A Probabilistic Graph Coupling View of Dimension Reduction

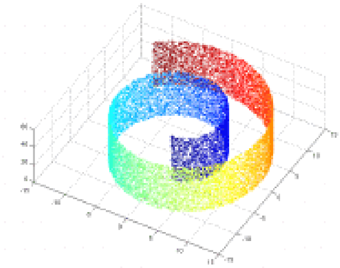
Hugues Van Assel[†], Thibault Espinasse [‡], Julien Chiquet[§], Franck Picard[†]

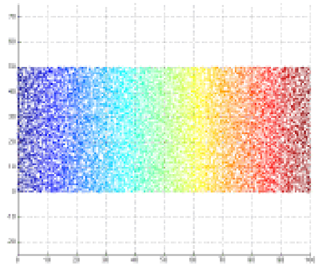
[†] Ecole Normale Supérieure de Lyon
[‡] Institut Camille Jordan, UCBL, Inria Dracula
[§] Université Paris-Saclay, AgroParisTech, INRAE

10/2021

Generalities about Dimension Reduction

- Input dataset $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ has intrinsic dimensionality $q^\star$.
- DR techniques transform $\boldsymbol{X}$ into a new dataset $\boldsymbol{Z} \in \mathbb{R}^{n \times q}$, while retaining the geometry of the data as much as possible.
- Neither the geometry of the data manifold, nor the intrinsic dimensionality $q^\star$ are known in practice (ill-posed problem).





Dimension Reduction

Spectral methods

Performs an eigendecomposition of a kernel matrix. These methods can be framed in the kernel PCA[1] framework:

- Linear : PCA, MDS
- Non-linear : Laplacian Eigenmaps, Isomap, LLE, Diffusion maps etc...

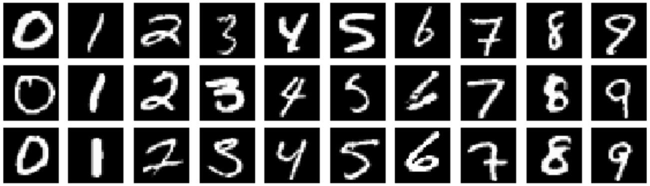
SNE-like methods

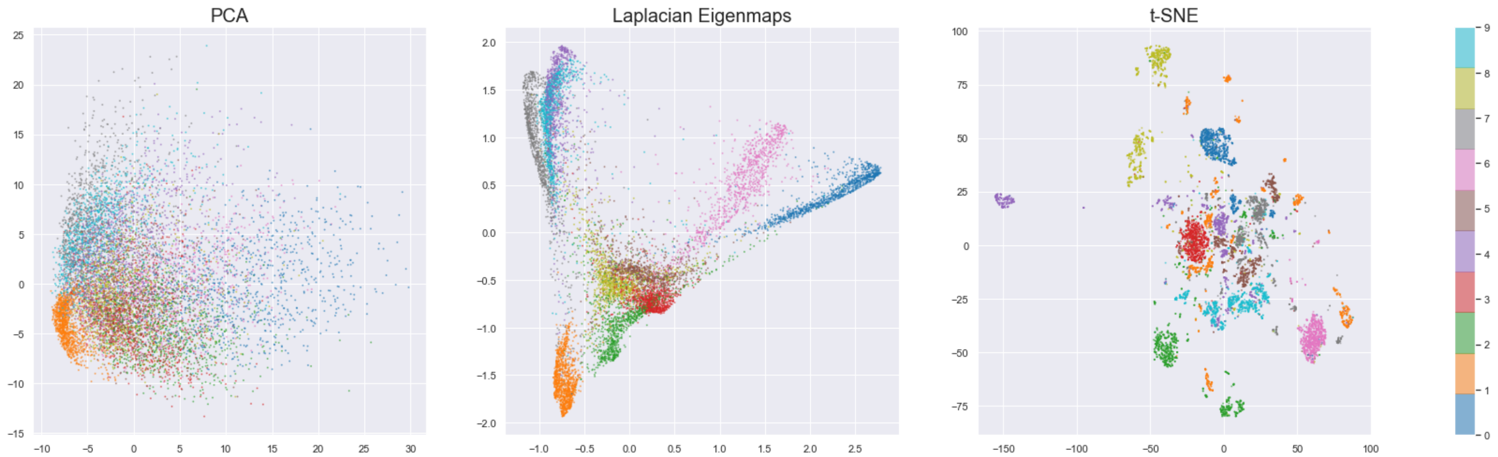
Defines similarities in both input and latent spaces and matches them through a non-convex loss optimized by gradient descent.

- SNE, t-SNE, UMAP, largeVis

[1] Ham et al. 2004.

MNIST experiments





SNE-like Methods

SNE-like methods slightly differ in the definition of the loss function:

Algorithm	Input Similarity	Latent Similarity	Loss Function
SNE	$P_{ij}^D = \frac{k_x(\mathbf{X}_i - \mathbf{X}_j)}{\sum_\ell k_x(\mathbf{X}_i - \mathbf{X}_\ell)}$	$Q_{ij}^D = \frac{k_z(\mathbf{Z}_i - \mathbf{Z}_j)}{\sum_\ell k_z(\mathbf{Z}_i - \mathbf{Z}_\ell)}$	$-\sum_{i \neq j} P_{ij}^D \log Q_{ij}^D$
Sym-SNE	$\overline{P}_{ij}^D = P_{ij}^D + P_{ji}^D$	$Q_{ij}^E = \frac{k_z(\mathbf{Z}_i - \mathbf{Z}_j)}{\sum_{\ell,t} k_z(\mathbf{Z}_\ell - \mathbf{Z}_t)}$	$-\sum_{i<j} \overline{P}_{ij}^D \log Q_{ij}^E$
LargeVis	$\overline{P}_{ij}^D = P_{ij}^D + P_{ji}^D$	$Q_{ij}^B = \frac{k_z(\mathbf{Z}_i - \mathbf{Z}_j)}{1 + k_z(\mathbf{Z}_i - \mathbf{Z}_j)}$	$-\sum_{i<j} \overline{P}_{ij}^D \log Q_{ij}^B + \left(2 - \overline{P}_{ij}^D\right) \log(1 - Q_{ij}^B)$
UMAP	$\widetilde{P}_{ij}^B = P_{ij}^B + P_{ji}^B - P_{ij}^B P_{ji}^B$	$Q_{ij}^B = \frac{k_z(\mathbf{Z}_i - \mathbf{Z}_j)}{1 + k_z(\mathbf{Z}_i - \mathbf{Z}_j)}$	$-\sum_{i<j} \widetilde{P}_{ij}^B \log Q_{ij}^B + \left(1 - \widetilde{P}_{ij}^B\right) \log(1 - Q_{ij}^B)$

