \|S^+ \odot \tilde{H}_l\|_1 + \|XX^\top - \lambda_3 W_1 \ldots W_l W_l^\top \ldots W_1^\top\|_F^2$$

Graph-based NMFs

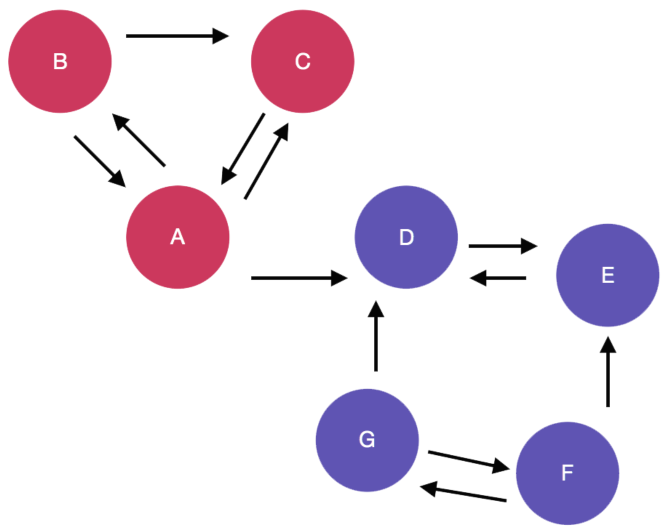
- Graph clustering
- Symmetric NMF
- Asymmetric NMF
- Interpretability of Asymmetric NMF
- Regularized Asymmetric NMFs

$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\boldsymbol{A} - \boldsymbol{W}\boldsymbol{H}\|_F^2, \quad s.t. \quad (\boldsymbol{W}, \boldsymbol{H}) \geq 0.$$



Graph clustering

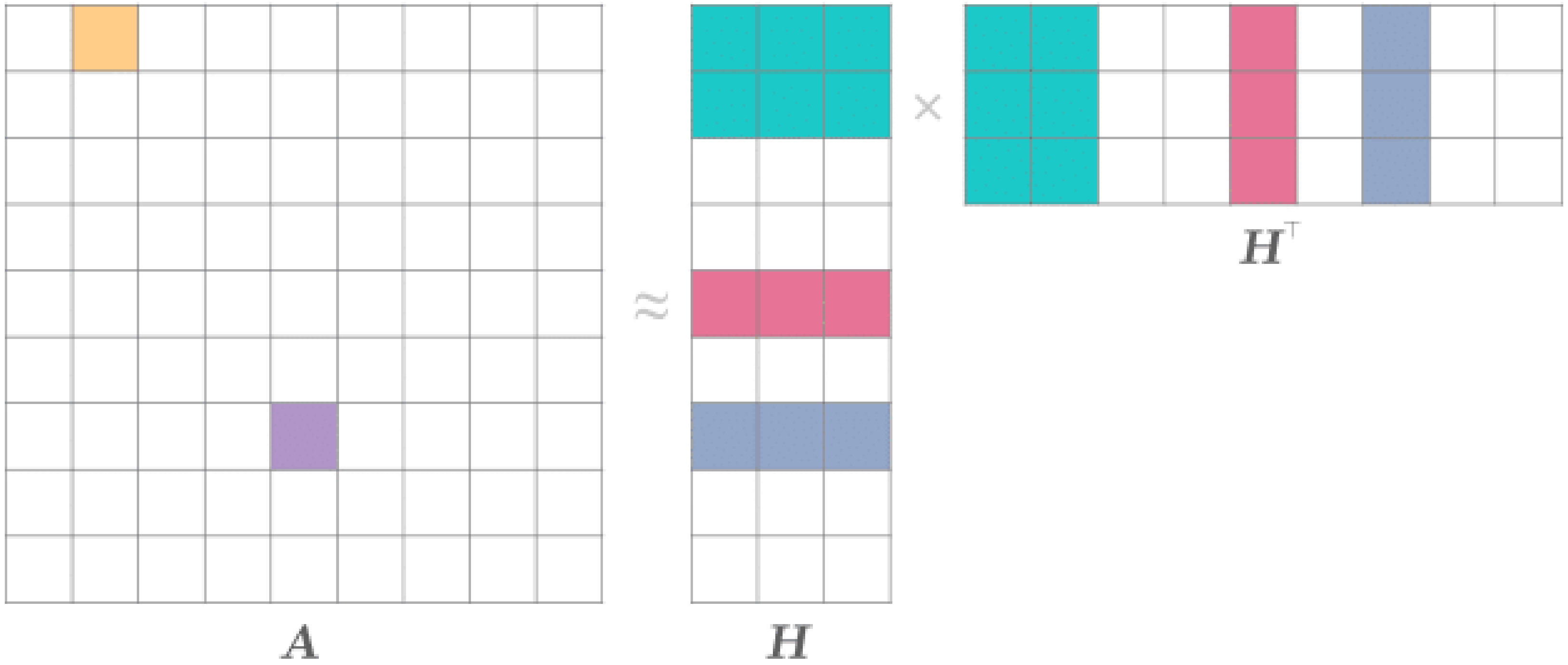
- **Definition:** Partitioning graph nodes into clusters with dense intra-cluster connections and sparse inter-cluster connections.
- **Purpose:** Identifying and grouping similar nodes to reveal the graph's structure.
- **Applications:** Used in social networks, biology, recommendation systems, and image segmentation.



Community discovery using nonnegative matrix factorization

Data Mining and Knowledge Discovery 2011

$$\min_{\boldsymbol{H}} \|\boldsymbol{A} - \boldsymbol{H}\boldsymbol{H}^\top\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(A_{ij} - \boldsymbol{h}^{(i)}\boldsymbol{h}^{(j)\top}\right)^2, \quad s.t. \quad (\boldsymbol{H}) \geq 0.$$

