

Link Prediction on Evolving Data using Tensor Factorization

Stephan Spiegel

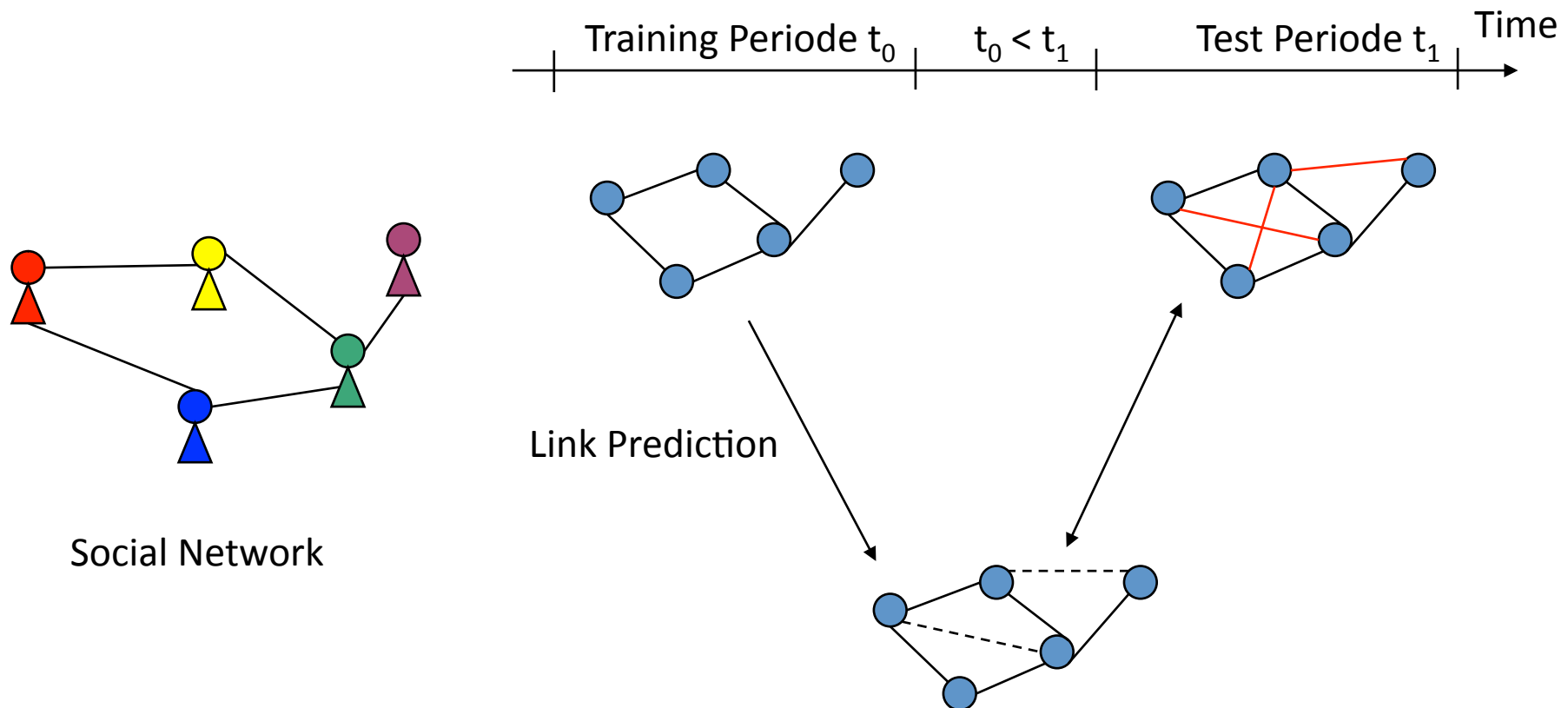
Technical University Berlin

Outline

- Link Prediction
- Singular Value Decomposition
- Tensor Factorization
- Link Prediction using Tensor Factorization
- Evaluation

Link Prediction

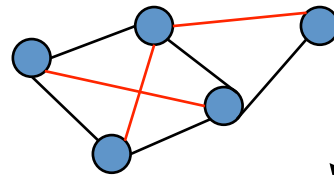
Link Prediction



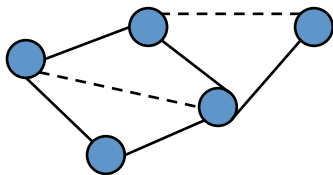
Kleinberg, J. and Liben-Nowell, D.: The Link Prediction Problem for Social Networks
Strogatz, S. H.: Exploring complex networks. In: Nature, 410, S. 268–276

Link Prediction

Correct Links at Time t_1

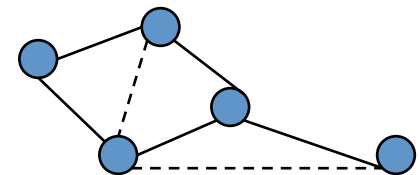


Link Prediction



precision = #correct links / #new links = $2/3$

Random Graph



precision = #correct links / #new links = $1/3$

Kleinberg, J. and Liben-Nowell, D.: The Link Prediction Problem for Social Networks
Strogatz, S. H.: Exploring complex networks. In: Nature, 410, S. 268–276

Link Prediction Algorithms

- Neighbor-Based
 - Common Neighbors
 - Preferential Attachment
- Link-Based
 - Katz
- Semantic
 - K-Nearest Neighbors
- Psychological
 - Theory of Balance
 - Theory of Status

Kleinberg, J. and Liben-Nowell, D.: The Link Prediction Problem for Social Networks

