Geometric and Statistical Approaches to Shallow and Deep Clustering

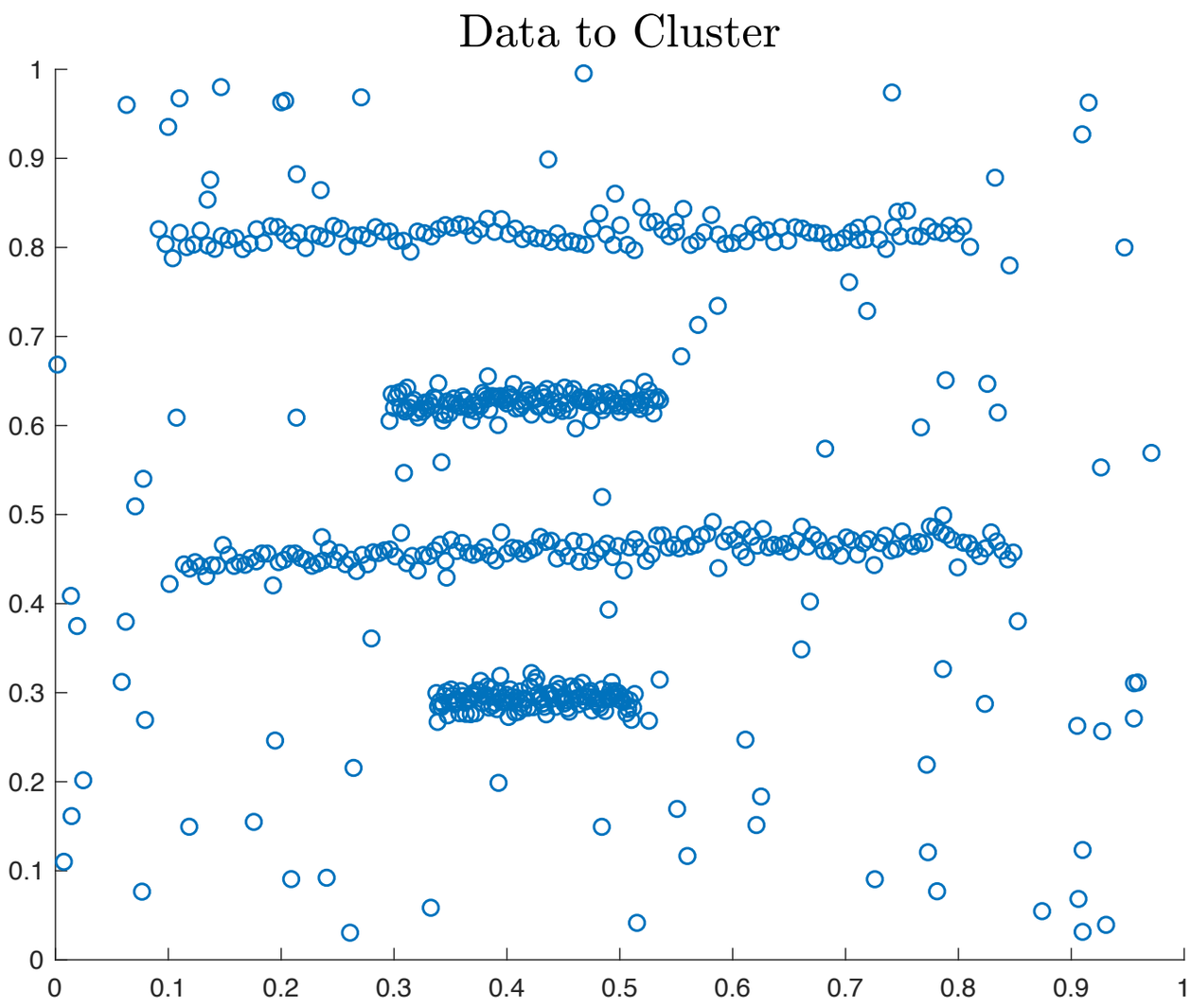
James M. Murphy
Department of Mathematics

November 5, 2021

Unsupervised Learning

Unsupervised learning: infer structure from data without access to *training data*, i.e. examples belonging to particular classes.

Clustering: unsupervised learning in which the goal is to label points as belonging to a given class.



$$x_1, \ldots, x_n \overset{iid}{\sim} \mu = \sum_{k=1}^{K} w_k \mu_k + \tilde{\mu}, \quad \sum_{k=1}^{K} w_k = 1$$

Labeling: Which x_j were generated from μ_k?
Number of Clusters: Can we estimate K?

Tufts UNIVERSITY

Standard Method: K-Means

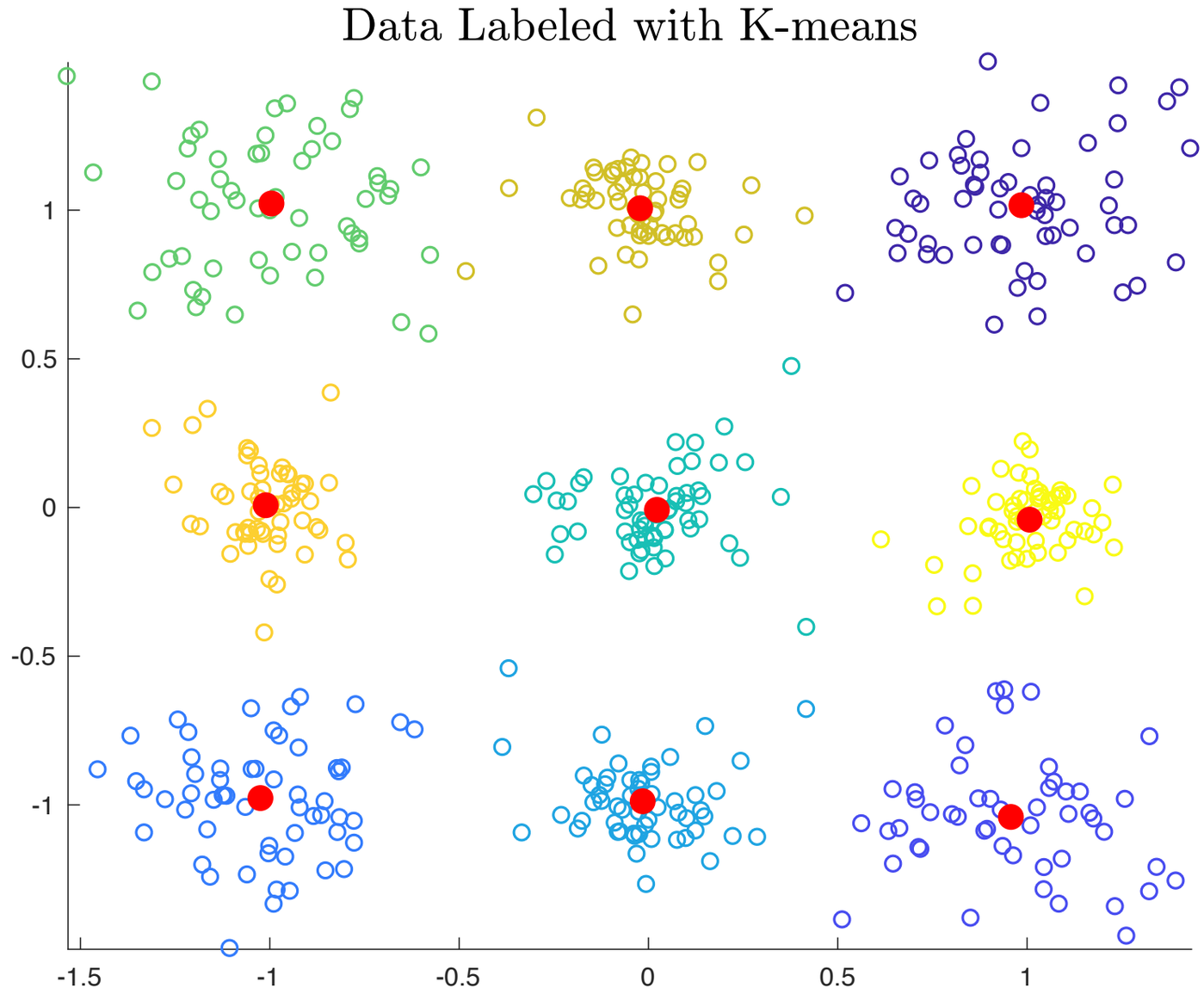
- **Idea**: find K centroids, then assign each point to its nearest centroid.
- Empirically good for same sized, spherical clusters.
- Guaranteed for certain Gaussians.
- Exact solution is NP-Hard to compute.
- Standard implementations involve non-convex optimization.
- Need to know K.

$$C^* = \underset{C=\{C_k\}_{k=1}^K}{\arg\min} \sum_{k=1}^{K} \sum_{x \in C_k} \|x - \bar{x}_k\|_2^2$$

Tufts UNIVERSITY

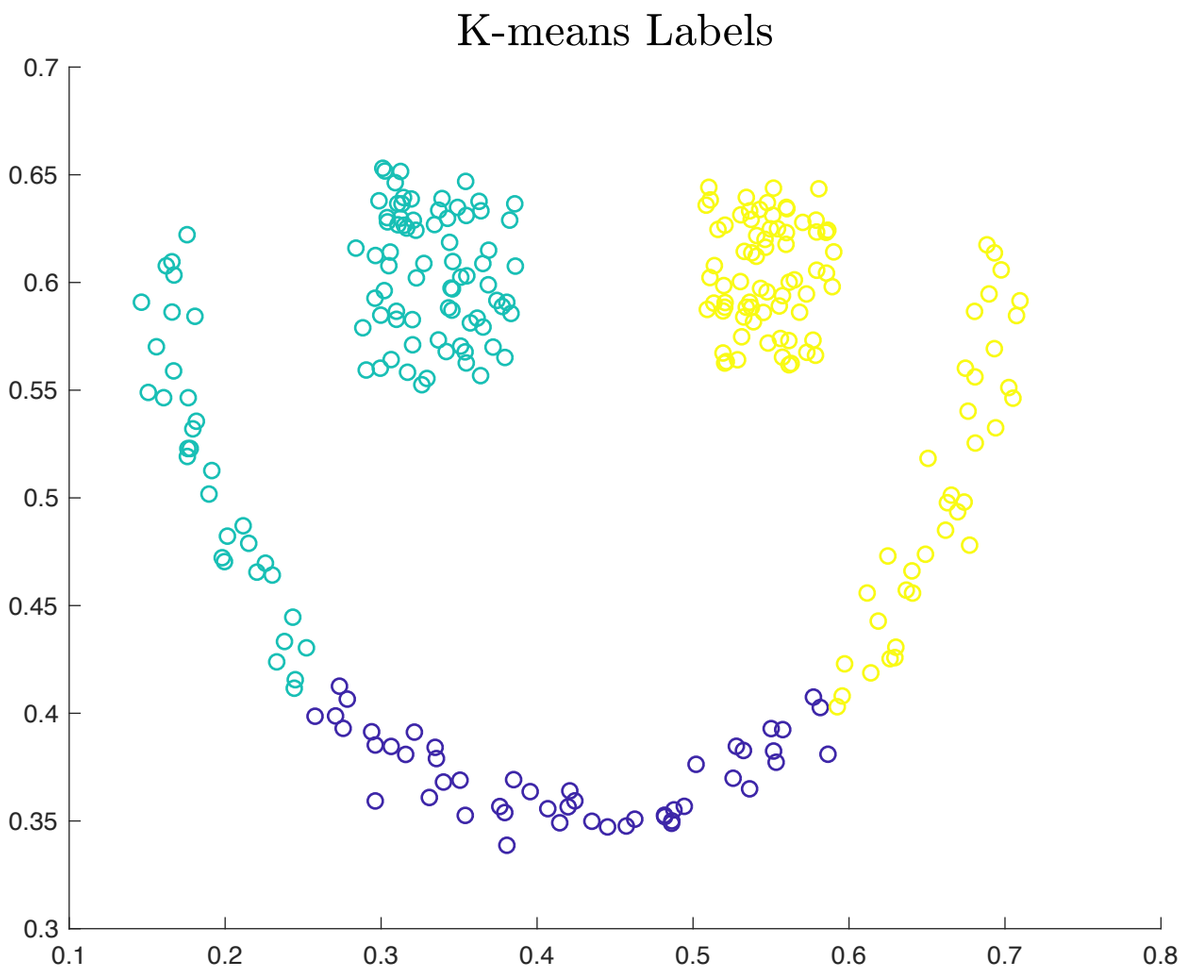
Standard Method: K-Means

- **Idea**: find K centroids, then assign each point to its nearest centroid.
- Empirically good for same sized, spherical clusters.
- Guaranteed for certain Gaussians.
- Exact solution is NP-Hard to compute.
- Standard implementations involve non-convex optimization.
- Need to know K.



$$C^* = \underset{C=\{C_k\}_{k=1}^K}{\arg\min} \sum_{k=1}^{K} \sum_{x \in C_k} \|x - \bar{x}_k\|_2^2$$

K-Means Often Fails



Problem: Some clusters are non-spherical!

