Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The Minimum Transfer Cost Principle for Model-Order Selection

Mario Frank

Morteza Haghir Chehreghani

Joachim M. Buhmann

Department of Computer Science, ETH Zurich

Outline

- The minimum Transfer Cost Principle
- Model order selection for truncated SVD
- Model order selection for correlation clustering
- Transfer costs for k-means clustering
- Conclusion

Motivation

- Given:
  - A set of N objects with the measurements $\mathbf{X}$.
  - Two sets of objects $\mathbf{O}^{(\mathrm{m})}, m \in \{1,2\}$ with corresponding measurements $\mathbf{X}^{(m)}$ generated from the same source.
  - A data **model** characterized by a cost function $R(\boldsymbol{s}, \mathbf{X}, k)$, where the solution $\boldsymbol{s}$ incorporates all relevant parameters.
- Question:
  - **What is the appropriate model order k?**

Transfer Costs: Intuition

- Cross-validation:
  - Good choice of the model-order based on a given dataset should also yield low costs on a second dataset.
- How to transfer a solution from $\{\mathbf{O}^{(1)}, \mathbf{X}^{(1)}\}$ to $\{\mathbf{O}^{(2)}, \mathbf{X}^{(2)}\}$?
- Classification:
  - Class labels make mapping obsolete.
- Unsupervised learning:
  - No labels are available.

Transfer costs for factorial models

- The cost function is: $$R(\boldsymbol{s}, \mathbf{X}, k) = \sum_{i=1}^{N} R_i(\boldsymbol{s}(i), \mathbf{x}_i, k)$$

- We define an object-wise mapping function ψ:

$$\psi: \; O^{(2)} \times \mathcal{X} \times \mathcal{X} \to O^{(1)}$$
$$\left(i', \mathbf{X}^{(1)}, \mathbf{X}^{(2)}\right) \mapsto \psi(i', \mathbf{X}^{(1)}, \mathbf{X}^{(2)})$$

  - ψ aligns each object in $\mathbf{O}^{(2)}$ with its nearest neighbor in $\mathbf{O}^{(1)}$.

- Transfer costs:

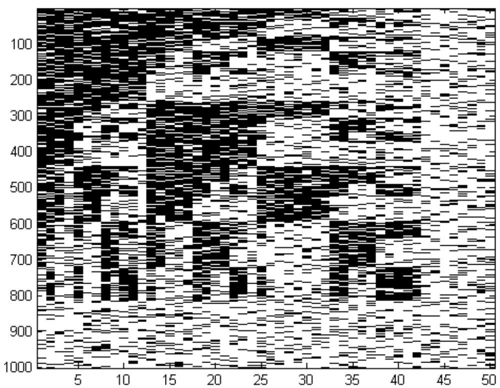
$$R^T(\boldsymbol{s}^{(1)}, \mathbf{X}^{(2)}, k) := \frac{1}{N_2} \sum_{i'=1}^{N_2} \sum_{i=1}^{N_1} R_{i'}(\boldsymbol{s}^{(1)}(i), \mathbf{x}_{i'}^{(2)}, k) \, \mathbb{I}_{\{\psi(i', \mathbf{X}^{(1)}, \mathbf{X}^{(2)})=i\}}$$

The Minimum Transfer Cost Principle

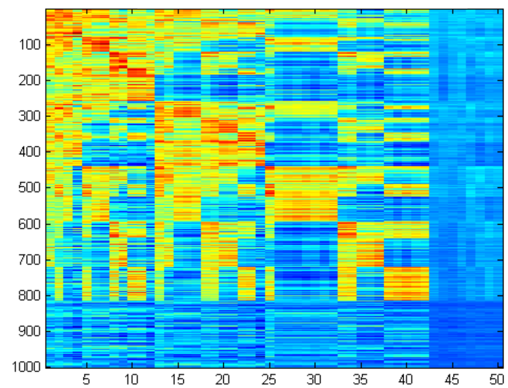
- The **minimum transfer cost** principle (MTC) selects the model order k with lowest transfer costs.
  - Too simple models underfit and achieve high costs on both datasets.
  - Too complex models overfit to the fluctuations of $\mathbf{X}^{(1)}$ which results in high costs on $\mathbf{X}^{(2)}$ where the fluctuations are different.

