Stochastic Neighbor Embedding as Graph Coupling

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

[†] Ecole Normale Supérieure de Lyon
[‡] Institut Camille Jordan
[§] AgroParisTech

10/2021

Outline

Dimension Reduction

$$\boldsymbol{X} \in \mathbb{R}^{n \times p} \rightarrow \boldsymbol{Z} \in \mathbb{R}^{n \times d}$$

with $p >> d$.

Spectral methods

Defines a kernel from a similarity, all related to kernel PCA[1]

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

Non-spectral methods

Defines similarities in both input and latent spaces and matches them

- SNE, t-SNE, UMAP, largeVis

[1] Ham et al. 2004.

Stochastic Neighbor Embedding[2]

$$\forall (i,j), \quad p_{j|i} = \frac{k_{\sigma_i}(X_i, X_j)}{\sum_j k_{\sigma_i}(X_i, X_j)}, \quad q_{j|i} = \frac{k_(Z_i, Z_j)}{\sum_j k(Z_i, Z_j)}$$

$$k_{\sigma_i}(X_i, X_j) = \exp\left(-\frac{1}{2}\frac{\|X_i - X_j\|^2}{\sigma_i^2}\right) \ \& \ k(X_i, X_j) = k_{\sigma_i = 1}(X_i, X_j)$$

SNE solves the following problem:

$$\min_{Z \in \mathbb{R}^{n \times d}} \quad \sum_i \mathrm{KL}(p_{\bullet|i} \| q_{\bullet|i})$$

where $\mathrm{KL}(p_{\bullet|i} \| q_{\bullet|i}) = \sum_{j \neq i} p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$

[2] G. E. Hinton and Roweis 2003.

Symmetric Stochastic Neighbor Embedding[3]

$$\forall (i,j), \quad \boldsymbol{p_{ij} = \frac{p_{j|i} + p_{i|j}}{2}}, \quad \boldsymbol{q_{ij} = \frac{k_(Z_i, Z_j)}{\sum_{k,\ell} k(Z_k, Z_\ell)}}$$

$p_{i|j}$ as in SNE & $k(Z_i, Z_j) = \exp\left(-\frac{1}{2}\|Z_i - Z_j\|_2^2\right)$

Symmetric SNE solves the following problem:

$$\min_{Z \in \mathbb{R}^{n \times d}} \quad \sum_{j<i} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

[3]Cook et al. 2007.

t-Distributed Stochastic Neighbor Embedding[4]

$$\forall (i,j), \quad p_{ij} = \frac{p_{j|i} + p_{i|j}}{2}, \quad q_{ij} = \frac{k_(Z_i, Z_j)}{\sum_{k,\ell} k(Z_k, Z_\ell)}$$

$p_{i|j}$ as in SNE & $\boldsymbol{k(Z_i, Z_j) = \left(1 + \|Z_i - Z_j\|_2^2\right)^{-1}}$ (crowding problem)

t-SNE solves the following problem:

$$\min_{Z \in \mathbb{R}^{n \times d}} \quad \sum_{i<j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

[4] van der Maaten and G. Hinton 2008.

Uniform Manifold Approximation and Projection[5]

$$\forall(i,j), \quad p_{j|i} = \exp\left(-\frac{\|X_i - X_j\|_2^2 - \rho_i}{\sigma_i}\right)$$

with $\rho_i = \min_{j \neq i} \|X_i - X_j\|^2$. Let us define

$$p_{ij} = p_{j|i} + p_{i|j} - p_{j|i}p_{i|j}$$

and:

$$\forall(i,j), \quad q_{ij} = \left(1 + a\|X_i - X_j\|_2^{2b}\right)^{-1}$$

UMAP solves the following problem:

$$\min_{Z \in \mathbb{R}^{n \times d}} \quad -\sum_{i<j} p_{ij} \log q_{ij} + (1 - p_{ij}) \log(1 - q_{ij})$$

[5] McInnes, Healy, and Melville 2018.

How t-SNE and UMAP perform[6]

- t-SNE and UMAP have an outstanding performance in cluster identification.
- but distance between clusters can't be interpreted.