Spectral Clustering I

Idea: embed data into a lower-dimensional space in a structure preserving way.

Input: $x_1, \ldots, x_n \subset \mathbb{R}^D$

Step 1: Build a *weight matrix*

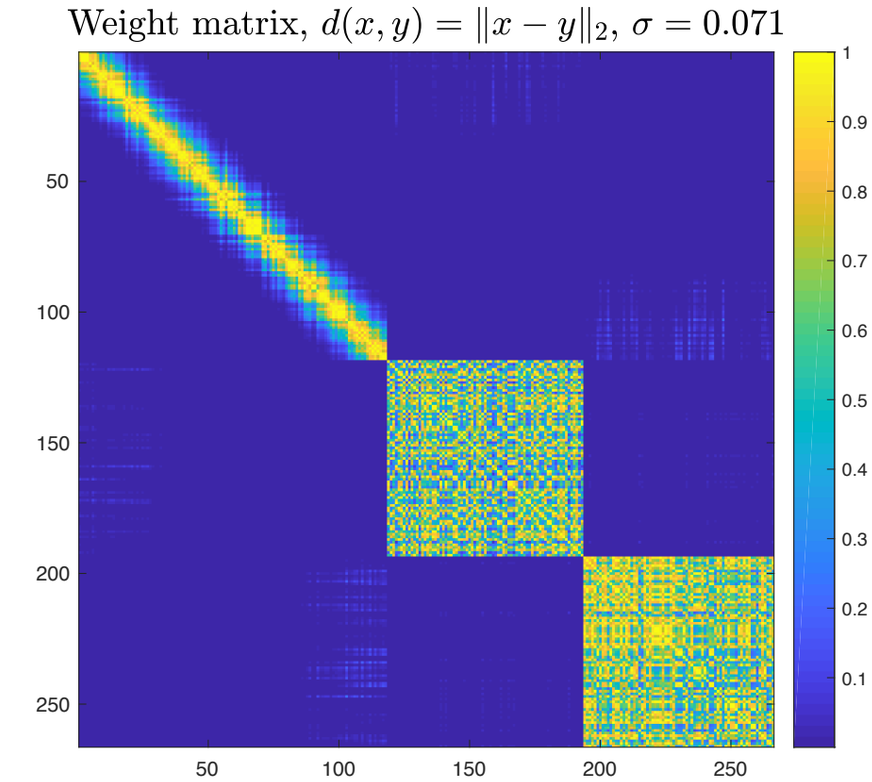
$$W_{ij} = e^{-d(x_i, x_j)^2/\sigma^2}$$

for some metric $d(\cdot, \cdot)$ and σ.

Step 2: Compute the *(graph) Laplacian*

$$L = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$$

$$D_{ii} = \sum_{j=1}^{n} W_{ij}; D_{ij} = 0, i \neq j.$$



Tufts UNIVERSITY

Spectral Clustering II

Step 3: Compute eigenvalues of L

$$0 \leq \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$$

and associated eigenvectors

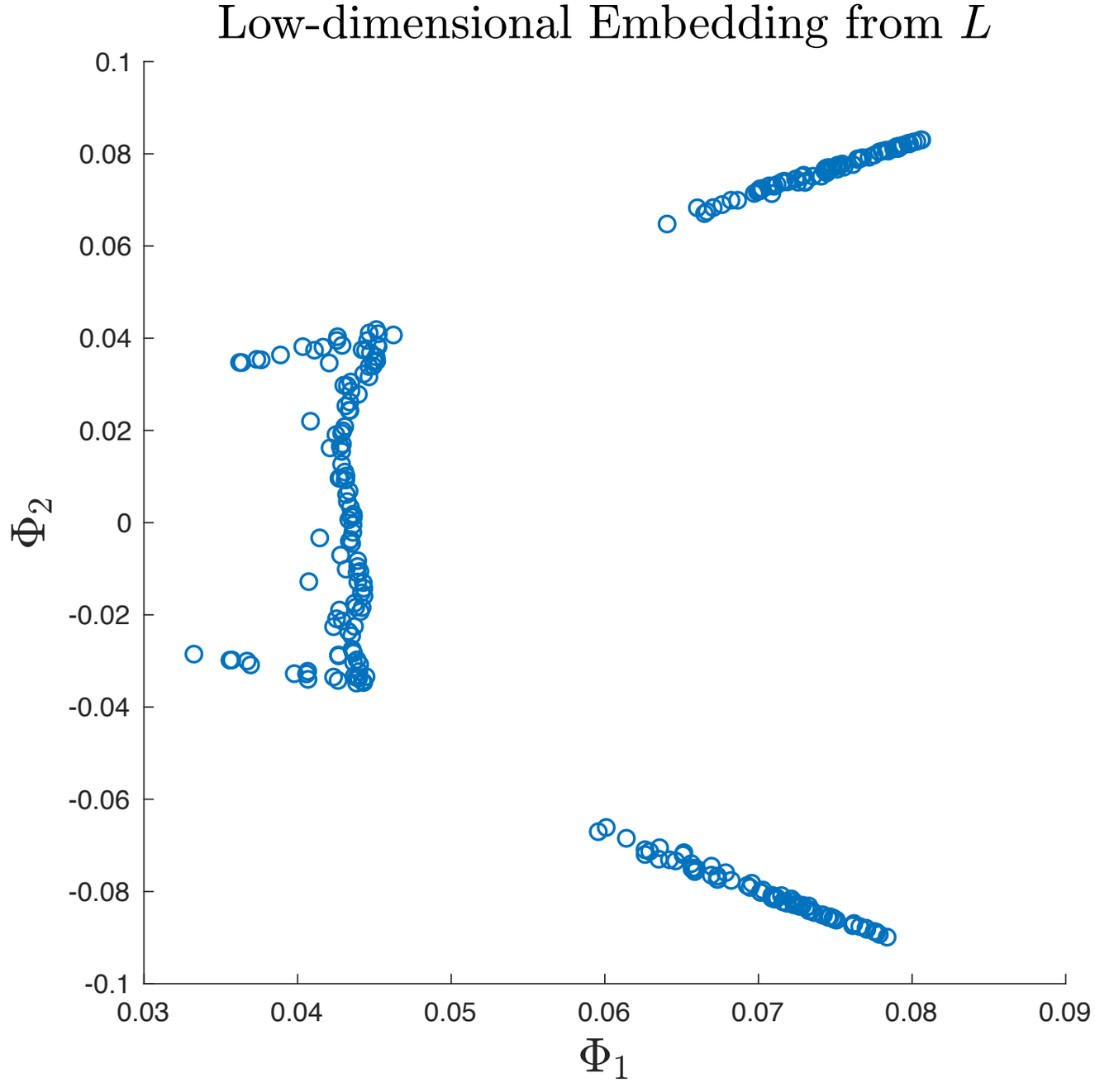
$$\Phi_1, ..., \Phi_n.$$

Step 4: Embed the data as

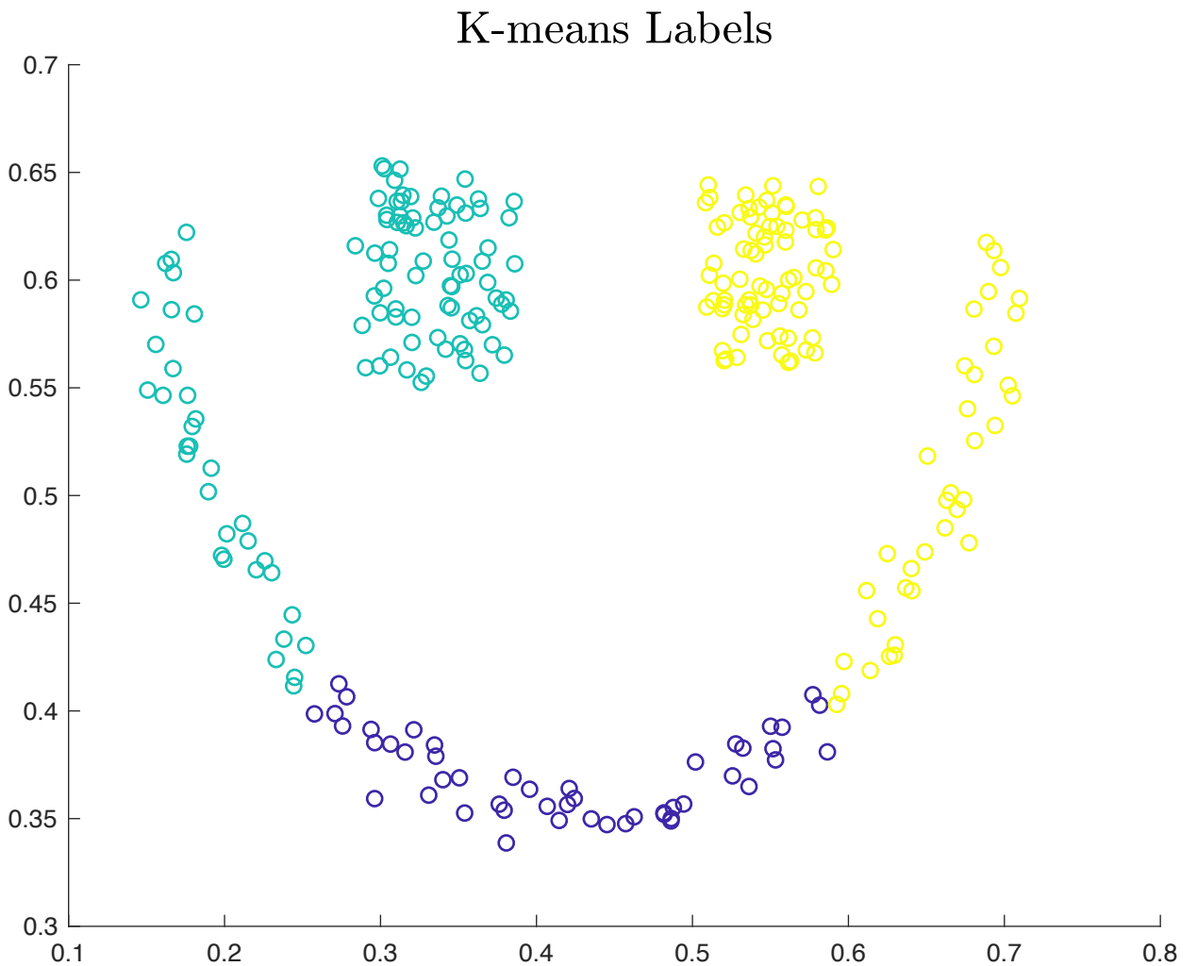
$$x_i \mapsto (\Phi_1(x_i), \ldots, \Phi_K(x_i))$$