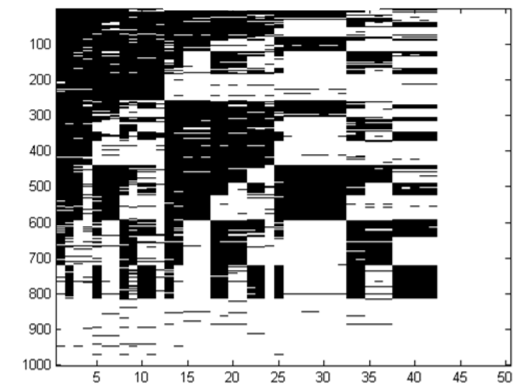
Denoising Matrices via truncated SVD

X = USV



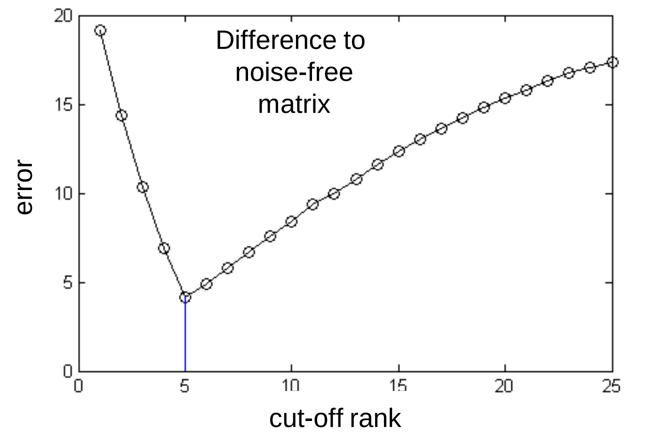
$X_5 = U_5 S_5 V_5$



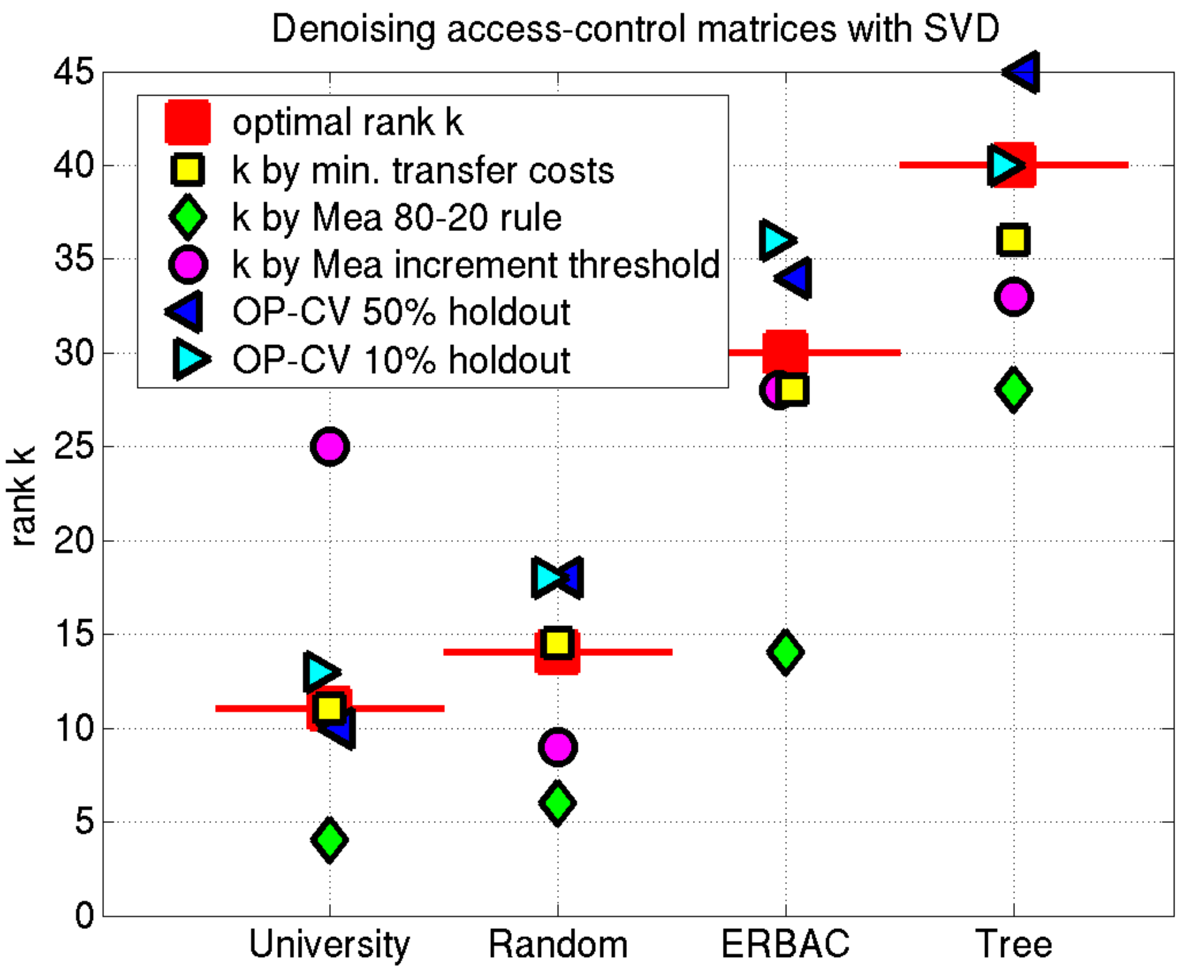


- Where to cut-off the spectrum?
- Error depends heavily on k.
- Transfer costs with nearest-n. mapping:

$$R^T(\boldsymbol{s}, \mathbf{X}, k) = \frac{1}{N_2} \left\| \psi_{NN}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \circ \mathbf{X}^{(2)} - \left(\mathbf{U}_{\mathbf{k}}^{(1)} \mathbf{S}_{\mathbf{k}}^{(1)} \mathbf{V}_{\mathbf{k}}^{(1)\mathbf{T}} \right) \right\|_2^2$$



Selecting the rank for truncated SVD



Correlation Clustering

- Formulation
  - Given: graph $G(\mathbf{O},\mathbf{X})$ with similarity matrix $\mathbf{X} := \{X_{ij}\} \in \{\pm 1\}^{\binom{N}{2}}$.
  - For the tuple $(\boldsymbol{s},\mathbf{X})$, the costs are

$$R(\boldsymbol{s},\mathbf{X},k) = -\frac{1}{2}\sum_{1\le u\le k}\sum_{(i,j)\in E_{u,u}}(X_{ij}-1)+\frac{1}{2}\sum_{1\le u\le k}\sum_{1\le v<u}\sum_{(i,j)\in E_{u,v}}(X_{ij}+1)$$

where, $E_{u,v} = \{(i,j) \in E\colon \boldsymbol{s}(i) = u \wedge \boldsymbol{s}(j) = v\}$.

