UCL
Université catholique de Louvain

miro
molecular imaging, radiotherapy & oncology

Multi-scale neighbor embedding:
Towards parameter-free dimensionality reduction

John A. Lee, Diego Peluffo, Michel Verleysen

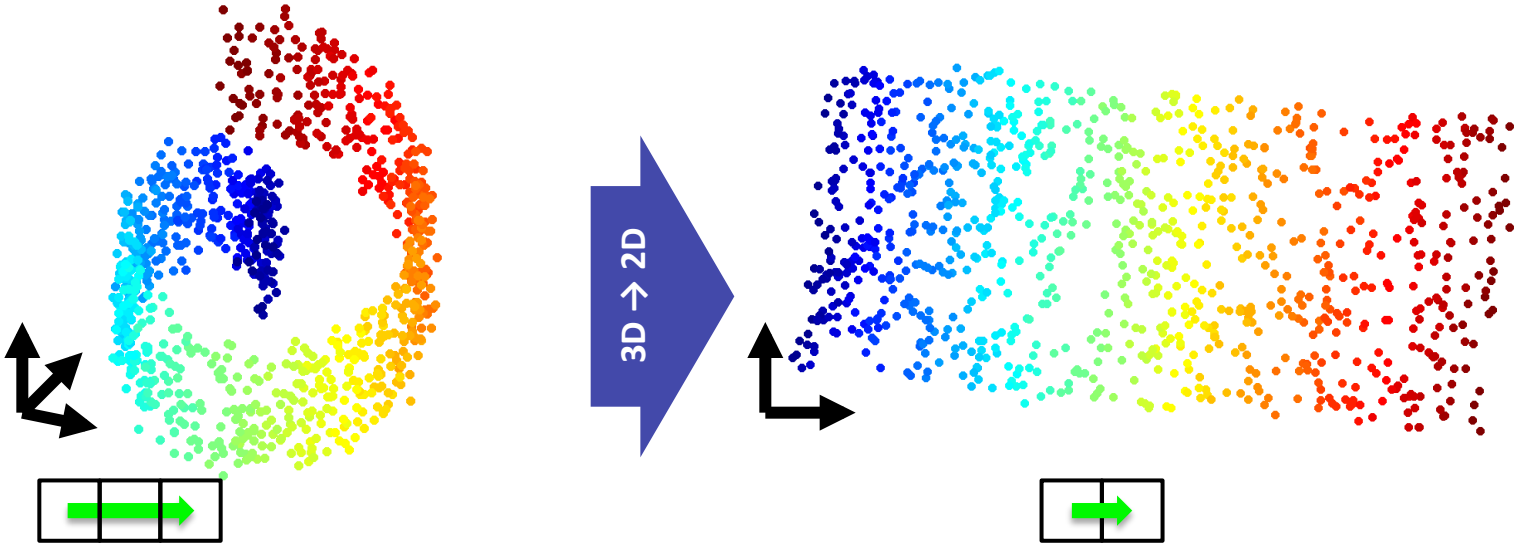
ESANN 2014, April 23-25, Bruges

Machine Learning Group

Dimensionality reduction

(a.k.a. (NL)DR, manifold learning, embedding, projection, …)

- Aims at representing high-dimensional (HD) data in low-dimensional (LD) spaces, while preserving structure
- Can be
  - Linear/nonlinear
  - Parametric/non-parametric
  - Supervised/semi-supervised/unsupervised

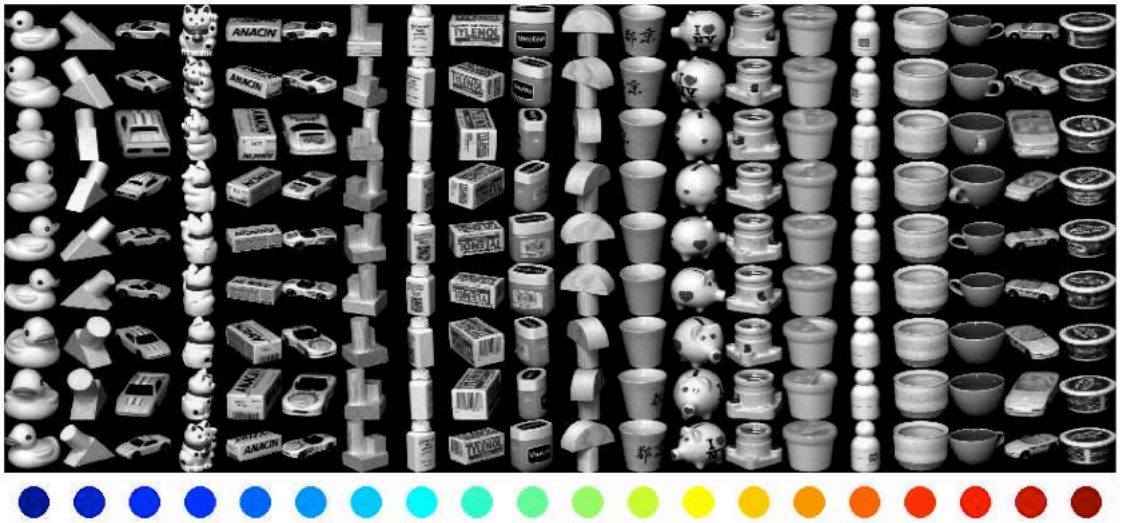


$$\Xi = [\boldsymbol{\xi}_i]_{1 \leq i \leq N}$$

$$\mathbf{X} = [\mathbf{x}_i]_{1 \leq i \leq N}$$

High-dimensional data

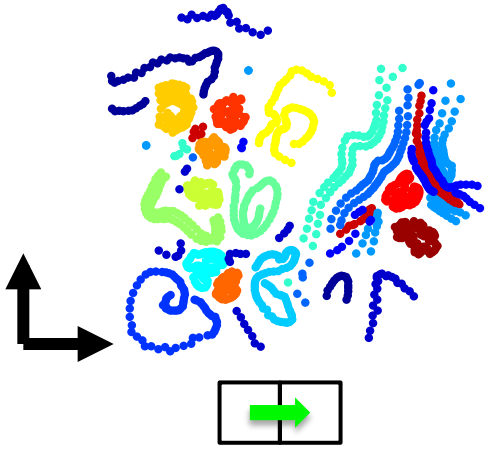
COIL-20 data set (1440 pictures of 20 rotated objects, 72 poses, every 5°)



Vectorised
128-by-128 images

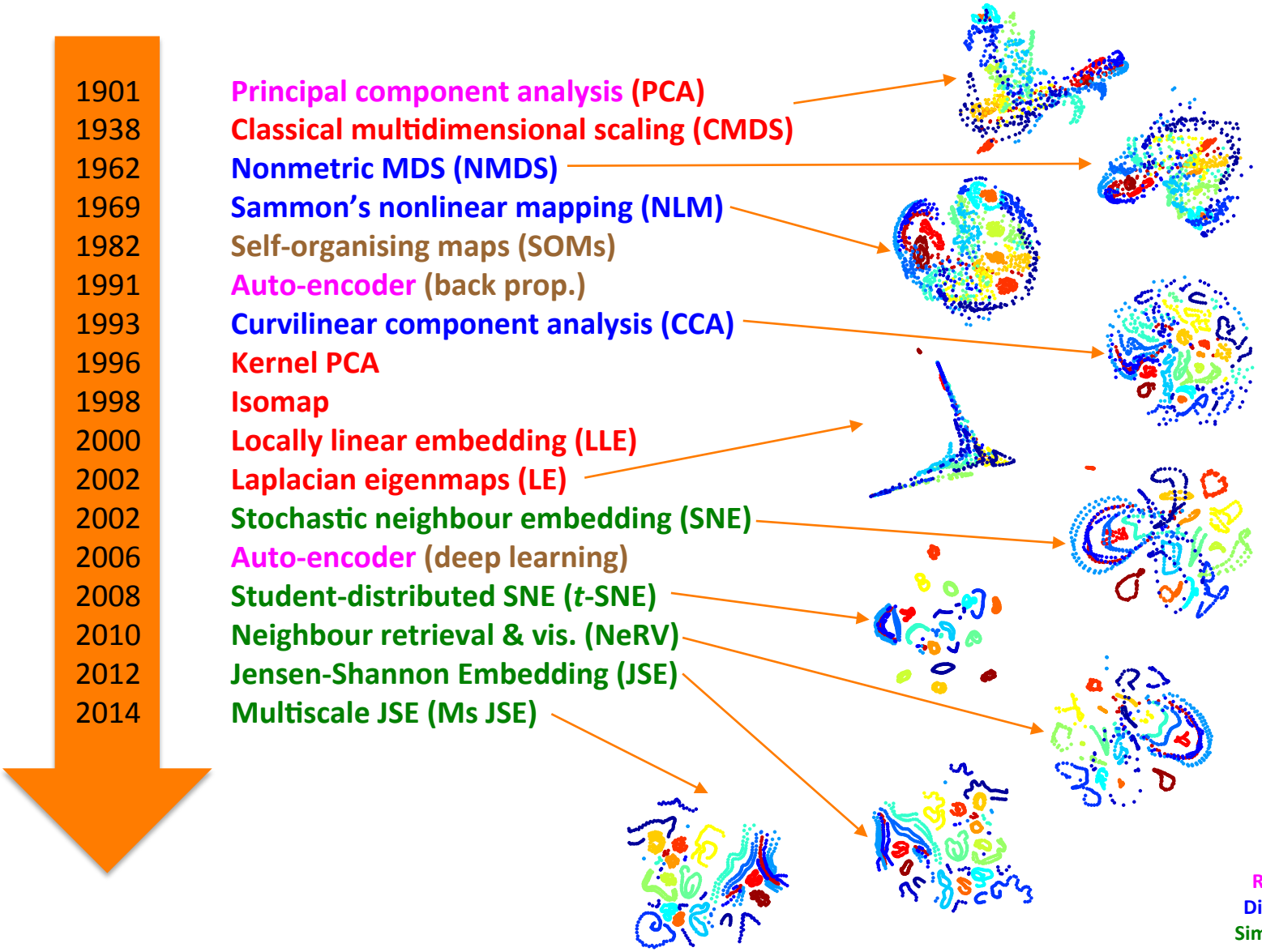
$$\Xi = [\boldsymbol{\xi}_i]_{1 \le i \le N}$$