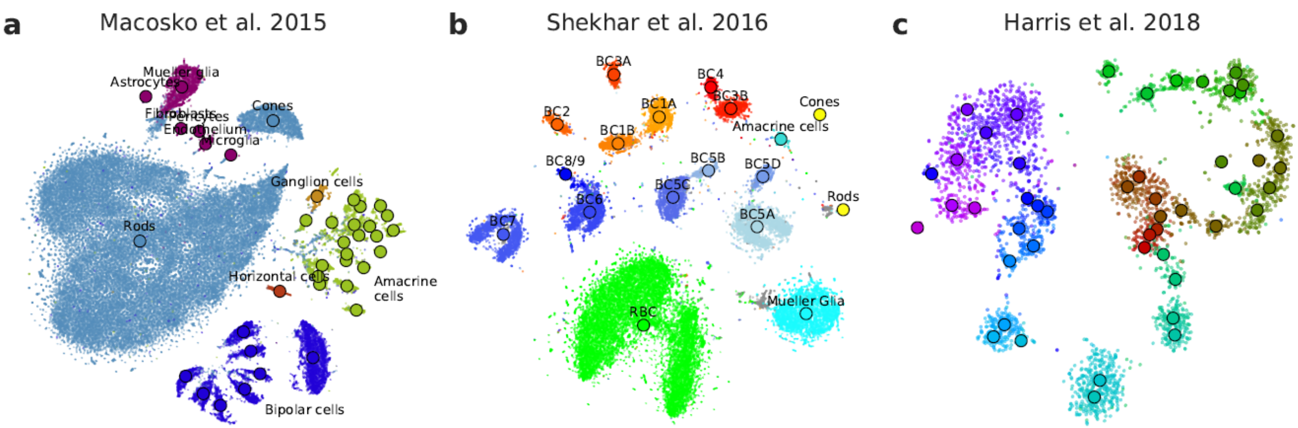
- All very good at identifying clusters.
- But the relative position of the embedded clusters can't be interpreted.

In practice

Table: Google scholar citations

SNE	t-SNE	UMAP
1672	22583	3287

Table: t-SNE on RNASeq data[2]



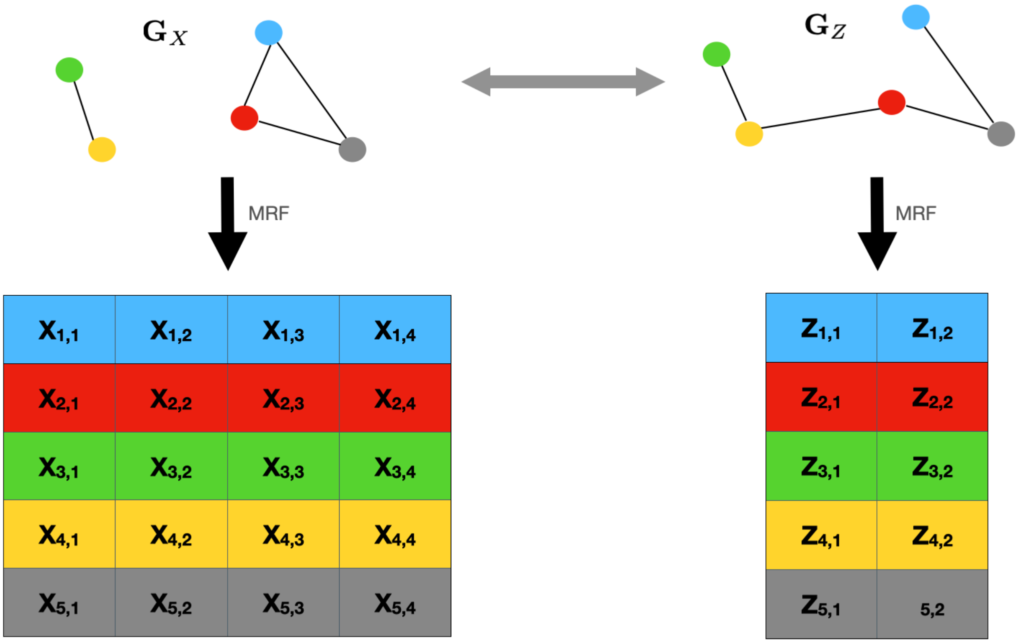
Outline

Overview of the Model

Our Model

$$\min_{\mathbf{Z}} \ \mathrm{KL}(\mathbb{P}(\mathbf{G}_X \mid \mathbf{X}) || \mathbb{P}(\mathbf{G}_Z \mid \mathbf{Z}))$$

$$\mathrm{KL}(P||Q) = \int p \log \frac{p}{q} d\lambda \quad \text{where} \quad dP = p d\lambda, \quad dQ = q d\lambda$$



General Idea

The KL amounts to minimizing the following cross-entropy:

$$\min_{\boldsymbol{Z}} \quad -\mathbb{E}_{\boldsymbol{G}_X \sim \mathbb{P}(\bullet|\boldsymbol{X})} \left[\log \mathbb{P}(\boldsymbol{G}_Z = \boldsymbol{G}_X | \boldsymbol{Z})\right]$$

Applying Bayes rule:

$$\mathbb{P}(\boldsymbol{G}_X|\boldsymbol{X}) \quad \propto \quad \underbrace{\mathbb{P}(\boldsymbol{X}|\boldsymbol{G}_X)}_{\text{Likelihood}} \underbrace{\mathbb{P}(\boldsymbol{G}_X)}_{\text{Prior}}$$

- We will see that the **likelihood takes the same form across all the DR methods (pairwise MRF)**.
- **What characterize each method are the priors** considered for the latent structuring graphs $\boldsymbol{G}_X$ and $\boldsymbol{G}_Z$.