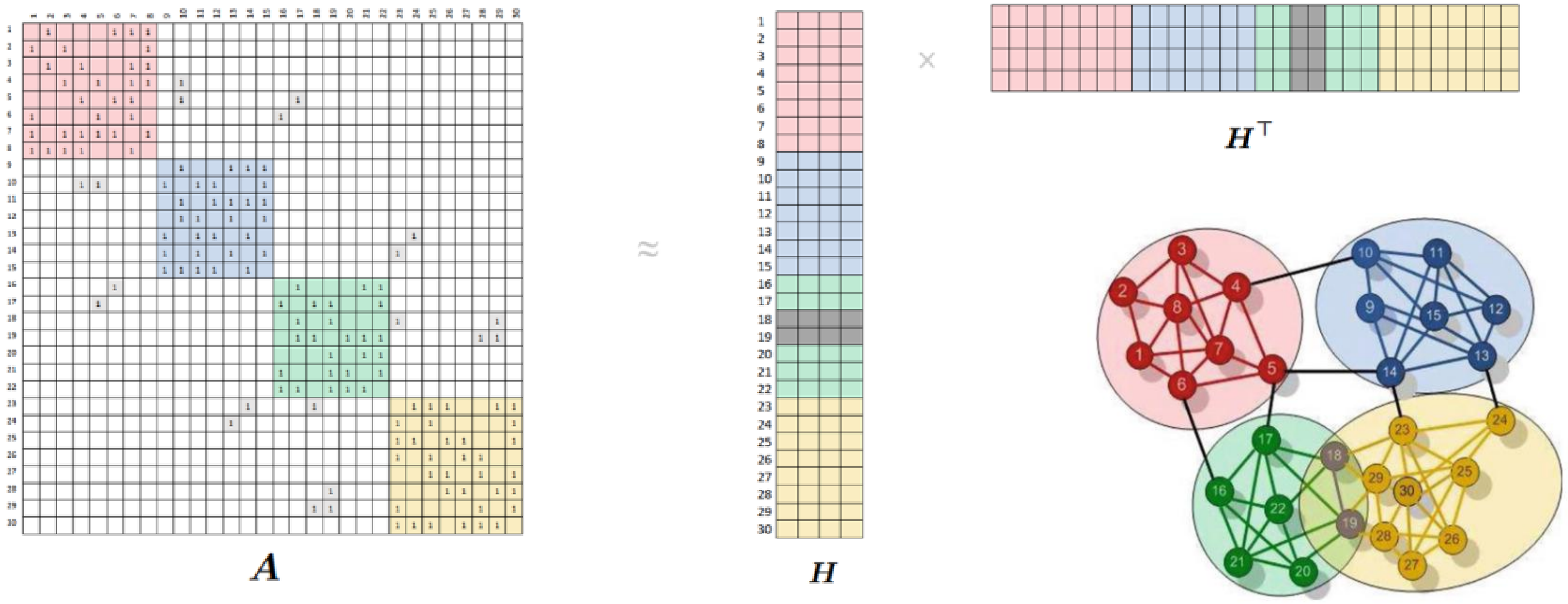


Symmetric NMF

Community discovery using nonnegative matrix factorization

Data Mining and Knowledge Discovery 2011

$$\min_{H} \| A - HH^\top \|_F^2$$

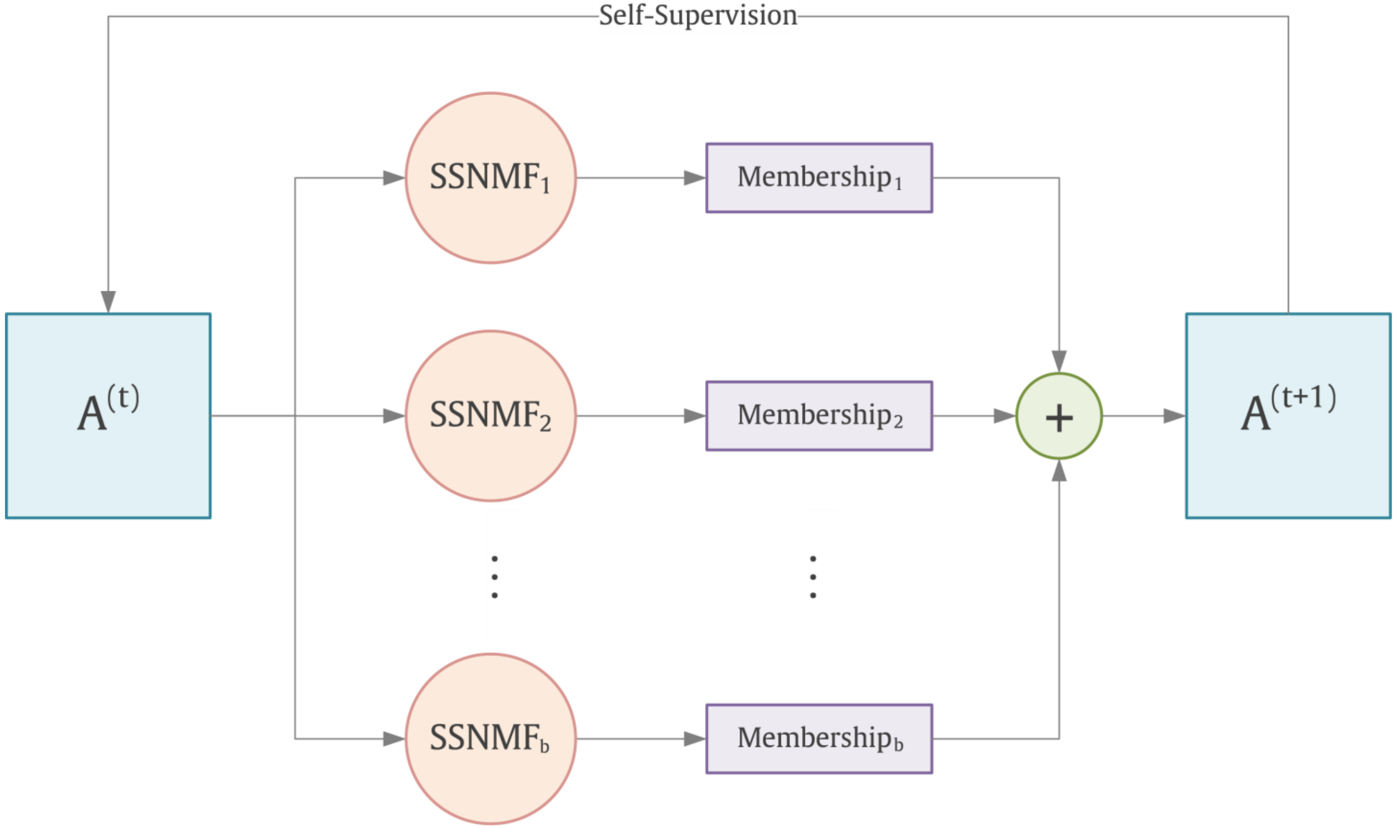


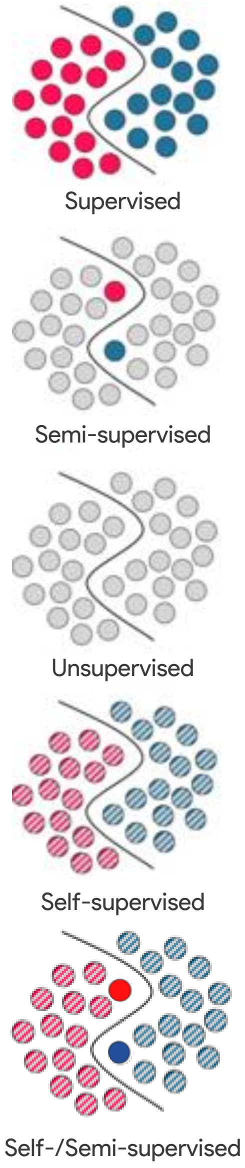
Self-supervised / Semi-supervised Learning

"S^4NMF: Self-Supervised Semi-Supervised Nonnegative Matrix Factorization for Data Clustering"

J. Chavoshinejad, **S.A. Seyedi** , F. Akhlaghian, and N. Salahian

Pattern Recognition, 2023





$$\min_{\alpha, V_m} \sum_{m=1}^{b} \alpha_m^{\tau} \left(\|A - V_m V_m^{\top}\|_F^2 + \lambda_1 \|D \odot (V_m V_m^{\top})\|_1 + \lambda_2 \|S \odot \tilde{V}_m\|_1 \right) \quad \text{s.t.} \quad \alpha \mathbf{1} = 1, \ \alpha, V_m \geq 0 \ \forall m,$$

Asymmetric NMF or Symmetric NMTF

Community discovery using nonnegative matrix factorization

Data Mining and Knowledge Discovery 2011

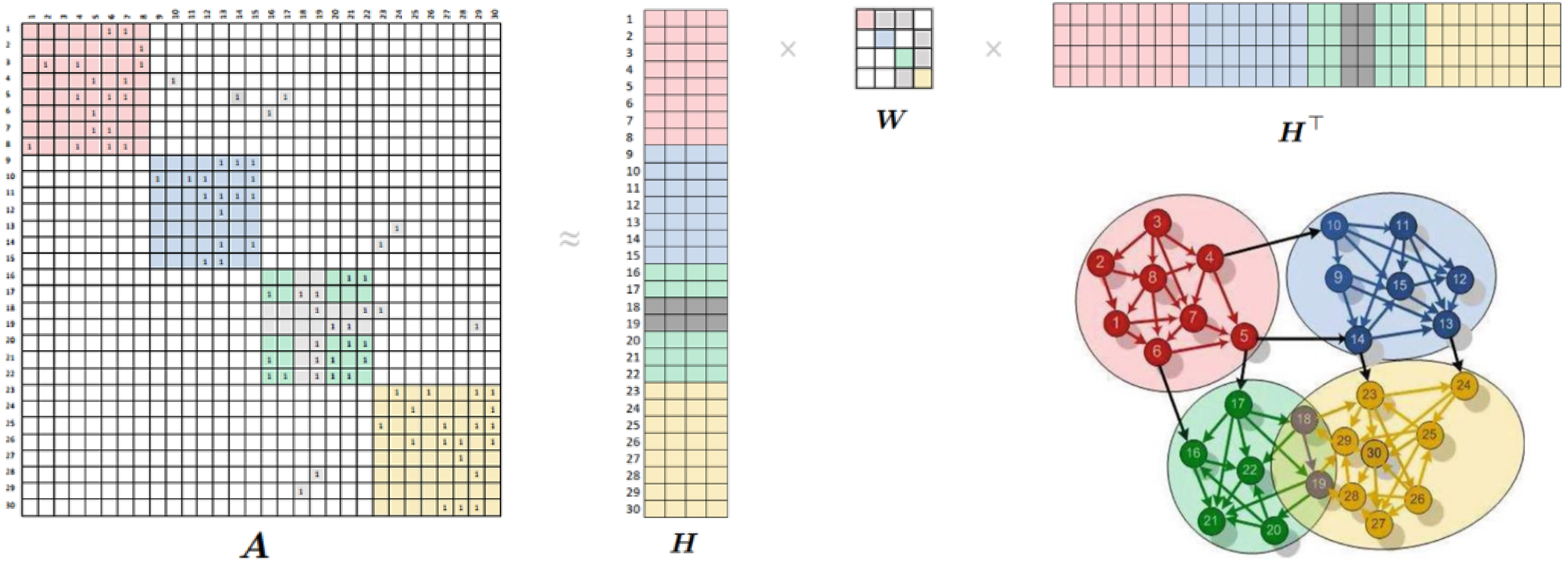
$$\min_{\boldsymbol{H}} \|\boldsymbol{A} - \boldsymbol{H}\boldsymbol{H}^\top\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(A_{ij} - \boldsymbol{h}^{(i)}\boldsymbol{h}^{(j)\top}\right)^2, \quad s.t. \quad (\boldsymbol{H}) \geq 0.$$