Lim, E., Liu, H., Ma, N., Nguyen, V. & Sun, A.: Trust relationship prediction using online product review data

Bauckhage, C., Kunegis, J. & Lommatzsch, A.: The Slashdot Zoo: Mining a social network with negative edges

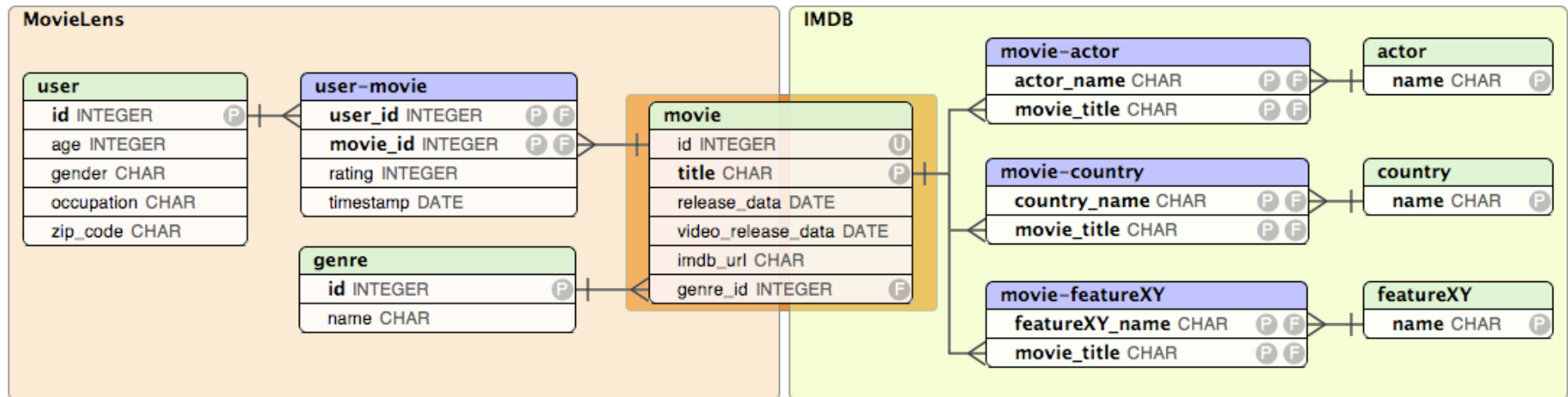
Huttenlocher, D., Kleinberg, J. & Leskovec, J.: Predicting positive and negative links in online social networks

Huttenlocher, D., Kleinberg, J. & Leskovec, J.: Signed networks in social media

Grossman, L.: How Computers Know What We Want - Before We Do

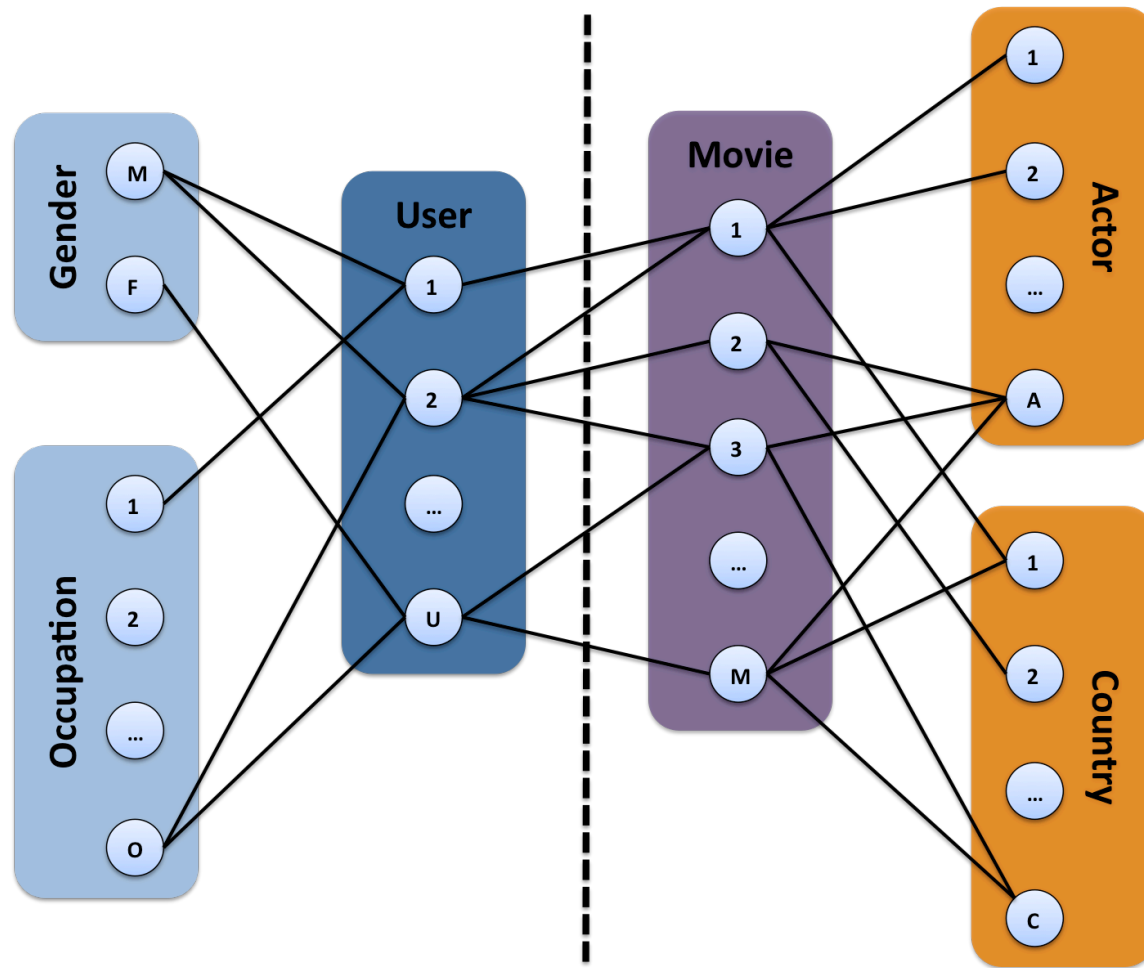
Link Prediction using Singular Value Decomposition