Pairwise Markov Random Field Likelihood

$$\mathbb{P}(\boldsymbol{X}|\boldsymbol{G}) \propto \prod_{i \overset{\boldsymbol{G}}{\sim} j} \Psi_{ij}(\boldsymbol{X}_i, \boldsymbol{X}_j)$$

Hammersley - Clifford theorem:[3]

If a probability density can be factorized over the cliques of $\boldsymbol{G}$ then it satisfies the Markov properties with respect to $\boldsymbol{G}$:

- two nodes that are not connected are conditionally independent given all other nodes.
- if A, B and C are disjoint subsets of nodes such that C separates A from B, then the distribution satisfies : $A \perp\!\!\!\perp B | C$.

[3] Besag 1974.

Gaussian MRF

Let $\boldsymbol{\Theta} \in \mathcal{S}^n_{++}(\mathbb{R})$. Consider the Gaussian potential:

- $\Psi_{ij}(\boldsymbol{X}_i, \boldsymbol{X}_j) = \exp(-\frac{1}{2}\Theta_{ij}\boldsymbol{X}_i\boldsymbol{X}_j^T)$

Then

$$\boldsymbol{X}|\boldsymbol{\Theta} \sim \mathcal{N}(0, \boldsymbol{\Theta}^{-1} \otimes \boldsymbol{I}_p)$$

Markov Properties

Conditional independence given by the zeros of $\boldsymbol{\Theta}$:

$$\boldsymbol{X}_i \perp \boldsymbol{X}_j \mid \boldsymbol{X} \backslash \{\boldsymbol{X}_i, \boldsymbol{X}_j\} \iff \boldsymbol{\Theta_{ij}} = 0$$

Dimension Reduction as Graph Coupling

Towards PCA as Graph Coupling

Starting from the Gaussian MRF with $\boldsymbol{\Theta}_X \in \mathcal{S}_{++}^n(\mathbb{R})$:

$$\boldsymbol{X} | \boldsymbol{\Theta}_X \sim \mathcal{N}(0, \boldsymbol{\Theta}_X^{-1} \otimes \boldsymbol{I}_p) \tag{1}$$

A natural prior for $\boldsymbol{\Theta}_X$ could be one that is conjugate to (1) *i.e.* the Wishart $\boldsymbol{\Theta}_X \sim \mathcal{W}(\nu, \boldsymbol{\Pi})$ defined as follows:

$$\mathbb{P}(\boldsymbol{\Theta}_X; \nu, \boldsymbol{\Pi}) \propto |\boldsymbol{\Theta}_X|^{\frac{\nu}{2}} e^{-\frac{1}{2}\langle \boldsymbol{\Pi}, \boldsymbol{\Theta}_X \rangle}$$

such that the posterior reads, choosing $\boldsymbol{\Pi} = \boldsymbol{I}_n$:

$$\boldsymbol{\Theta}_X | \boldsymbol{X} \sim \mathcal{W}(\nu + p, (\boldsymbol{I}_n + \boldsymbol{X}\boldsymbol{X}^T)^{-1})$$

PCA as Graph Coupling

Let $\nu \geq n$, $\boldsymbol{\Theta}_x \sim \mathcal{W}(\nu, \boldsymbol{I}_n)$ and $\boldsymbol{\Theta}_z \sim \mathcal{W}(\nu + p - q, \boldsymbol{I}_n)$. If $\boldsymbol{\Theta}_x$ and $\boldsymbol{\Theta}_z$ structure the rows of respectively $\boldsymbol{X}$ and $\boldsymbol{Z}$ such that:

$$\boldsymbol{X} | \boldsymbol{\Theta}_x \sim \mathcal{N}(0, \boldsymbol{\Theta}_x^{-1} \otimes \boldsymbol{I}_p)$$
$$\boldsymbol{Z} | \boldsymbol{\Theta}_z \sim \mathcal{N}(0, \boldsymbol{\Theta}_z^{-1} \otimes \boldsymbol{I}_q)$$

Then the solution of the precision coupling problem:

$$\min_{\boldsymbol{Z} \in \mathbb{R}^{n \times q}} \text{KL}(\mathbb{P}(\boldsymbol{\Theta}_x | \boldsymbol{X}) || \mathbb{P}(\boldsymbol{\Theta}_z | \boldsymbol{Z}))$$

is a PCA embedding of $\boldsymbol{X}$ with q components.

Considering the SVD $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$, the above coupling is solved for $\boldsymbol{Z}^* = \boldsymbol{U}\boldsymbol{S}_{[:q]}$ (q principal components).