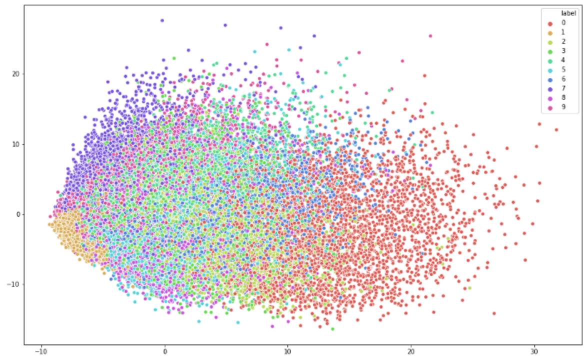


Figure 1: PCA on mnist

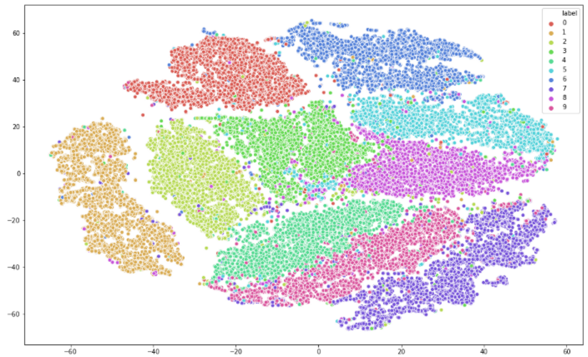


Figure 2: t-SNE on mnist

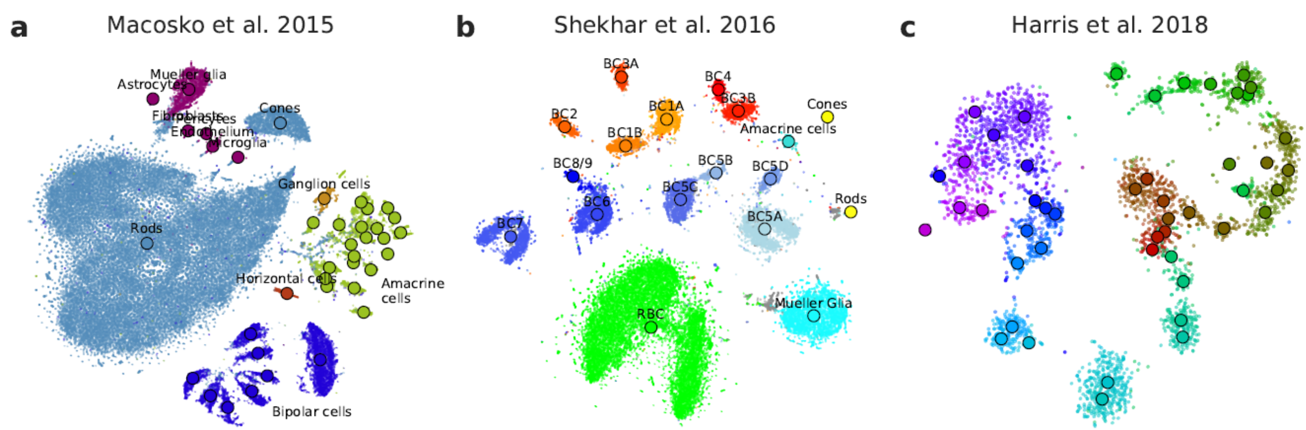
[6] Xia et al. 2021.

In practice

Table 1: Google scholar citations

SNE	t-SNE	UMAP
1672	22583	3287

Table 2: t-SNE on RNASeq data[7]



[7] Kobak and Berens 2019.

A graph coupling perspective

$$\begin{tikzcd}
W_X \arrow[rr, leftrightarrow, dotted] \arrow[d] & & W_Z \arrow[d] \\
X & & Z
\end{tikzcd}$$

Similarity matching $\iff$ Graph coupling

$$\mathcal{S}_W = \left\{ \boldsymbol{W} \in \mathbb{N}^{n \times n} \text{ s.t } \forall i, W_{ii} = 0 \right\}$$

Pairwise Markov random field

We will consider likelihoods of the form, where k is symmetric and takes positive values:

$$\mathbb{P}_k(\boldsymbol{X}|\boldsymbol{W}) \quad \propto \quad \prod_{ij} k(X_i, X_j)^{W_{ij}}$$