$MD \rightarrow 2D$

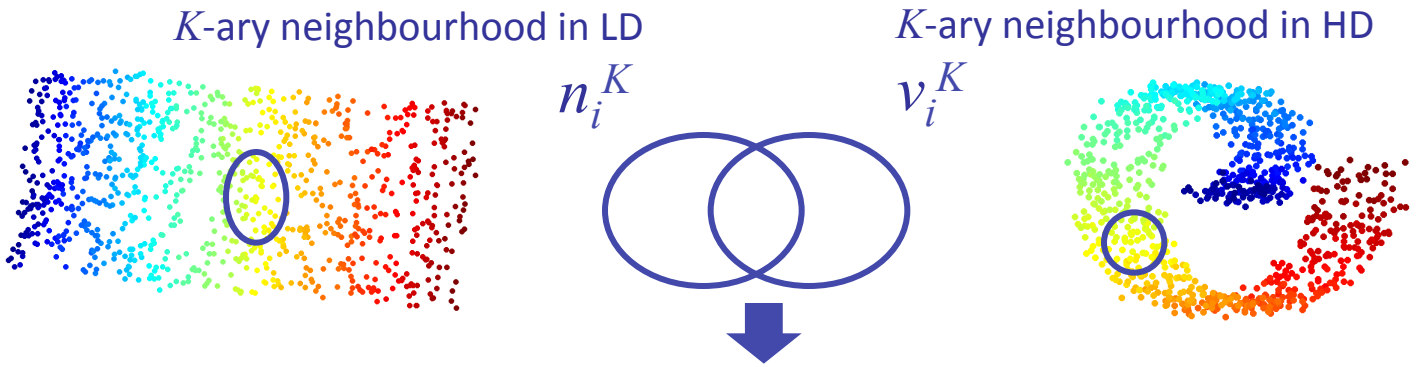


$$\mathbf{X} = [\mathbf{x}_i]_{1 \le i \le N}$$

NLDR through time…



Multi-scale quality assessment

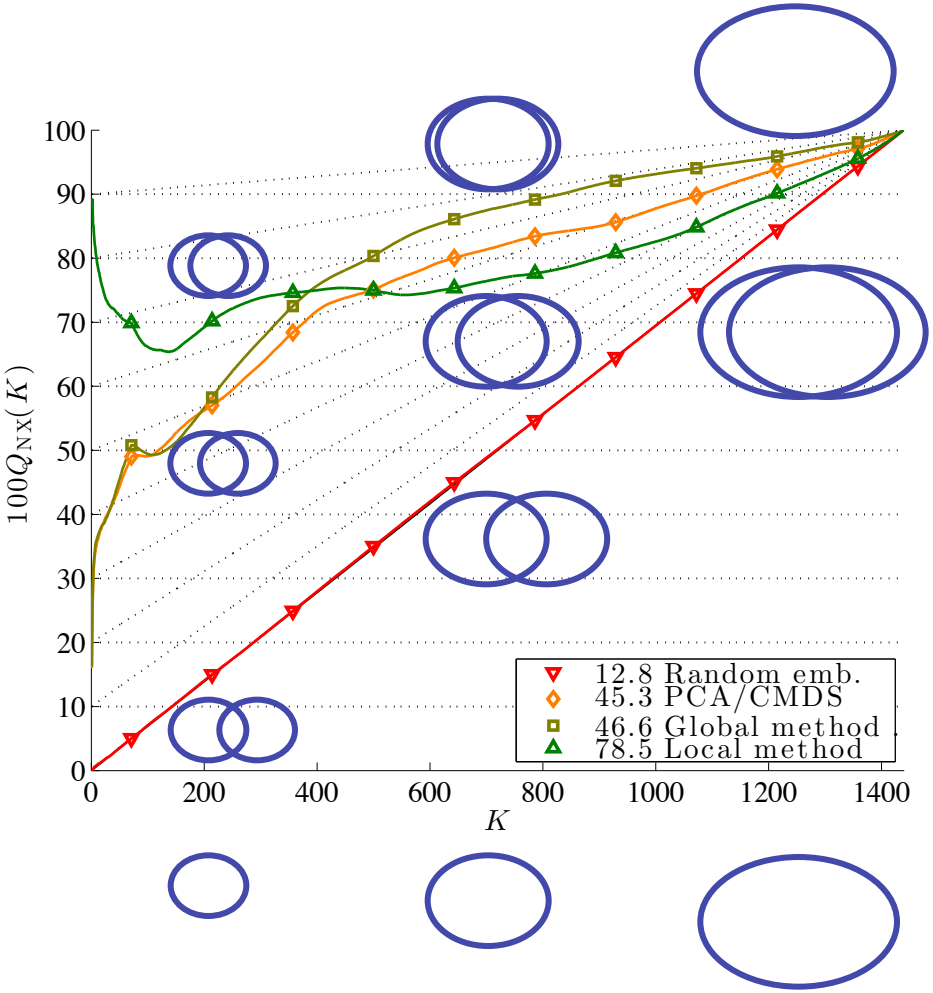


$$Q_{\mathrm{NX}}(K) = \sum_{i=1}^{N} \frac{|\nu_i^K \cap n_i^K|}{KN}$$

Average agreement of the K-ary neighbourhoods

Multi-scale quality assessment

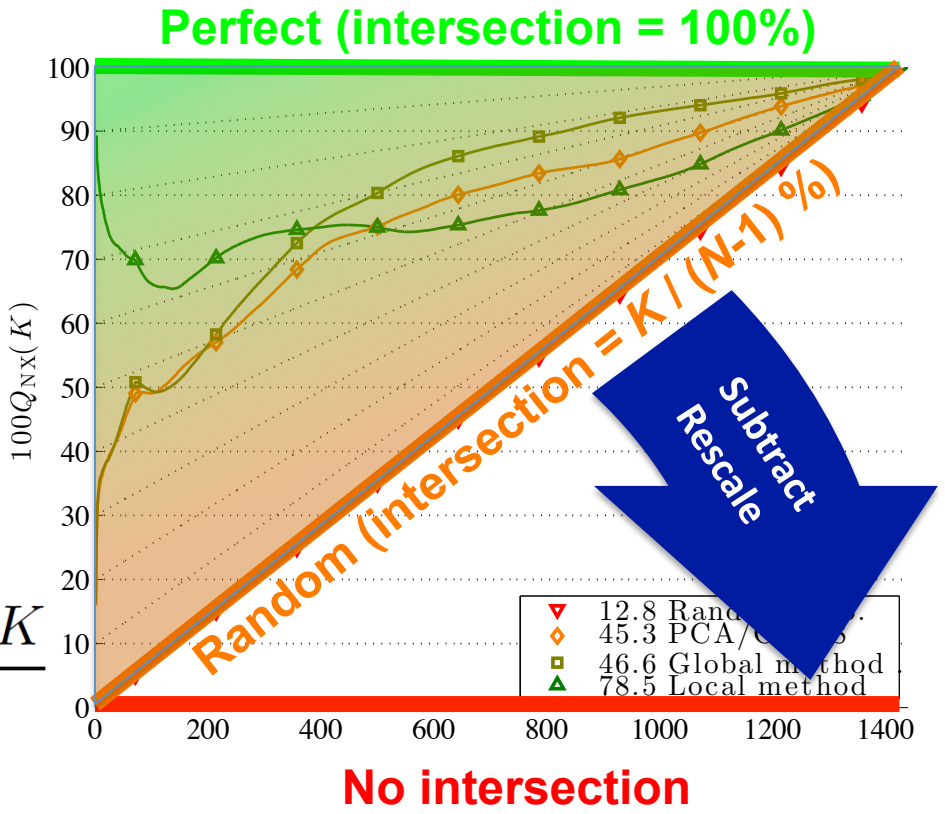
$$Q_{\mathrm{NX}}(K) = \sum_{i=1}^{N} \frac{|\nu_i^K \cap n_i^K|}{KN}$$