Gaussian MRF with Laplacian precision

Let us consider the Gaussian kernel:

$$k(\boldsymbol{x}) = \exp\left(-\|\boldsymbol{x}\|_2^2\right)$$

Graph Laplacian

We define the map $L : \mathbb{R}_+^{n \times n} \to \mathcal{S}_+^n(\mathbb{R})$ such that for $(i,j) \in [n]^2$:

$$L(\boldsymbol{W})_{ij} = \begin{cases} -W_{ij} & \text{if } i \neq j \\ \sum_j W_{ij} & \text{otherwise} \end{cases}$$

one has, $\forall \boldsymbol{W} \in S_W$, with the notation $\overline{\boldsymbol{W}} = \boldsymbol{W} + \boldsymbol{W}^T$:

$$\sum_{i,j=1}^{n} W_{ij} \|\boldsymbol{X}_i - \boldsymbol{X}_j\|_2^2 = \operatorname{tr}\left(\boldsymbol{X}^T L(\overline{\boldsymbol{W}}) \boldsymbol{X}\right)$$

Gaussian Markov random field

We recover an improper multivariate Gaussian:

$$\begin{aligned} \mathbb{P}(\boldsymbol{X}|\boldsymbol{W}) &\propto \prod_{ij} k(\boldsymbol{X}_i - \boldsymbol{X}_j)^{W_{ij}} \\ &\propto \exp\left(-\frac{1}{2}\sum_{i,j=1}^{n} W_{ij}\|X_i - X_j\|_2^2\right) \\ &= \frac{|\boldsymbol{L}|_\star^{p/2}}{(2\pi)^{\frac{p\,\mathrm{r}(\boldsymbol{L})}{2}}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{L}\boldsymbol{X}\right)\right) \end{aligned}$$

where $\boldsymbol{L} = L(\overline{\boldsymbol{W}})$.
Hence $\boldsymbol{X}|\boldsymbol{W} \sim \mathcal{N}(0, \boldsymbol{L}^\dagger \otimes \boldsymbol{I_p})$.

Rank deficiency of L

Null space of L[4]

Let $(C_1, ..., C_R)$ be a partition of $[n]$ corresponding to the connected components (CCs) of $\overline{\boldsymbol{W}}$. The null space of $\boldsymbol{L} = L(\overline{\boldsymbol{W}})$ is spanned by the orthonormal vectors $\{\boldsymbol{U}_r\}_{r \in [R]}$ such that for $r \in [R]$,
$\boldsymbol{U}_r = \left(n_r^{-1/2} \mathbb{1}_{i \in C_r} \right)_{i \in [n]}$ with $n_r = \mathrm{Card}(C_r)$.

In particular, $\mathrm{rank}(\boldsymbol{L}) = n - \#\{\text{CCs of } \overline{\boldsymbol{W}}\}$.

$\mathcal{N}(0, \boldsymbol{L}^\dagger \otimes \boldsymbol{I_p})$ only well defines a probability on $(\ker \boldsymbol{L})^\perp \otimes \mathbb{R}^p$
i.e. is improper on $\mathbb{R}^n \otimes \mathbb{R}^p$.

[4] Chung 1997.

Degenerate MRF : generalization

Let k be even and positive, we now consider:

$$\mathbb{P}(\boldsymbol{X}|\boldsymbol{W}) \quad \propto \quad \prod_{ij} k(\boldsymbol{X}_i - \boldsymbol{X}_j)^{W_{ij}}$$

where $\boldsymbol{W} \in \mathbb{N}^{n \times n}$ (amounts to $\Psi_{ij}(\boldsymbol{X}_i, \boldsymbol{X}_j) = k(\boldsymbol{X}_i - \boldsymbol{X}_j)^{W_{ij}}$).
The above is a pairwise Markov random field associated to $\overline{\boldsymbol{W}} = \boldsymbol{W} + \boldsymbol{W}^T$, indeed since k is even:

$$\mathbb{P}(\boldsymbol{X}|\boldsymbol{W}) \quad = \quad \mathcal{C}_k(\boldsymbol{W})^{-1} \prod_{i<j} k(\boldsymbol{X}_i - \boldsymbol{X}_j)^{\overline{W}_{ij}}$$

where $\mathcal{C}_k(\boldsymbol{W}) = \int_{\mathcal{X}} \prod_{ij} k(\boldsymbol{X}_i - \boldsymbol{X}_j)^{W_{ij}} d\boldsymbol{X}$

Extension to other kernel functions

We would like to go beyond the Gaussian kernel, as heavy-tails kernels have been shown to be sometimes more efficient in DR (*e.g.* student kernel in tSNE[5]).

Shift-Invariant Pariwise MRF integrability

If k is $\mathbb{R}^p$-integrable and bounded above, then $\boldsymbol{X} \mapsto \prod_{ij} k(\boldsymbol{X}_i - \boldsymbol{X}_j)^{W_{ij}}$ is integrable on $(\ker \boldsymbol{L})^\perp \otimes \mathbb{R}^p$.

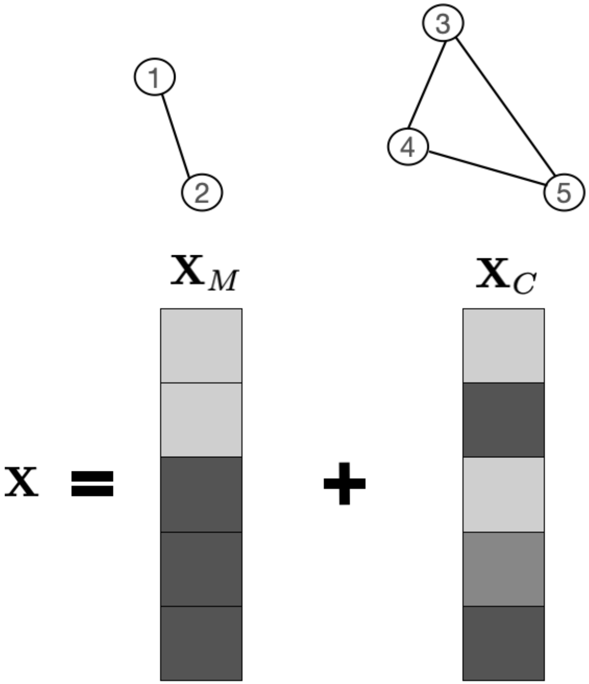
[5] van der Maaten and Hinton 2008.