$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\boldsymbol{A} - \boldsymbol{H}\boldsymbol{W}\boldsymbol{H}^\top\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\left(A_{ij} - \boldsymbol{h}^{(i)}\boldsymbol{W}\boldsymbol{h}^{(j)\top}\right)^2, \quad s.t. \quad (\boldsymbol{W},\boldsymbol{H}) \geq 0.$$

Asymmetric NMF or Symmetric NMTF

Community discovery using nonnegative matrix factorization

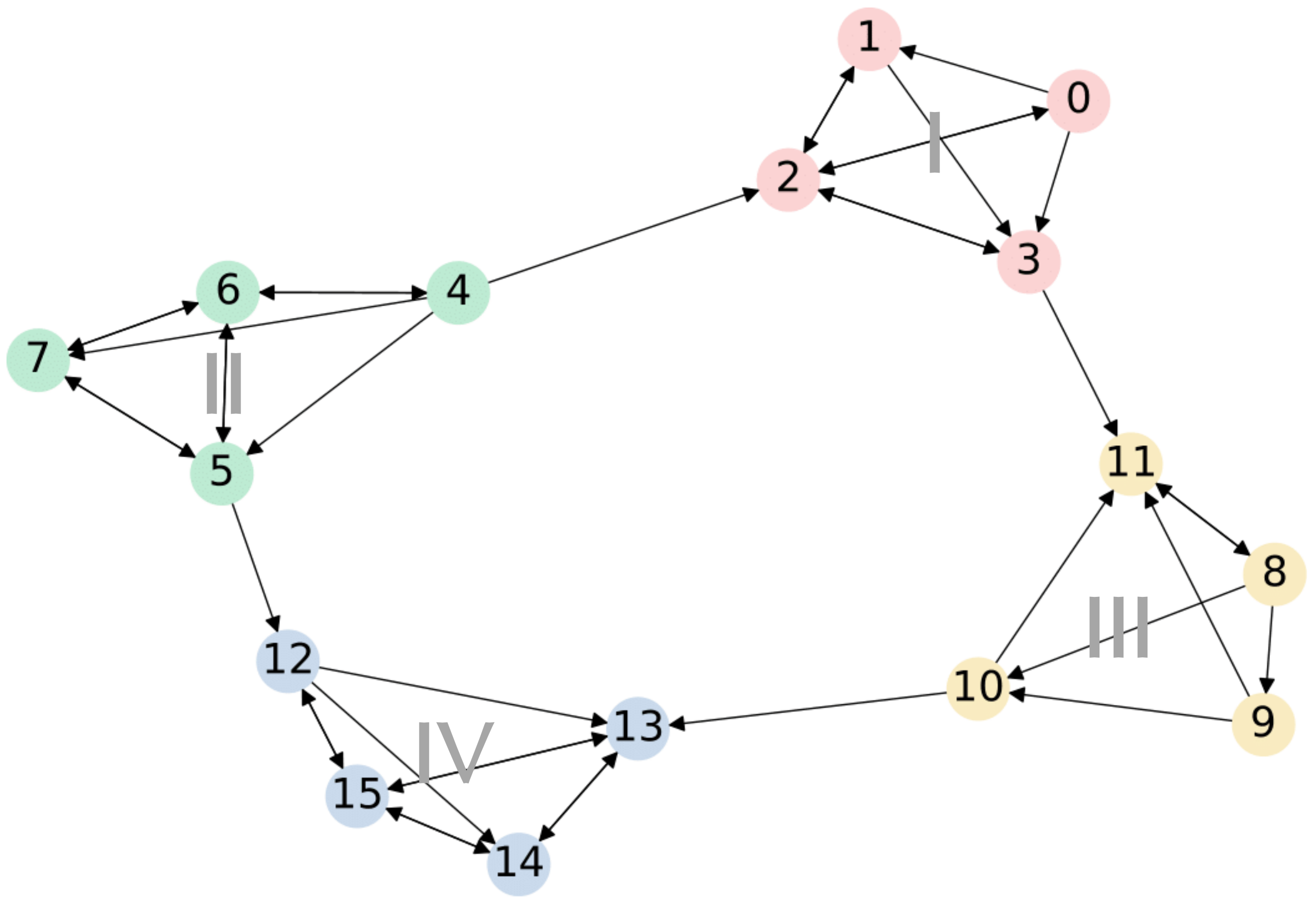
Data Mining and Knowledge Discovery 2011



Community discovery using nonnegative matrix factorization

Data Mining and Knowledge Discovery 2011

$$\min_{W,H} \|A - HWH^\top\|_F^2$$



	I	II	III	IV
0	3e-01	3e-06	2e-06	5e-25
1	3e-01	1e-06	2e-05	2e-25
2	4e-01	3e-02	6e-04	9e-13
3	3e-01	3e-07	8e-02	5e-29
4	6e-02	3e-01	7e-18	2e-11
5	8e-14	4e-01	8e-07	4e-02
6	2e-12	4e-01	3e-08	2e-07
7	7e-17	4e-01	8e-08	2e-06
8	2e-17	0e+00	4e-01	1e-08
9	2e-19	0e+00	3e-01	9e-07
10	3e-18	1e-19	3e-01	6e-02
11	3e-02	2e-50	4e-01	4e-07
12	1e-14	6e-02	2e-08	3e-01
13	3e-30	9e-12	5e-02	4e-01
14	5e-12	1e-11	2e-08	4e-01
15	1e-12	4e-07	2e-08	4e-01

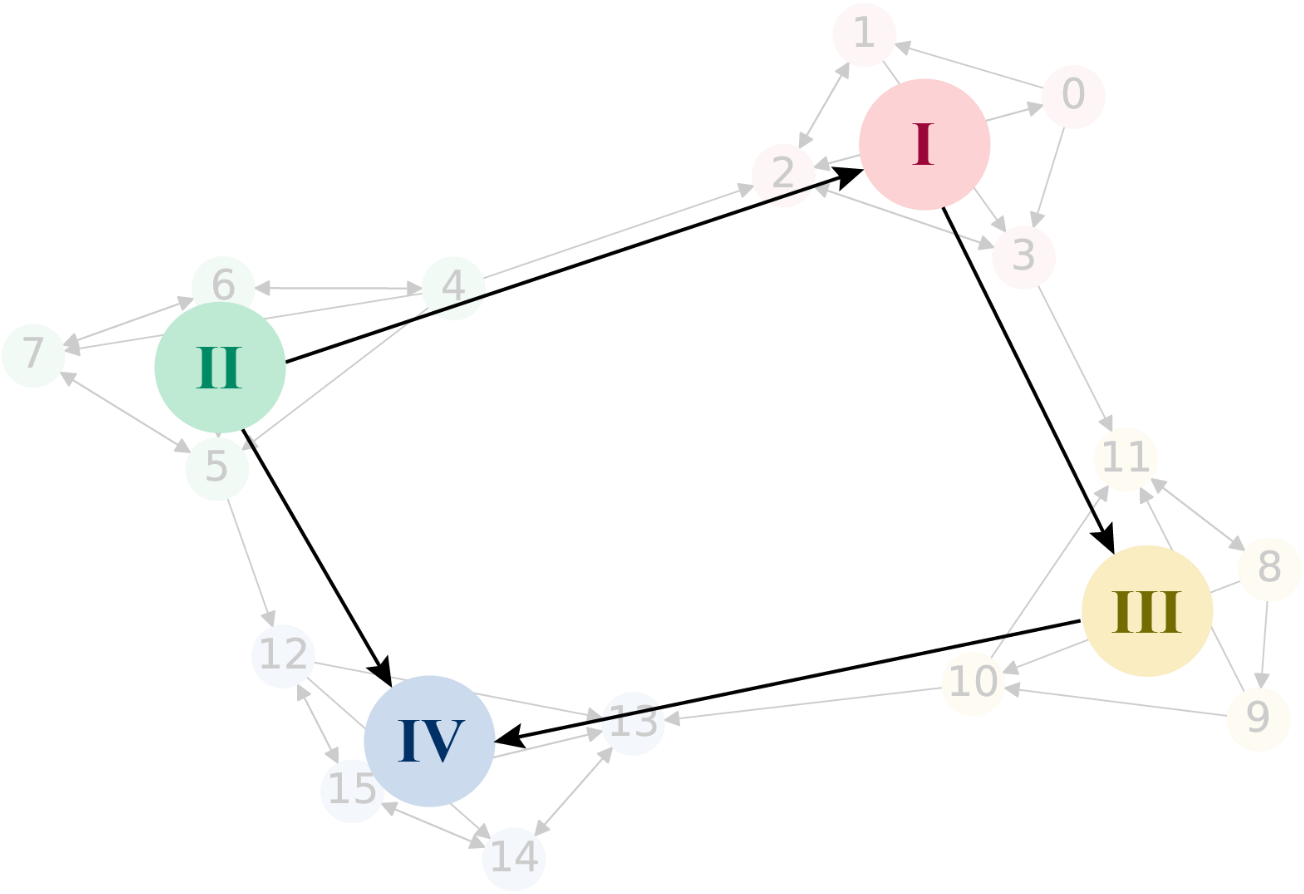
$\boldsymbol{H}$

Interpretability of Asymmetric NMF

Community discovery using nonnegative matrix factorization

Data Mining and Knowledge Discovery 2011

$$\min_{W,H} \| A - HWH^\top \|_F^2$$



	I	II	III	IV
I	5.51	6e-24	2e-01	5e-74
II	9e-02	4.17	6e-38	1e-01
III	1e-31	2e-51	2.92	1e-01
IV	2e-70	6e-44	1e-31	4.10