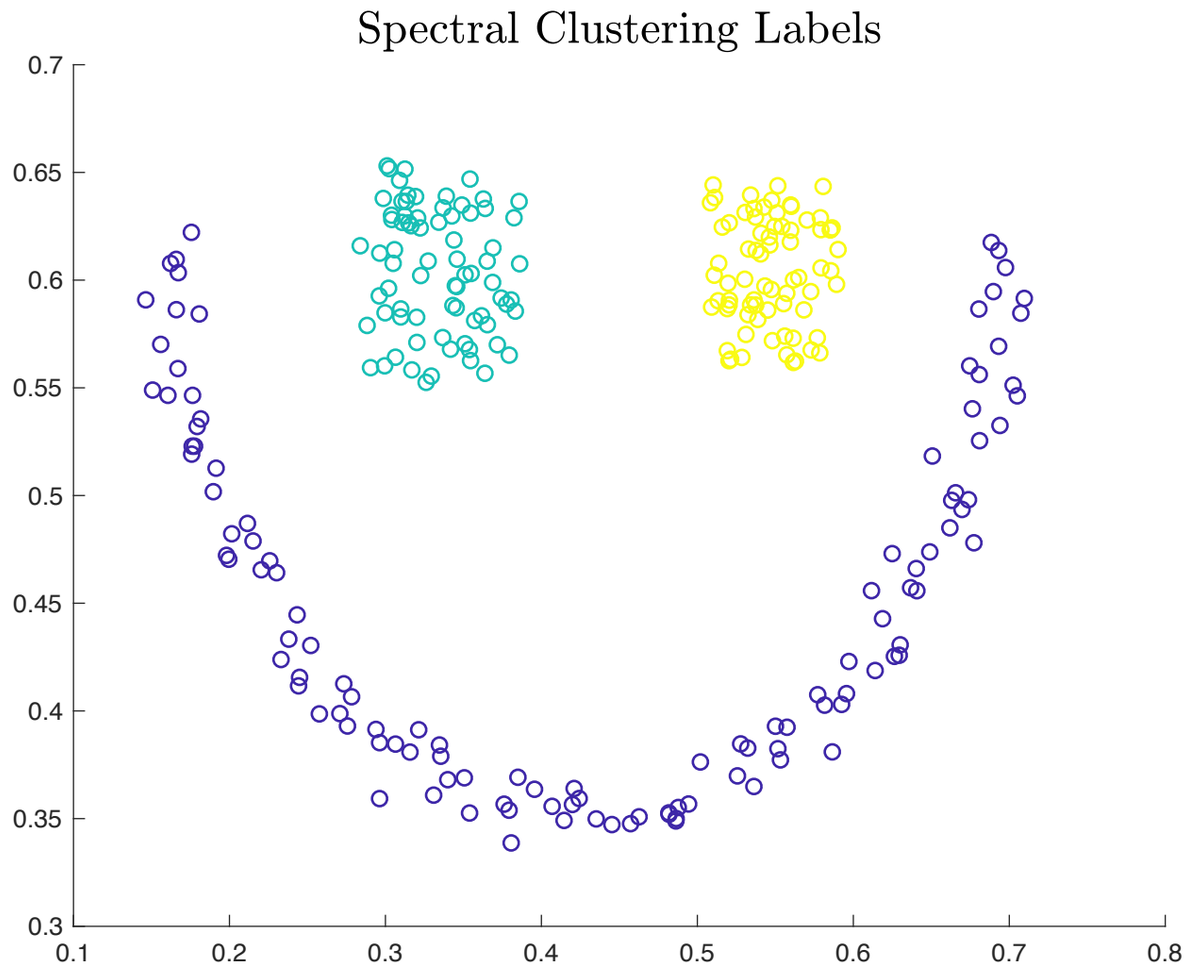
then run K-means. Note

$$\Phi_j(x_i) := \Phi_j(i).$$



K-Means v. Spectral Clustering





- Spectral clustering (with a "good" σ) succeeds where K-means fails!

- Theoretical estimates are limited, particularly for estimating the number of clusters. Common heuristic: $K \approx \arg\max_k \lambda_{k+1} - \lambda_k$.

Data-Dependent LLPD Metric

Definition. *For a discrete set $X = \{x_i\}_{i=1}^n \subset \mathbb{R}^D$, let $\mathcal{G}$ be the graph on X with edges given by the Euclidean distance between points. For $x_i, x_s \in X$, let $\mathcal{P}(x_i, x_s)$ denote the space of paths connecting x_i, x_s in $\mathcal{G}$. The* ***longest leg path distance (LLPD)*** *between x_i, x_s is:*

$$d_{\ell\ell}(x_i, x_s) = \min_{\{y_j\}_{j=1}^L \in \mathcal{P}(x_i, x_s)} \max_{j=1,2,\ldots,L-1} \|y_{j+1} - y_j\|_2,$$

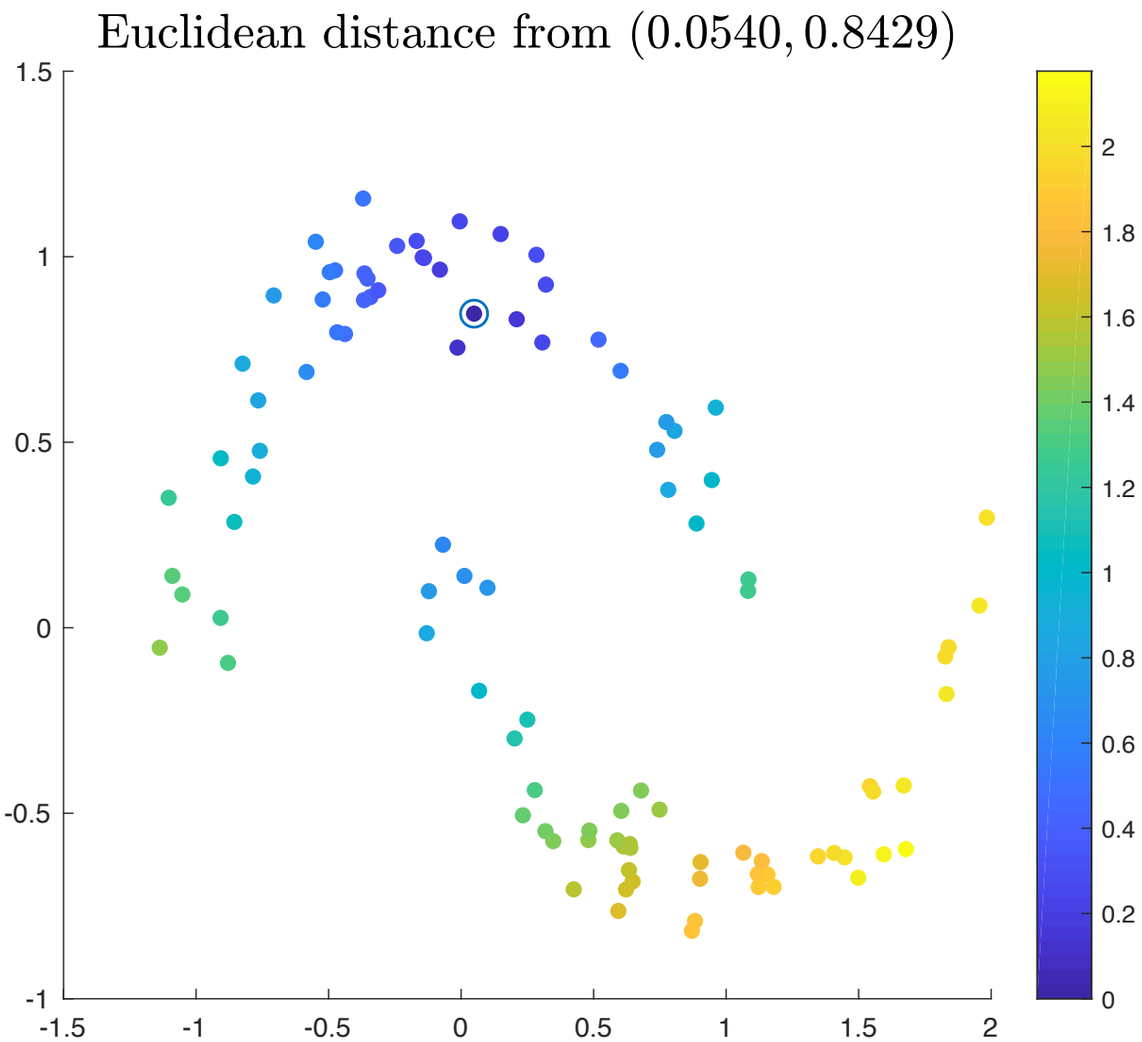
- The distance between points x, y is the minimum over all paths between x, y of the longest edge in the path.

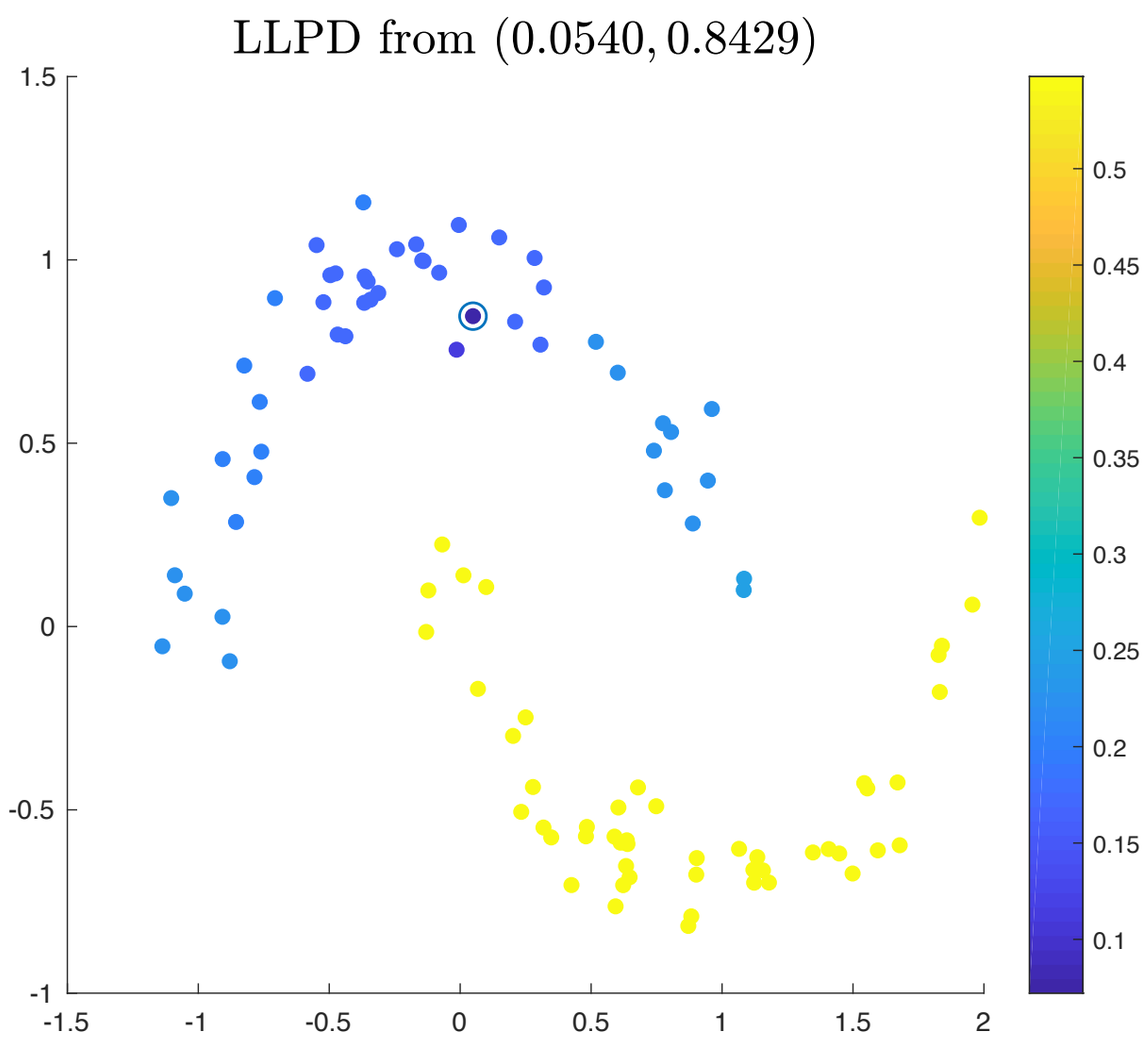
- Depending on the data X, this distance changes!

- $\mathcal{G}$ could be a complete graph (all points connected to all points) or a connected NN graph.