Limit of proper distributions

Let $\boldsymbol{X}_M = \mathrm{Proj}_{(\ker \boldsymbol{L}) \otimes \mathbb{R}^p}(\boldsymbol{X})$ and $\boldsymbol{X}_C = \mathrm{Proj}_{(\ker \boldsymbol{L})^\perp \otimes \mathbb{R}^p}(\boldsymbol{X})$.

- $\boldsymbol{X}_M$ is the mean of $\boldsymbol{X}$ on the CCs of $\overline{\boldsymbol{W}}$.
- $\boldsymbol{X}_C$ is centered on the CCs of $\overline{\boldsymbol{W}}$.

$\boldsymbol{X}_C$ is structured by the model unlike $\boldsymbol{X}_M$ which is taken from a distribution with infinite variance.



Diffuse Model for $\boldsymbol{X}_M$

We will consider the following distribution:

$$\mathbb{P}^{\varepsilon}(\boldsymbol{X}_M|\boldsymbol{W}) \propto f^{\varepsilon}(\boldsymbol{X}_M, \boldsymbol{W})$$

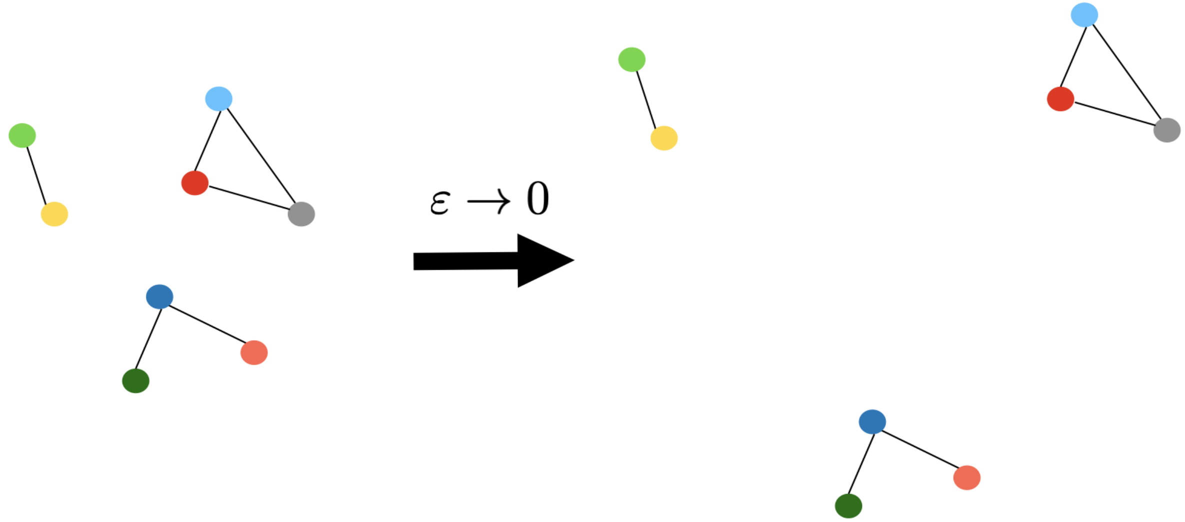
where $\forall \varepsilon > 0, f^{\varepsilon}(\cdot, \boldsymbol{W})$ is integrable on $(\ker \boldsymbol{L}) \otimes \mathbb{R}^p$ and $f^{\varepsilon}(\cdot, \boldsymbol{W}) \xrightarrow[\varepsilon \to 0]{} 1$ almost everywhere.

The likelihood is constructed as the product measure:

$$\mathbb{P}(\boldsymbol{X}|\boldsymbol{W}) = \mathbb{P}(\boldsymbol{X}_c|\boldsymbol{W}) \times \mathbb{P}^{\varepsilon}(\boldsymbol{X}_M|\boldsymbol{W})$$
$$\xrightarrow{\varepsilon \to 0} \propto \prod_{ij} k(\boldsymbol{X}_i - \boldsymbol{X}_j)^{W_{ij}}$$

Variability at the diffusion limit

$$\mathbb{V}^{\varepsilon}(\boldsymbol{X}_M | \boldsymbol{W}) \xrightarrow{\varepsilon \to 0} \infty$$



The shift-invariant pairwise MRF has a **clustering effect**.

Graph Priors (Gaussian kernel)

Definition (Laplacian Wishart distribution)

Let $\mathbf{\Pi} \in \mathbb{R}^{n \times n}$, $\nu \in \mathbb{R}$. For $\boldsymbol{W} \in \mathcal{S}_W$ we introduce the *Laplacian Wishart* distribution, denoted by $\boldsymbol{W} \sim \mathcal{LW}(\nu, \mathbf{\Pi})$:

$$\mathbb{P}(\boldsymbol{W}; \nu, \mathbf{\Pi}) \propto |L(\boldsymbol{W})|_\star^{\nu/2} e^{-\frac{1}{2}\langle \mathbf{\Pi}, \boldsymbol{W} \rangle} \Omega_{\mathcal{P}}(\boldsymbol{W}) \tag{2}$$

where $\Omega_B(\boldsymbol{W}) = \prod_{ij} \mathbb{1}_{W_{ij} \leq 1}$, $\Omega_D(\boldsymbol{W}) = \prod_i \mathbb{1}_{W_{i+}=1}$ and $\Omega_E(\boldsymbol{W}) = \mathbb{1}_{W_{++}=n} \prod_{ij}(W_{ij}!)^{-1}$ and $|\cdot|_\star$ is the pseudo determinant.

Generalization to other kernels is also possible.

