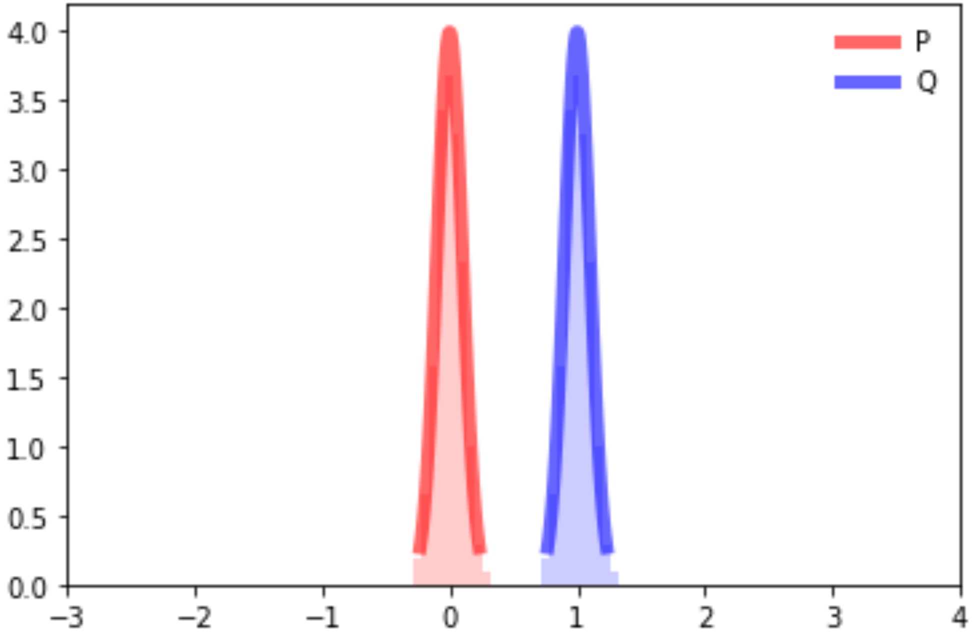
Diffusion Earth Mover's Distance and Distribution Embeddings

Alexander Tong*, Guillaume Huguet*, Amine Natik*, Kincaid MacDonald, Manik Kuchroo, Ronald Coifman, Guy Wolf**, Smita Krishnaswamy**

ICML 2021

* Denotes Equal Contribution

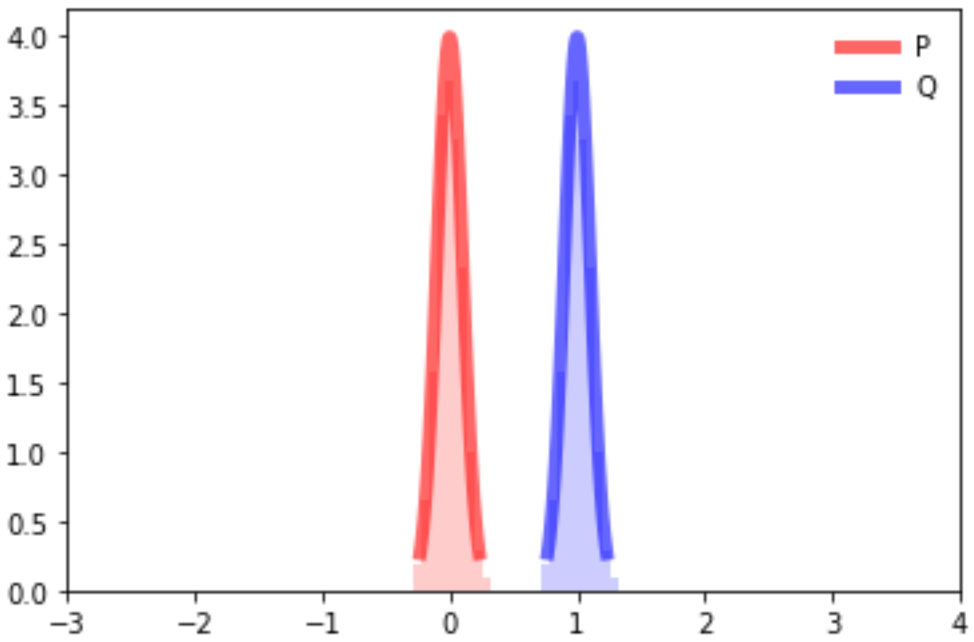
Distances between probability distributions