ER-Diagram



Hydra: A Hybrid Recommender System. Stephan Spiegel, Jerome Kunegis and Fang Li. Workshop on Complex Networks in Information and Knowledge Management, 2009.

Bipartite Graph



Hydra: A Hybrid Recommender System. Stephan Spiegel, Jerome Kunegis and Fang Li. Workshop on Complex Networks in Information and Knowledge Management, 2009.

Feature Matrix

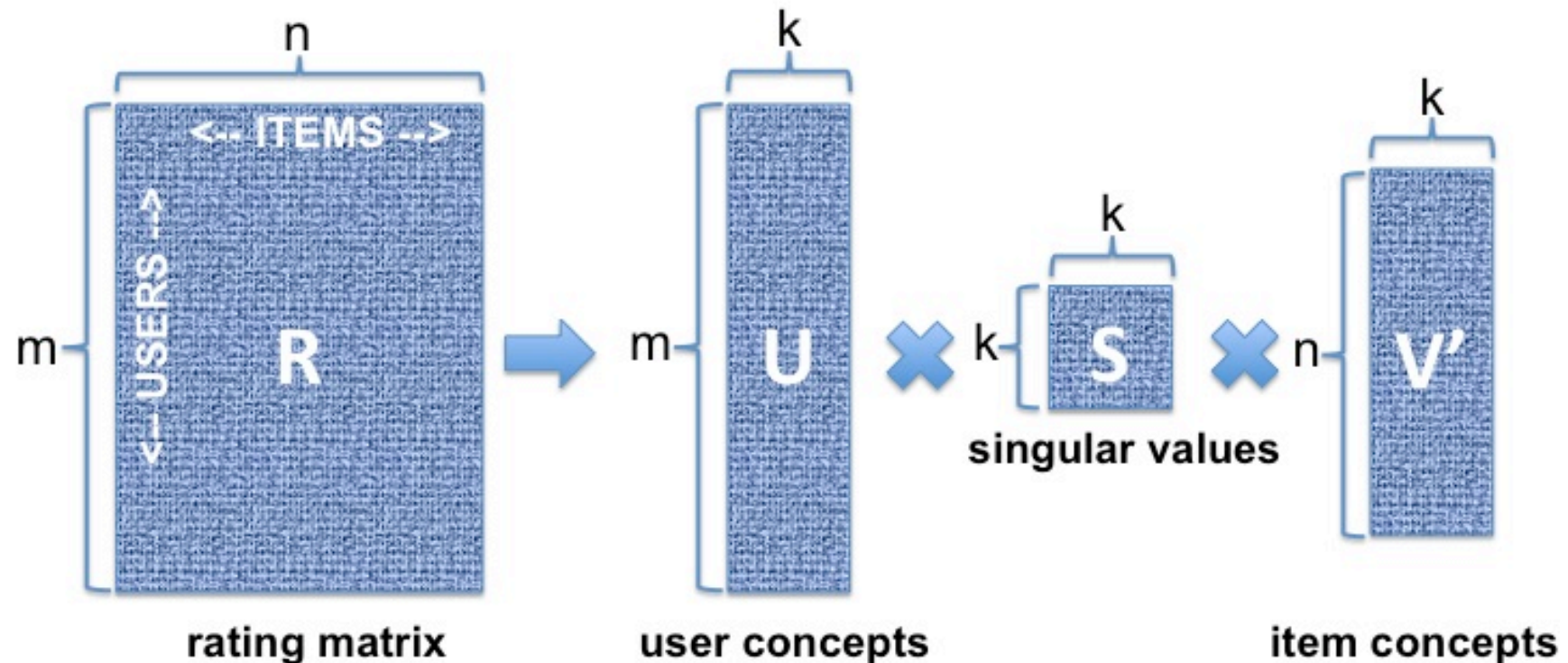
MovieLens User Features

		Movie					Occupation					
		M_1	M_2	M_3	...	M_m	O_1	O_2	...	O_o		
IMDB Item Features	User	U_1	1	0	0	...	0	1	0	...	0	
		U_2	1	1	1	...	0	0	0	...	1	
	
		U_u	0	0	1	...	1	0	0	...	1	
	Country	C_1	1	0	0	...	1	unrated item				
C_2		0	1	0	...	0						
...							
C_c		0	0	1	...	1						

rated item

Hydra: A Hybrid Recommender System. Stephan Spiegel, Jerome Kunegis and Fang Li. Workshop on Complex Networks in Information and Knowledge Management, 2009.