$\boldsymbol{W}$

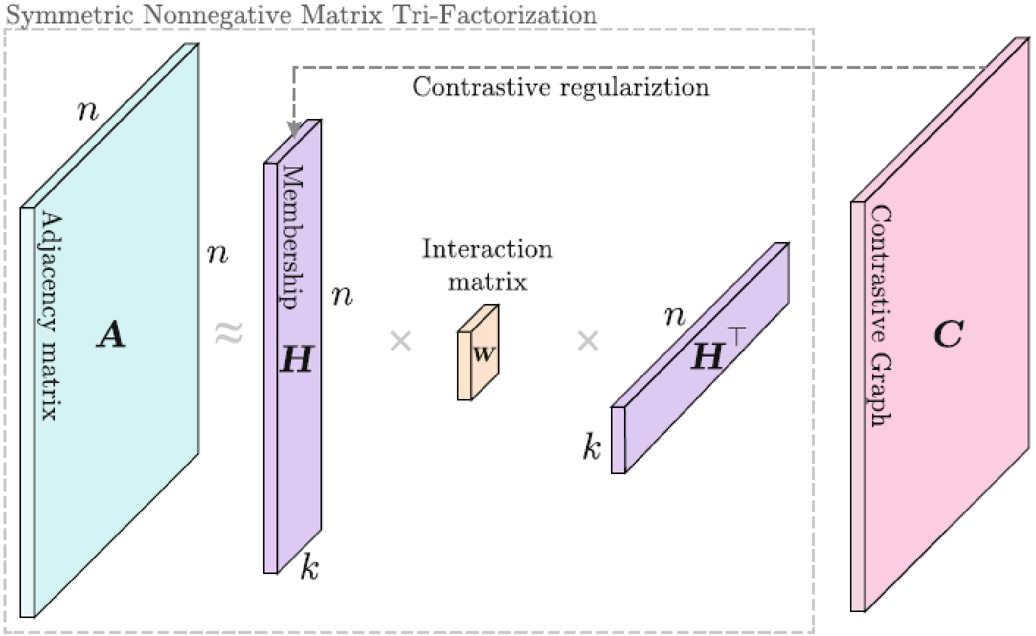
Regularized Asymmetric NMF

Regularized Asymmetric Semi-nonnegative Matrix Factorization for Directed Graph Clustering
R. Abdollahi, A. Seyedi, M. Noorimehr
ICCKE 2020

$$\min_{\boldsymbol{W},\boldsymbol{H}} \|\boldsymbol{A}_{\pm} - \boldsymbol{H}\boldsymbol{W}_{\pm}\boldsymbol{H}^\top\|_F^2 + \lambda \mathrm{Tr}(\boldsymbol{H}^\top \boldsymbol{L}'\boldsymbol{H}), \quad \text{s.t. } \boldsymbol{H} \geq 0$$

Towards Cohesion-Fairness Harmony - Contrastive Regularization in Individual Fair Graph Clustering
Ghodsi, A. Seyedi , and E. Ntoutsi
Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2024

$$\min_{\boldsymbol{H},\boldsymbol{W}\geq 0} \|\boldsymbol{A} - \boldsymbol{H}\boldsymbol{W}\boldsymbol{H}^\top\|_F^2 + \lambda \mathrm{Tr}(\boldsymbol{H}^\top \boldsymbol{L}\boldsymbol{H}),$$



Deep Asymmetric NMF for Directed/Undirected graph clustering

- Deep Symmetric NMF
- Multi-layer Asymmetric NMF
- Deep Asymmetric NMF
- Interpretability of Deep Asymmetric NMF

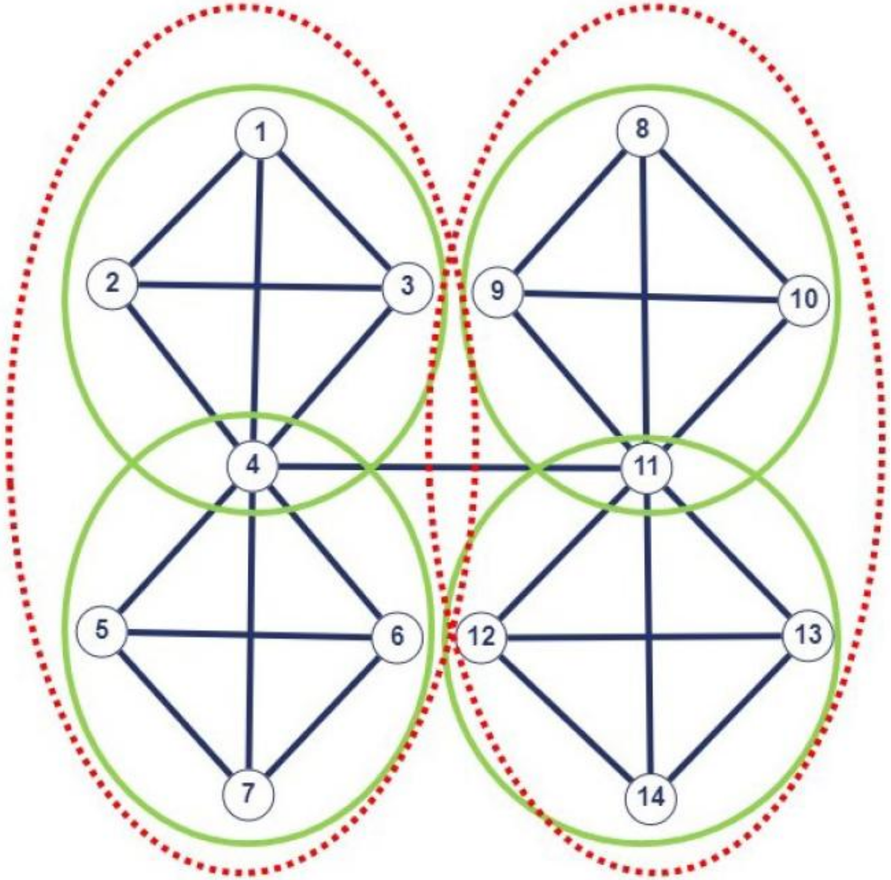
Deep Symmetric Matrix Factorization

P. De Handschutter, N. Gillis, and W. Blekic

European Signal Processing Conference (EUSIPCO) 2023

$$\min_{\substack{W \in \mathbb{R}_+^{n \times r} \\ H \in \mathbb{R}_+^{r \times n}}} \|X - WH\|_F^2 + \mu \|W - H^T\|_F^2.$$

$$\mathcal{L}_{DSNMF} = \frac{1}{2}(\|X - W_1H_1\|_F^2 + \mu_1\|W_1 - H_1^T\|_F^2 + \lambda_1 \\ (\|W_1 - W_2H_2\|_F^2 + \mu_2\|W_2 - (H_2H_1)^T\|_F^2) + \cdots + \lambda_{L-1} \\ (\|W_{L-1} - W_LH_L\|_F^2 + \mu_L\|W_L - (H_LH_{L-1}\cdots H_1)^T\|_F^2)).$$



Pattern Recognition 148 (2024) 110179

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Deep asymmetric nonnegative matrix factorization for graph clustering

Akram Hajiveiseh, Seyed Amjad Seyedi, Fardin Akhlaghian Tab *

Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran

ARTICLE INFO

Keywords:
Nonnegative matrix factorization
Deep learning
Graph clustering
Directed graph