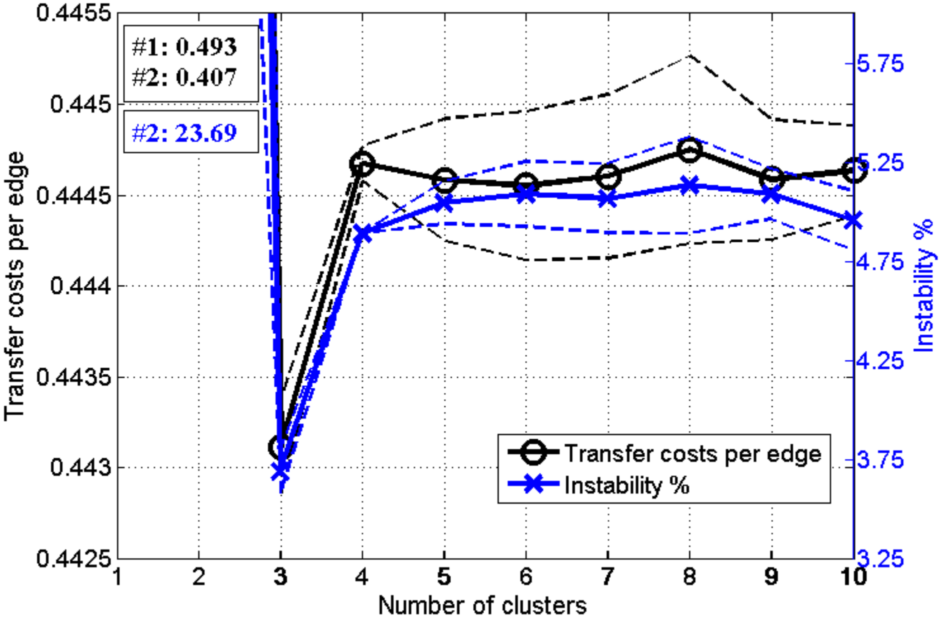
- The cluster index of object i' from $\mathbf{O}^{(2)}$ is determined by

$$\boldsymbol{s}^{(1)}(i') = \arg\min_{1\le v\le k} H(i',\boldsymbol{s}_v^{(1)}), \quad \text{with}$$

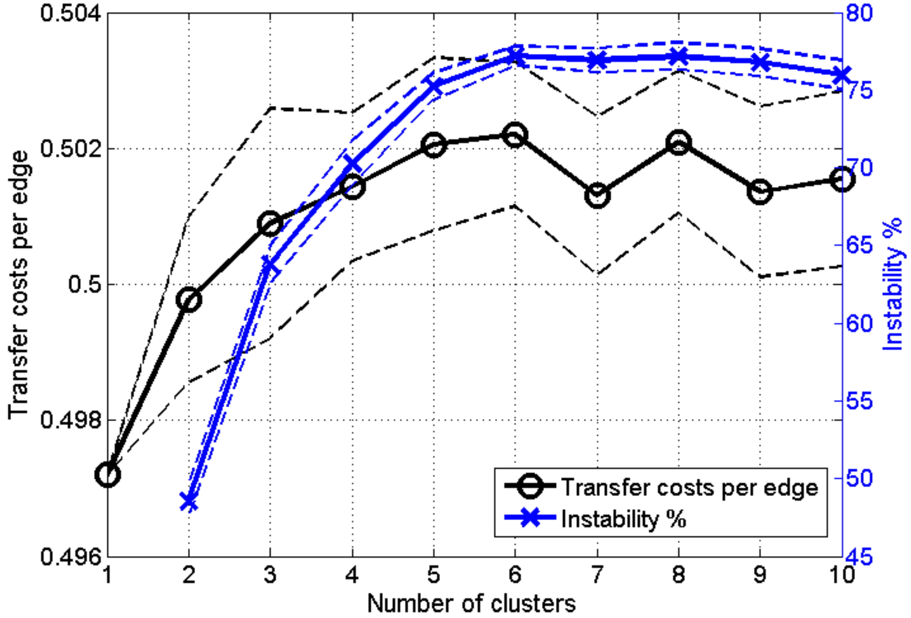
$$H(i',\boldsymbol{s}_v^{(1)}) = -\frac{1}{2}\sum_{j\in\boldsymbol{s}_v}(X_{ij}-1)+\frac{1}{2}\sum_{1\le u\le k,u\ne v}\sum_{j\in\boldsymbol{s}_u}(X_{ij}+1)$$