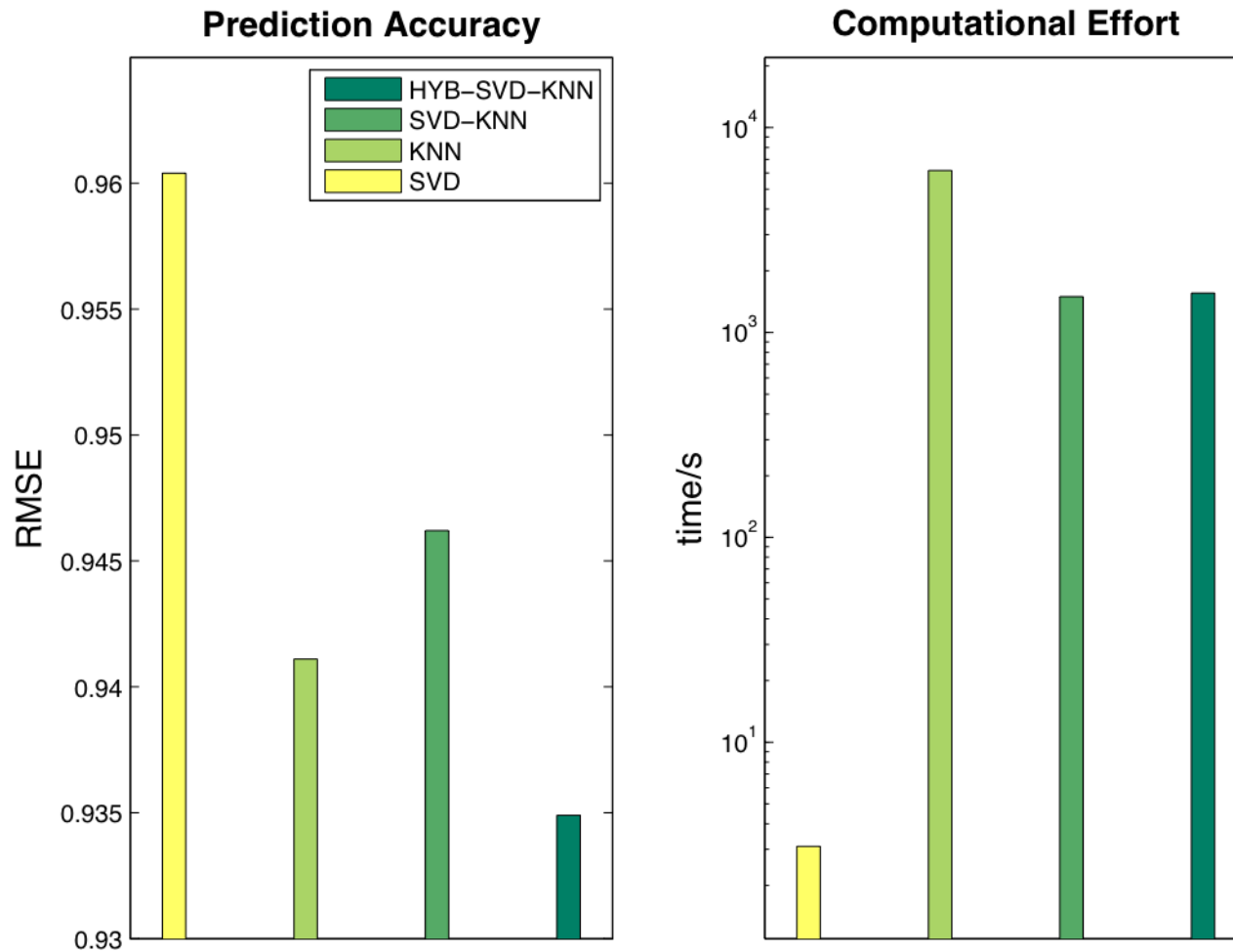
Singular Value Decomposition



$$R_{m \times n} = U_{m \times k} \cdot S_{k \times k} \cdot V_{k \times n}^T$$

Hydra: A Hybrid Recommender System. Stephan Spiegel, Jerome Kunegis and Fang Li. Workshop on Complex Networks in Information and Knowledge Management, 2009.

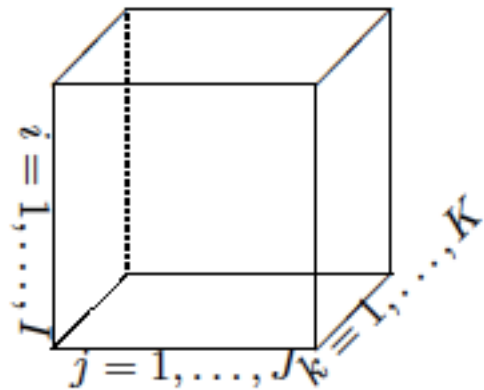
Evaluation



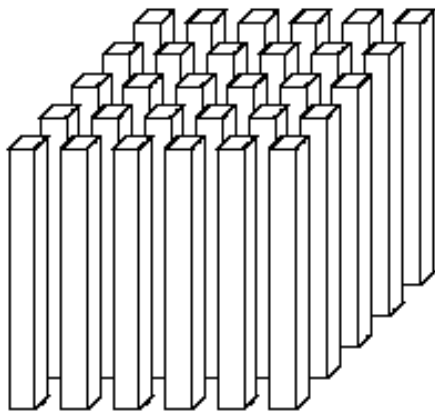
Hydra: A Hybrid Recommender System. Stephan Spiegel, Jerome Kunegis and Fang Li. Workshop on Complex Networks in Information and Knowledge Management, 2009.