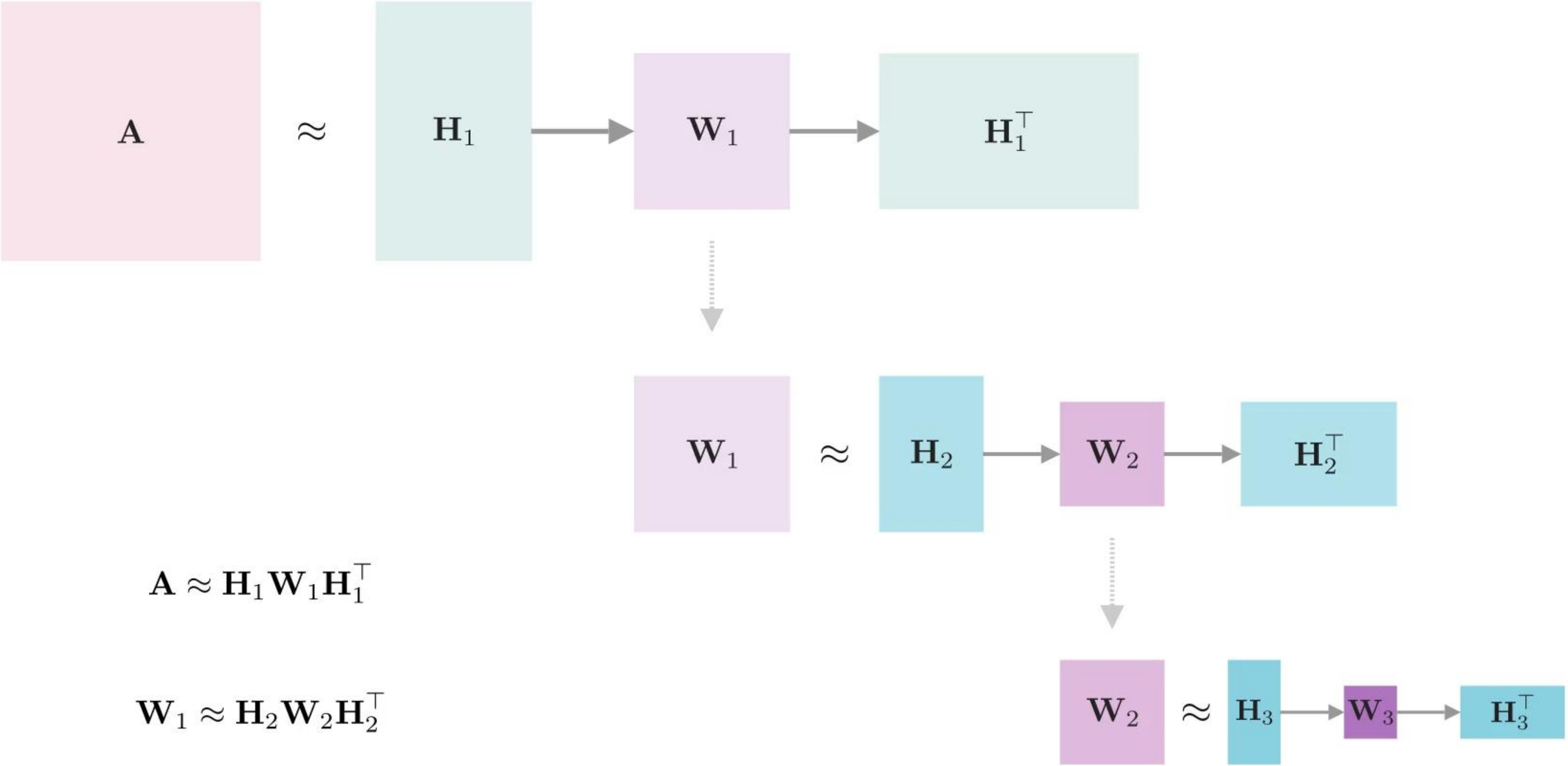
ABSTRACT

Graph clustering is a fundamental technique in machine learning that has widespread applications in various fields. Deep Nonnegative Matrix Factorization (DNMF) was recently emerged to cope with the extraction of several layers of features, and it has been demonstrated to achieve remarkable results on unsupervised tasks. While DNMF has been applied for analyzing graphs, the effectiveness of the current DNMF approaches for graph clustering is generally unsatisfactory: these methods are intrinsically data representation models, and their objective functions do not capture cluster structures, also ignores direction which is crucial in the directed graph clustering problems. To overcome these downsides, this paper proposes a graph-specific DNMF model based on the Asymmetric NMF which can handle undirected and directed graphs. Inspired by

* Corresponding author.
E-mail addresses: akram.hajiveiseh@uok.ac.ir (A. Hajiveiseh), amjadseyedi@uok.ac.ir (S.A. Seyedi), f.akhlaghian@uok.ac.ir (F. Akhlaghian Tab).

https://doi.org/10.1016/j.patcog.2023.110179
Received 19 March 2023; Received in revised form 29 September 2023; Accepted 29 November 2023
Available online 2 December 2023