Graph Posterior

Posterior limit

Let k be an integrable upper bounded function, $\boldsymbol{K} = (k(\boldsymbol{X}_i - \boldsymbol{X}_j))_{(i,j)\in[n]^2}$ and $\mathcal{P} \in \{B, D, E\}$.
If $\boldsymbol{W} \sim \mathcal{LW}(\cdot\,; 1, 1)$ then, assuming the pairwise MRF structure when $\varepsilon \to 0$:

$$\boldsymbol{W}|\boldsymbol{X} \sim \mathbb{P}^{\star}_{\mathcal{P}}(\cdot\,; \boldsymbol{K})\,.$$

If $\boldsymbol{W} \sim \mathbb{P}^{\star}_{\mathcal{P}}(\cdot\,; \boldsymbol{K})$ then:

- if $\mathcal{P} = B$, $\forall (i,j) \in [n]^2$, $W_{ij} \overset{\perp\!\!\!\perp}{\sim} \mathcal{B}\left(K_{ij}/(1 + K_{ij})\right)$.
- if $\mathcal{P} = D$, $\forall i \in [n]$, $\boldsymbol{W}_i \overset{\perp\!\!\!\perp}{\sim} \mathcal{M}\left(1, \boldsymbol{K}_i/K_{i+}\right)$.
- if $\mathcal{P} = E$, $\boldsymbol{W} \sim \mathcal{M}\left(n, \boldsymbol{K}/K_{++}\right)$.

Retrieving SNE-like Methods

For $(\mathcal{P}_X, \mathcal{P}_Z) \in \{B, D, E\}^2$, we can retrieve the losses of SNE-like methods as $\text{KL}(\mathbb{P}^{\star}_{\mathcal{P}_X}(\cdot\,;\boldsymbol{K}_X) || \mathbb{P}^{\star}_{\mathcal{P}_Z}(\cdot\,;\boldsymbol{K}_Z))$.

$\mathcal{P}_Z, \mathcal{P}_X$	B	D	E
B	UMAP		
D	LargeVis	SNE	Sym-SNE

Algorithm	Input Similarity	Latent Similarity	Loss Function
SNE	$P^D_{ij} = \frac{k_X(\boldsymbol{X}_i - \boldsymbol{X}_j)}{\sum_\ell k_X(\boldsymbol{X}_i - \boldsymbol{X}_\ell)}$	$Q^D_{ij} = \frac{k_Z(\boldsymbol{Z}_i - \boldsymbol{Z}_j)}{\sum_\ell k_Z(\boldsymbol{Z}_i - \boldsymbol{Z}_\ell)}$	$-\sum_{i \neq j} P^D_{ij} \log Q^D_{ij}$
Sym-SNE	$\overline{P}^D_{ij} = P^D_{ij} + P^D_{ji}$	$Q^E_{ij} = \frac{k_Z(\boldsymbol{Z}_i - \boldsymbol{Z}_j)}{\sum_{\ell, t} k_Z(\boldsymbol{Z}_\ell - \boldsymbol{Z}_t)}$	$-\sum_{i<j} \overline{P}^D_{ij} \log Q^E_{ij}$
LargeVis	$\overline{P}^D_{ij} = P^D_{ij} + P^D_{ji}$	$Q^B_{ij} = \frac{k_Z(\boldsymbol{Z}_i - \boldsymbol{Z}_j)}{1 + k_Z(\boldsymbol{Z}_i - \boldsymbol{Z}_j)}$	$-\sum_{i<j} \overline{P}^D_{ij} \log Q^B_{ij} + \left(2 - \overline{P}^D_{ij}\right) \log(1 - Q^B_{ij})$
UMAP	$\widetilde{P}^B_{ij} = P^B_{ij} + P^B_{ji} - P^B_{ij} P^B_{ji}$	$Q^B_{ij} = \frac{k_Z(\boldsymbol{Z}_i - \boldsymbol{Z}_j)}{1 + k_Z(\boldsymbol{Z}_i - \boldsymbol{Z}_j)}$	$-\sum_{i<j} \widetilde{P}^B_{ij} \log Q^B_{ij} + \left(1 - \widetilde{P}^B_{ij}\right) \log(1 - Q^B_{ij})$

Retrieving Laplacian Eigenmaps

Laplacian Eigenmaps as Graph Coupling

Let $\boldsymbol{W}_X \sim \mathcal{LW}(\cdot\,; 1, 1)$. Let $\nu > 0$, $\boldsymbol{\Theta}_Z \sim \mathcal{W}(\nu, \boldsymbol{I}_n)$. If $\boldsymbol{W}_X$ and $\boldsymbol{\Theta}_Z$ structure the rows of respectively $\boldsymbol{X}$ and $\boldsymbol{Z}$ such that:

$$\mathbb{P}(\boldsymbol{X}|\boldsymbol{W}_X) \propto \prod_{ij} k(\boldsymbol{X}_i - \boldsymbol{X}_j)^{W_{X,ij}}$$

$$\boldsymbol{Z}|\boldsymbol{\Theta}_Z \sim \mathcal{N}(0, \boldsymbol{\Theta}_Z^{-1} \otimes \boldsymbol{I}_q)$$

Then the solution of the precision coupling problem:

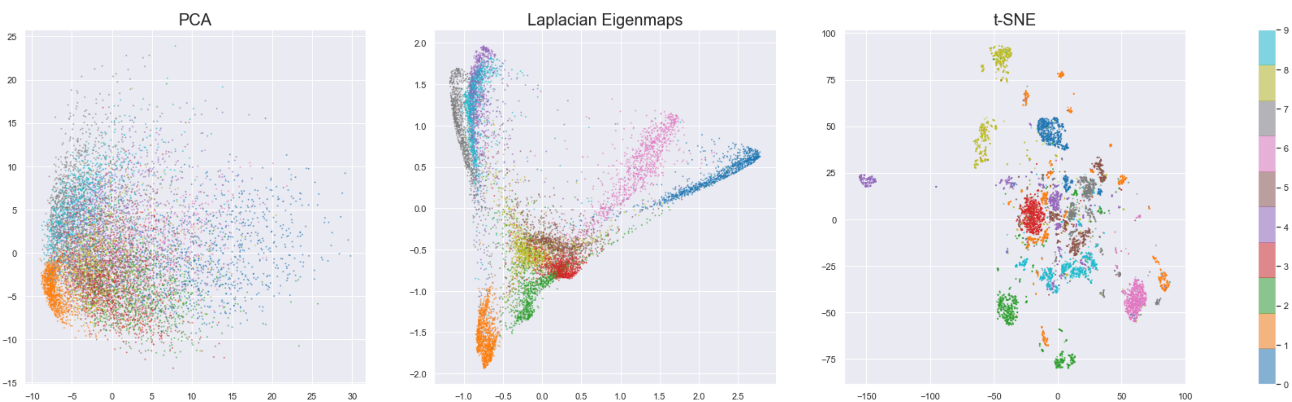
$$\min_{\boldsymbol{Z} \in \mathbb{R}^{n \times q}} \mathrm{KL}(\mathbb{P}(L(\overline{\boldsymbol{W}}_X)|\boldsymbol{X}) || \mathbb{P}(\boldsymbol{\Theta}_Z|\boldsymbol{Z}))$$

is a Laplacian Eigenmaps embedding of $\boldsymbol{X}$ with q components.

Effect of the degeneracy

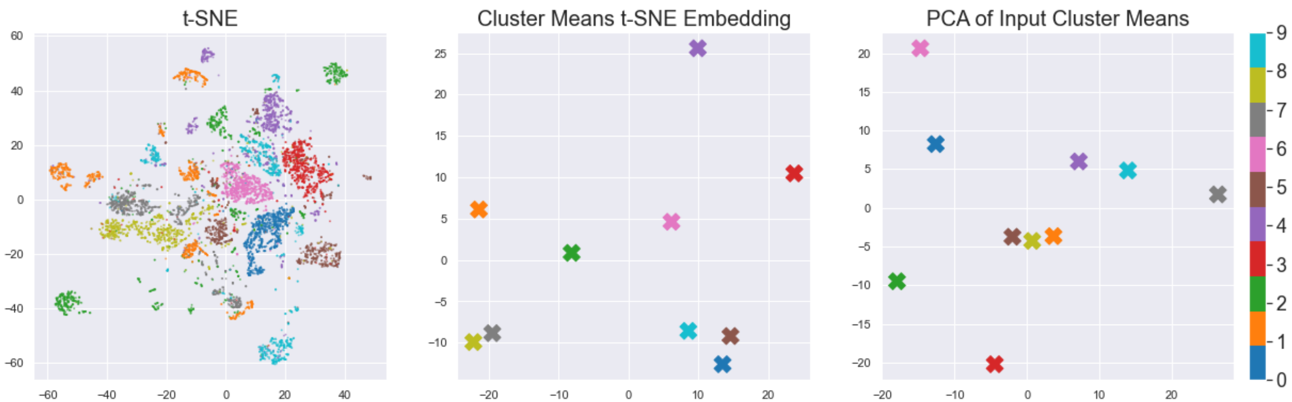
	PCA	LE	SNE
Graph model on X			
Graph model on Z			

Full-rank structure   Degeneracy (clustering effect)



Recovering Large Scale Structure in SNE

Large Scale Deficiency



Left: t-SNE embeddings initialized with i.i.d $\mathcal{N}(0, 1)$ coordinates. Middle: using these t-SNE embeddings, mean coordinates for each digit. Right: matrix of mean input coordinates for each of the 10 digits on MNIST embedded using PCA.

Towards Positioning Clusters

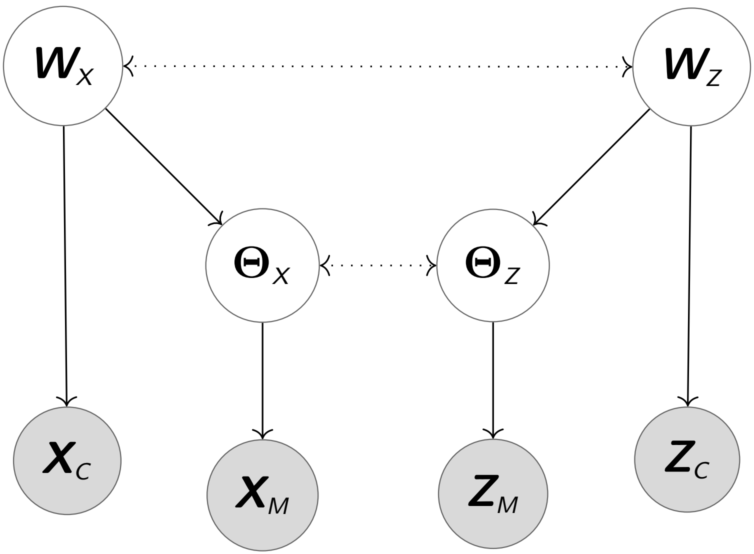


Figure: Plain directed arrows represent conditional dependencies while dotted arrows represent the coupling links. In addition to the objective considered previously between $\boldsymbol{W}_X$ and $\boldsymbol{W}_Z$, we consider a coupling between $\boldsymbol{\Theta}_X$ and $\boldsymbol{\Theta}_Z$ to structure the CCs' positions in the embeddings.