The above is a pairwise Markov random field associated to $\frac{\boldsymbol{W}+\boldsymbol{W^T}}{2}$, indeed since k is symmetric:

$$\begin{aligned}
\mathbb{P}_k(\boldsymbol{X}|\boldsymbol{W}) \quad &= \quad \mathcal{Z}_k(\boldsymbol{W})^{-1} \exp\left(\sum_{i,j} W_{ij} \log k(X_i, X_j)\right) \\
&= \quad \mathcal{Z}_k(\boldsymbol{W})^{-1} \exp\left(\sum_{i,j} \frac{W_{ij} + W_{ji}}{2} \log k(X_i, X_j)\right)
\end{aligned}$$

where $\mathcal{Z}_k(\boldsymbol{W}) = \int_{\mathcal{X}} \prod_{i \neq j} k(X_i, X_j)^{W_{ij}} dx$

Conditional Independence properties

$\overline{\boldsymbol{W}} = \frac{\boldsymbol{W}+\boldsymbol{W}^{\boldsymbol{T}}}{2}$ encodes conditional independence properties (Hammersley - Clifford)[8]:

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

[8] Besag 1974.

Gaussian Markov random field

Let us consider the Gaussian kernel with point specific width σ_i:

$$k_{\sigma_i}(X_i, X_j) = \exp\left(-\frac{\|X_i - X_j\|_2^2}{2\sigma_i^2}\right)$$

Introducing the unnormalized graph Laplacian:

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

one has, $\forall \boldsymbol{W} \in S_W$:

$$\sum_{i,j=1}^{n} W_{ij}\|X_i - X_j\|_2^2 = \operatorname{tr}\left(\boldsymbol{X}^T \phi_L(\boldsymbol{W} + \boldsymbol{W}^{\boldsymbol{T}})\boldsymbol{X}\right)$$

Gaussian Markov random field

We recover an improper multivariate Gaussian:

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

where $L = \phi_L(\boldsymbol{\Sigma}^{-1}\boldsymbol{W} + \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{\Sigma}^{-1})$ and $\boldsymbol{\Sigma} = \operatorname{diag}\left((\sigma_i^2)_{1\le i\le n}\right)$.
Hence $\boldsymbol{X}|\boldsymbol{W} \sim \mathcal{MN}(\boldsymbol{0}, \boldsymbol{L}^{-1}, \boldsymbol{I_p})$

Limit of proper distributions

Let $\boldsymbol{W} \in \mathcal{S}_W$, $\tilde{\boldsymbol{W}} = \boldsymbol{\Sigma}^{-1}\boldsymbol{W} + \boldsymbol{W}^T\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{L} = \boldsymbol{\phi_L}(\tilde{\boldsymbol{W}})$, let $\boldsymbol{C}$ be the number of connected components of $\tilde{\boldsymbol{W}}$. Consider the hierarchical model:

$$\begin{array}{rcl} \boldsymbol{M} & \sim & \mathcal{MN}(0, \gamma^{-2}\boldsymbol{I_{\#C}}, \boldsymbol{I_p}) \\ \tilde{M}_i & = & M_i \text{ if } X_i \in C_i \\ \boldsymbol{X} & \sim & \underbrace{\mathcal{MN}(\tilde{\boldsymbol{M}}, \boldsymbol{L^{-1}}, \boldsymbol{I_p})}_{\mathbb{P}_k^\gamma(\bullet|\boldsymbol{W})} \end{array}$$

$\mathbb{P}_k^\gamma(\boldsymbol{X}|\boldsymbol{W})$ is a proper multivariate Gaussian and one has: if $\boldsymbol{X}^\gamma \sim \mathbb{P}_k^\gamma(\bullet|\boldsymbol{W})$ and $\boldsymbol{X} \sim \mathbb{P}_k(\bullet|\boldsymbol{W})$:

$$\boldsymbol{X}^\gamma \xrightarrow{\mathcal{D}} \boldsymbol{X} \text{ when } \gamma \to 0$$