Tensor Factorization

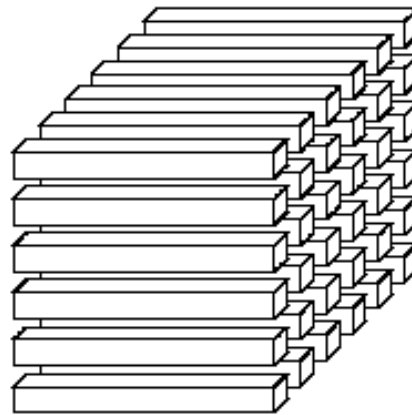
Tensor Factorization - Preliminaries



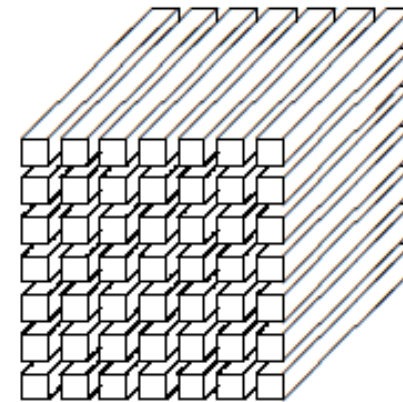
A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$



(a) Mode-1 (column) fibers:
 $\mathbf{X}_{:jk}$

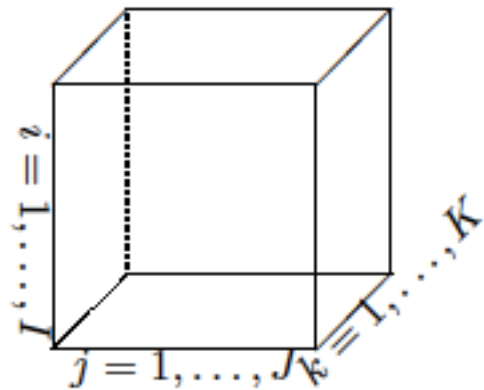


(b) Mode-2 (row) fibers:
 $\mathbf{X}_{i:k}$

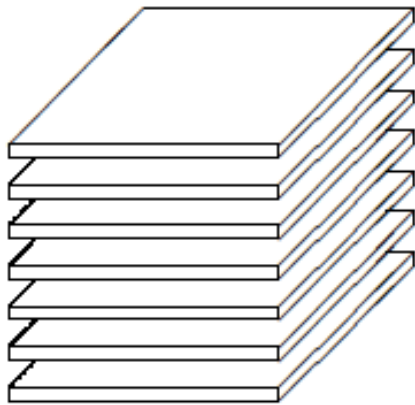


(c) Mode-3 (tube) fibers:
 \mathbf{X}_{ij}

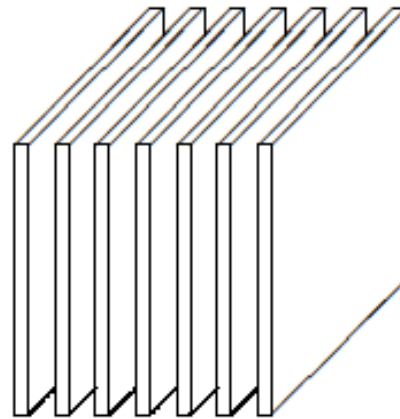
Tensor Factorization - Preliminaries



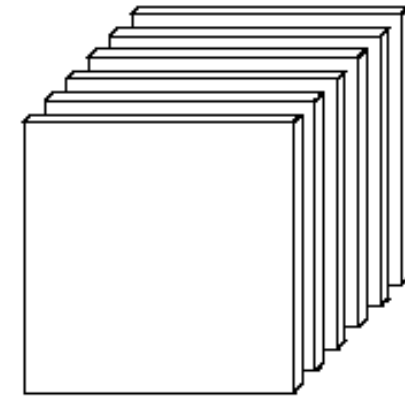
A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$



(a) Horizontal slices: $\mathbf{X}_{i::}$



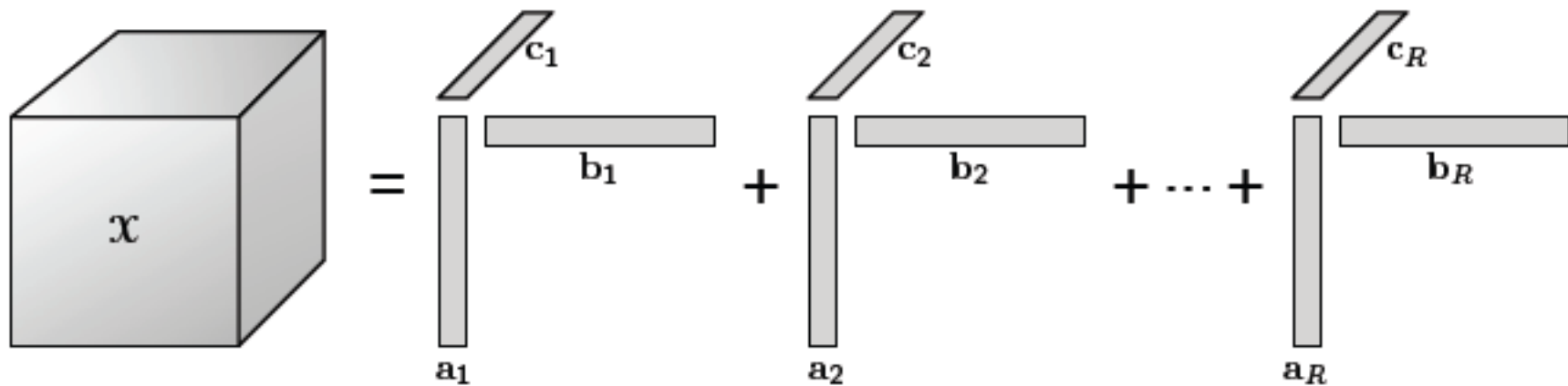
(b) Lateral slices: $\mathbf{X}_{:j:}$



(c) Frontal slices: $\mathbf{X}_{::k}$ (or \mathbf{X}_k)