Hierarchical Graph Coupling

Let $\mathcal{P}_X \in \{B, D, E\}$, k_X is a valid kernel and $\nu_X \geq n$

$$\boldsymbol{W}_X \sim \mathbb{P}^{\varepsilon}_{\mathcal{P}_X, k_X}(\cdot\,; 1, 1) \quad (3)$$

$$\boldsymbol{X}_C | \boldsymbol{W}_X \sim \mathbb{P}_{k_X}(\cdot \,| \boldsymbol{W}_X) \quad (4)$$

$$\boldsymbol{\Theta}_X | \boldsymbol{W}_X \sim \mathcal{W}(\nu_X, \boldsymbol{I}_R) \quad (5)$$

$$\boldsymbol{X}_M | \boldsymbol{\Theta}_X \sim \mathcal{N}\left(0, \left(\varepsilon \boldsymbol{U}_{[:R]} \boldsymbol{\Theta}_X \boldsymbol{U}_{[R]}^T\right)^{-1} \otimes \boldsymbol{I}_p\right) \quad (6)$$

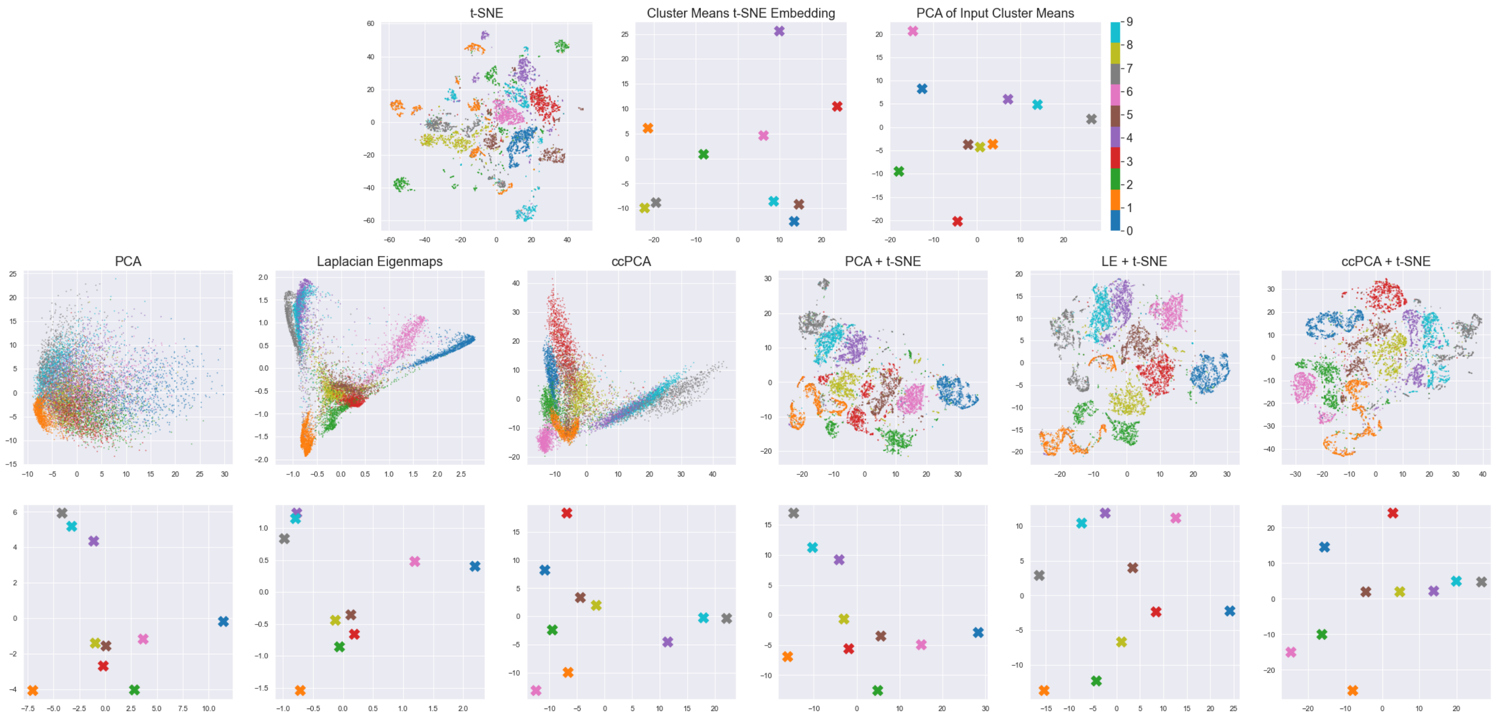
$\boldsymbol{U}_{[R]}$ are the eigenvectors associated to the Laplacian null-space of $\overline{\boldsymbol{W}}_X$. Given a graph $\boldsymbol{W}_X$, the idea is to structure the CCs' relative positions with a full-rank Gaussian model. The same model is considered for $\boldsymbol{W}_Z$, $\boldsymbol{\Theta}_Z$ and $\boldsymbol{Z}$.

Hierarchical Graph Coupling Inference

Inference in this model is performed with a heuristic consisting of two steps:

- Solve the coupling problem between $\boldsymbol{\Theta}_X$ and $\boldsymbol{\Theta}_Z$ with a PCA embedding of $\mathbb{E}_{\mathbb{P}_{\mathcal{P}_X}(\cdot;\boldsymbol{K}_X)}\left[\boldsymbol{U}_{[:R]}\boldsymbol{U}_{[R]}^T\right]\boldsymbol{X}$ (ccPCA).
- Solve the coupling problem between $\boldsymbol{W}_X$ and $\boldsymbol{W}_Z$ by running the associated DR algorithm.

ccPCA in Action



the end

Thank you!

Biblio I

- Julian Besag. "Spatial interaction and the statistical analysis of lattice systems". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2 (1974), pp. 192–225.
- F. R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.
- Jihun Ham et al. "A kernel view of the dimensionality reduction of manifolds". In: *Proceedings of the twenty-first international conference on Machine learning*. 2004, p. 47.
- Dmitry Kobak and Philipp Berens. "The art of using t-SNE for single-cell transcriptomics". In: *Nature communications* 10.1 (2019), pp. 1–14.

Biblio II

L.J.P. van der Maaten and G.E. Hinton. "Visualizing High-Dimensional Data Using t-SNE". English. In: *Journal of Machine Learning Research* 9.nov (2008). Pagination: 27, pp. 2579–2605. ISSN: 1532-4435.