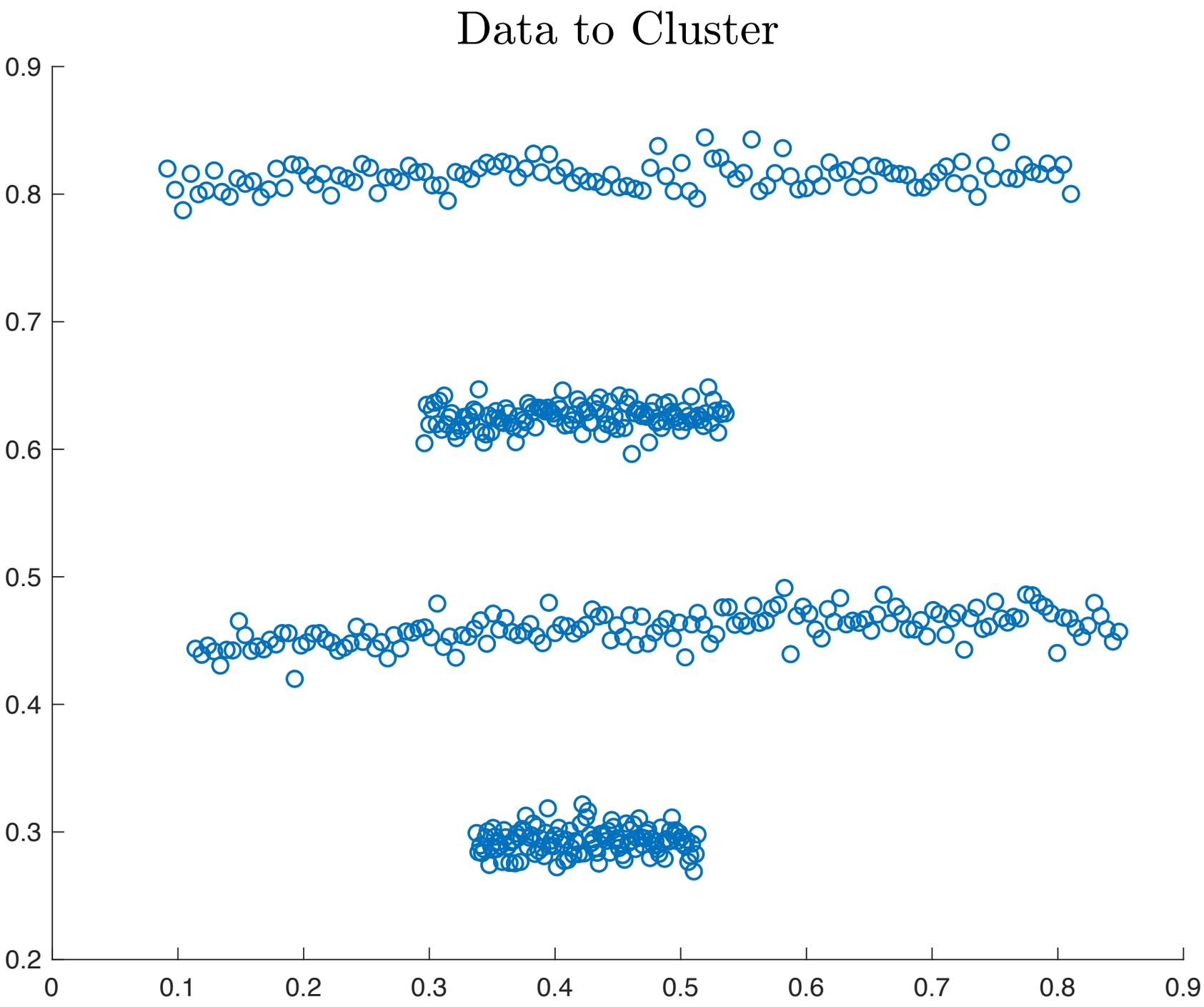
- Looks hard to compute. We have a fast approximation, so this turns out not to be an obstacle.

Euclidean Distance versus LLPD



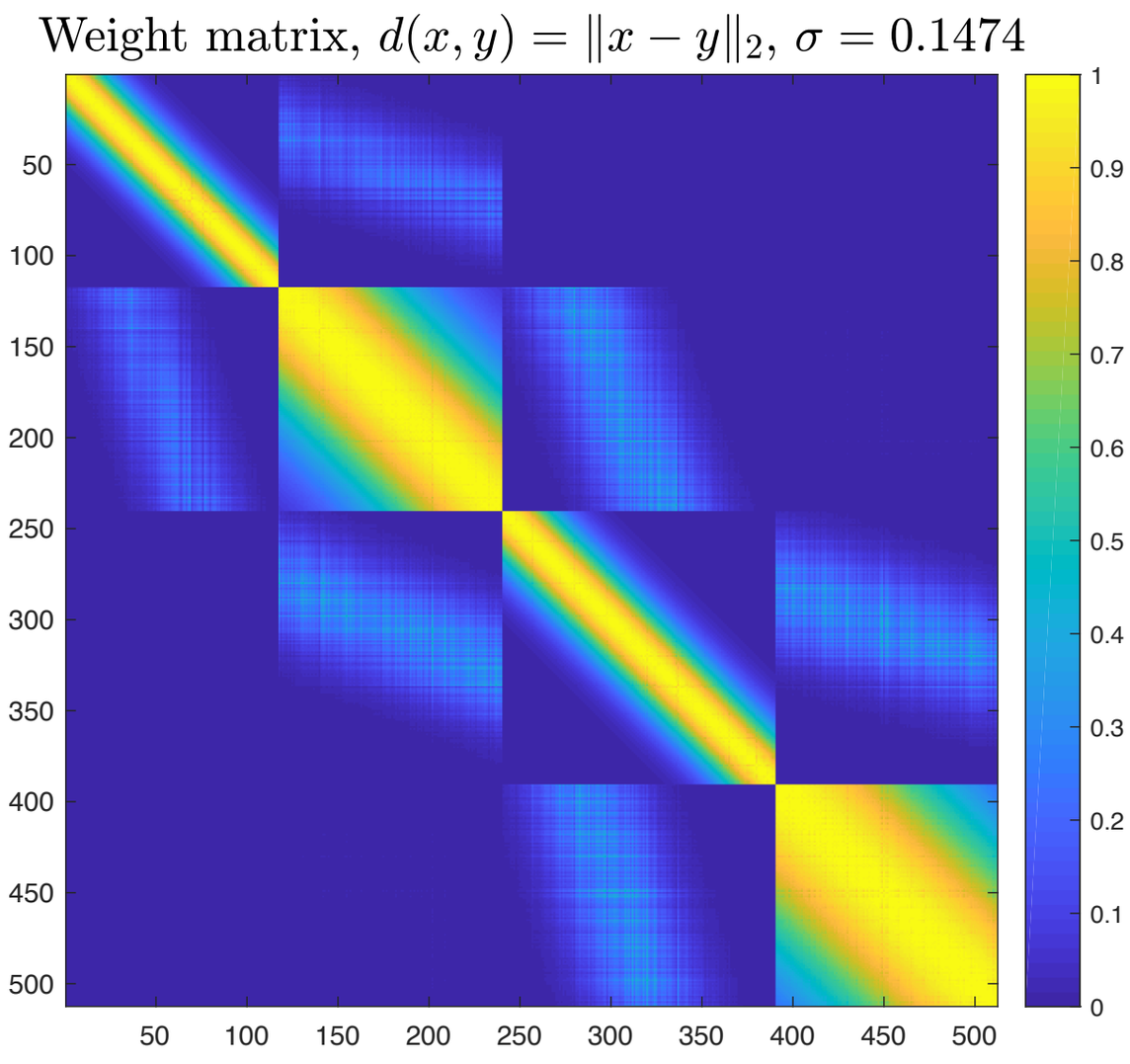


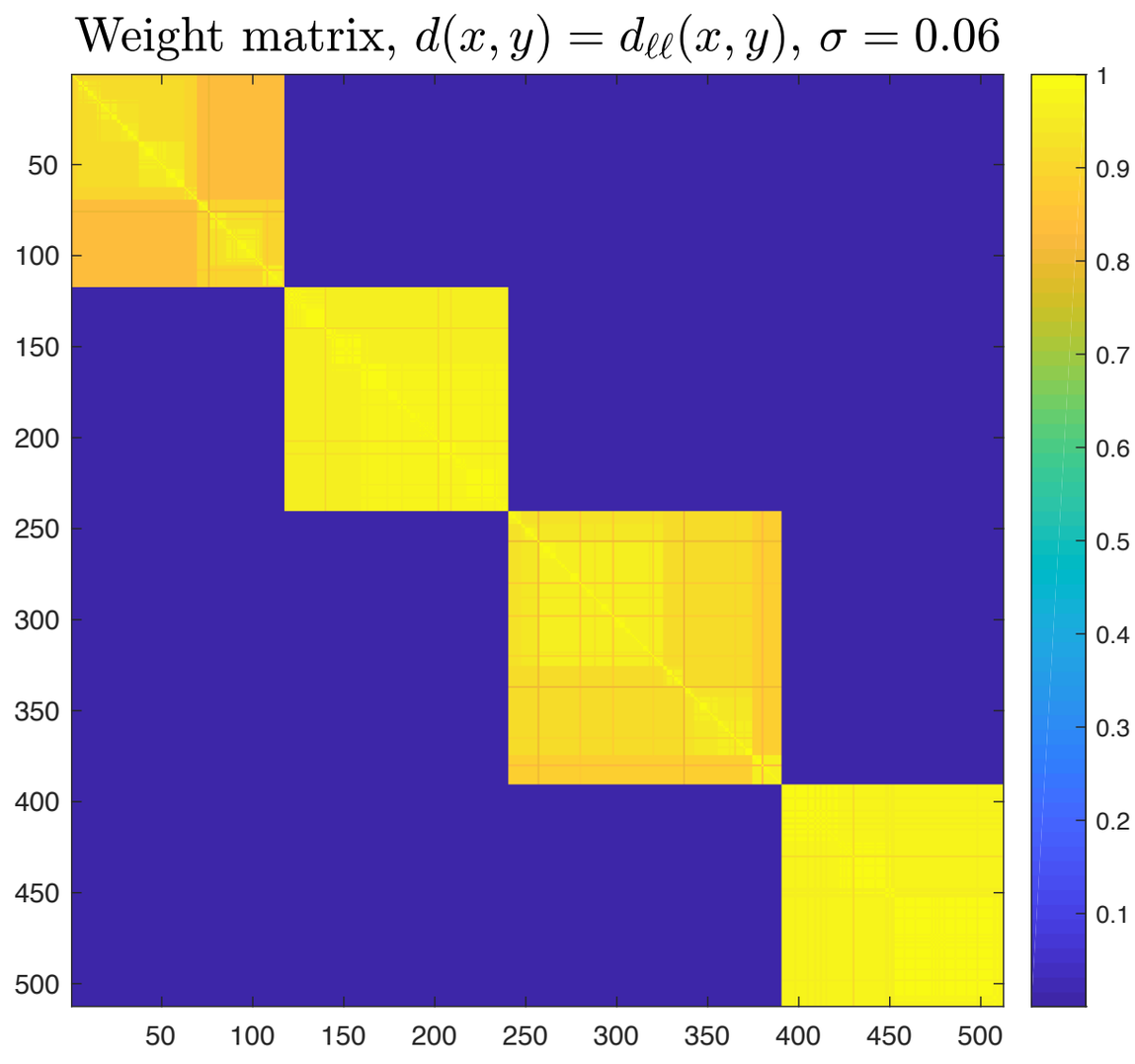
Data Well-Suited for LLPD



LLPD Weight Matrix

- For our simple "four lines" data, there is a big difference between Euclidean distance (data independent) and LLPD (data dependent).
- The LLPD weight matrix has block-constant structure.