Total-Variation Distance (TV)

$$\mathrm{TV}(P,Q) = \frac{1}{2}\|P - Q\|_1$$

Distances between probability distributions

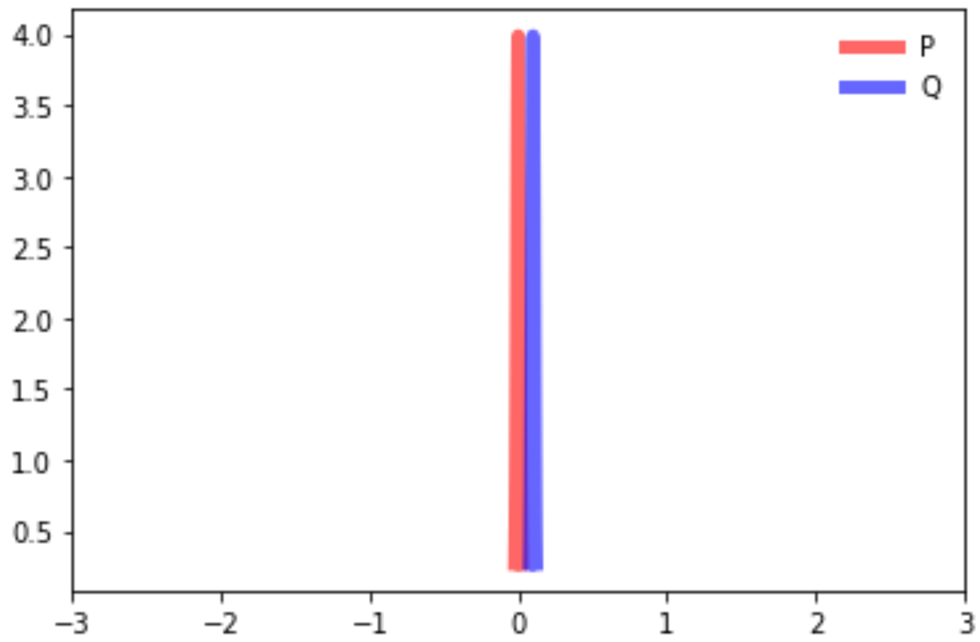


Total-Variation Distance (TV)

$$\mathrm{TV}(P,Q) = \frac{1}{2}\|P-Q\|_1$$

$$\boxed{\mathrm{TV}(P,Q) = 1}$$

Distances between probability distributions

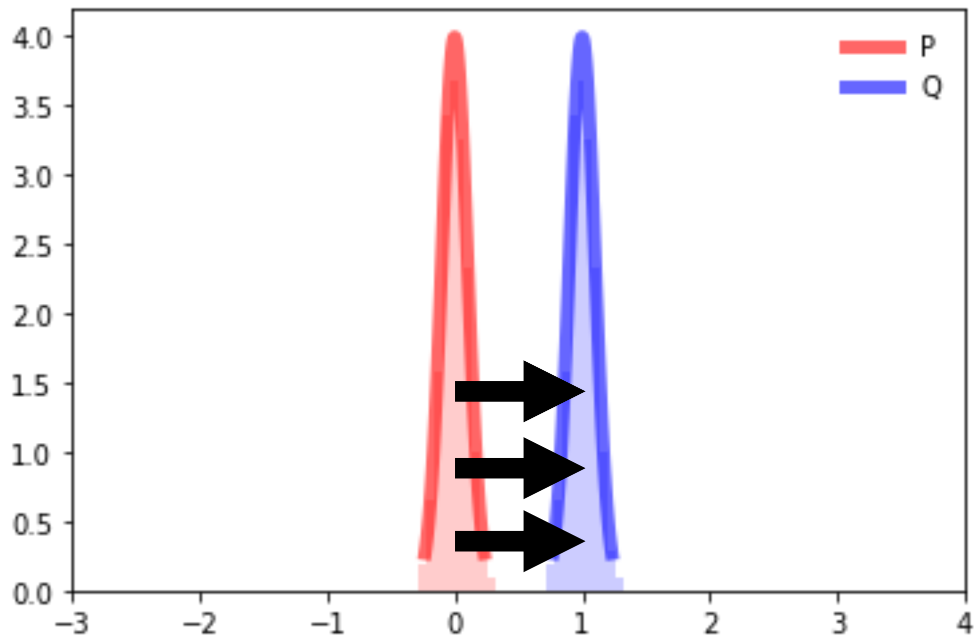


Total-Variation Distance (TV)