Multi-scale quality assessment

$$Q_{\mathrm{NX}}(K) = \sum_{i=1}^{N} \frac{|\nu_i^K \cap n_i^K|}{KN}$$

$$R_{\mathrm{NX}}(K) = \frac{(N-1)Q_{\mathrm{NX}}(K) - K}{N-1-K}$$

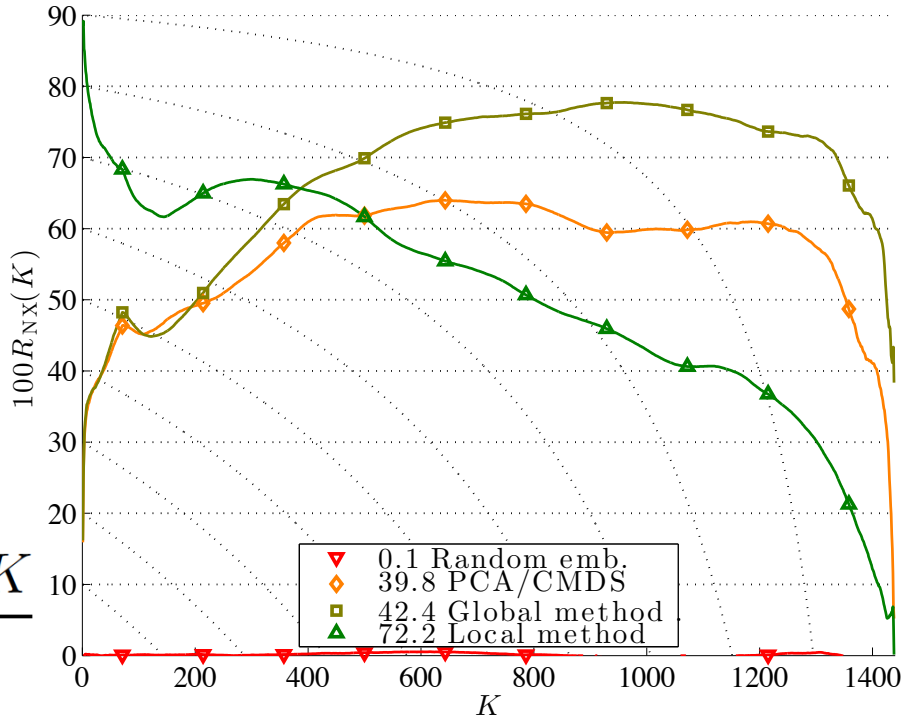


Relative quality between a perfect embedding and a random one

Multi-scale quality assessment

$$Q_{\mathrm{NX}}(K) = \sum_{i=1}^{N} \frac{|\nu_i^K \cap n_i^K|}{KN}$$

$$R_{\mathrm{NX}}(K) = \frac{(N-1)Q_{\mathrm{NX}}(K) - K}{N-1-K}$$

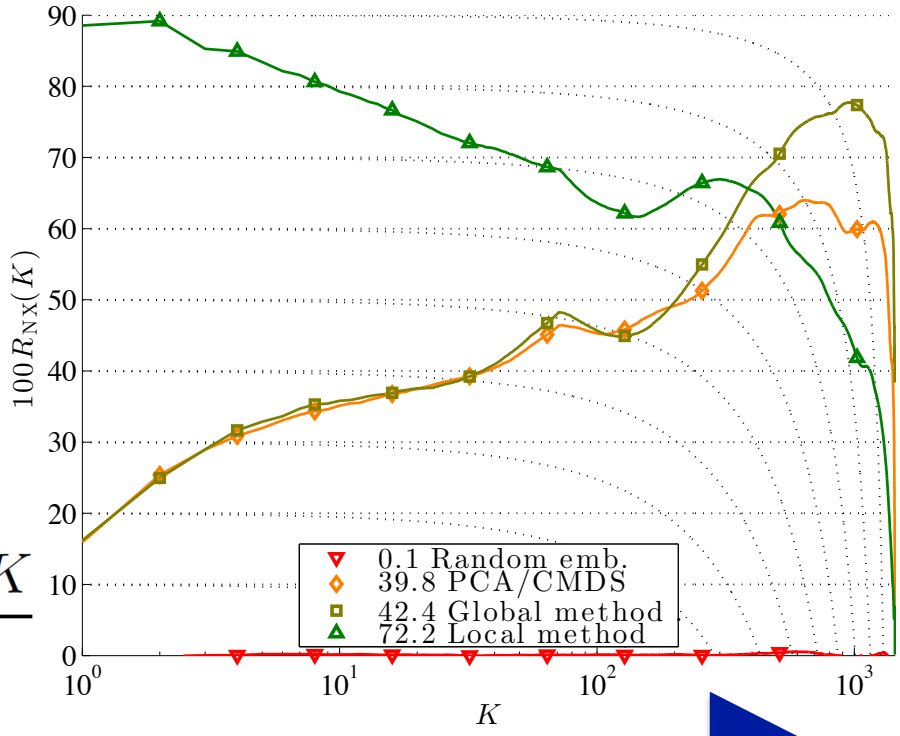


Relative quality between a perfect embedding and a random one

Multi-scale quality assessment

$$Q_{\mathrm{NX}}(K) = \sum_{i=1}^{N} \frac{|\nu_i^K \cap n_i^K|}{KN}$$

$$R_{\mathrm{NX}}(K) = \frac{(N-1)Q_{\mathrm{NX}}(K) - K}{N-1-K}$$



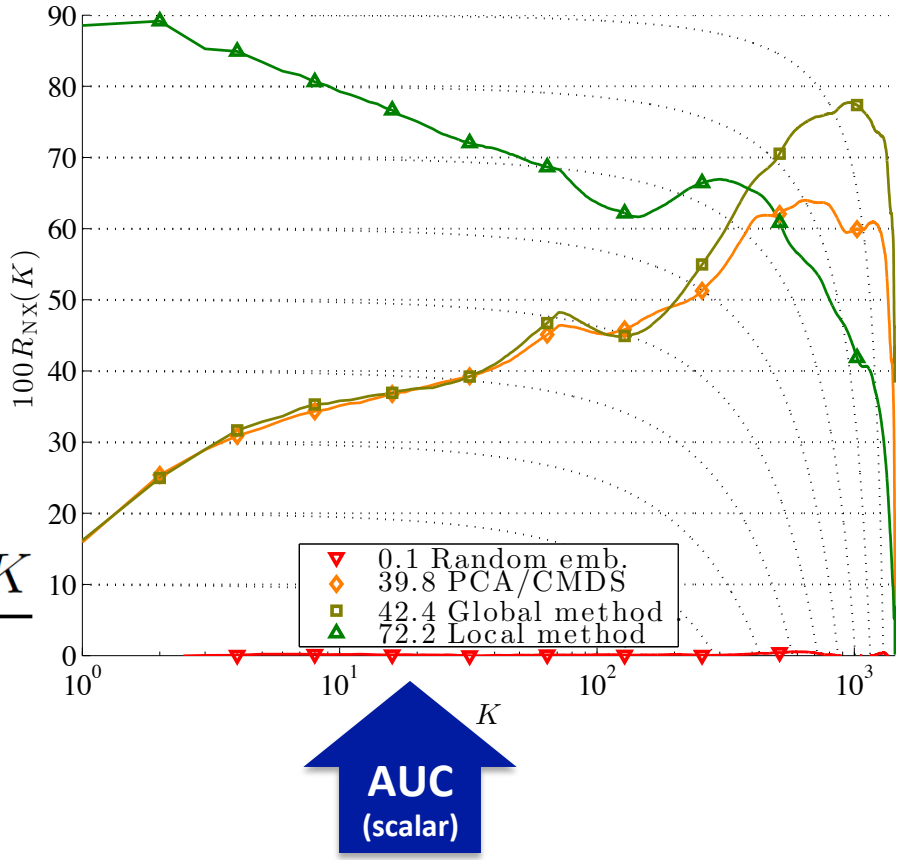
Log scale

Exponential relationship between the size K and radius r of a K-ary neighborhood in a uniform P-dimensional distribution: K proportional to r^P

Multi-scale quality assessment

$$R_{\mathrm{NX}}(K) = \frac{(N-1)Q_{\mathrm{NX}}(K) - K}{N-1-K}$$

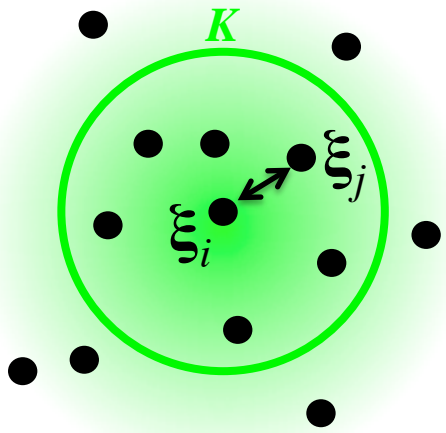
$$\mathrm{AUC} = \frac{\sum_{K=1}^{N-2} R_{\mathrm{NX}}(K)/K}{\sum_{K=1}^{N-2} 1/K}$$