Low Dimensional, Large Noise (LDLN) Model

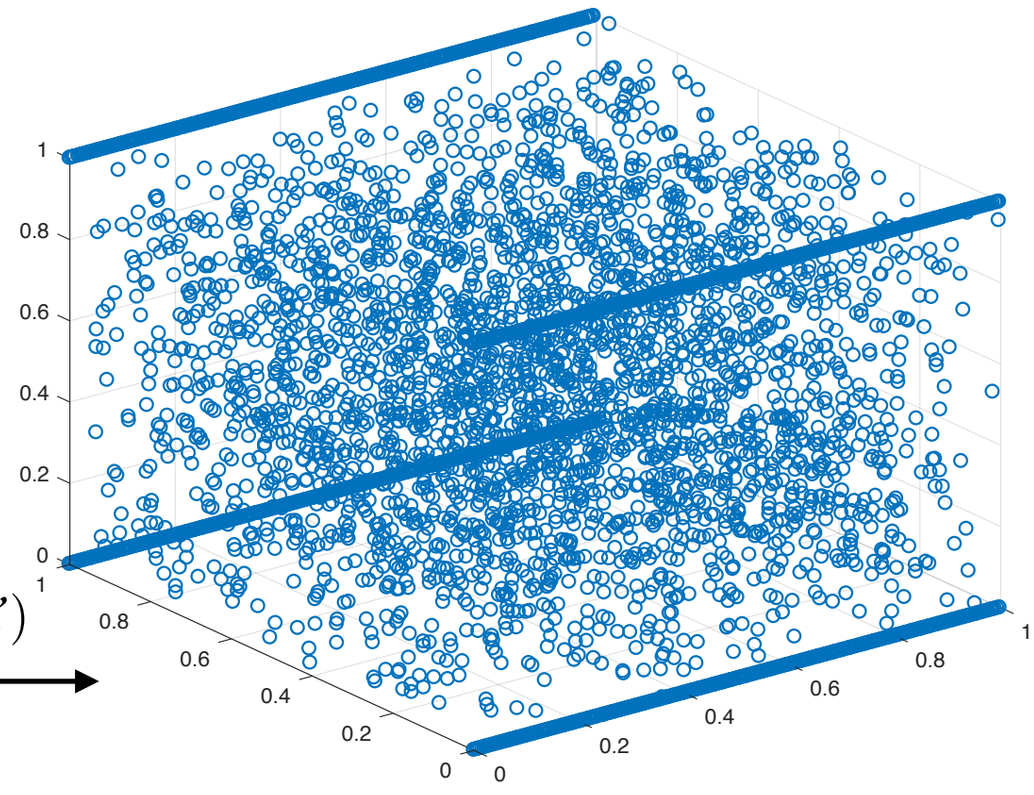
Definition. *A set $S \subset \mathbb{R}^D$ is an element of $\mathcal{S}_d(\kappa, \epsilon_0)$ for some $\kappa \geq 1$ if it has finite d-dimensional Hausdorff measure, denoted by $\mathcal{H}^d$, is connected, and for some $\epsilon_0 > 0$, it satisfies the following geometric condition:*

$$\forall x \in S, \quad \forall \epsilon \in (0, \epsilon_0), \qquad \kappa^{-1}\epsilon^d \leq \frac{\mathcal{H}^d(S \cap B_\epsilon(x))}{\mathcal{H}^d(B_1(0))} \leq \kappa\epsilon^d \,.$$

Low-dimensional

$$\mathcal{X}_1, \ldots, \mathcal{X}_K \subset \mathcal{X} \subset \mathbb{R}^D$$
$$\mathcal{X}_1, \ldots, \mathcal{X}_K \in \mathcal{S}_d(\kappa, \epsilon_0)$$
$$\delta = \min_{k \neq k'} \operatorname{dist}(\mathcal{X}_k, \mathcal{X}_{k'})$$

n_i i.i.d. draws from Unif$(\mathcal{X}_i)$



Large noise

$$\tilde{\mathcal{X}} = \mathcal{X} \setminus (\mathcal{X}_1 \cup \ldots \cup \mathcal{X}_K)$$

$\tilde{n}$ i.i.d. draws from Unif$(\tilde{\mathcal{X}})$

$$n = n_1 + \ldots + n_K + \tilde{n}$$
$$n_{\min} = \min_{1 \leq k \leq K} n_k$$

Tufts UNIVERSITY

Nearest Neighbors in LLPD and Denoising

- In the LDLN model, points within clusters all have comparable distances, and points from different clusters are well separated.

- We denoise points by removing all points whose distance to their $k_{\text{nse}}{}^{th}$ nearest neighbor exceeds some threshold θ .

- k_{nse}, θ are parameters.

- This analysis, based on percolation theory, proves the weight matrix is nearly block constant.

Performance Guarantees

Theorem. *(Little, Maggioni, **M.**) Under the LDLN data model and assumptions, suppose that the cardinality $\tilde{n}$ of the noise set is such that*

$$\tilde{n} \leq \left(\frac{C_2}{C_1}\right)^{\frac{k_{nse}D}{k_{nse}+1}} n_{min}^{\frac{D}{d+1}\left(\frac{k_{nse}}{k_{nse}+1}\right)}.$$

Let $f_\sigma(x) = e^{-x^2/\sigma^2}$ be the Gaussian kernel and assume $k_{nse} = O(1)$ and $\frac{\min_i n_i}{n_{max}} = O(1)$. If n_{min} is large enough and θ, σ satisfy

$$C_1 n_{min}^{-\frac{1}{d+1}} \leq \theta \leq C_2 \tilde{n}^{-\left(\frac{k_{nse}+1}{k_{nse}}\right)\frac{1}{D}} \tag{1}$$

$$C_3\theta \leq \sigma \leq C_4\delta \tag{2}$$

then with high probability the graph Laplacian L on the denoised LDLN data X_N satisfies:

(i) the largest gap in the eigenvalues of L is $\lambda_{K+1} - \lambda_K$.

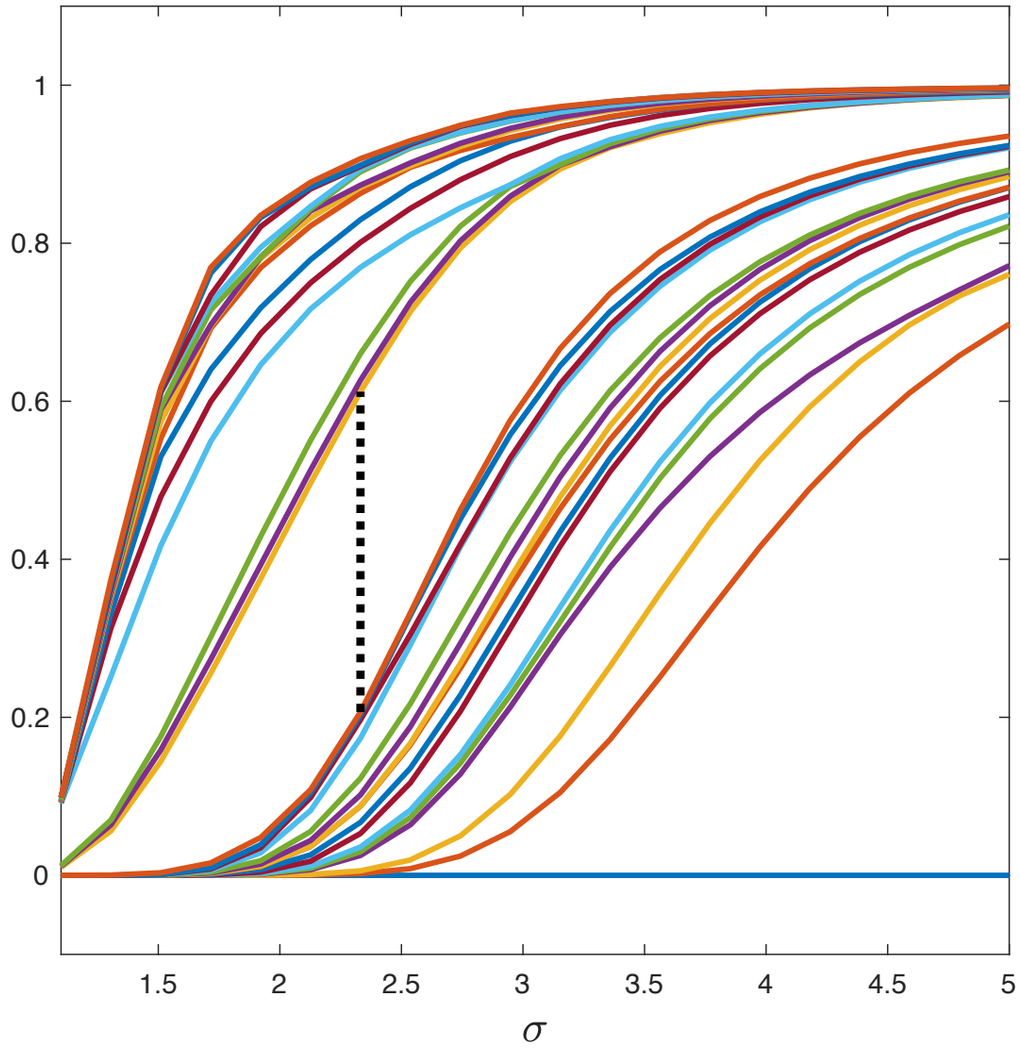
(ii) spectral clustering with L with K principal eigenvectors achieves perfect accuracy on X_N.

The constants $\{C_i\}_{I=1}^4$ depend on geometric quantities but do not depend on $n_1, \ldots, n_K, \tilde{n}, \theta, \sigma$.

Columbia Object Image Library (COIL)

COIL 16 Classes

Multiscale Eigenvalues for LLPD SC



- 16 classes, ambient dimensionality 1024, about 100 samples per class.

- LLPD spectral clustering achieve 99+% accuracy, and correctly identifies that there are 16 classes.

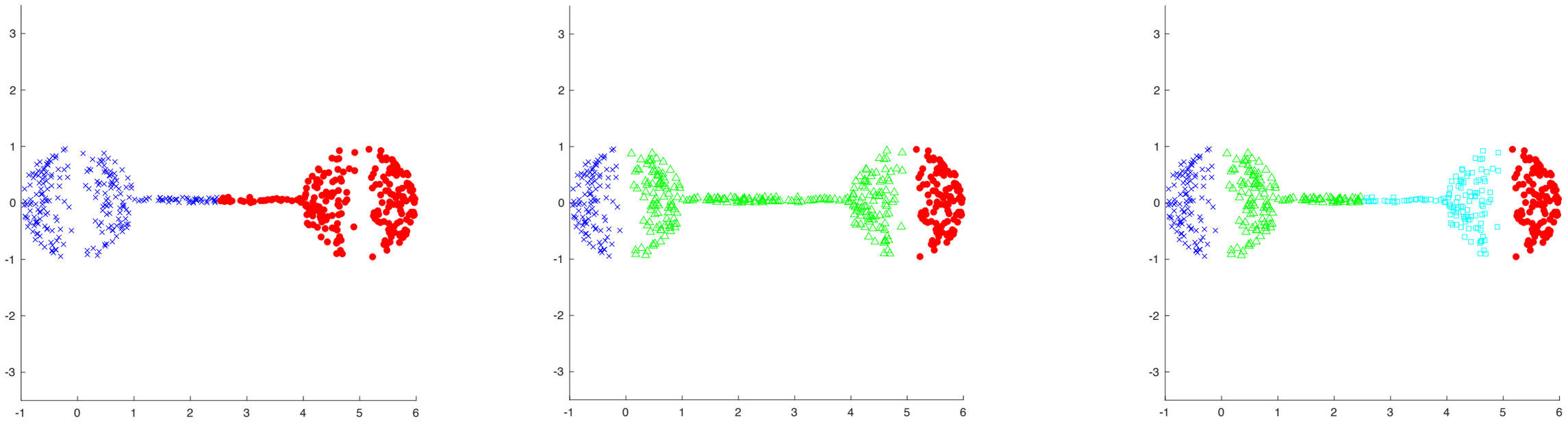
Tufts UNIVERSITY

Incorporating Geometry?