$$\mathrm{TV}(P, Q) = \frac{1}{2}\|P - Q\|_1$$

$$\boxed{\mathrm{TV}(P, Q) = 1}$$

Optimal Transport – The Earth Mover's Distance



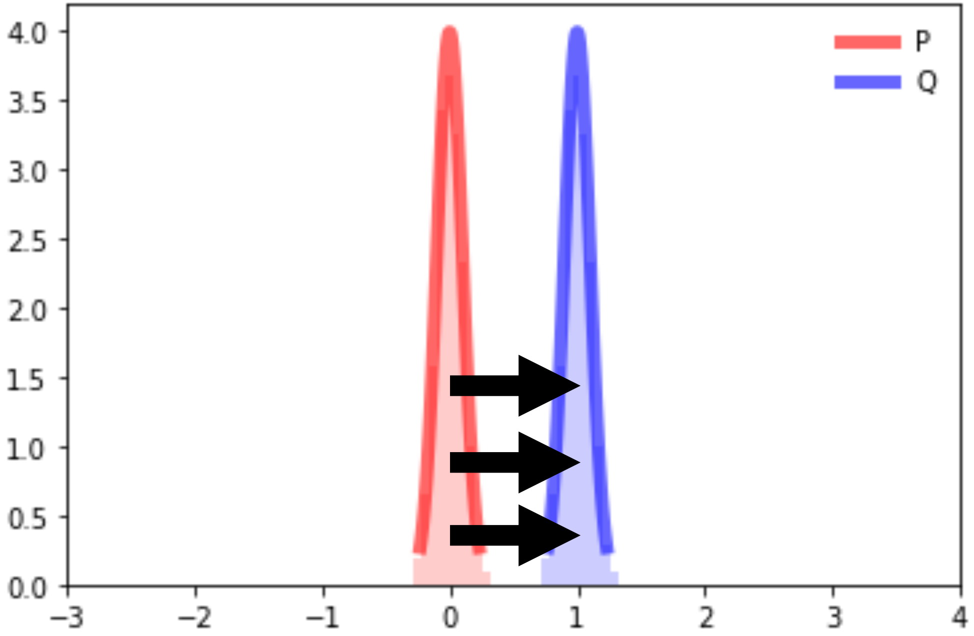
Wasserstein Distance:

$$W_d(P,Q) = \inf_{\pi \in \Pi(P,Q)} \int d(x,y)\pi(dx,dy)$$

Ground "cost" or distance:

$$d(x,y) = \|x-y\|_2$$

Optimal Transport – The Earth Mover's Distance



Wasserstein Distance:

$$W_d(P,Q) = \inf_{\pi \in \Pi(P,Q)} \int d(x,y)\pi(dx,dy)$$

Ground "cost" or distance:

$$d(x,y) = \|x-y\|_2$$

Kantorovich-Rubenstein Dual:

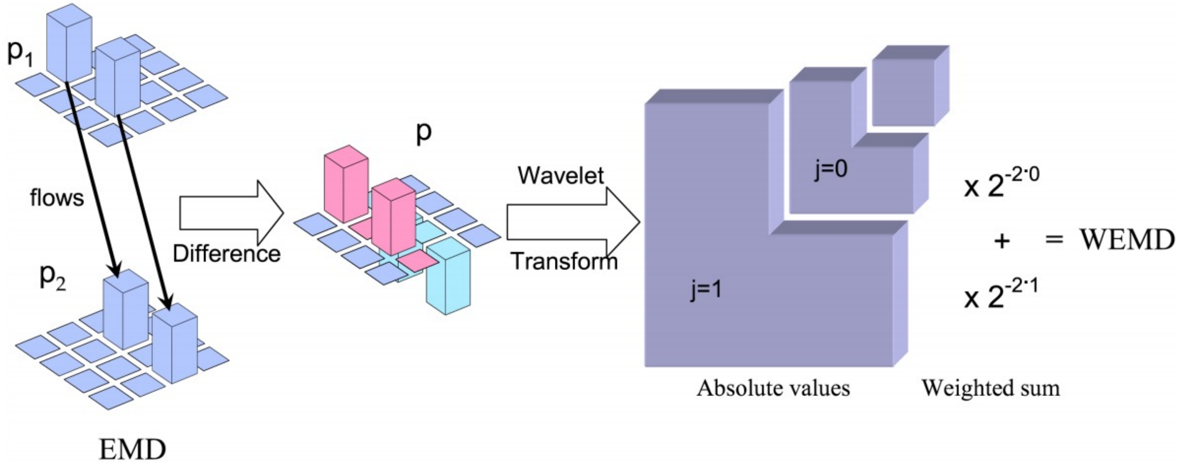
$$W_d(P,Q) = \sup_{f:|f(x)-f(y)| \leq d(x,y)} \int f(x)(P(x)-Q(x))dx$$

Computing EMD with the Dual Form

- In the wavelet domain, there is an explicit form for the witness function f

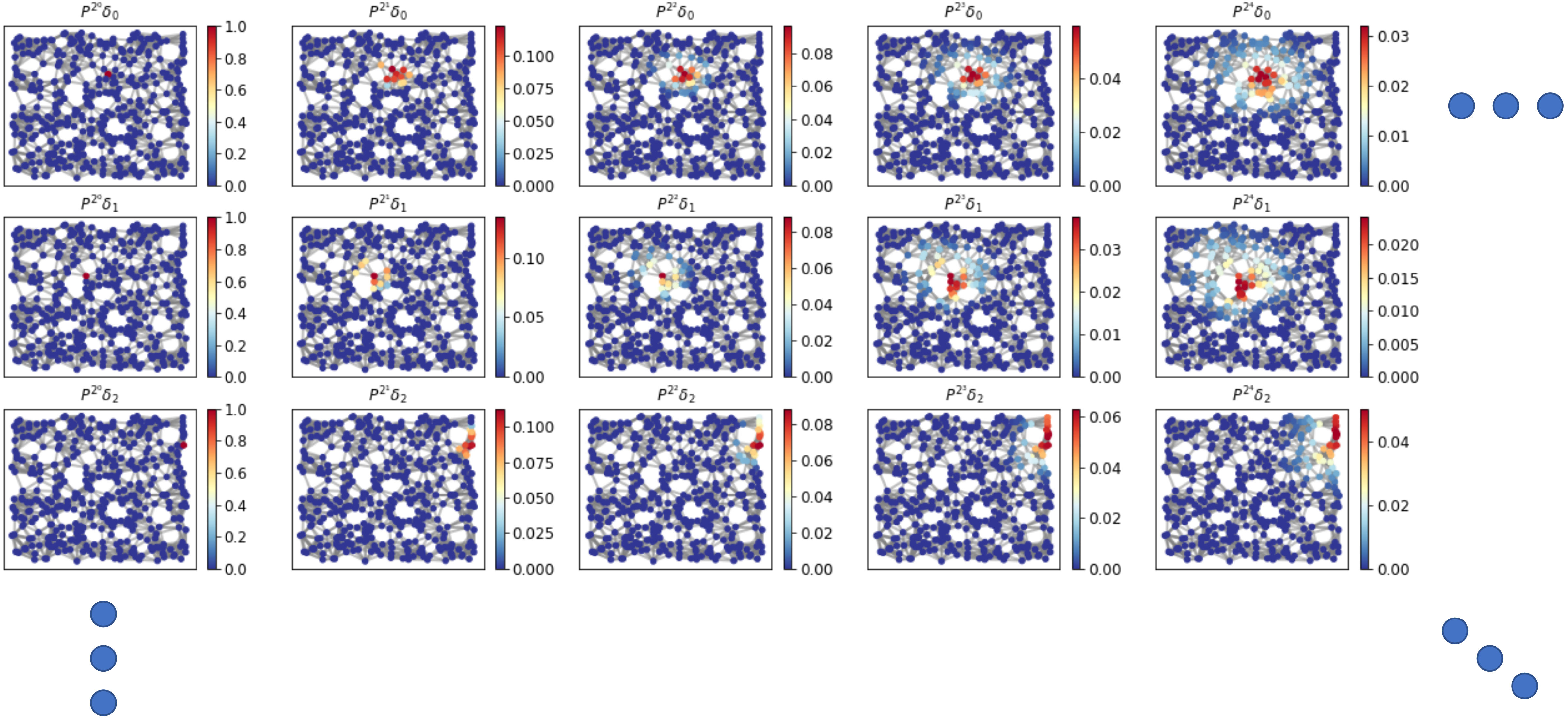
$$d(p)_{wemd} = \sum_{\lambda} 2^{-j(1+n/2)} |p_\lambda|$$