CANDECOMP/Parafac Decomposition

$$\mathcal{X} \approx \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$$

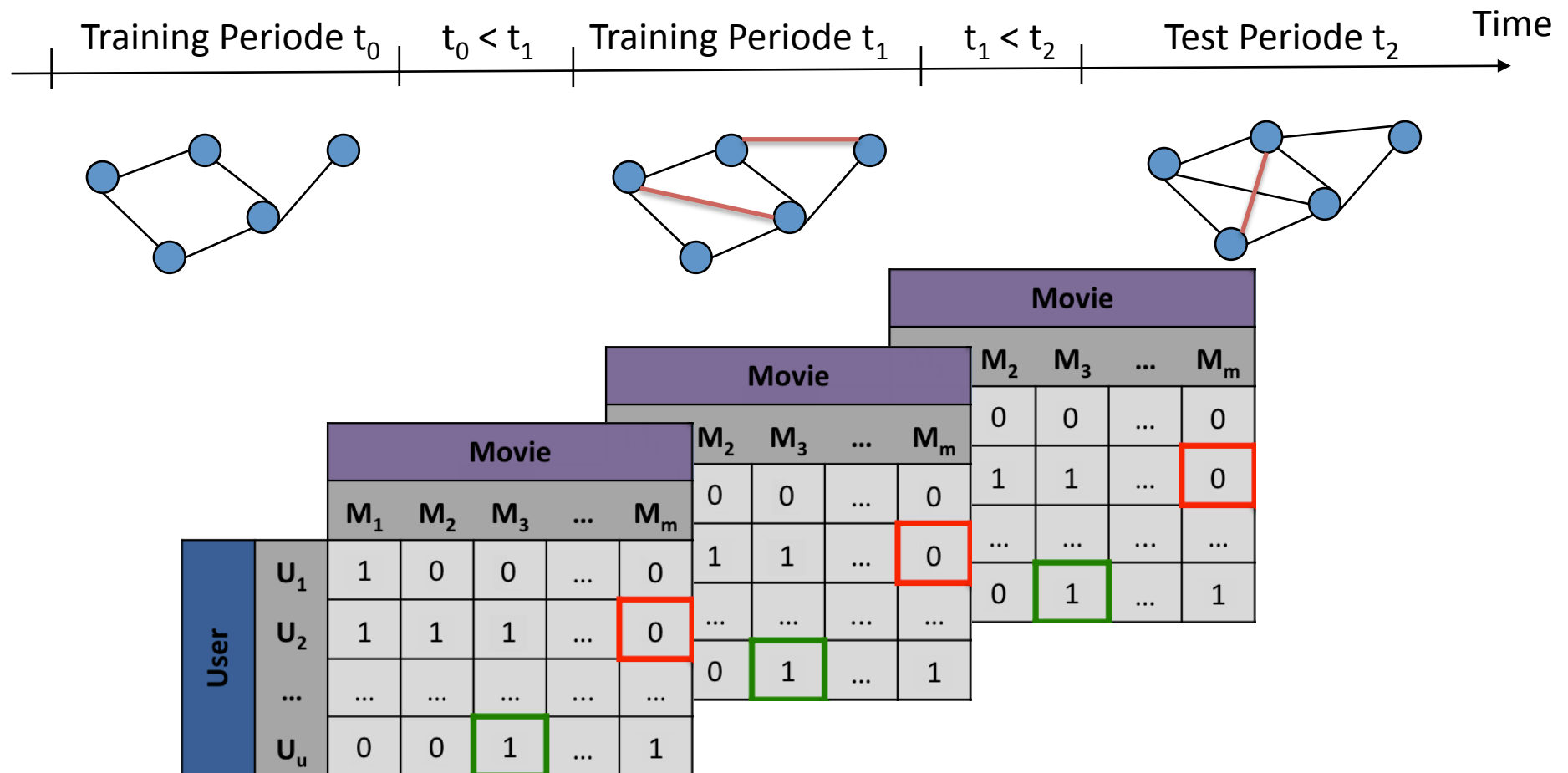


$$x_{ijk} \approx \sum_{r=1}^R a_{ir} b_{jr} c_{kr}$$

$$\mathcal{X} \approx \sum_{r=1}^R \lambda_r \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r = [[\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]].$$