For $p \in [1, \infty)$ and for $x, y \in \mathcal{X}$, the *(discrete) p-weighted shortest path distance (PWSPD)* from x to y is:

$$\ell_p(x, y) = \min_{\pi = \{x_i\}_{i=1}^{L}} \left(\sum_{i=1}^{L-1} \|x_i - x_{i+1}\|^p \right)^{\frac{1}{p}},$$

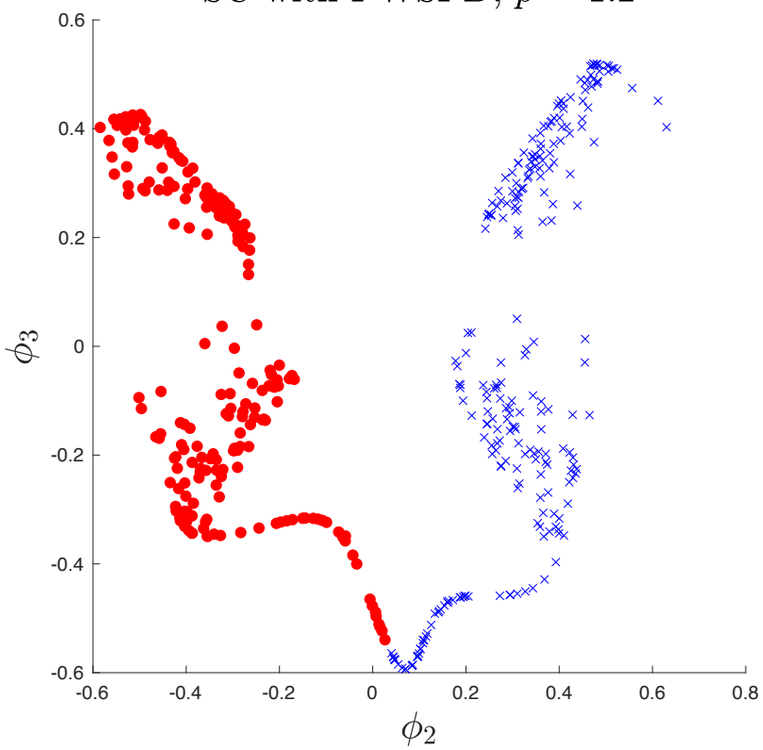
where π is a path consisting of data points in $\mathcal{X}$ with $x_1 = x$ and $x_L = y$.

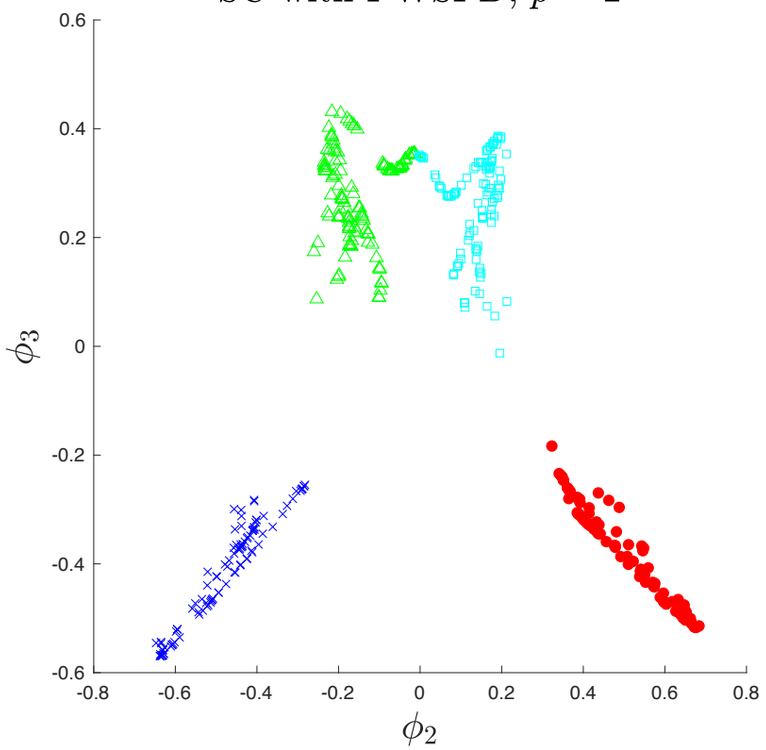


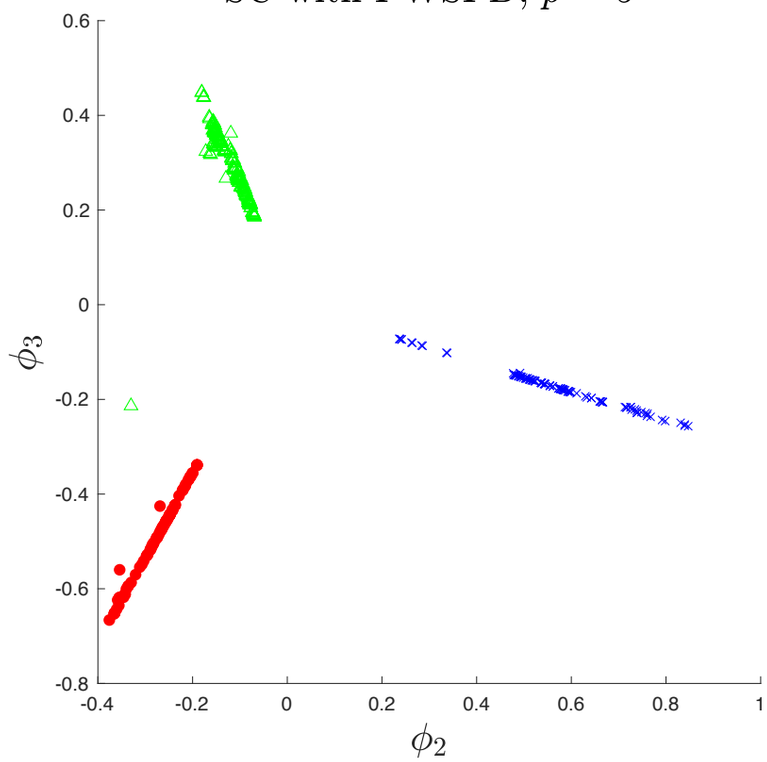
How to balance density and geometry when both are salient?

Tufts UNIVERSITY

Role of p







- As p changes, the embedding changes!
- Can we build a cluster model that balances geometry and density?
- Need to understand continuum versions of PWSPD and associated Riemannian metrics.

Tufts UNIVERSITY

Hyperspectral Images (HSI)

High dimensional images:

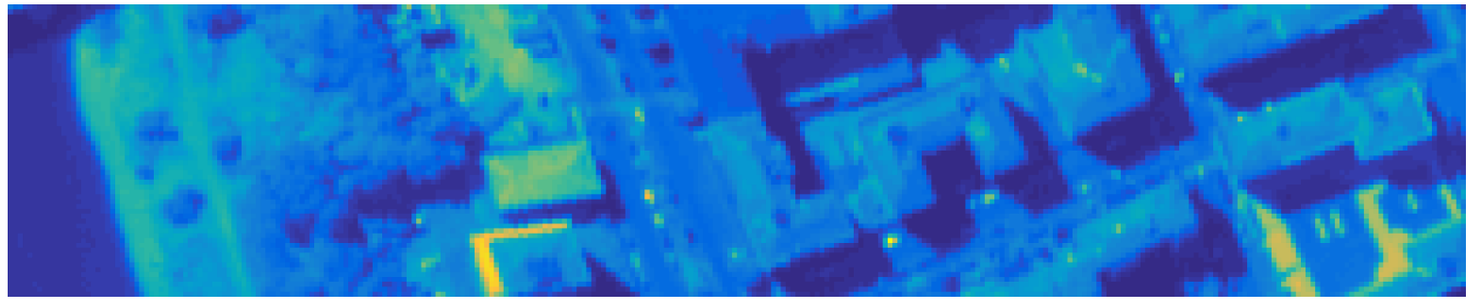
$$M \times N \times D$$

Spatial Spectral

The spectral bands are localized at certain electromagnetic frequencies, allowing for precise differentiation of materials in scenes

Learning Problems: clustering, anomaly detection, active learning, classification, segmentation, compression…

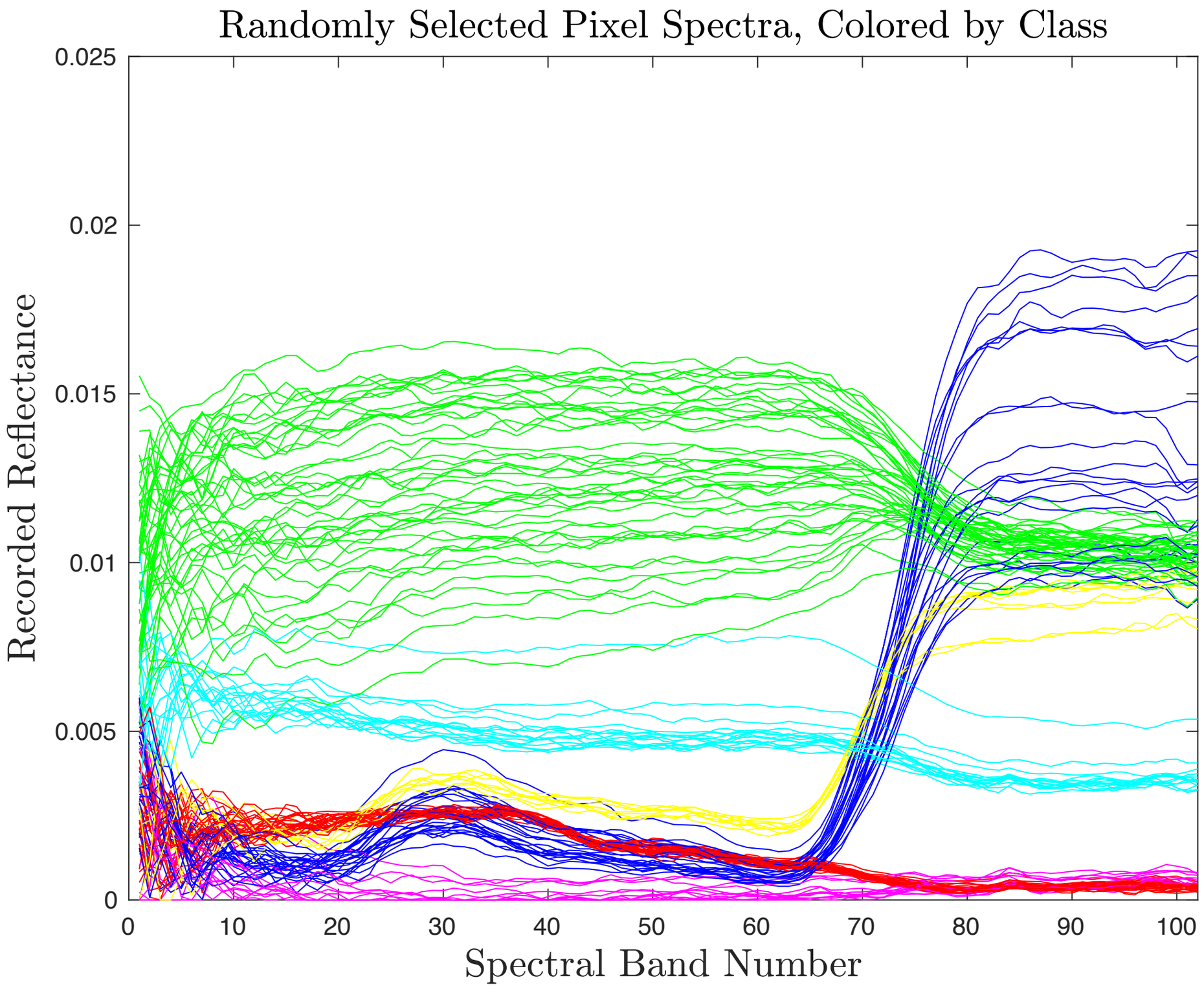
Applications: land cover change, water quality evaluation, precision agriculture, pollution tracking, defense and security…



Subset of Pavia dataset, sum of all bands (sum across D)