- Take the difference of two histograms and use a wavelet basis to represent them
- Avoids pairwise distance matrices

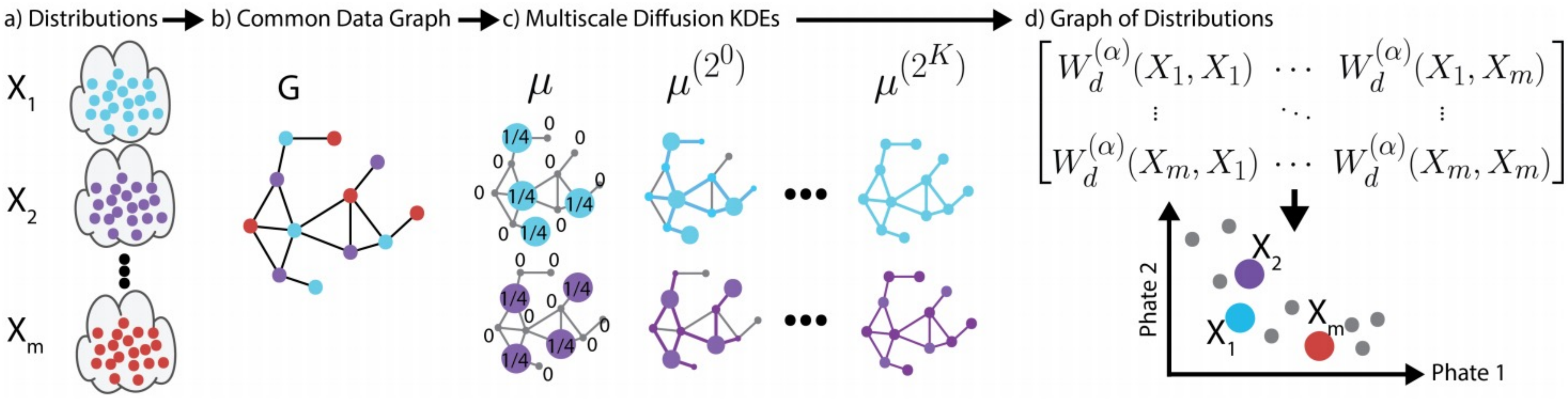


Shirdhonkar and Jacobs 2008

Diffusion on a Graphs



Diffusion EMD



Use multi-scale density estimates to compute a wavelet EMD on a common data graph

$$\mathrm{WEMD}_\alpha(\mu,\nu) := \sum_j 2^{-j(\alpha+1/2)} \sum_k |\langle \mu-\nu, \psi_{j,k}\rangle|$$

Diffusion EMD

Diffusion EMD between two datasets X_i, X_j supported on a graph with Diffusion operator $\mathbf{P}_K$

$$\mathrm{DEMD}_{\alpha,K}(X_i,X_j) := \sum_{k=0} \|T_{\alpha,k}(X_i) - T_{\alpha,k}(X_j)\|_1; \quad 0<\alpha<1/2$$

$$T_{\alpha,k}(X_i) := \begin{cases} 2^{-(K-k-1)\alpha}(\mu_i^{(2^{k+1})} - \mu_i^{(2^k)}) & k<K \\ \mu_i^{(2^K)} & k=K \end{cases}$$

$$\mu_i^{(t)} := \frac{1}{n_i}\mathbf{P}^t \mathbf{1}_{X_i}$$

$$\mathrm{WEMD}_\alpha(\mu,\nu) := \sum_j 2^{-j(\alpha+1/2)} \sum_k |\langle \mu - \nu, \psi_{j,k}\rangle|$$