Multi-layer Asymmetric NMF

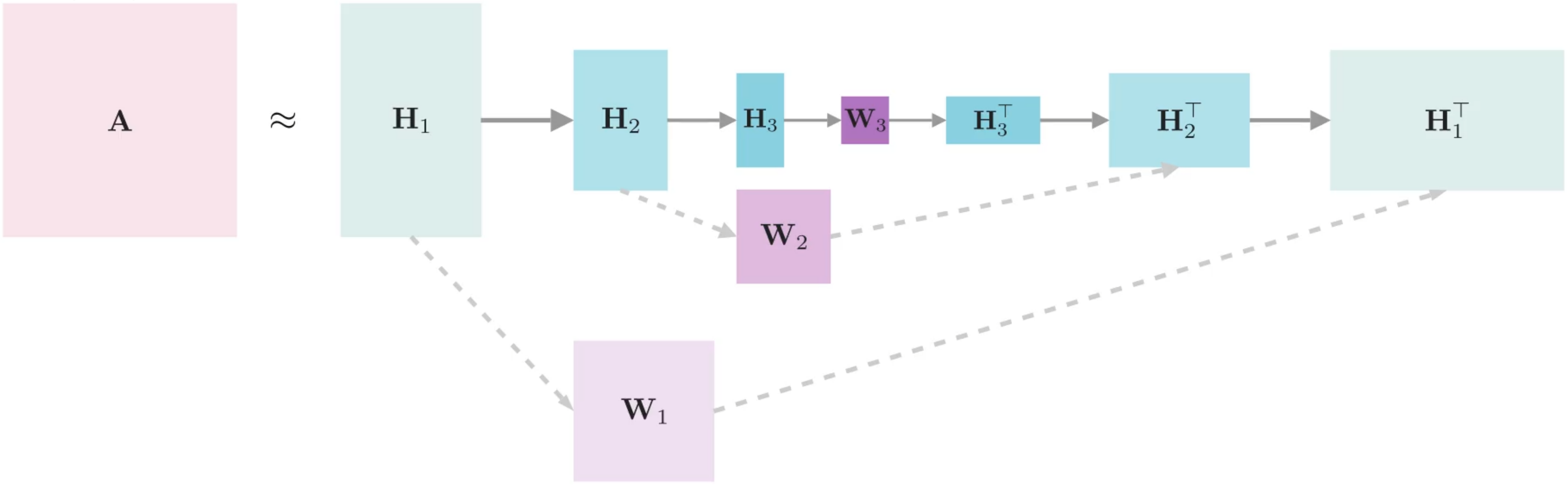


$$\mathbf{A} \approx \mathbf{H}_1 \mathbf{W}_1 \mathbf{H}_1^\top$$

$$\mathbf{W}_1 \approx \mathbf{H}_2 \mathbf{W}_2 \mathbf{H}_2^\top$$

$$\mathbf{W}_2 \approx \mathbf{H}_3 \mathbf{W}_3 \mathbf{H}_3^\top$$

Deep Asymmetric NMF

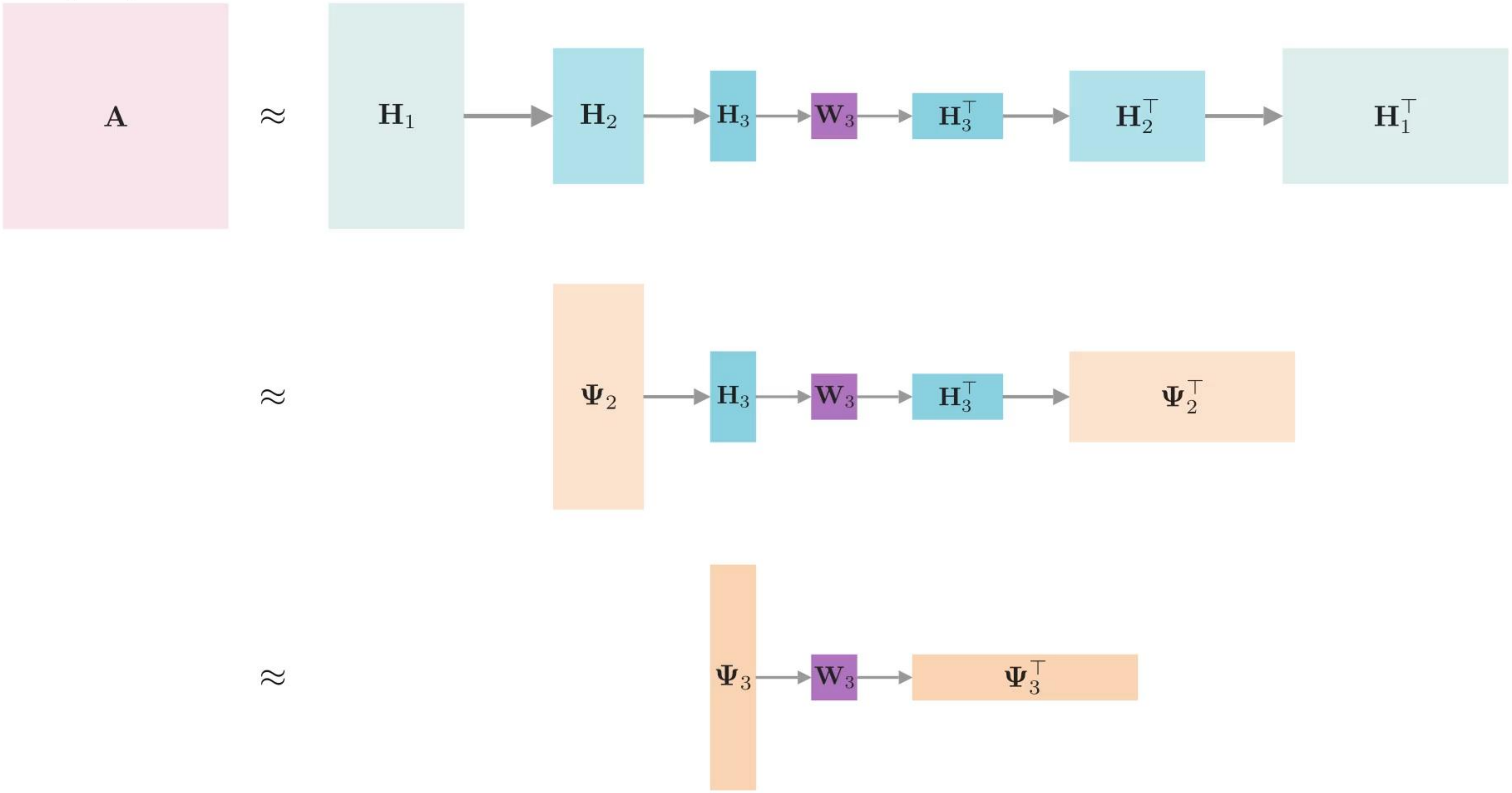


$$\mathbf{A} \approx \mathbf{H}_1\mathbf{H}_2\mathbf{H}_3\mathbf{W}_3\mathbf{H}_3^\top\mathbf{H}_2^\top\mathbf{H}_1^\top$$

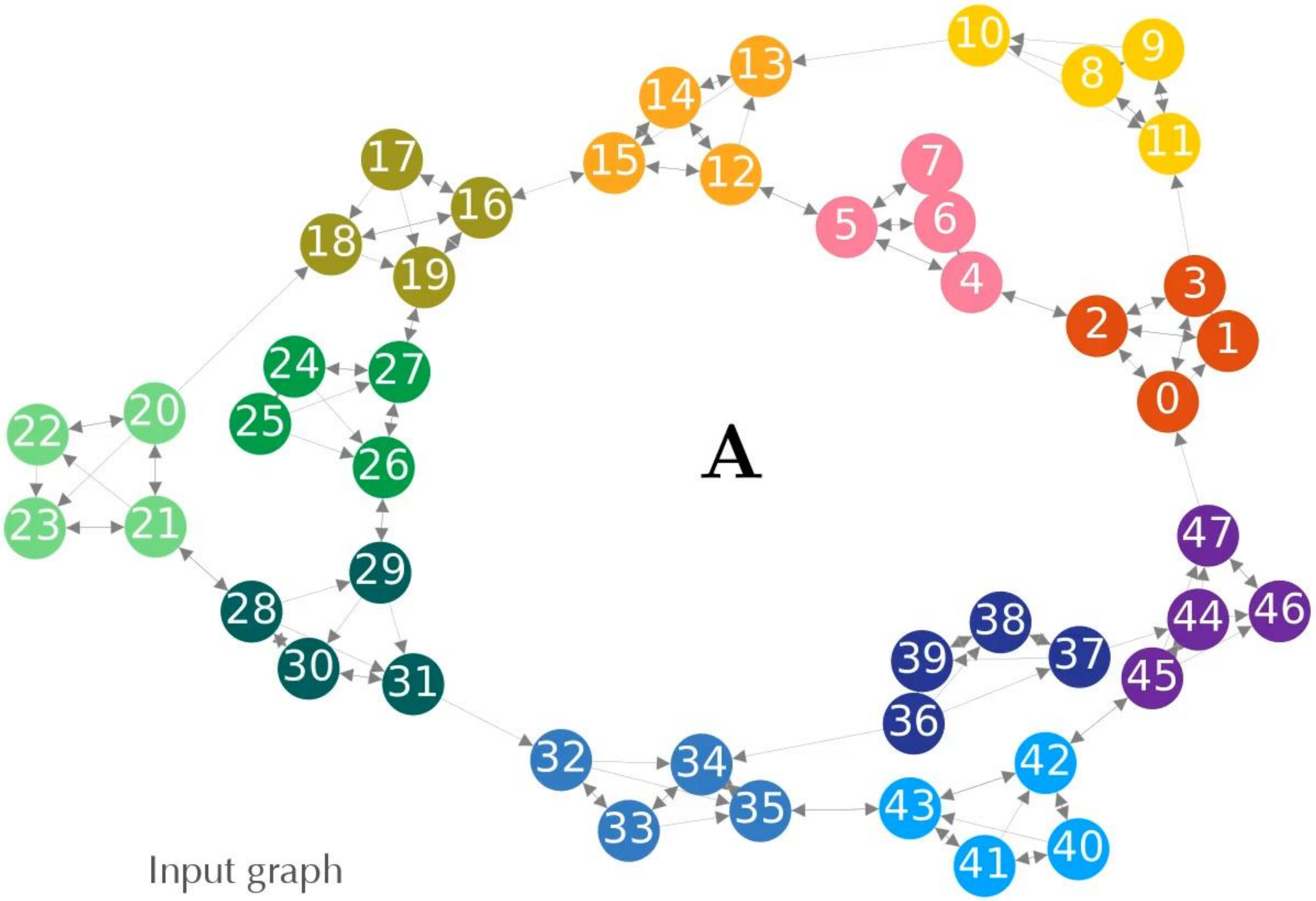
$$\mathbf{W}_1 \approx \mathbf{H}_2\mathbf{H}_3\mathbf{W}_3\mathbf{H}_3^\top\mathbf{H}_2^\top$$