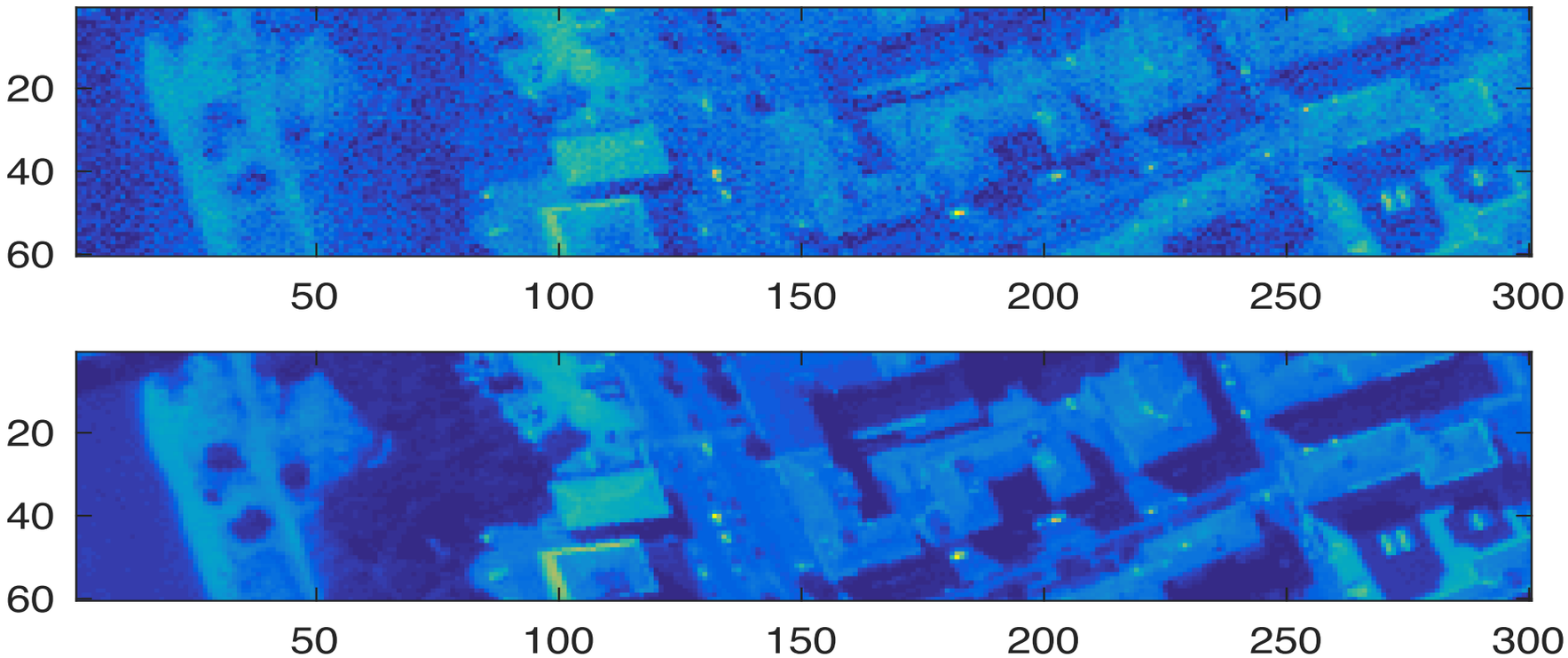
Ground Truth

Each Pixel is a High-Dimensional Vector



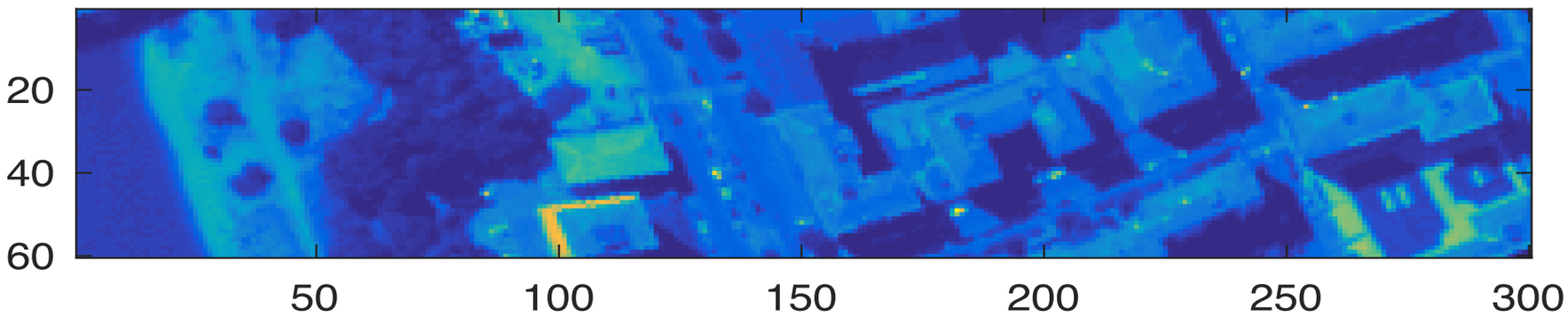
- Large within-class variation.
- Significant between-class overlap in some dimensions.

Individual Spectral Bands

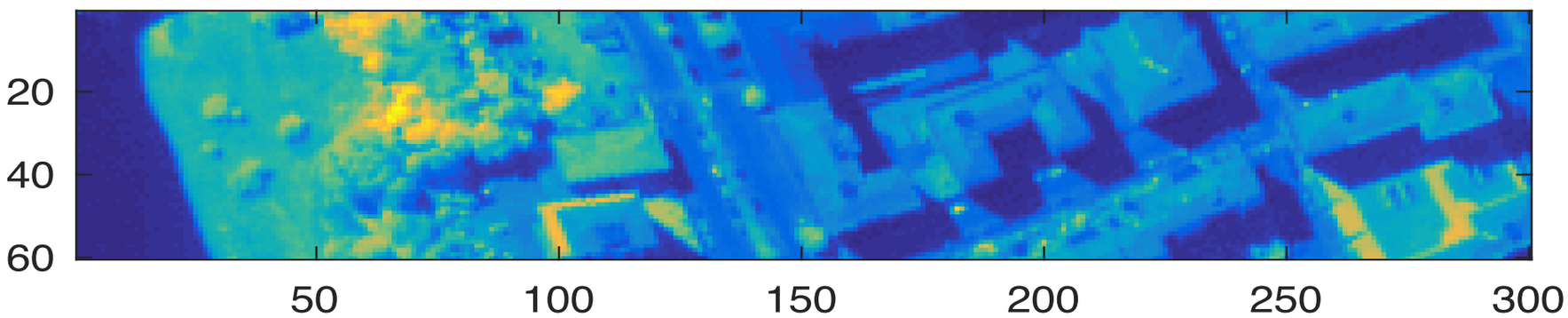


Band 2

Band 20



Band 40



Band 102

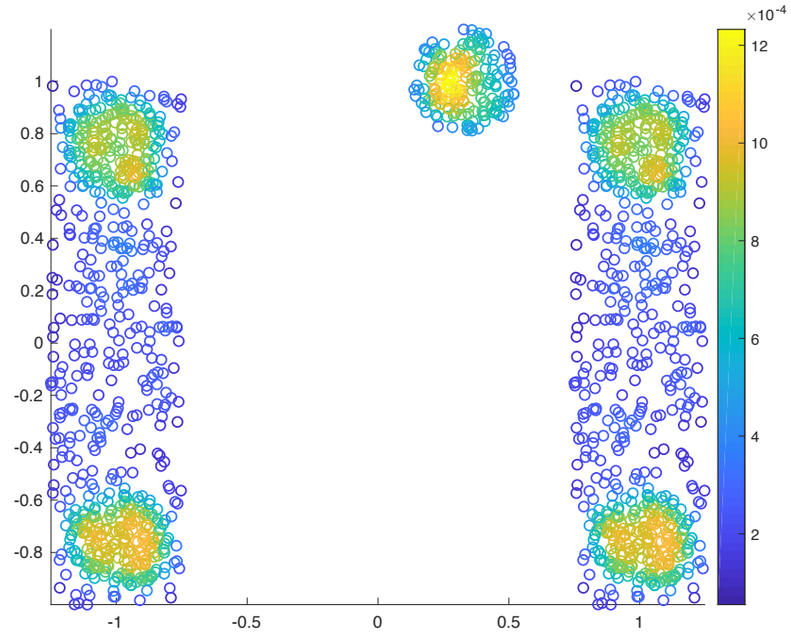
Tufts UNIVERSITY

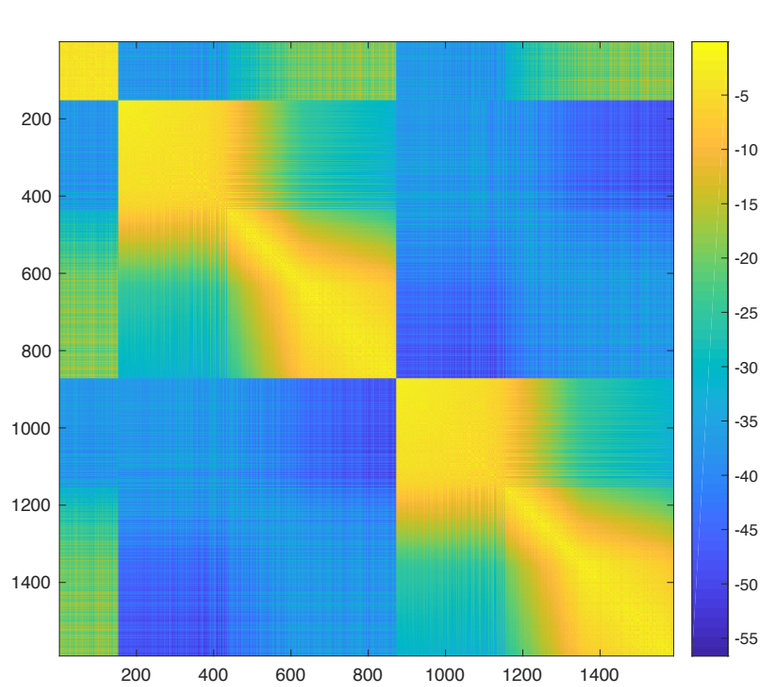
Incorporating Nonlinear Geometry

Learn nonlinear geometry with a diffusion process

$$P_{ij} = \frac{W_{ij}}{\sum_{\ell=1}^{n} W_{i\ell}}$$

$$W_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|_2^2}{\sigma}}, & x_i \in NN_k(x_j), \\ 0, & \text{else.} \end{cases}$$





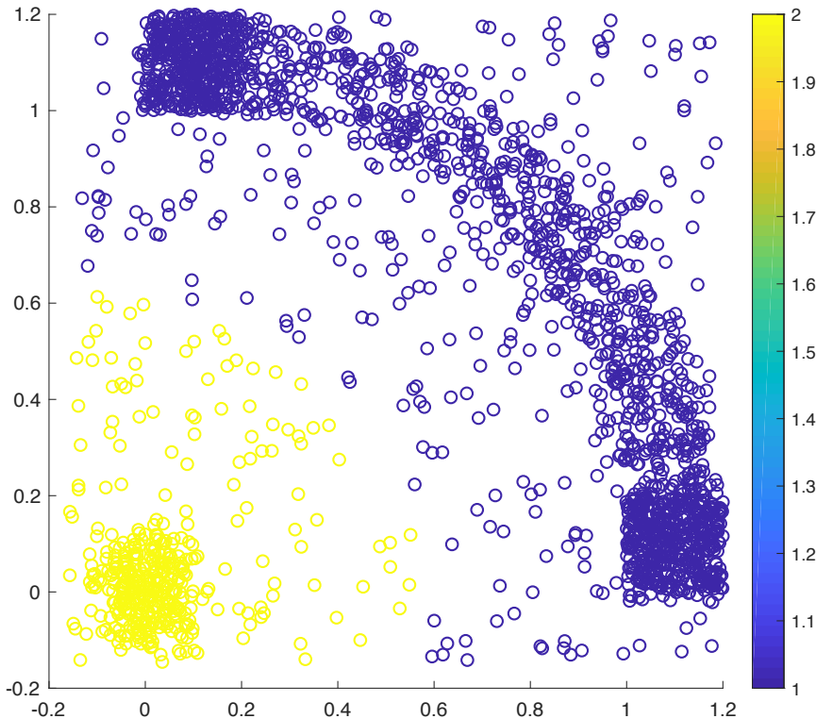
Diffusion Distances:

$$d_t(x_i, x_j) = \sqrt{\sum_{\ell=1}^{n} (P_{i\ell}^t - P_{j\ell}^t)^2 \frac{1}{\pi_\ell}}$$

Spectral Formulation

$$d_t(x_i, x_j) = \sqrt{\sum_{\ell=1}^{n} \lambda_\ell^{2t} (\Phi_\ell(i) - \Phi_\ell(j))^2}$$