Posterior graph coupling

Applying Bayes rule:

$$\mathbb{P}_k(\boldsymbol{W}|\boldsymbol{X}) \quad \propto \quad \mathbb{P}_k(\boldsymbol{X}|\boldsymbol{W})\mathbb{P}_k(\boldsymbol{W})$$

where $\mathbb{P}_k(\boldsymbol{X}|\boldsymbol{W}) = \mathcal{Z}_k(\boldsymbol{W})^{-1} \prod_{ij} k(X_i, X_j)^{W_{ij}}$.
We will focus on coupling the posteriors:

$$\min_{Z \in \mathbb{R}^{n \times d}} \mathrm{KL}(\boldsymbol{W_X}|\boldsymbol{X} \| \boldsymbol{W_Z}|\boldsymbol{Z}) = \min_{Z \in \mathbb{R}^{n \times d}} \mathrm{H}_{\boldsymbol{W_X}|\boldsymbol{X}}(\boldsymbol{W_Z}|\boldsymbol{Z})$$

where $\mathrm{H}_{\boldsymbol{W_X}|\boldsymbol{X}}(\boldsymbol{W_Z}|\boldsymbol{Z}) = -\mathbb{E}_{\boldsymbol{W} \sim \mathbb{P}(\bullet|\boldsymbol{X})} \left[\log \mathbb{P}(\boldsymbol{W_Z} = \boldsymbol{W} \mid \boldsymbol{Z})\right]$.

Let $\boldsymbol{K_X} = (k(X_i, X_j))_{1 \leq i,j \leq n}$ and $\boldsymbol{K_Z} = (k(Z_i, Z_j))_{1 \leq i,j \leq n}$.

Heterogeneous multivariate Bernoulli

$\boldsymbol{W} \in \mathcal{S}_W$ is said to follow a heterogeneous multivariate Bernoulli distribution, with parameters $\boldsymbol{\pi} \in \mathbb{R}_+^{n \times n}$ and $\alpha \in \mathbb{R}$ if $\boldsymbol{W} \sim \mathbb{P}_k^{\mathcal{B}}(\bullet; \boldsymbol{\pi}, \alpha)$ with :

$$\mathbb{P}_k^{\mathcal{B}}(\boldsymbol{W}; \boldsymbol{\pi}, \alpha) \quad \propto \quad \mathcal{Z}_k(\boldsymbol{W})^{\alpha} \prod_{i,j} \pi_{ij}^{W_{ij}} \mathbb{1}_{\sum_k W_{ij} \leq 1}$$

Remark

Taking $\alpha = 0$ leads to independence between the edges:

$$\forall (i,j), \quad W_{ij} \sim \mathcal{B}\left(\frac{\pi_{ij}}{1 + \pi_{ij}}\right)$$

Posterior

If $\boldsymbol{W} \sim \mathbb{P}_k^{\mathcal{B}}(\bullet; \boldsymbol{\pi}, \alpha)$, then $\boldsymbol{W} | \boldsymbol{X} \sim \mathbb{P}_k^{\mathcal{B}}(\bullet; \boldsymbol{\pi} \odot \boldsymbol{K}_{\boldsymbol{X}}, \alpha - 1)$

Cross Entropy Bernoulli priors

Let k_X and k_Z be two valid pairwise potentials, let $\boldsymbol{\pi}_{\boldsymbol{X}} \in \mathbb{R}_+^{n\times n}$ and $\boldsymbol{\pi}_{\boldsymbol{Z}} \in \mathbb{R}_+^{n\times n}$.
Let $\boldsymbol{W}_{\boldsymbol{X}} \sim \mathbb{P}_{k_X}^{\mathcal{B}}(\bullet; \boldsymbol{\pi}_{\boldsymbol{X}}, \alpha = 1)$ and $\boldsymbol{W}_{\boldsymbol{Z}} \sim \mathbb{P}_{k_Z}^{\mathcal{B}}(\bullet; \boldsymbol{\pi}_{\boldsymbol{Z}}, \alpha = 1)$.
Then the cross entropy $\mathrm{H}_{\boldsymbol{W}_{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{W}_{\boldsymbol{Z}}|\boldsymbol{Z})$ writes:

$$\begin{aligned}\mathrm{H}_{\boldsymbol{W}_{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{W}_{\boldsymbol{Z}}|\boldsymbol{Z}) &= -\sum_{i\neq j} P^{\mathcal{B}}_{\pi_X,k_X}(\boldsymbol{X})_{ij} \log Q^{\mathcal{B}}_{\pi_Z,k_Z}(\boldsymbol{Z})_{ij} \\ &+ \left(1 - P^{\mathcal{B}}_{\pi_X,k_X}(\boldsymbol{X})_{ij}\right) \log\left(1 - Q^{\mathcal{B}}_{\pi_Z,k_Z}(\boldsymbol{Z})_{ij}\right)\end{aligned}$$