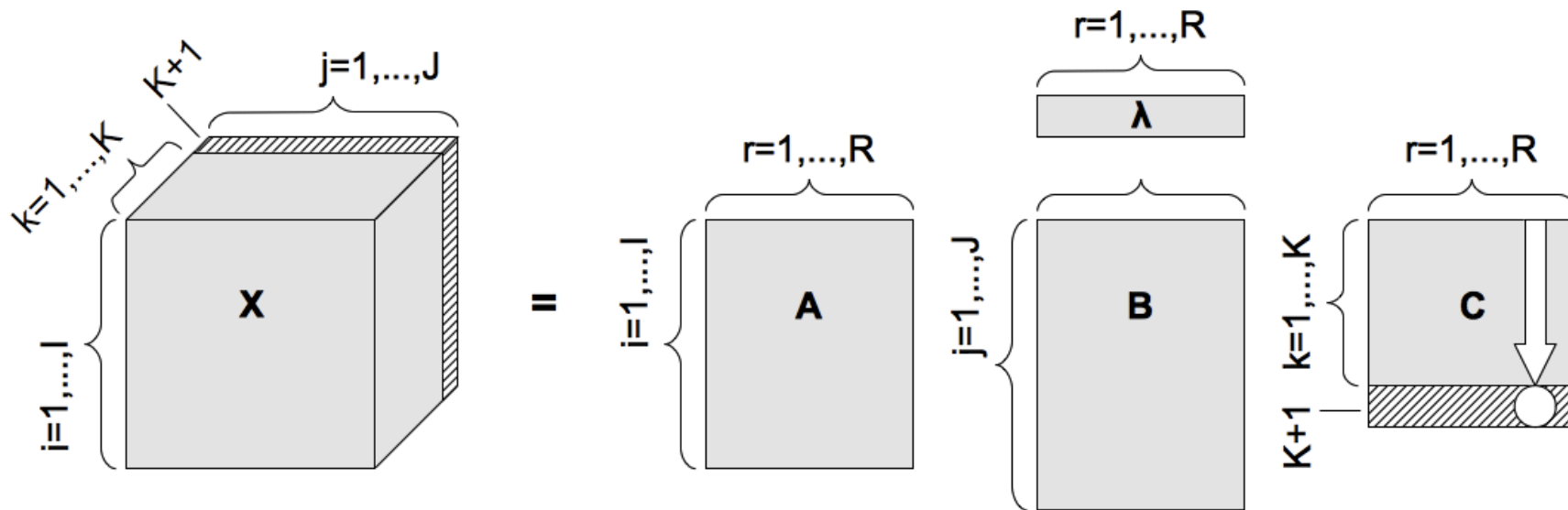
Link Prediction using Tensor Factorization

Link Prediction on Evolving Data



Kleinberg, J. and Liben-Nowell, D.: The Link Prediction Problem for Social Networks
 Strogatz, S. H.: Exploring complex networks. In: Nature, 410, S. 268–276

Tensor Factorization / Parafac

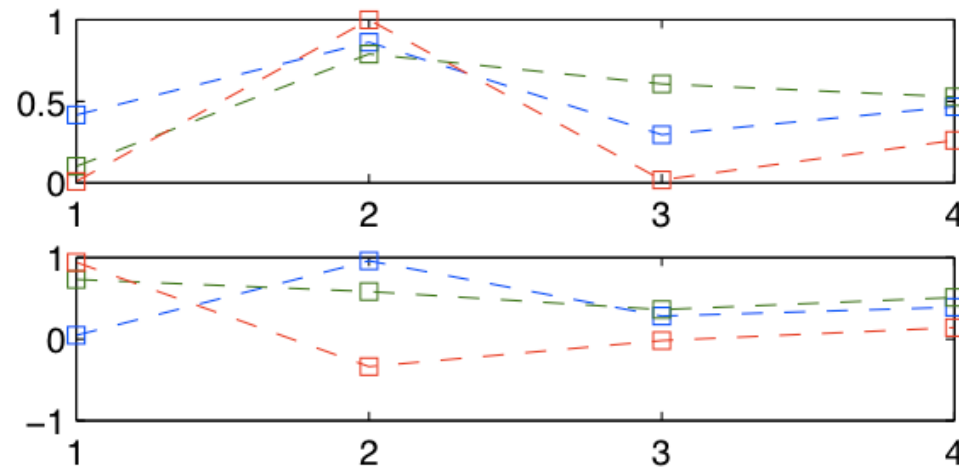


$$X \in \mathbb{R}^{I \times J \times (K+1)} \Rightarrow \lambda \in \mathbb{R}^R; A \in \mathbb{R}^{I \times R}; B \in \mathbb{R}^{J \times R}; C \in \mathbb{R}^{(K+1) \times R}$$

$$x_{i,j,K+1} \approx \sum_{r=1}^R \lambda_r (a_{i,r} \cdot b_{j,r} \cdot c_{K+1,r})$$

Link Prediction on Evolving Data using Tensor Factorization. Stephan Spiegel, Jan Clausen and Jerome Kunegis. Pacific Asian Conference in Knowledge Discovery and Data Mining, 2011.

Exponential Smoothing



$$s_1 = x_0$$
$$s_{t+1} = \alpha x_t + (1 - \alpha)s_t$$

Fig. 2. Temporal Trends in Time Factor Matrix of Wiki-Edit-Ar and Facebook-Links dataset, where the 4th point is extrapolated by Exponential Smoothing