Diffusion EMD

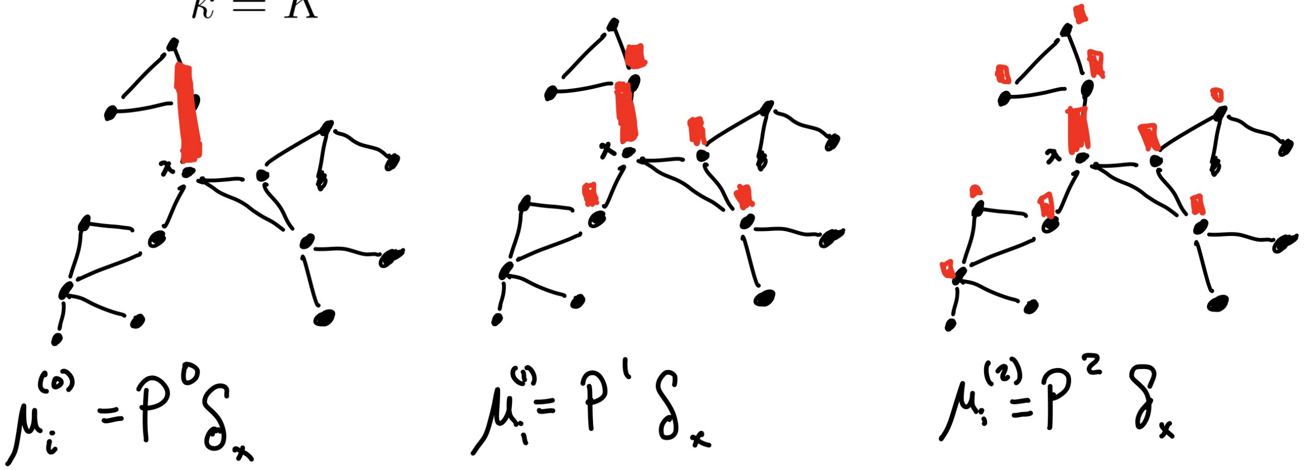
Diffusion EMD between two datasets X_i, X_j supported on a graph with Diffusion operator **P**

$$\mathrm{DEMD}_{\alpha,K}(X_i, X_j) := \sum_{k=0}^{K} \|T_{\alpha,k}(X_i) - T_{\alpha,k}(X_j)\|_1; \quad 0 < \alpha < 1/2$$

$$T_{\alpha,k}(X_i) := \begin{cases} 2^{-(K-k-1)\alpha}(\mu_i^{(2^{k+1})} - \mu_i^{(2^k)}) & k < K \\ \mu_i^{(2^K)} & k = K \end{cases}$$

$$\mu_i^{(t)} := \frac{1}{n_i}\mathbf{P}^t \mathbf{1}_{X_i}$$

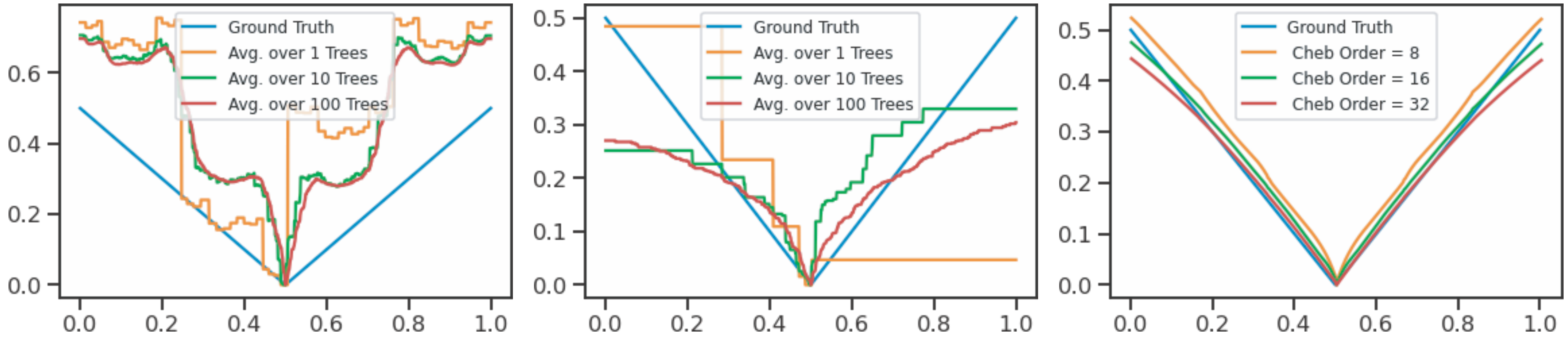
We show equivalence to an earth mover's distance with a geodesic ground distance as the number of samples increases:



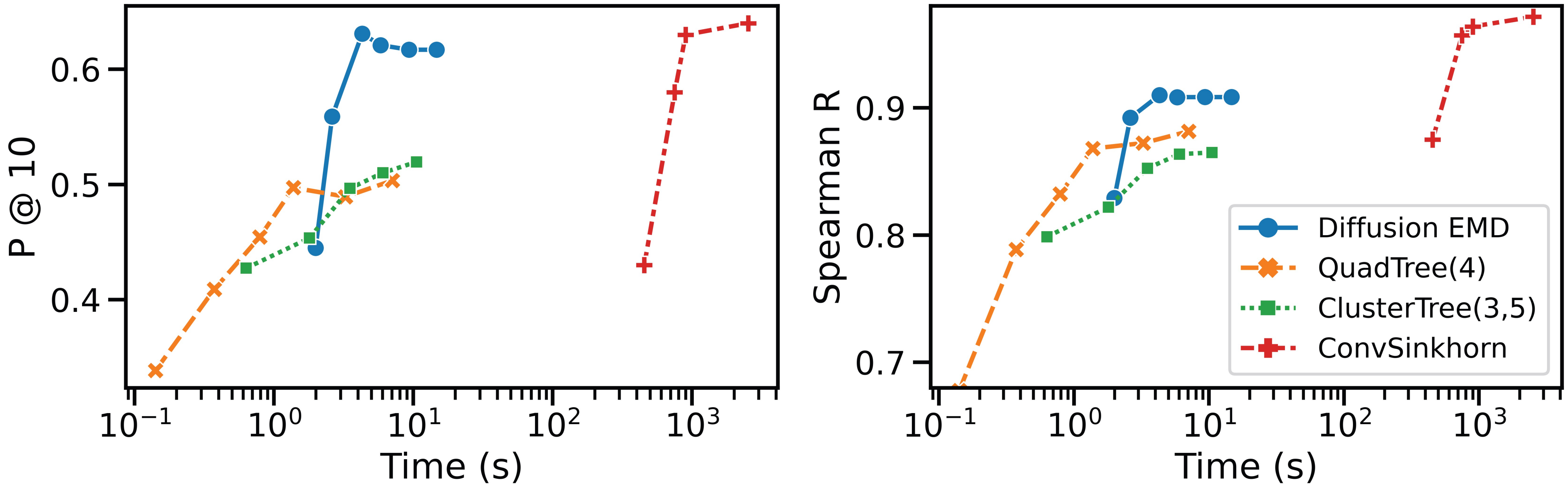
$$\lim_{X_i \to \mu_i, X_j \to \mu_j} \mathrm{DEMD}_{\alpha,K}(X_i, X_j) \simeq \mathrm{EMD}(\mu_i, \mu_j) \text{ with a geodesic ground distance } d_{\mathcal{M}}^{2\alpha}$$