where $P^{\mathcal{B}}_{\pi_X,k_X}(\boldsymbol{X})_{ij} = \frac{\pi_{X,ij} k_X(X_i,X_j)}{1+\pi_{X,ij} k_X(X_i,X_j)}$ and
$Q^{\mathcal{B}}_{\pi_Z,k_Z}(\boldsymbol{Z})_{ij} = \frac{\pi_{Z,ij} k_Z(Z_i,Z_j)}{1+\pi_{Z,ij} k_Z(Z_i,Z_j)}$.

Probabilistic model of UMAP

UMAP can be recovered with:

- $\boldsymbol{\pi_X} = \boldsymbol{\pi_Z} = \mathbf{1}$
- $k_X(X_i, X_j) = \left(\exp\left(\frac{\|X_i - X_j\|_2^2 - \rho_i}{2\sigma_i^2}\right) - 1\right)^{-1}$
  $\implies P^{\mathcal{B}}_{\pi_X, k_X}(\boldsymbol{X})_{ij} = \exp\left(-\frac{\|X_i - X_j\|_2^2 - \rho_i}{2\sigma_i^2}\right)$
- $k_Z(Z_i, Z_j) = a^{-1}\|Z_i - Z_j\|_2^{-2b}$
  $\implies Q^{\mathcal{B}}_{\pi_Z, k_Z}(\boldsymbol{Z})_{ij} = \left(1 + a\|Z_i - Z_j\|_2^{2b}\right)^{-1}$
- Considering the coupling with the graph $\overline{\boldsymbol{W}} = \mathbb{1}_{\boldsymbol{W}+\boldsymbol{W}^T \geq \mathbf{1}}$, one has: $\overline{P^{\mathcal{B}}_{\pi_X, k_X}}(\boldsymbol{X})_{ij} =$
  $P^{\mathcal{B}}_{\pi_X, k_X}(\boldsymbol{X})_{ij} + P^{\mathcal{B}}_{\pi_X, k_X}(\boldsymbol{X})_{ji} - P^{\mathcal{B}}_{\pi_X, k_X}(\boldsymbol{X})_{ij} P^{\mathcal{B}}_{\pi_X, k_X}(\boldsymbol{X})_{ji}$
  by independence between the edges in the posterior.

Heterogeneous multivariate unitary outdegree

$\boldsymbol{W} \in \mathcal{S}_W$ is said to follow a heterogeneous multivariate unitary outdegree distribution, with parameters $\boldsymbol{\pi} \in \mathbb{R}_+^{n \times n}$ and $\alpha \in \mathbb{R}$ if $\boldsymbol{W} \sim \mathbb{P}_k^{\mathcal{D}}(\bullet; \boldsymbol{\pi}, \alpha)$ with:

$$\mathbb{P}_k^{\mathcal{D}}(\boldsymbol{W}; \boldsymbol{\pi}, \alpha) \quad \propto \quad \mathcal{Z}_k(\boldsymbol{W})^{\alpha} \prod_{i \neq j} \pi_{ij}^{W_{ij}} \mathbb{1}_{\sum_k W_{ik} = 1}$$

Remark

Taking $\alpha = 0$ leads to independence between the rows:

$$\forall i, \quad W_{i\bullet} \sim \mathcal{M}\left(1; \left(\frac{\pi_{ij}}{\sum_\ell \pi_{i\ell}}\right)_{1 \leq j \leq n}\right)$$

Posterior

If $\boldsymbol{W} \sim \mathbb{P}_k^{\mathcal{D}}(\bullet; \boldsymbol{\pi}, \alpha)$, then $\boldsymbol{W} | \boldsymbol{X} \sim \mathbb{P}_k^{\mathcal{D}}(\bullet; \boldsymbol{\pi} \odot \boldsymbol{K}_{\boldsymbol{X}}, \alpha - 1)$