$$\mathbf{W}_2 \approx \mathbf{H}_3\mathbf{W}_3\mathbf{H}_3^\top$$

Deep Asymmetric NMF

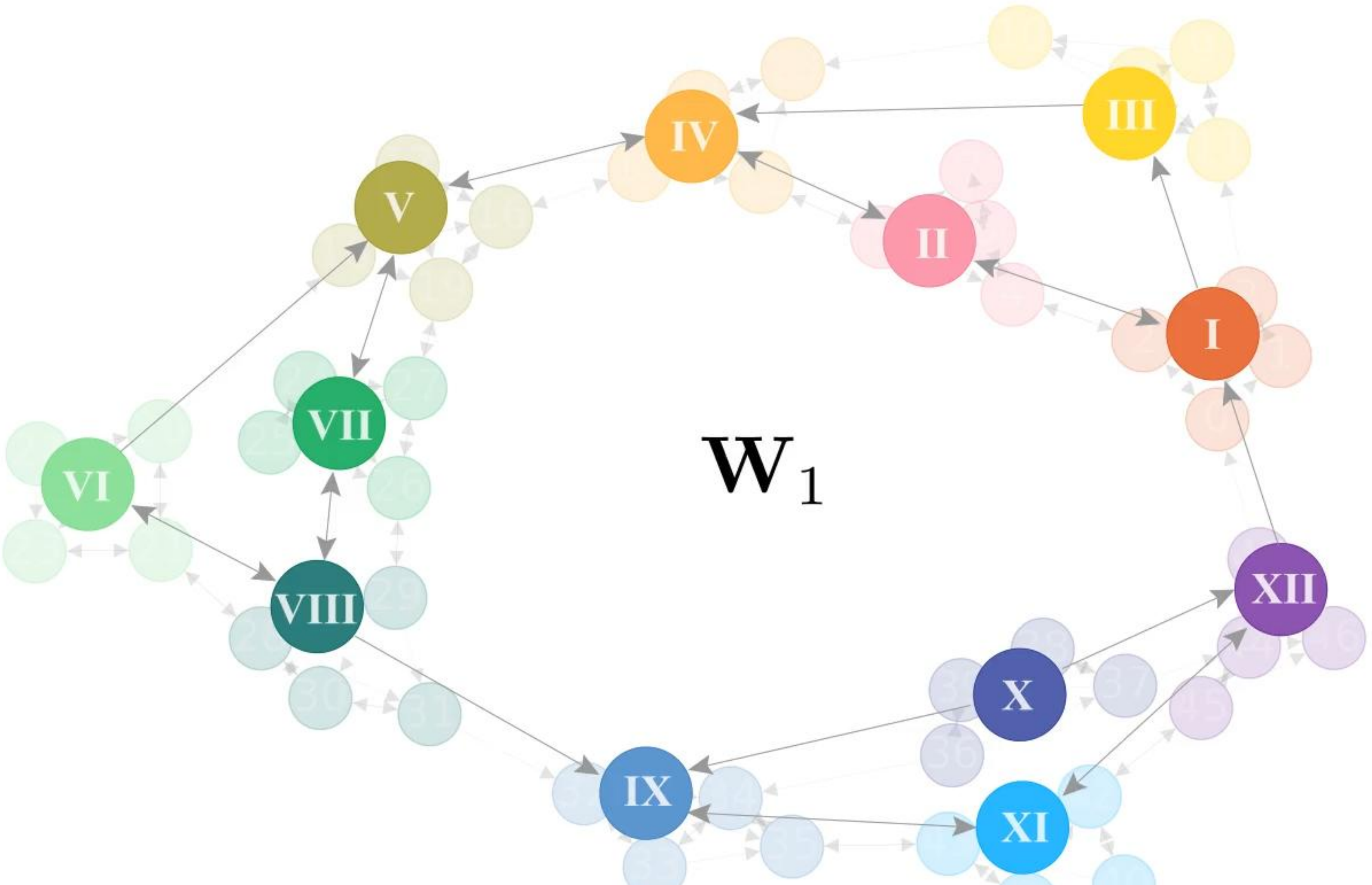


Interpretability of Deep Asymmetric NMF



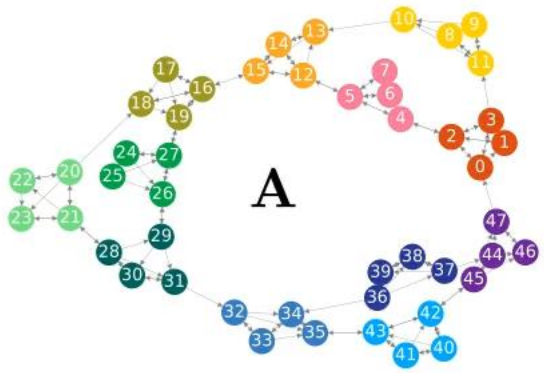
Input graph

Interpretability of Deep Asymmetric NMF

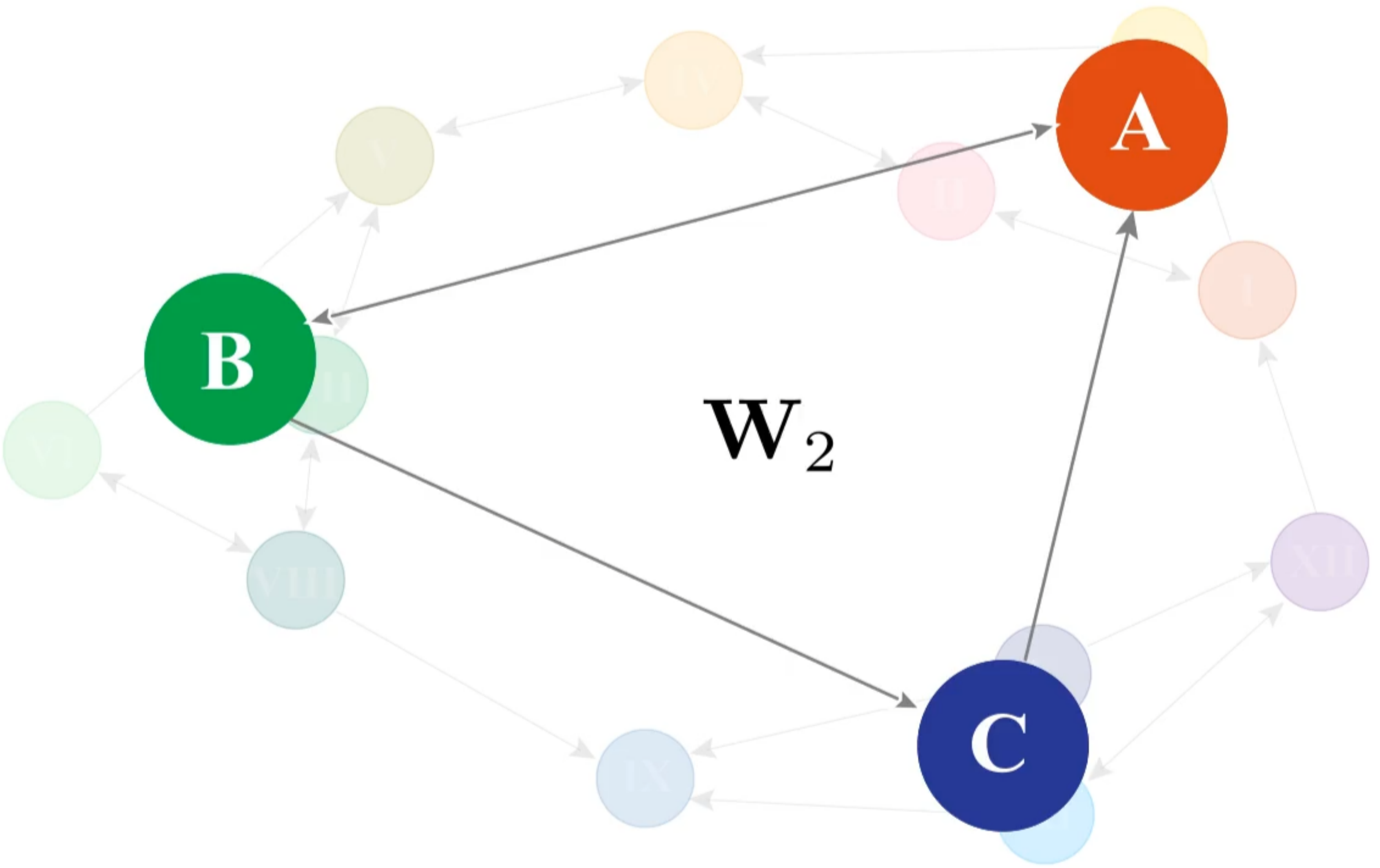


First-level summarized graph

Input graph

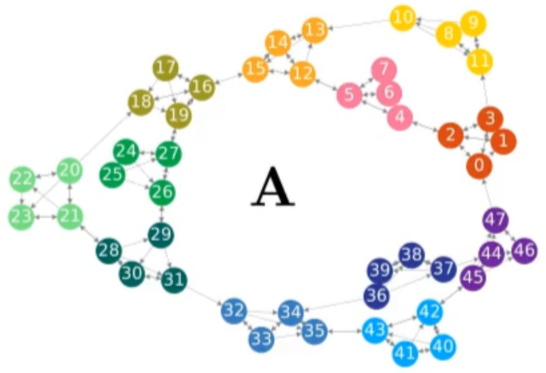


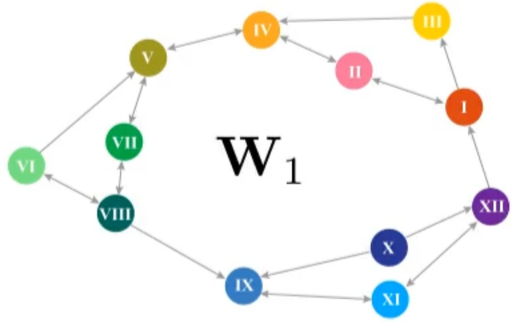
Interpretability of Deep Asymmetric NMF



Second-level summarized graph
(cluster-level)

Input graph





First-level summarized graph

Graph regularization

"Deep Asymmetric Nonnegative Matrix Factorization for Graph Clustering"

A. Hajiveiseh, **S.A. Seyedi**, and F. Akhlaghian

Pattern Recognition, 2024

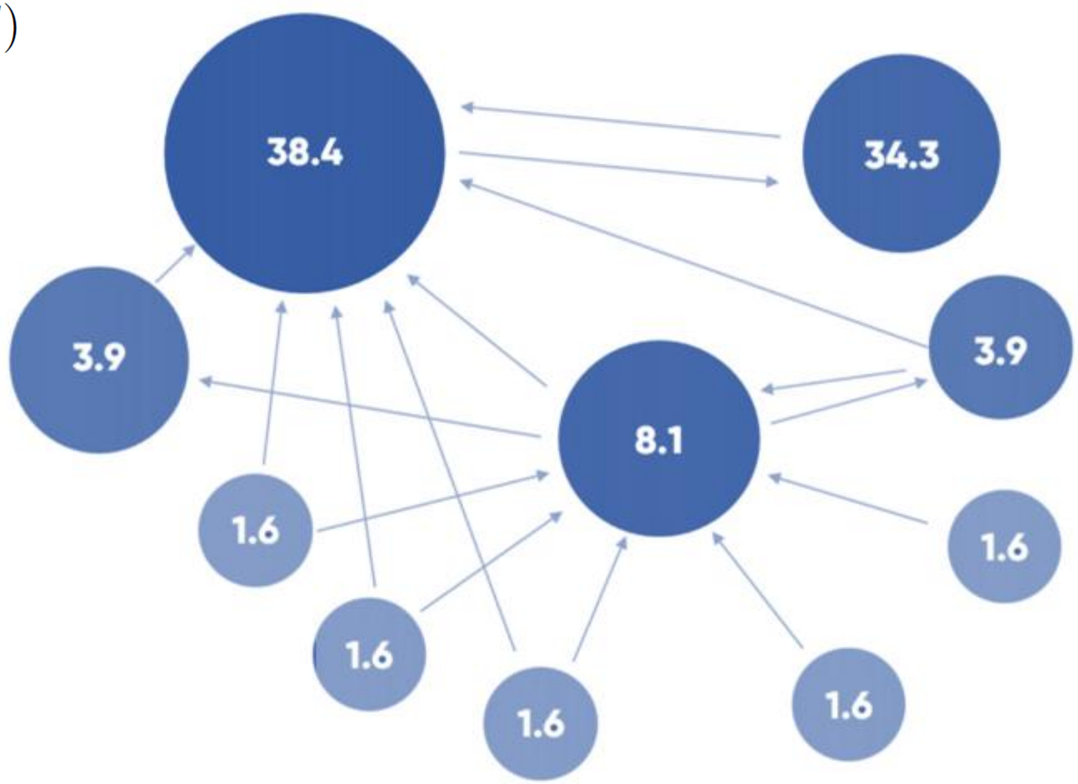
$$\min_{\boldsymbol{H}_i,\boldsymbol{W}_p} \mathcal{L} = \| \boldsymbol{A_S} - \boldsymbol{H}_1 \ldots \boldsymbol{H}_p \boldsymbol{W}_p \boldsymbol{H}_p^\top \ldots \boldsymbol{H}_1^\top \|_F^2 + \lambda \mathcal{R}(\boldsymbol{C})$$