Student *t*-distributed

Stochastic neighbour embedding

1. Choose size K of neighbourhoods in HD space



$$\delta_{ij} = \|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j\|_2$$

2. Convert hard neighbourhoods into soft ones

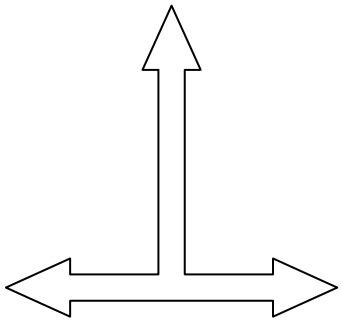
$$\sigma_{ij} = \frac{\exp(-\delta_{ij}^2/(2\lambda_i^2))}{\sum_{k,k\neq i}\exp(-\delta_{ik}^2/(2\lambda_i^2))}$$

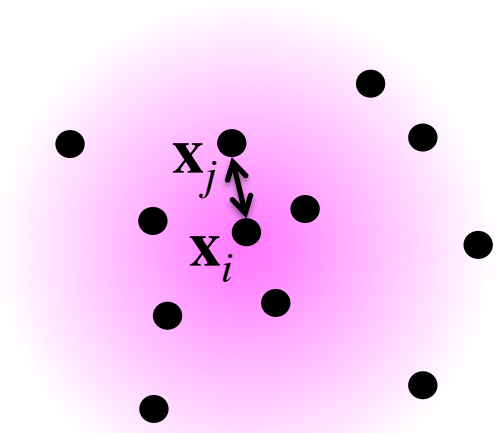
3. Adjust all bandwidths (same entropies for all i)

$$\log(K) = -\sum_{j=1}^{N}\sigma_{ij}\log\sigma_{ij}$$

5. Minimise KL divergences (for all i)

$$D_{\mathrm{KL}}(\boldsymbol{\sigma}_i\|\mathbf{s}_i) = \sum_{j=1}^{N}\sigma_{ij}\log(\sigma_{ij}/s_{ij})$$





$$d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$$

4. Define soft neighbourhoods in LD space

~~$$s_{ij} = \frac{\exp(-d_{ij}^2/2)}{\sum_{k,k\neq i}\exp(-d_{ik}^2/2)}$$~~

(with unit bandwidths)

$$s_{ij} = \frac{(1+d_{ij}^2)^{-1}}{\sum_{k,l,k\neq l}(1+d_{kl}^2)^{-1}}$$

Beyond Kullback-Leibler...

Neighbourhood retrieval and visualisation (NeRV)

Type 1 mixture of KL divergences — Venna et al., JMLR 2010

$$D^{\beta}_{\mathrm{KLs1}}(\boldsymbol{\sigma}_i \| \mathbf{s}_i) = (1-\beta) D_{\mathrm{KL}}(\boldsymbol{\sigma}_i \| \mathbf{s}_i) + \beta D_{\mathrm{KL}}(\mathbf{s}_i \| \boldsymbol{\sigma}_i)$$

Jensen-Shannon embedding (JSE, 'Jessie')

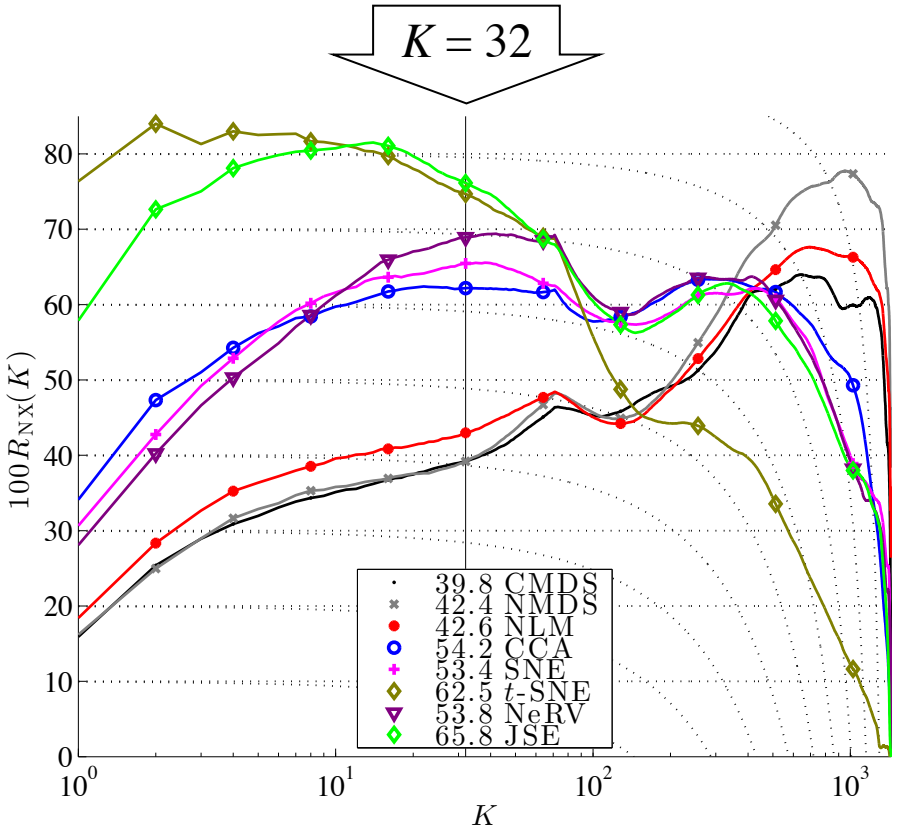
Type 2 mixture of KL divergences — Lee et al., ESANN 2012, Neurocomputing 2013

$$D^{\beta}_{\mathrm{KLs2}}(\boldsymbol{\sigma}_i \| \mathbf{s}_i) = (1-\beta) D_{\mathrm{KL}}(\boldsymbol{\sigma}_i \| \mathbf{z}_i) + \beta D_{\mathrm{KL}}(\mathbf{s}_i \| \mathbf{z}_i)$$

$$\text{where } \mathbf{z}_i = (1-\beta)\boldsymbol{\sigma}_i + \beta \mathbf{s}_i$$

Doing Better…

but not perfect yet!



For the specified value of K,
JSE succeeds best in lifting the curve…

➔ **Use several values of K !**

Multi-scale JSE

- $K_l = 2, 4, \ldots, 2^{L_{\max}-l+1}$ with $1 \leq l \leq L \leq L_{\max} \leq \log(N/4)$
- Multi-scale similarities

  Non-weighted average of single-scale similarities

$$\sigma_{ij} = \frac{1}{L}\sum_{l=1}^{L}\sigma_{ijl} \qquad \sigma_{ijl} = \frac{\exp(-\pi_{il}\delta_{ij}/2)}{\sum_{k,k\neq i}\exp(-\pi_{il}\delta_{ik}/2)}$$

$$s_{ij} = \frac{1}{L}\sum_{l=1}^{L}s_{ijl} \qquad s_{ijl} = \frac{\exp(-p_{il}d_{ij}/2)}{\sum_{k,k\neq i}\exp(-p_{il}d_{ik}/2)}$$

- Sequential identification of the bandwidths in HD space

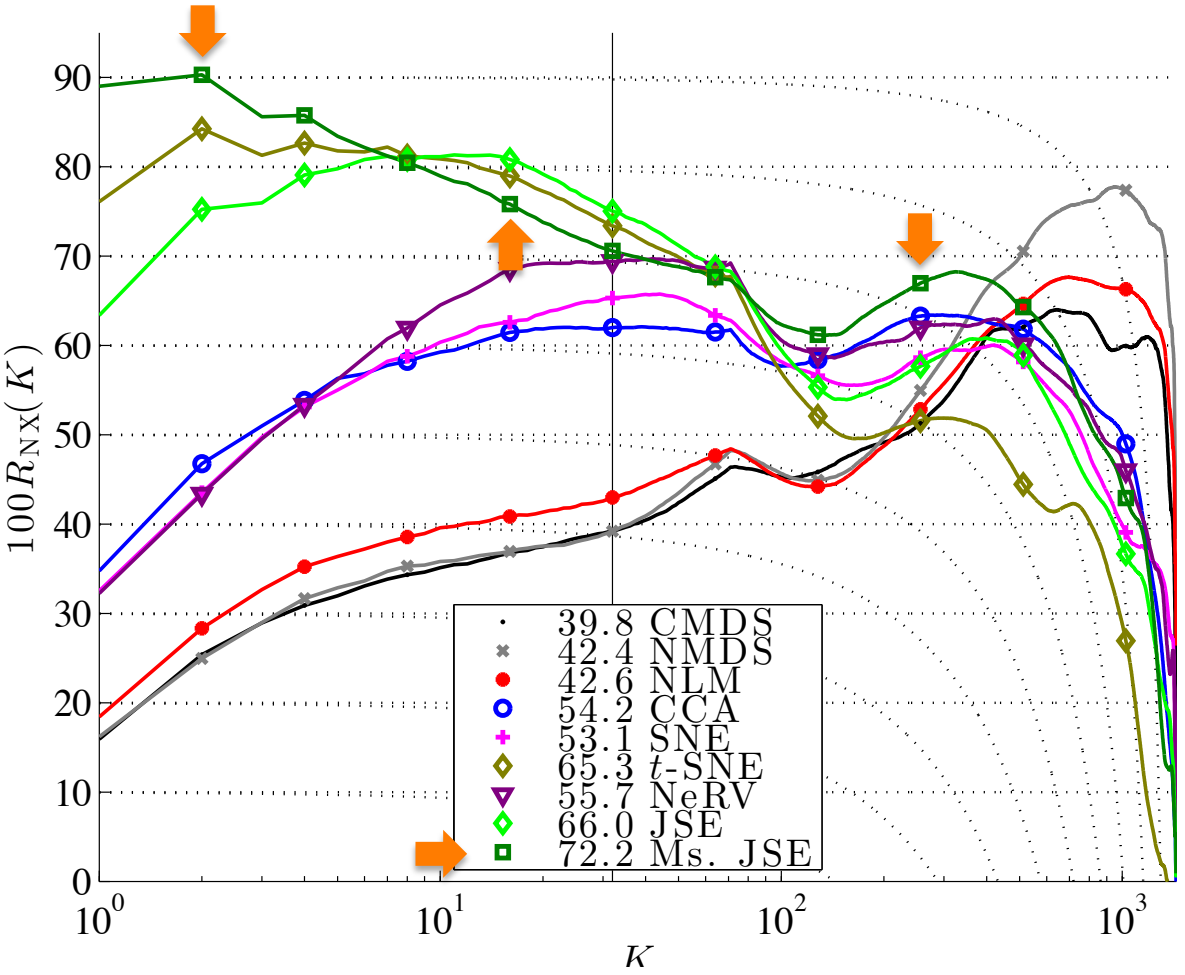
  (Usual entropy equalisation, backward from $2^{L_{\max}}$ to 2)
- A priori bandwidths in LD space: $p_{il} = K_l^{-2/P}$