Cross Entropy unitary outdegree priors

Let k_X and k_Z be two valid pairwise potentials, let $\boldsymbol{\pi}_{\boldsymbol{X}} \in \mathbb{R}_+^{n\times n}$ and $\boldsymbol{\pi}_{\boldsymbol{Z}} \in \mathbb{R}_+^{n\times n}$.
Let $\boldsymbol{W}_{\boldsymbol{X}} \sim \mathbb{P}_{k_X}^{\mathcal{D}}(\bullet; \boldsymbol{\pi}_{\boldsymbol{X}}, \alpha = 1)$ and $\boldsymbol{W}_{\boldsymbol{Z}} \sim \mathbb{P}_{k_Z}^{\mathcal{D}}(\bullet; \boldsymbol{\pi}_{\boldsymbol{Z}}, \alpha = 1)$.
Then the cross entropy $\mathrm{H}_{\boldsymbol{W}_{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{W}_{\boldsymbol{Z}}|\boldsymbol{Z})$ writes:

$$\mathrm{H}_{\boldsymbol{W}_{\boldsymbol{X}}|\boldsymbol{X}}(\boldsymbol{W}_{\boldsymbol{Z}}|\boldsymbol{Z}) \quad = \quad -\sum_{i\neq j} P^{\mathcal{D}}_{\pi_X, k_X}(\boldsymbol{X})_{ij} \log Q^{\mathcal{D}}_{\pi_Z, k_Z}(\boldsymbol{Z})_{ij}$$

where $P^{\mathcal{D}}_{\pi_X, k_X}(\boldsymbol{X})_{ij} = \frac{\pi_{X,ij} k_X(X_i, X_j)}{\sum_\ell \pi_{X,i\ell} k_X(X_i, X_\ell)}$ and
$Q^{\mathcal{D}}_{\pi_Z, k_Z}(\boldsymbol{Z})_{ij} = \frac{\pi_{Z,ij} k_Z(Z_i, Z_j)}{\sum_\ell \pi_{Z,i\ell} k_Z(Z_i, Z_\ell)}$.

Probabilistic model of SNE

SNE can be recovered with:

- $\boldsymbol{\pi_X} = \boldsymbol{\pi_Z} = \mathbf{1}$
- $k_X(X_i, X_j) = \exp\left(-\frac{1}{2}\frac{\|X_i - X_j\|_2^2}{\sigma_i^2}\right)$

$$\implies P^{\mathcal{D}}_{\pi_X, k_X}(\boldsymbol{X})_{ij} = \frac{\exp\left(-\frac{\|X_i - X_j\|_2^2}{2\sigma_i^2}\right)}{\sum_\ell \exp\left(-\frac{\|X_i - X_\ell\|_2^2}{2\sigma_i^2}\right)}$$

- $k_Z(Z_i, Z_j) = \exp\left(-\frac{1}{2}\|Z_i - Z_j\|^2\right)$

$$\implies Q^{\mathcal{D}}_{\pi_Z, k_Z}(\boldsymbol{Z})_{ij} = \frac{\exp\left(-\frac{1}{2}\|Z_i - Z_j\|_2^2\right)}{\sum_\ell \exp\left(-\frac{1}{2}\|Z_i - Z_\ell\|_2^2\right)}$$

Heterogeneous multivariate fixed number of edges

$\boldsymbol{W} \in \mathcal{S}_W$ is said to follow a heterogeneous multivariate fixed number of edges distribution, with parameters $\boldsymbol{\pi} \in \mathbb{R}_+^{n\times n}$, $\alpha \in \mathbb{R}$ and $N_e \in \mathbb{N}$ if $\boldsymbol{W} \sim \mathbb{P}_k^{N_e}(\bullet; \boldsymbol{\pi}, \alpha)$ with:

$$\mathbb{P}_k^{N_e}(\boldsymbol{W}; \boldsymbol{\pi}, \alpha) \quad \propto \quad \mathcal{Z}_k(\boldsymbol{W})^{\alpha} \frac{N_e!}{\prod_{i\neq j} W_{ij}!} \prod_{i\neq j} \pi_{ij}^{W_{ij}} \mathbb{1}_{\sum_{k,\ell} W_{k,\ell}=N_e}$$