$$\text{s.t.} \quad \boldsymbol{W}_p \geq 0, \boldsymbol{H}_i \geq 0, \forall i = 1, 2, \ldots, p.$$

Google PageRank

$$c_i = \frac{1}{n}(1-\rho) + \rho \sum_{j=1}^{n} \frac{A_{i,j}}{K_j^{out}} c_j$$

$$C_{i,j} = \begin{cases} c_i, & \text{if } A_{i,j} \neq 0, \\ 0, & \text{if } A_{i,j} = 0. \end{cases}$$



Graph regularization

"Deep Asymmetric Nonnegative Matrix Factorization for Graph Clustering"
A. Hajiveiseh, **S.A. Seyedi**, and F. Akhlaghian
Pattern Recognition, 2024

$$\mathcal{R} = \sum_{i=1}^{n}\sum_{j=1}^{n}\|\psi_i - \psi_j\|^2 C_{i,j} = 2\mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{D}\boldsymbol{\Psi}^\top) - \mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{C}\boldsymbol{\Psi}^\top) - \mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{C}^\top\boldsymbol{\Psi}^\top)$$

$$\boldsymbol{\Psi} = \prod_{i=1}^{p} \boldsymbol{H}_i$$

$$\min_{\boldsymbol{H}_i,\boldsymbol{W}_p} \|\boldsymbol{A}_{\boldsymbol{S}} - \boldsymbol{H}_1 \ldots \boldsymbol{H}_p \boldsymbol{W}_p \boldsymbol{H}_p^\top \ldots \boldsymbol{H}_1^\top\|_F^2 + \lambda[2\mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{D}\boldsymbol{\Psi}^\top) - \mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{C}\boldsymbol{\Psi}^\top) - \mathrm{Tr}(\boldsymbol{\Psi}\boldsymbol{C}^\top\boldsymbol{\Psi}^\top)]$$

$$\text{s.t.} \quad \boldsymbol{H}_i, \boldsymbol{W}_p \geq 0, \ \forall i \in \{1, \ldots, p\}.$$

Algorithm

Input: The adjacency matrix of graph $\mathcal{G}$, $\boldsymbol{A}$; layer size of each layer, r_i; scale parameter η; regularization parameter λ; dumping factor $\rho = 0.85$;
Output: $\boldsymbol{W}_i$ $(1 \le i < p)$, $\boldsymbol{H}_i$ $(1 \le i < p)$, and the cluster matrix $\boldsymbol{\Psi}$;

Constructing the second-order similarity matrix $\boldsymbol{S}$ by (11);
Constructing the input graph $\boldsymbol{A_S}$ by $\boldsymbol{A_S} = \boldsymbol{A} + \eta \boldsymbol{S}$;
Constructing the influence score matrix $\boldsymbol{C}$ by (19);
▷ **Pre-training process:**
$\boldsymbol{W}_1, \boldsymbol{H}_1 \leftarrow \text{ShallowAsNMF}(\boldsymbol{A_S}, r_1)$;
for $i = 2$ **to** p **do**
    $\boldsymbol{W}_i, \boldsymbol{H}_i \leftarrow \text{ShallowAsNMF}(\boldsymbol{W}_{i-1}, r_i)$;
end for
▷ **Fine-tuning process:**
while convergence not reached **do**
    for $i = 1$ **to** p **do**
        $\boldsymbol{\Psi}_{i-1} \leftarrow \prod_{\tau=1}^{i-1} \boldsymbol{H}_\tau (\boldsymbol{\Psi}_0 \leftarrow \mathbf{I})$;
        $\boldsymbol{\Phi}_{i+1} \leftarrow \prod_{\tau=i+1}^{p} \boldsymbol{H}_\tau (\boldsymbol{\Phi}_{p+1} \leftarrow \mathbf{I})$;
        Update $\boldsymbol{H}_i$ by $\boldsymbol{H}_i \leftarrow \boldsymbol{H}_i \odot \left[\frac{\boldsymbol{\Psi}_{i-1}^\top (\boldsymbol{A}^\top \boldsymbol{\Psi} \boldsymbol{W}_p + \boldsymbol{A} \boldsymbol{\Psi} \boldsymbol{W}_p^\top + \lambda \boldsymbol{C} \boldsymbol{\Psi} + \lambda \boldsymbol{C}^\top \boldsymbol{\Psi}) \boldsymbol{\Phi}_{i+1}^\top}{\boldsymbol{\Psi}_{i-1}^\top (\boldsymbol{\Psi} \boldsymbol{W}_p^\top \boldsymbol{\Psi}^\top \boldsymbol{\Psi} \boldsymbol{W}_p + \boldsymbol{\Psi} \boldsymbol{W}_p \boldsymbol{\Psi}^\top \boldsymbol{\Psi} \boldsymbol{W}_p^\top + 2\lambda \boldsymbol{D} \boldsymbol{\Psi}) \boldsymbol{\Phi}_{i+1}^\top} \right]^{\frac{1}{4}}$;
        $\boldsymbol{\Psi}_i \leftarrow \boldsymbol{\Psi}_{i-1} \boldsymbol{H}_i$;
        Update $\boldsymbol{W}_i$ by $\boldsymbol{W}_i \leftarrow \boldsymbol{W}_i \odot \frac{\boldsymbol{\Psi}_i^\top \boldsymbol{A} \boldsymbol{\Psi}_i}{\boldsymbol{\Psi}_i^\top \boldsymbol{\Psi}_i \boldsymbol{W}_i \boldsymbol{\Psi}_i^\top \boldsymbol{\Psi}_i}$ $(i < p$, optional$)$ or by $\boldsymbol{W}_p \leftarrow \boldsymbol{W}_p \odot \frac{\boldsymbol{\Psi}^\top \boldsymbol{A} \boldsymbol{\Psi}}{\boldsymbol{\Psi}^\top \boldsymbol{\Psi} \boldsymbol{W}_p \boldsymbol{\Psi}^\top \boldsymbol{\Psi}}$ $(i = p)$;
    end for
end while

Brain Networks Analysis

- Studies functional and structural brain connections using fMRI data.
- Helps understand brain functions and neurological diseases by uncovering patterns in networks.

Why Deep NMF?

- Scalability & Flexibility: Handles high-dimensional brain data and discovers key network features.
- Enhanced Interpretability: Nonnegative factorizations offer clearer insights, ideal for understanding complex brain regions and relationships.

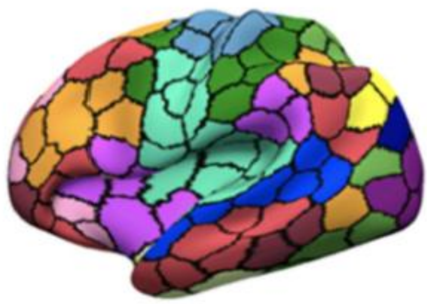
Deep NMF's Role in Brain Network Analysis

- Dimensionality Reduction: Learns low-dimensional representations to identify key regions and patterns linked to diseases or cognitive functions.
- Unsupervised Learning: Discovers hidden patterns and new brain network structures without labeled data.

NeuroGraph datasets (HCP-Task, HCP-Gender, and HCP-Age)

- The dataset $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ consists of n individual samples.
- Each sample A_i is represented by a symmetric correlation matrix $A_i \in \mathbb{R}^{m \times m}$, where m is the number of Regions of Interest (ROIs).

Group-Level Atlas Summarizing Regions of Interest (ROI)



Thank You

amjadseyedi.github.io