  (Exponential relationship between the size and radius of a K-ary neighborhood in a uniform P-dimensional distribution)
- Multi-scale minimisation of JS divergences

  (From $L = 1$ to $L = L_{\max}$, limited memory BFGS)

Results for COIL-20

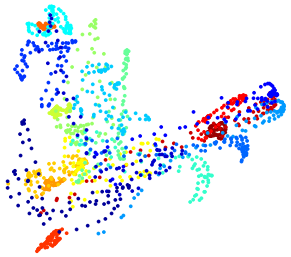
quality assessment



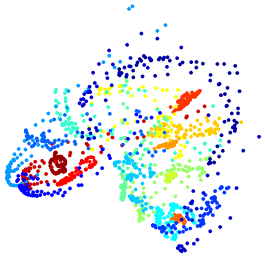
Results for COIL-20

embeddings

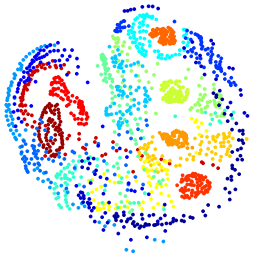
CMDS



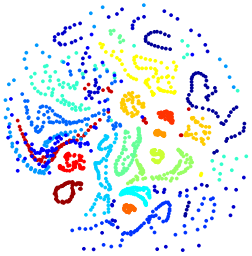
NMDS



NLM



CCA



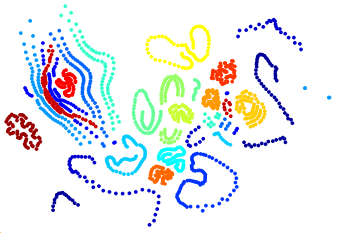
SNE

t-SNE

NeRV

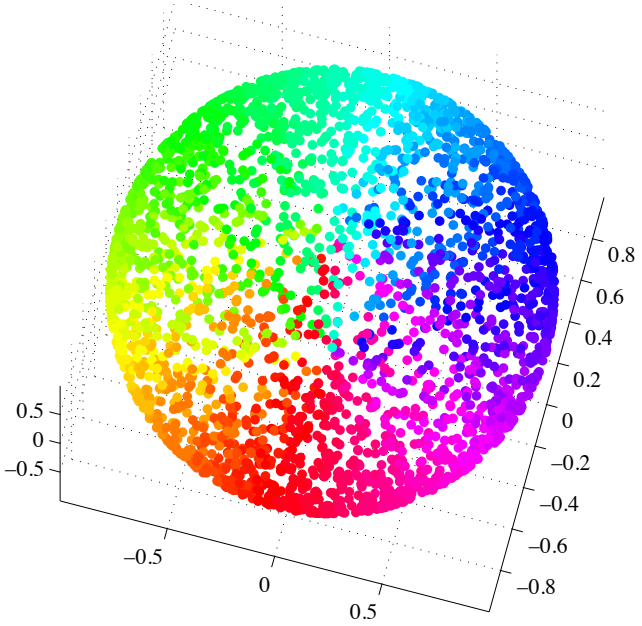
JSE

Ms. JSE

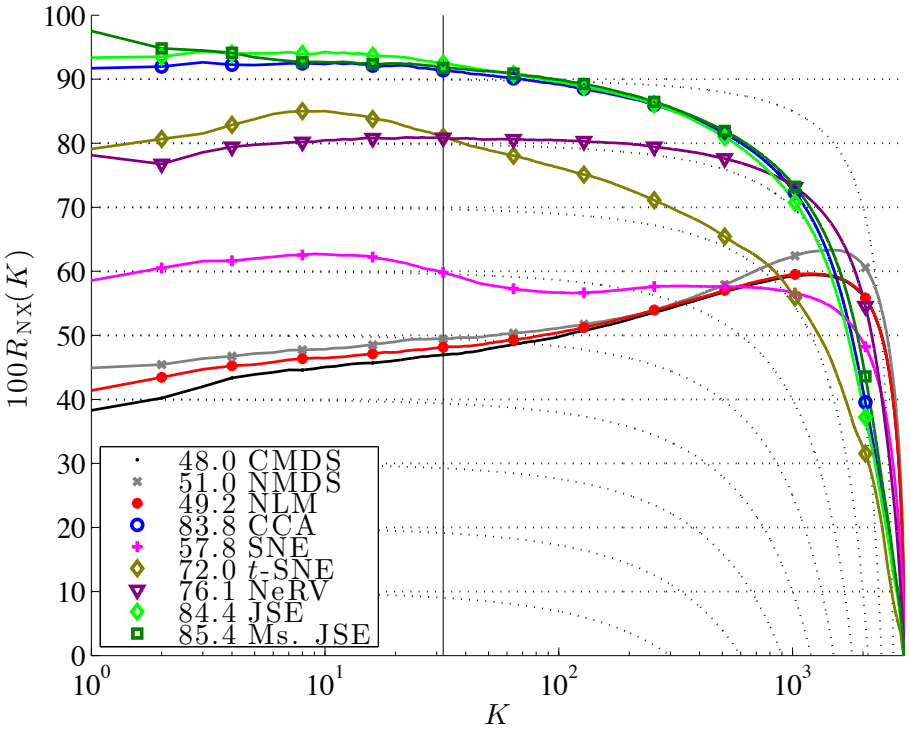


Results for 3D Sphere

quality assessment

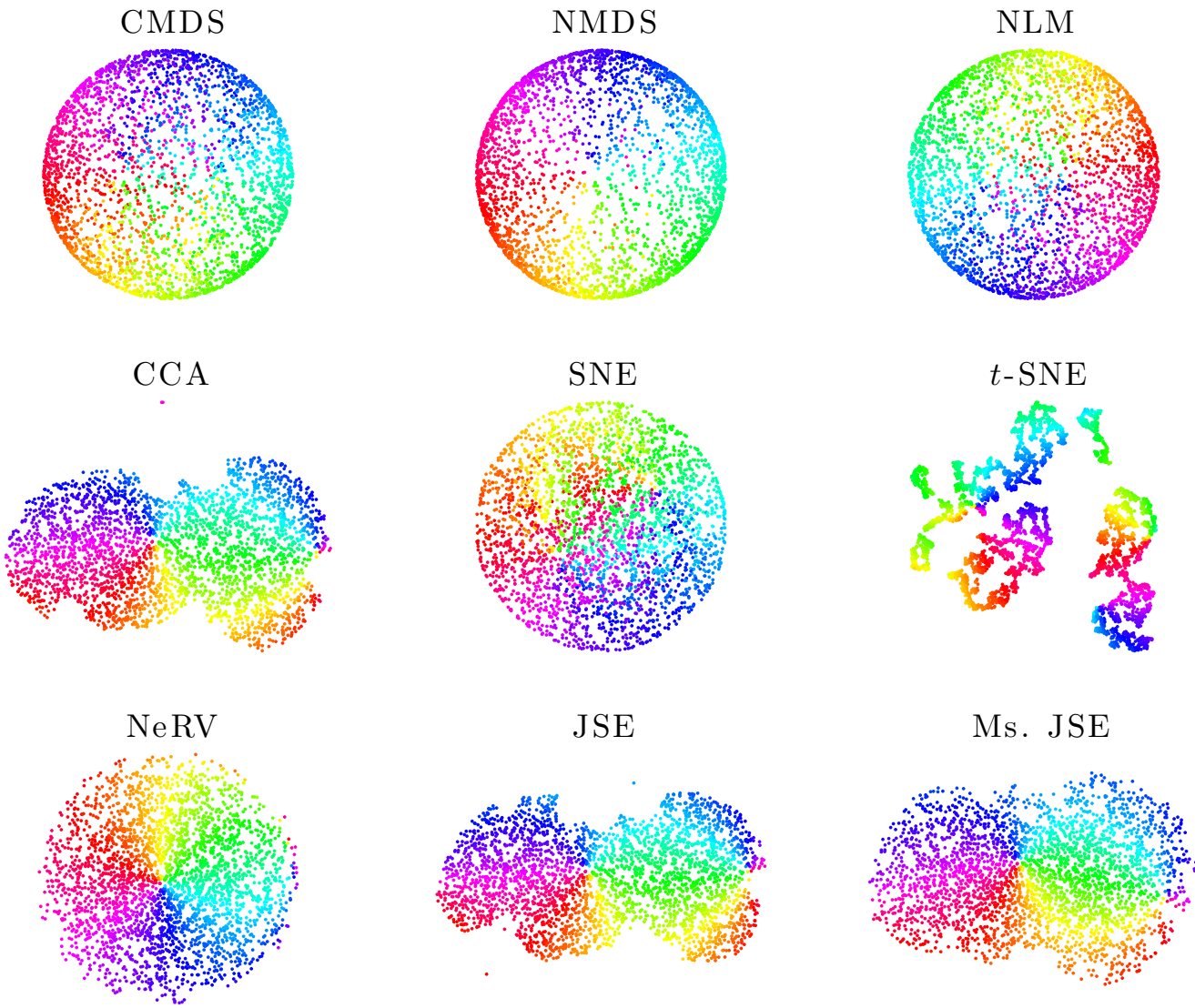


3 dimensions, $N = 3000$



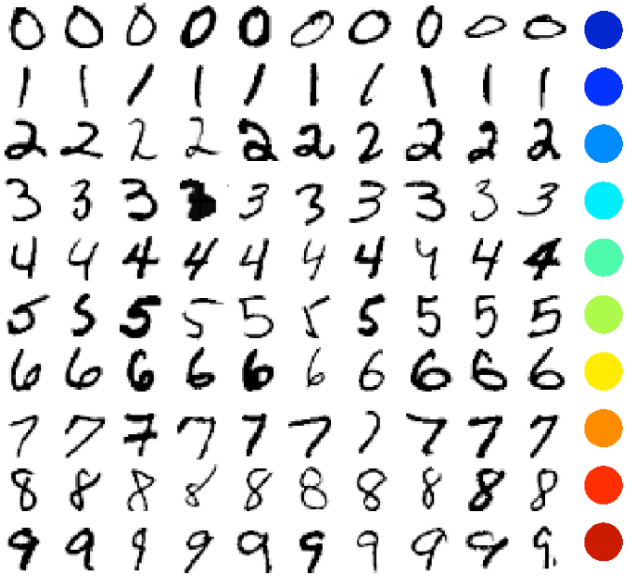
Results for 3D Sphere

embeddings