Remark

Taking $\alpha = 0$ leads to the multinomial distribution:

$$\boldsymbol{W} \sim \mathcal{M}\left(N_e; \mathrm{Vect}\left(\left(\frac{\pi_{ij}}{\sum_{k,\ell}\pi_{k\ell}}\right)_{1\leq i,j\leq n}\right)\right)$$

Posterior

If $\boldsymbol{W} \sim \mathbb{P}_k^{N_e}(\bullet; \boldsymbol{\pi}, \alpha)$, then $\boldsymbol{W}|\boldsymbol{X} \sim \mathbb{P}_k^{N_e}(\bullet; \boldsymbol{\pi} \odot \boldsymbol{K}_{\boldsymbol{X}}, \alpha - 1)$

Cross Entropy between unitary outdegree prior and n edges prior

Let k_X and k_Z be two valid pairwise potentials, let $\boldsymbol{\pi_X} \in \mathbb{R}_+^{n\times n}$ and $\boldsymbol{\pi_Z} \in \mathbb{R}_+^{n\times n}$.
Let $\boldsymbol{W_X} \sim \mathbb{P}_{k_X}^{\mathcal{D}}(\bullet; \boldsymbol{\pi_X}, \alpha = 1)$ and $\boldsymbol{W_Z} \sim \mathbb{P}_{k_Z}^{N_e=n}(\bullet; \boldsymbol{\pi_Z}, \alpha = 1)$.
Then the cross entropy $\mathrm{H}_{\boldsymbol{W_X}|\boldsymbol{X}}(\boldsymbol{W_Z}|\boldsymbol{Z})$ writes:

$$\mathrm{H}_{\boldsymbol{W_X}|\boldsymbol{X}}(\boldsymbol{W_Z}|\boldsymbol{Z}) = -\sum_{i<j} \overline{P_{\pi_X,k_X}^{\mathcal{D}}}(\boldsymbol{X})_{ij} \log Q_{\pi_Z,k_Z}^{N_e}(\boldsymbol{Z})_{ij}$$

where $\overline{P_{\pi_X,k_X}^{\mathcal{D}}}(\boldsymbol{X})_{ij} = \frac{P_{\pi_X,k_X}^{\mathcal{D}}(\boldsymbol{X})_{ij} + P_{\pi_X,k_X}^{\mathcal{D}}(\boldsymbol{X})_{ji}}{2}$ and
$Q_{\pi_Z,k_Z}^{N_e}(\boldsymbol{Z})_{ij} = \frac{\pi_{Z,ij} k_Z(Z_i,Z_j)}{\sum_{k,\ell} \pi_{Z,k\ell} k_Z(Z_k,Z_\ell)}$.

Probabilistic model of Symmetric SNE

Symmetric SNE can be recovered with:

- $\boldsymbol{\pi_X} = \boldsymbol{\pi_Z} = \mathbf{1}$
- $k_X(X_i, X_j) = \exp\left(-\frac{1}{2}\frac{\|X_i - X_j\|_2^2}{\sigma_i^2}\right)$
- for Gaussian symmetric SNE:
  $k_Z(Z_i, Z_j) = \exp\left(-\frac{1}{2}\|Z_i - Z_j\|_2^2\right)$
  $\implies Q^{N_e}_{\pi_X, k_X}(\boldsymbol{X})_{ij} = \frac{\exp\left(-\frac{1}{2}\|Z_i - Z_j\|_2^2\right)}{\sum_{k,\ell}\exp\left(-\frac{1}{2}\|Z_k - Z_\ell\|_2^2\right)}$
- for t-SNE:
  $k_Z(Z_i, Z_j) = \left(1 + \|Z_i - Z_j\|_2^2\right)^{-1}$
  $\implies Q^{N_e}_{\pi_Z, k_Z}(\boldsymbol{Z})_{ij} = \frac{\left(1 + \|Z_i - Z_j\|_2^2\right)^{-1}}{\sum_{k,\ell}\left(1 + \|Z_k - Z_\ell\|_2^2\right)^{-1}}$

Future directions

How to leverage our probabilistic model?

- Better understanding of the methods:
  - Model selection
  - Clearly state the limits in terms of between-clusters variance
  - Properties of the embedding found by graph coupling
- Towards a better algorithm:
  - Position the connected components
  - Extension to other graph models and Monte Carlo or variational estimation
  - Improved kernel width estimation