Finding the number of clusters

- Generate correlation data with varying noise level η



$\eta = 0.80$

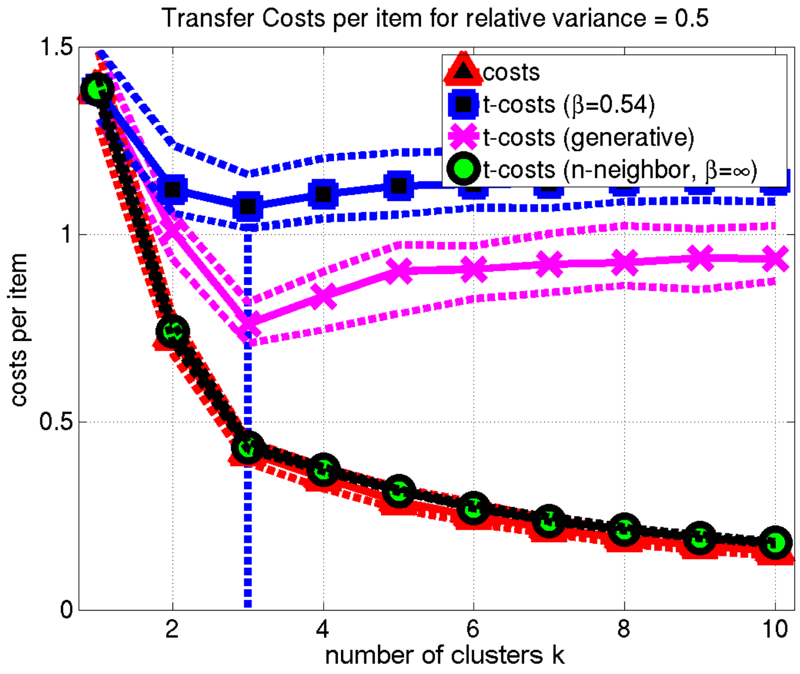


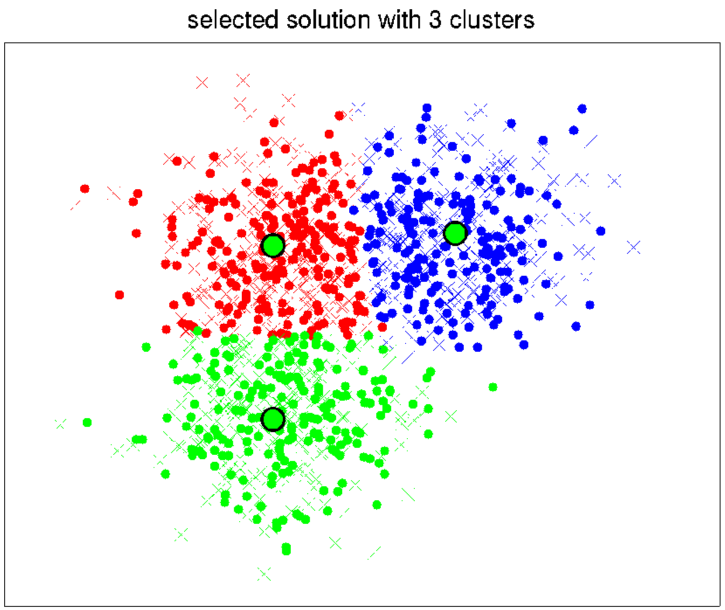
$\eta = 0.95$

Transfer costs for k-means

- Goal: challenge the criterion with a **model mismatch**.
- Experiment:
  - Generate two datasets from a **mixture of Gaussians**.
  - Cluster with **k-means**.
  - Select number of clusters with minimum transfer costs.
- Model mismatch: **no variance for k-means** → vector quantization preferred.