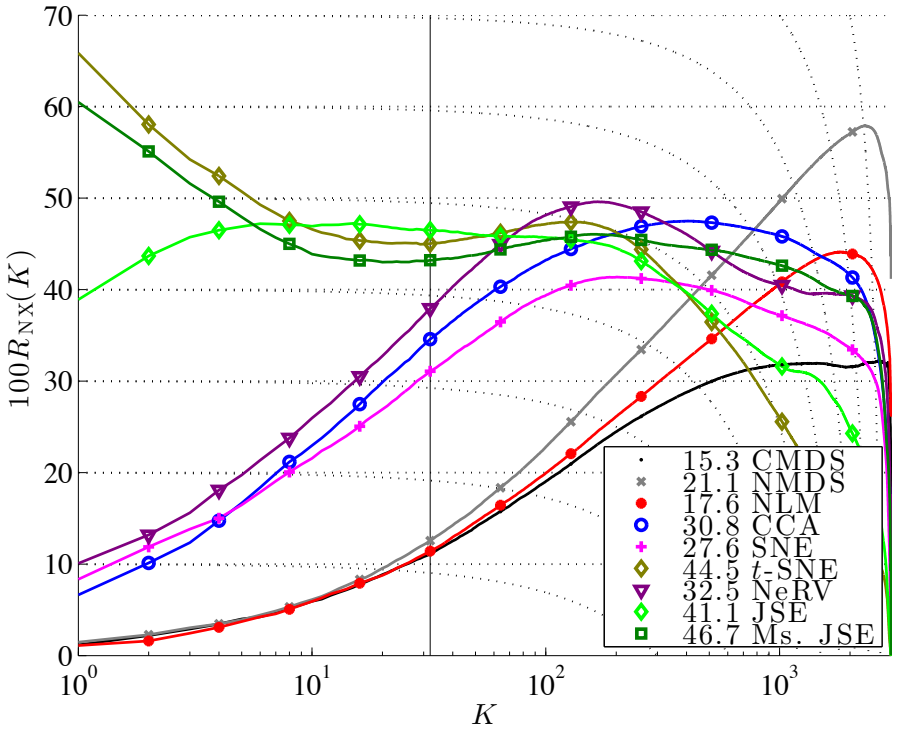


Results for MNIST digits

quality assessment

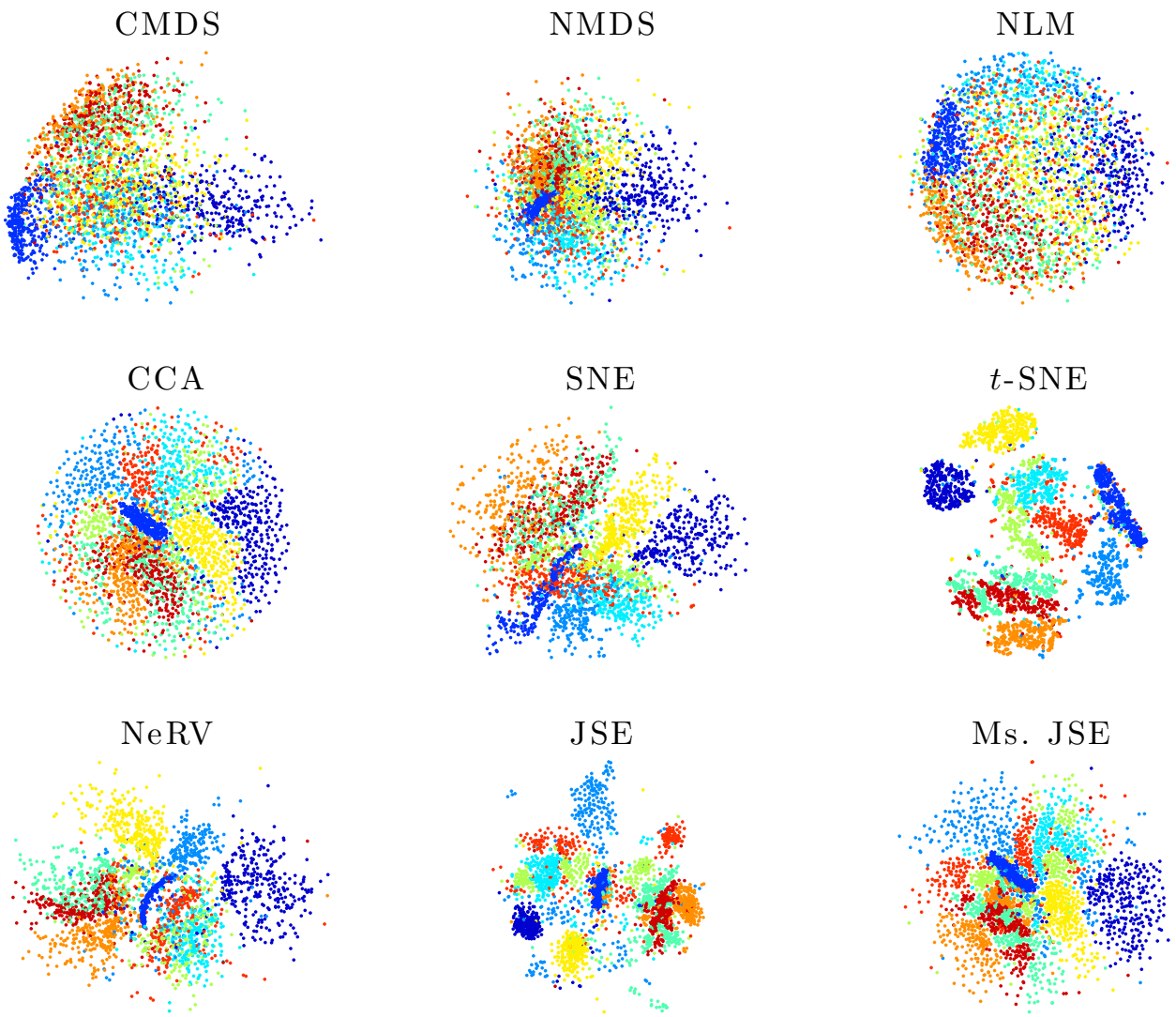


784 dimensions, $N = 3000$



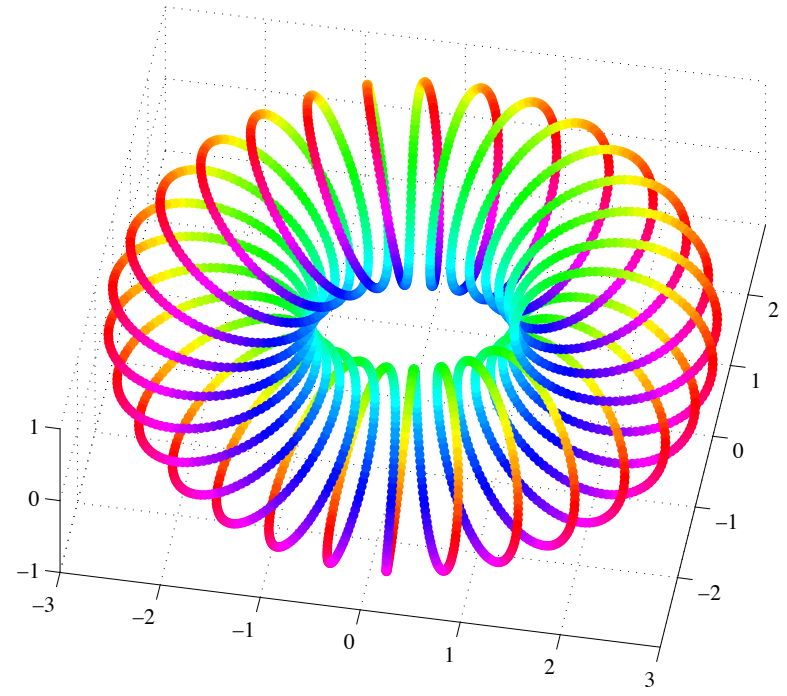
Results for MNIST digits

embeddings

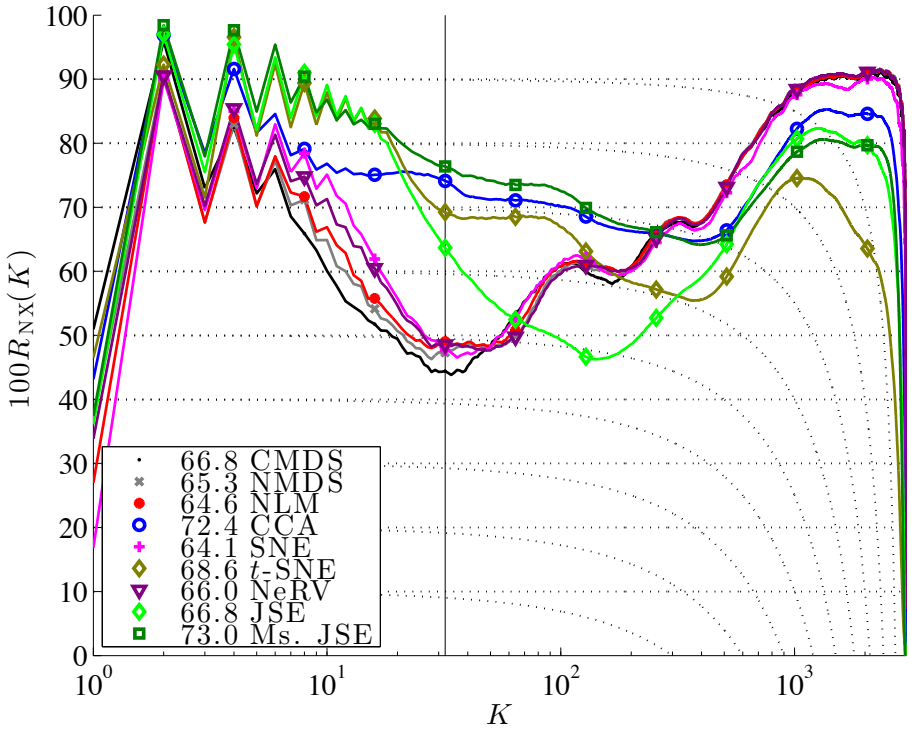


Results for Toroidal String

quality assessment



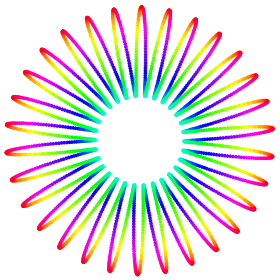
3 dimensions, $N = 3000$



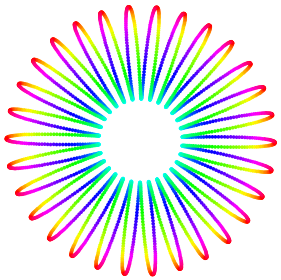
Results for Toroidal String

embeddings

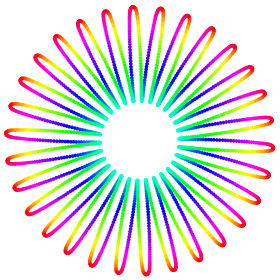
CMDS



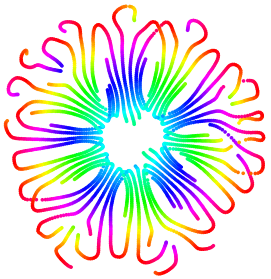
NMDS



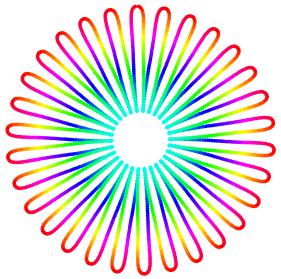
NLM



CCA

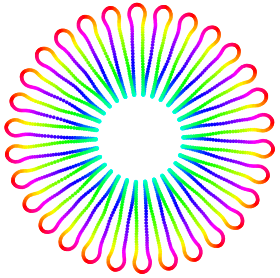


SNE

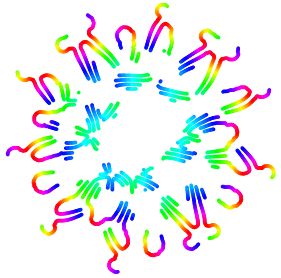


t-SNE

NeRV



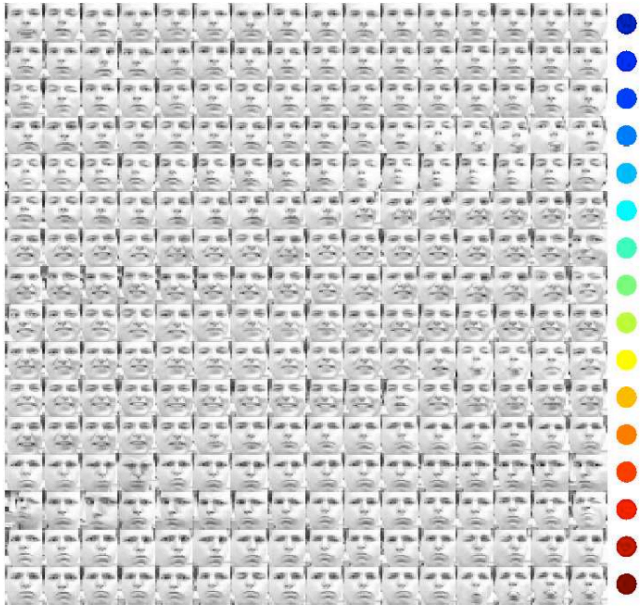
JSE



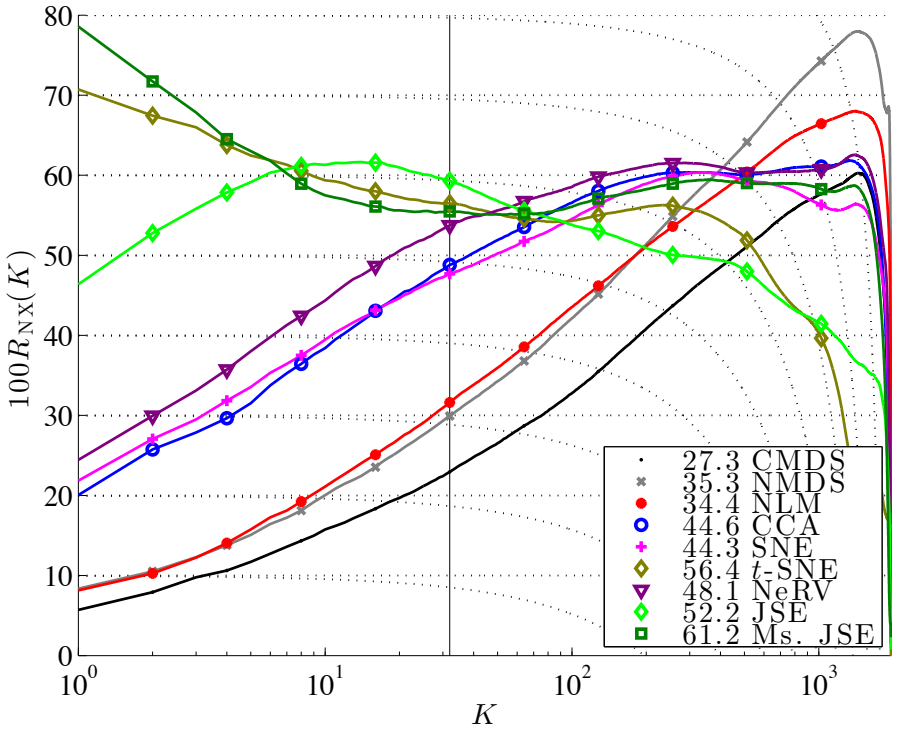
Ms. JSE

Results for B. Frey's faces