Diffusion EMD uses “soft” density estimates so recreates ground distances better

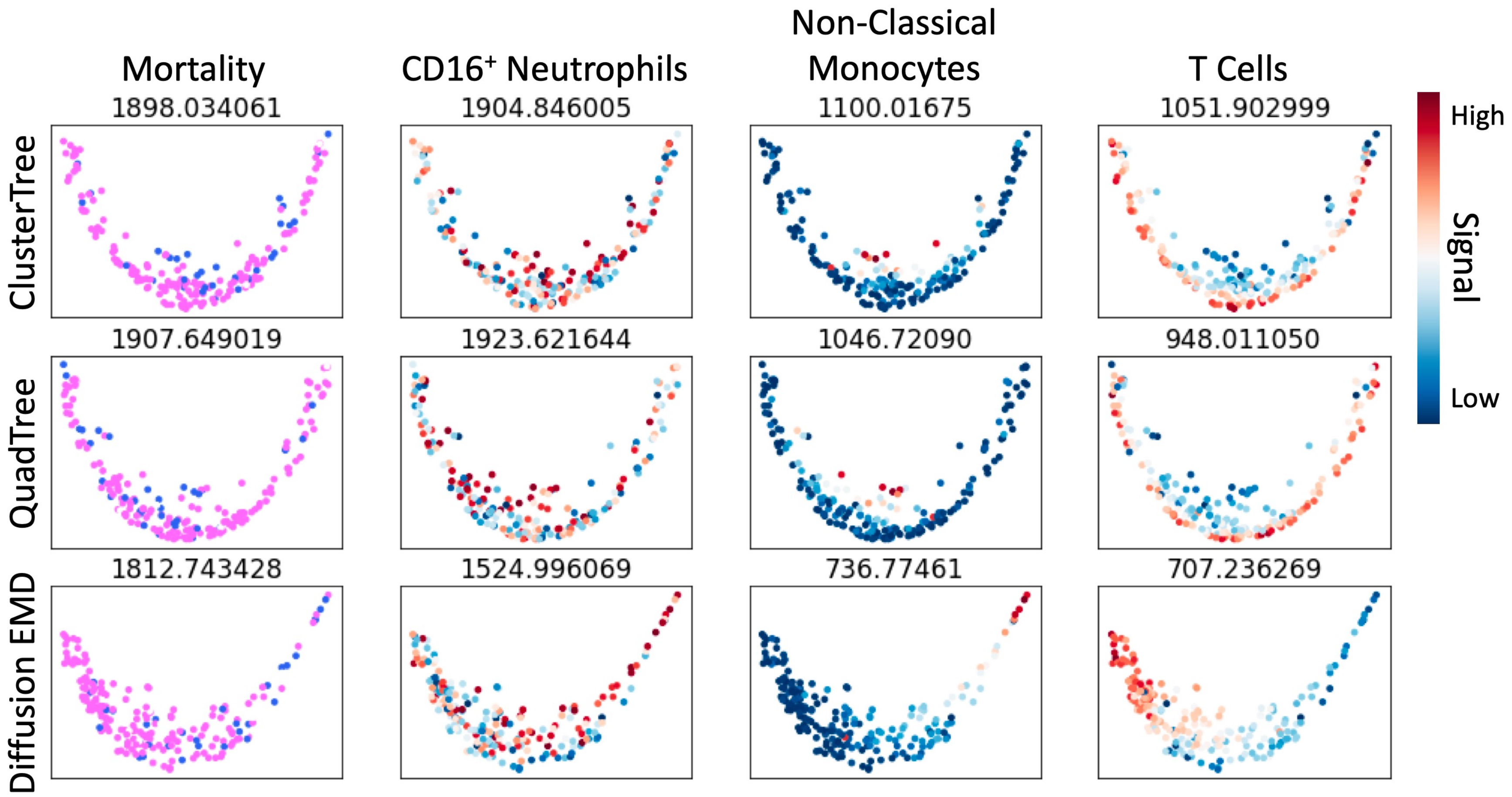
Cluster Tree | Quad Tree | Diffusion EMD



Diffusion EMD is more accurate for the same computation budget



Diffusion EMD can be used to organize patients according to their single-cell data



Summary

Diffusion EMD embeds the distributions → vectors such that L^1 between vectors is equivalent to EMD between distributions

- Uses a geodesic ground distance, scaling with intrinsic dimensionality
- Avoids constructing pairwise distance matrices
- For nearest EMD-neighbors avoids calculating all pairwise EMDs

Thanks!

Code: https://github.com/KrishnaswamyLab/DiffusionEMD

Paper: https://arxiv.org/abs/2102.12833

Lab Website: https://www.krishnaswamylab.org