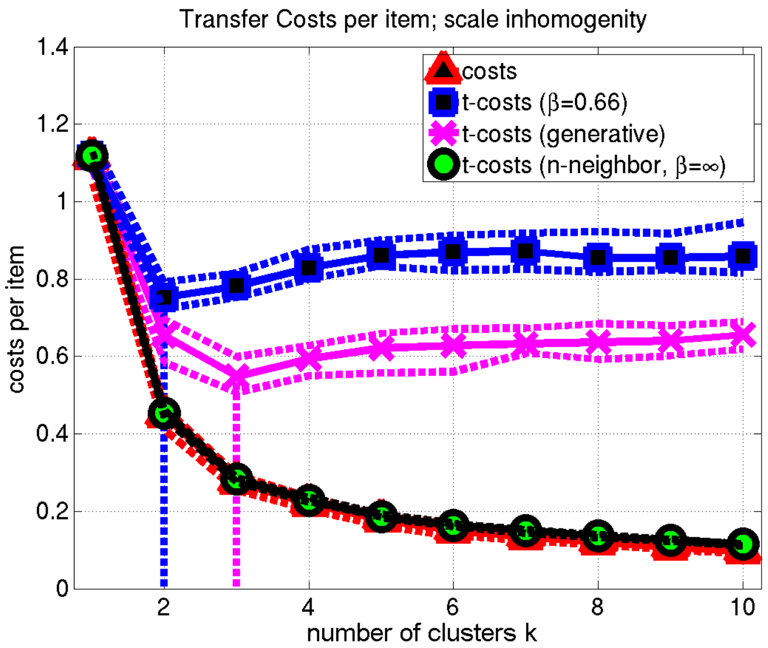
k-means: easy case

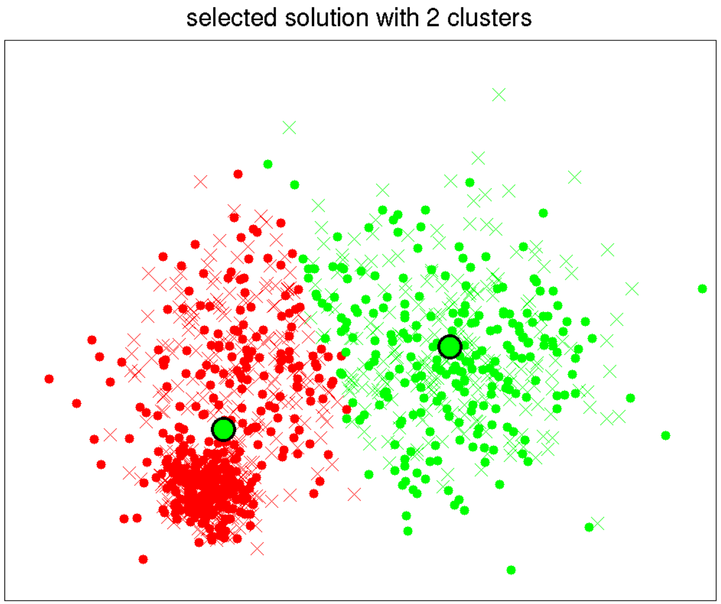




- Discrete mapping leads to **monotonically decreasing t-costs**!
- **Soft mapping** selects true number of clusters.

k-means: scale inhomogenity





- Soft mapping temperature serves as a **`global' variance**.
- Gaussian mixture models are required if variance changes locally.

Conclusion

- MTC: an easily applicable method for model-order selection in unsupervised scenarios.
- Demonstration of model-order selection in
  - Gaussian Mixture Models
  - Truncated SVD: denoising of images and Boolean matrices
  - Boolean matrix factorization (role mining)
  - Correlation clustering
- Soft mapping for model mismatch
  - *k*-means case study