quality assessment



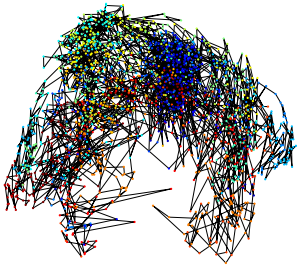
560 dimensions, $N = 1965$



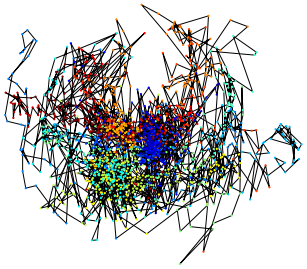
Results for B. Frey's faces

embeddings

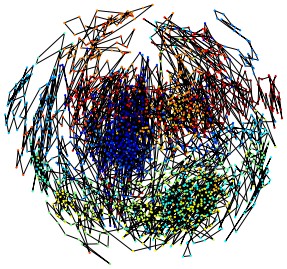
CMDS



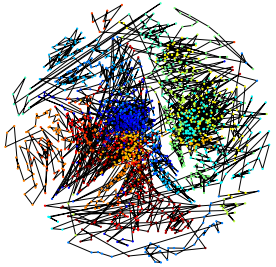
NMDS



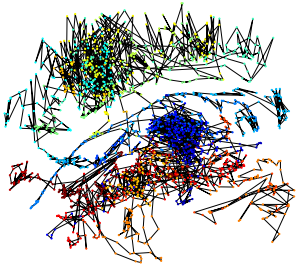
NLM



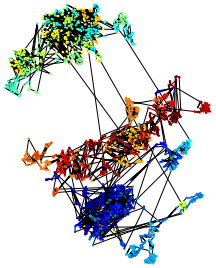
CCA



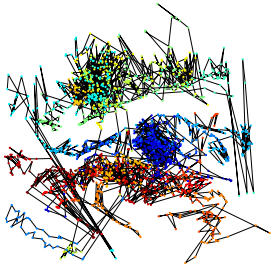
SNE



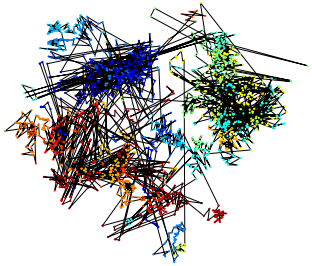
t-SNE



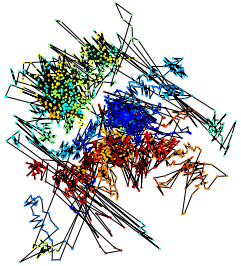
NeRV



JSE

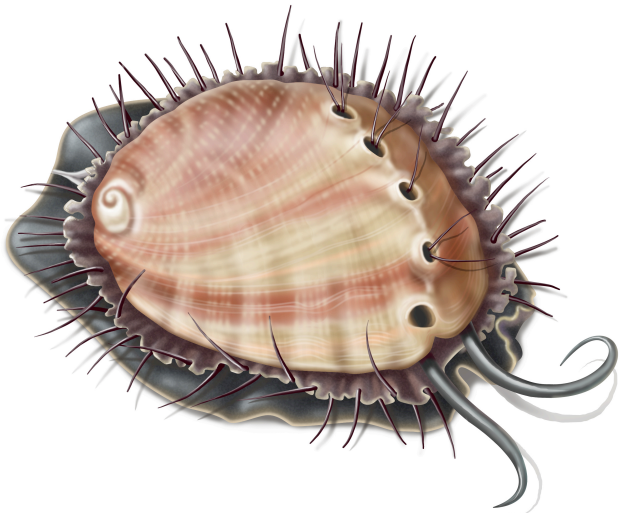


Ms. JSE

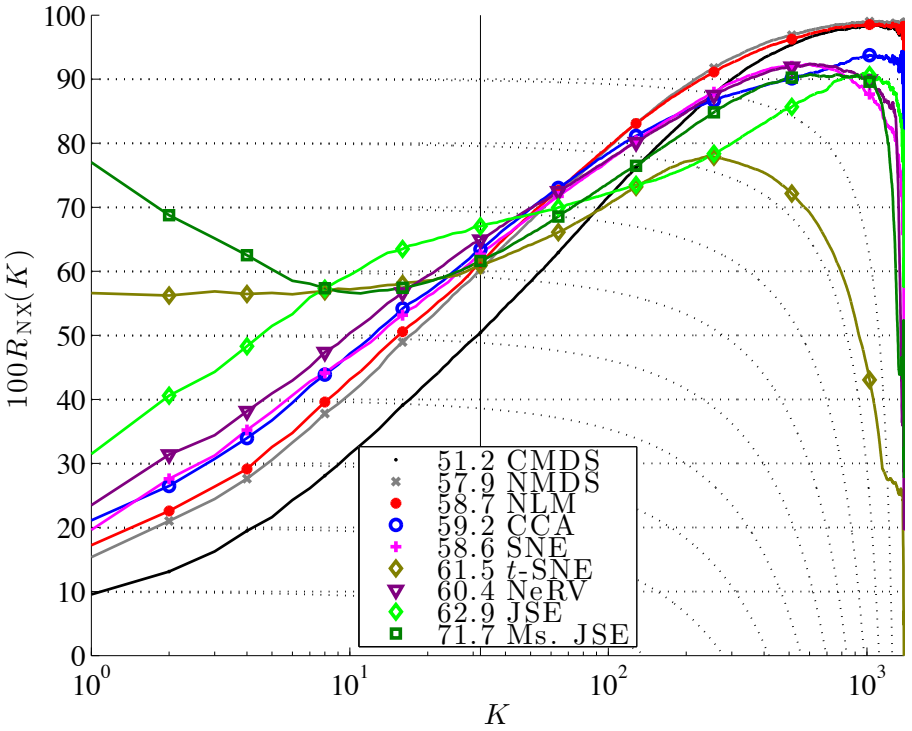


Results for Abalone

quality assessment



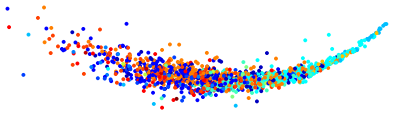
7 dimensions, $N = 1393$



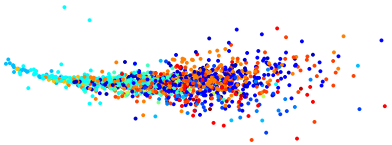
Results for Abalone

quality assessment

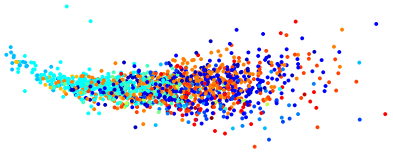