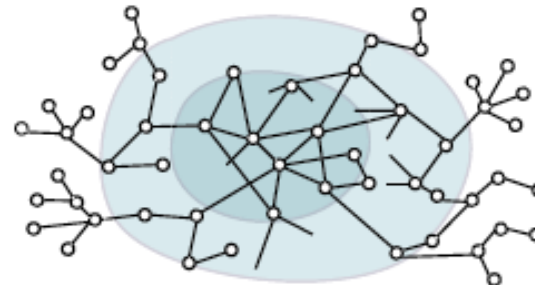
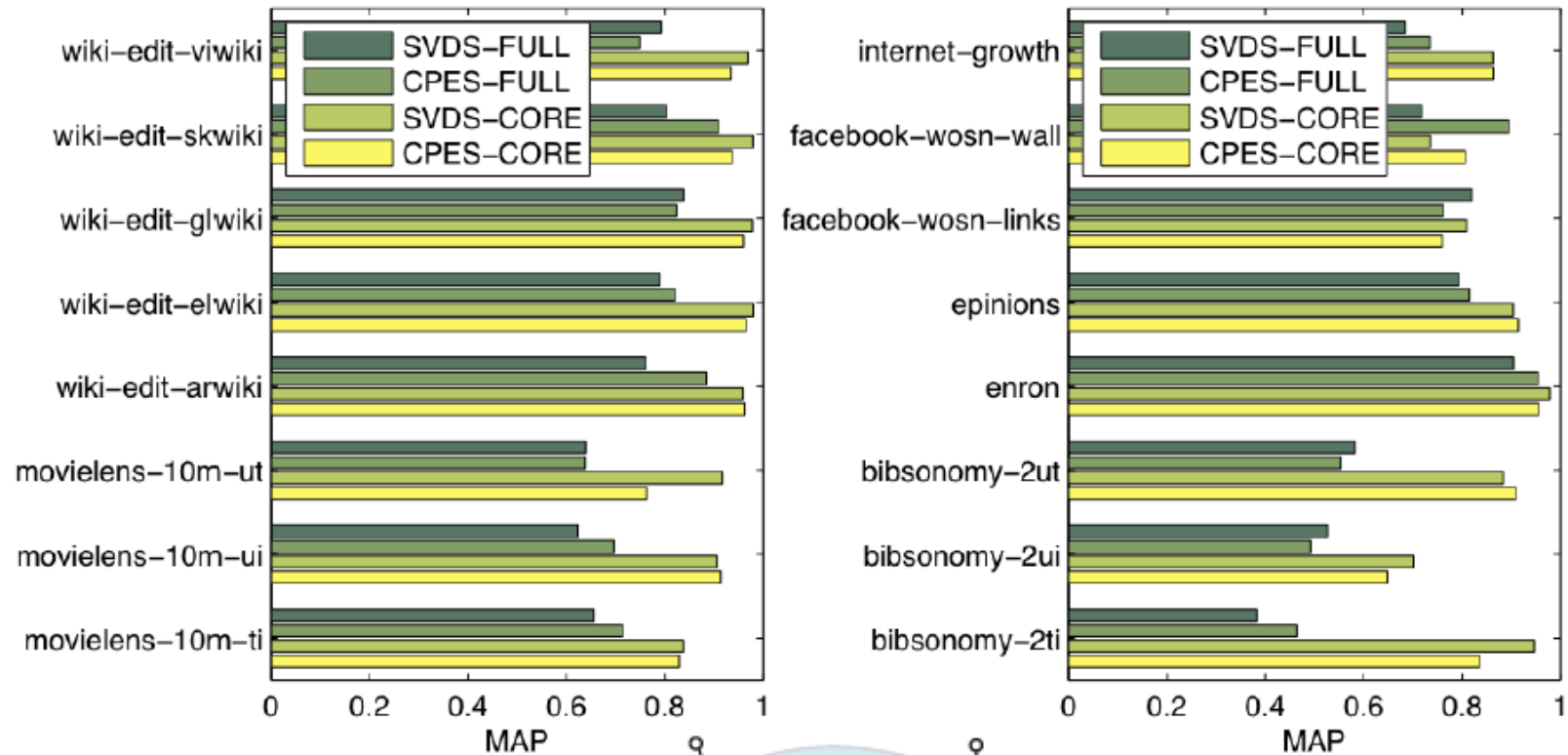
Examined Data Sets

DATASET	STRUCTURE	#ENTRIES	#MODE1	#MODE2	#SLICES
wiki-edit-viwiki	$user \times page \times time$	2262679	13766	303867	~ 4
wiki-edit-skwiki	$user \times page \times time$	2526392	7229	215638	~ 6
wiki-edit-glwiki	$user \times page \times time$	1315066	2850	91594	~ 7
wiki-edit-elwiki	$user \times page \times time$	1569075	8049	97149	~ 8
wiki-edit-arwiki	$user \times page \times time$	4000735	25692	510033	~ 4
movielens-10m-ut	$user \times tags \times time$	95580	2795	12553	~ 4
movielens-10m-ui	$user \times item \times time$	95580	3097	6367	~ 8
movielens-10m-ti	$tags \times item \times time$	95580	12775	6190	~ 4
internet-growth	$page \times page \times time$	104824	20689	20689	~ 3
facebook-wosn-wall	$user \times user \times time$	876993	30839	30839	~ 14
facebook-wosn-links	$user \times user \times time$	1545686	57356	57356	~ 13
epinions	$user \times user \times time$	19793847	91596	91596	~ 5
enron	$user \times user \times time$	1149884	64145	64145	~ 9
bibsonomy-2ut	$user \times tags \times time$	2555080	4804	167963	~ 8
bibsonomy-2ui	$user \times item \times time$	2555080	1345	335335	~ 4
bibsonomy-2ti	$tags \times item \times time$	2555080	155264	571768	~ 2

$$\#slices = \frac{\#entries}{2 * \max(\#mode1, \#mode2)}$$

Link Prediction on Evolving Data using Tensor Factorization. Stephan Spiegel, Jan Clausen and Jerome Kunegis. Pacific Asian Conference in Knowledge Discovery and Data Mining, 2011.

Examined Data Sets



Link Prediction on Evolving Data using Tensor Factorization. Stephan Spiegel, Jan Clausen and Jerome Kunegis. Pacific Asian Conference in Knowledge Discovery and Data Mining, 2011.

Tensor Factorization

- References

- "Applied Multiway Data Analysis" - Pieter M. Kroonenberg - Wiley & Sons
- "Unsupervised Multiway Data Analysis: A Literature Study" - Evrim Acar, Bulent Yener - TKDE'09
- "Tensor Decompositions and Applications" - Tamara G. Kolda, Brett W. Bader - SIAM'09

- Tools

- "Matlab Tensor Toolbox" - Sandia National Laboratories