CMDS



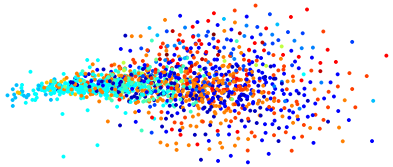
NMDS



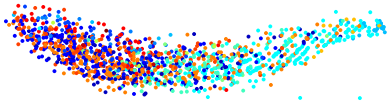
NLM



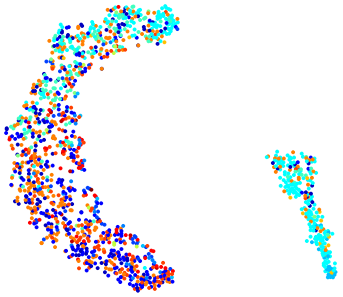
CCA



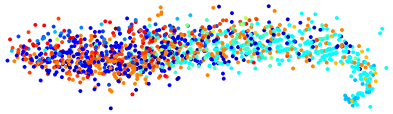
SNE



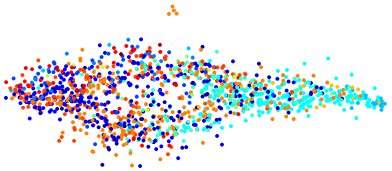
t-SNE



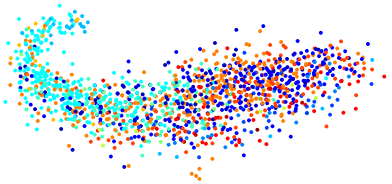
NeRV



JSE



Ms. JSE



Conclusions

and perspectives…

- Multi-scale approach
  - ☹ Slightly higher time complexity: $O(N^2 \log N)$
  - ☺ Parameter-free
  - ☺ Performance
    - Small scales: as good as *t*-SNE, often better
    - Large scales: better than *t*-SNE and NeRV
    - All scales: best results all around!
- M-code available
  compatible with gpuArray (fast!)
- Lower-complexity implementation in the future
  but multi-scale similarities are not sparse…

Thank you for your attention

Any question? Here or later… john.lee@uclouvain.be