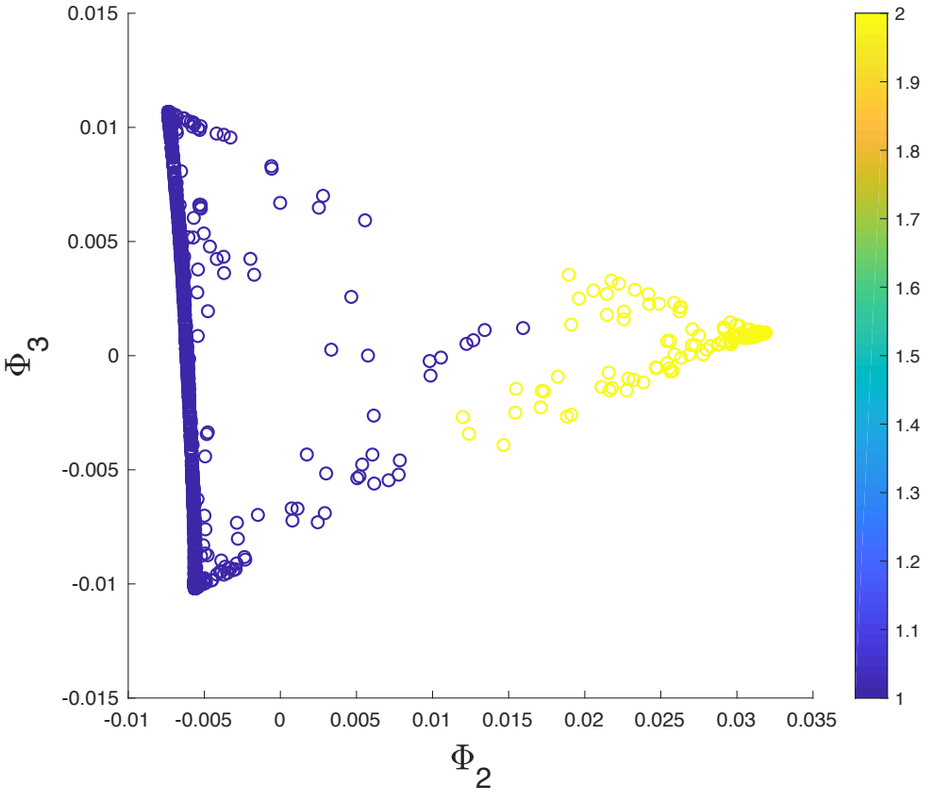
$\{\lambda_\ell, \Phi_\ell\}_{\ell=1}^{n}$ Spectral decomposition of P



Convert to

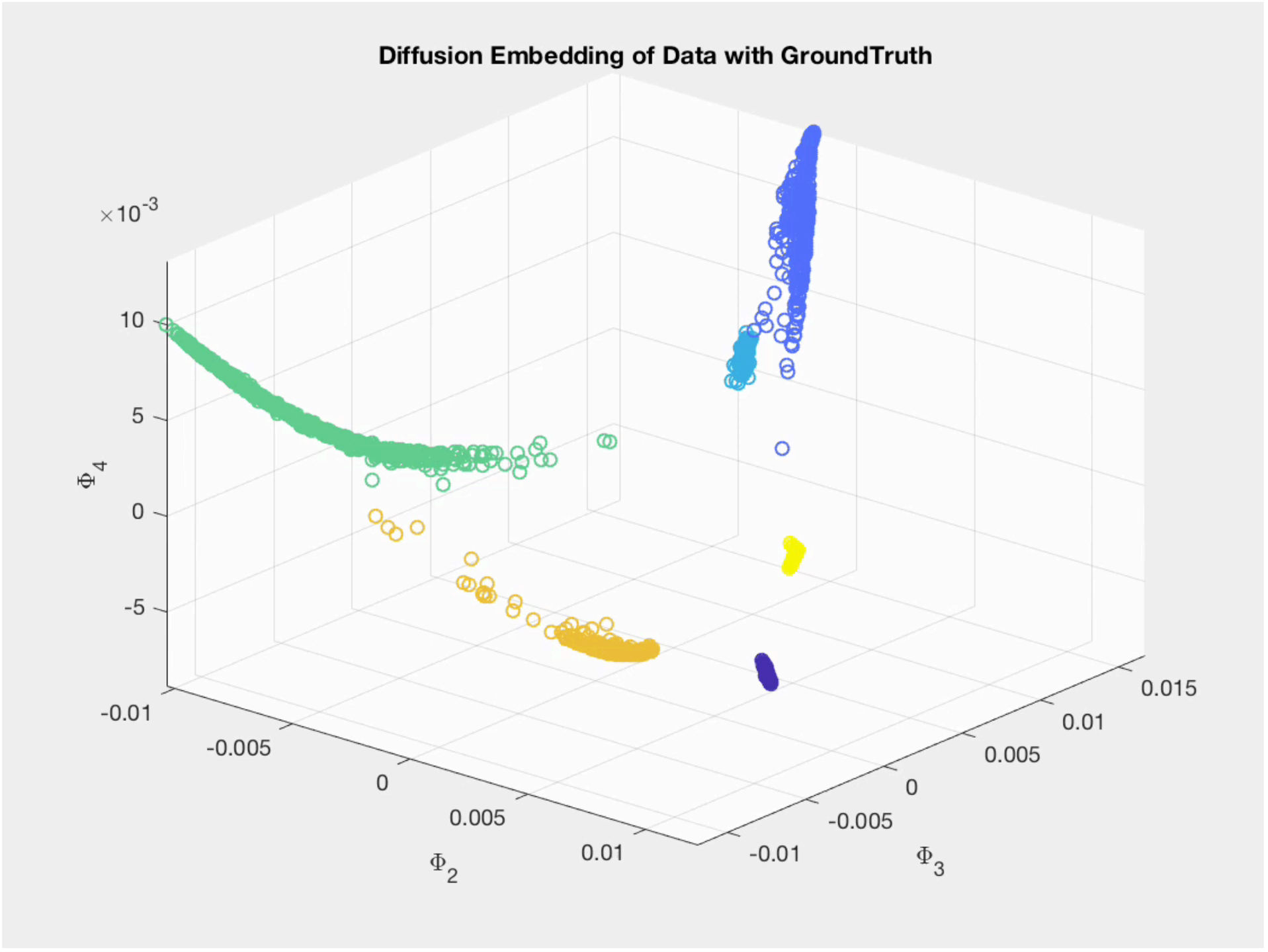
(Φ_2, Φ_3)

Coordinates



Tufts UNIVERSITY

Exploiting Nonlinear Structure: Diffusion Maps



Learning by Unsupervised Nonlinear Diffusion (LUND)

1.) **Compute empirical density:**

$$p_0(x_i) = \sum_{x_j \in NN_k(x_i)} e^{\frac{-\|x_i - x_j\|_2^2}{\sigma^2}}$$

$$p(x_i) = p_0(x_i) / \sum_{j=1}^{n} p_0(x_j)$$

2.) **Find points that are d_t -far from higher density points:**

$$\tilde{\rho}_t(x_i) = \begin{cases} \min\limits_{\{p(x_j) \geq p(x_i)\}} d_t(x_i, x_j), & x_i \neq \arg\max\limits_i p(x_i), \\ \max\limits_{x_j} d_t(x_i, x_j), & x_i = \arg\max\limits_i p(x_i). \end{cases}$$

$$\rho_t(x_i) = \tilde{\rho}_t(x_i) / \max_{x_j} \tilde{\rho}_t(x_j)$$

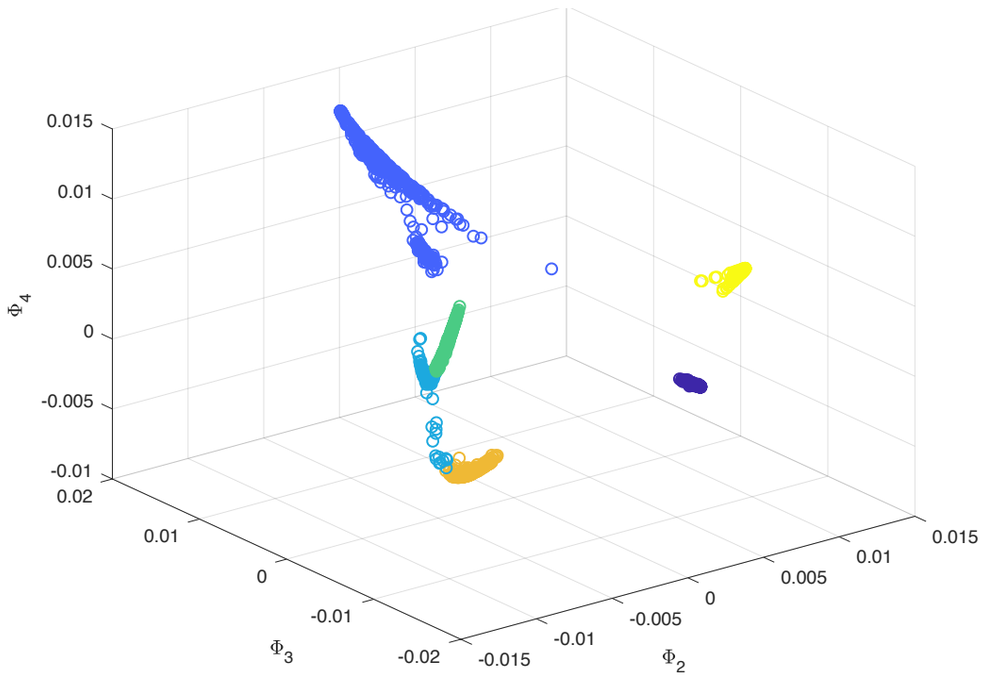
3.) **Estimate modes as maximizers of:**

$$\boxed{\mathcal{D}_t(x_i) = p(x_i)\rho_t(x_i)}$$

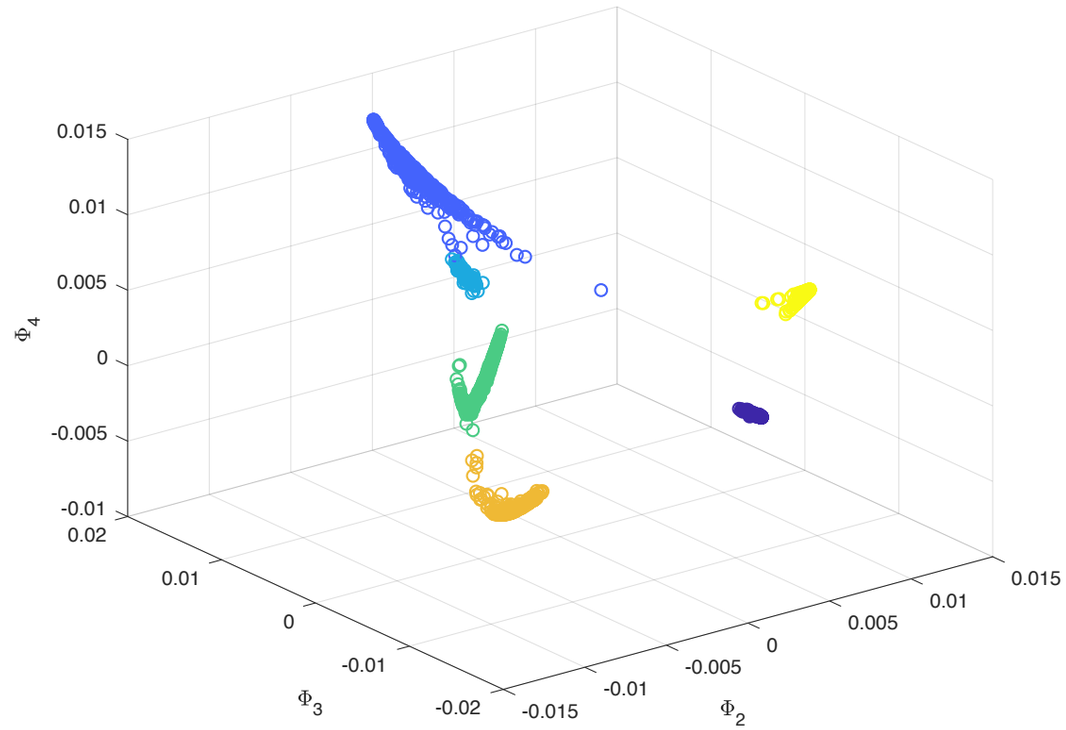
Learning by Unsupervised Nonlinear Diffusion (LUND)

Assign points the label of d_t-nearest neighbor of higher density.

LUND Labels



Ground Truth



With fast nearest-neighbor look-ups, complexity is $O(n \log(n) D C^d)$

D — ambient dimension

d — intrinsic dimension

n — number of data points

Mathematical Guarantees

Let $X = \bigcup_{k=1}^{K} X_k$ be the latent clusters in the data.

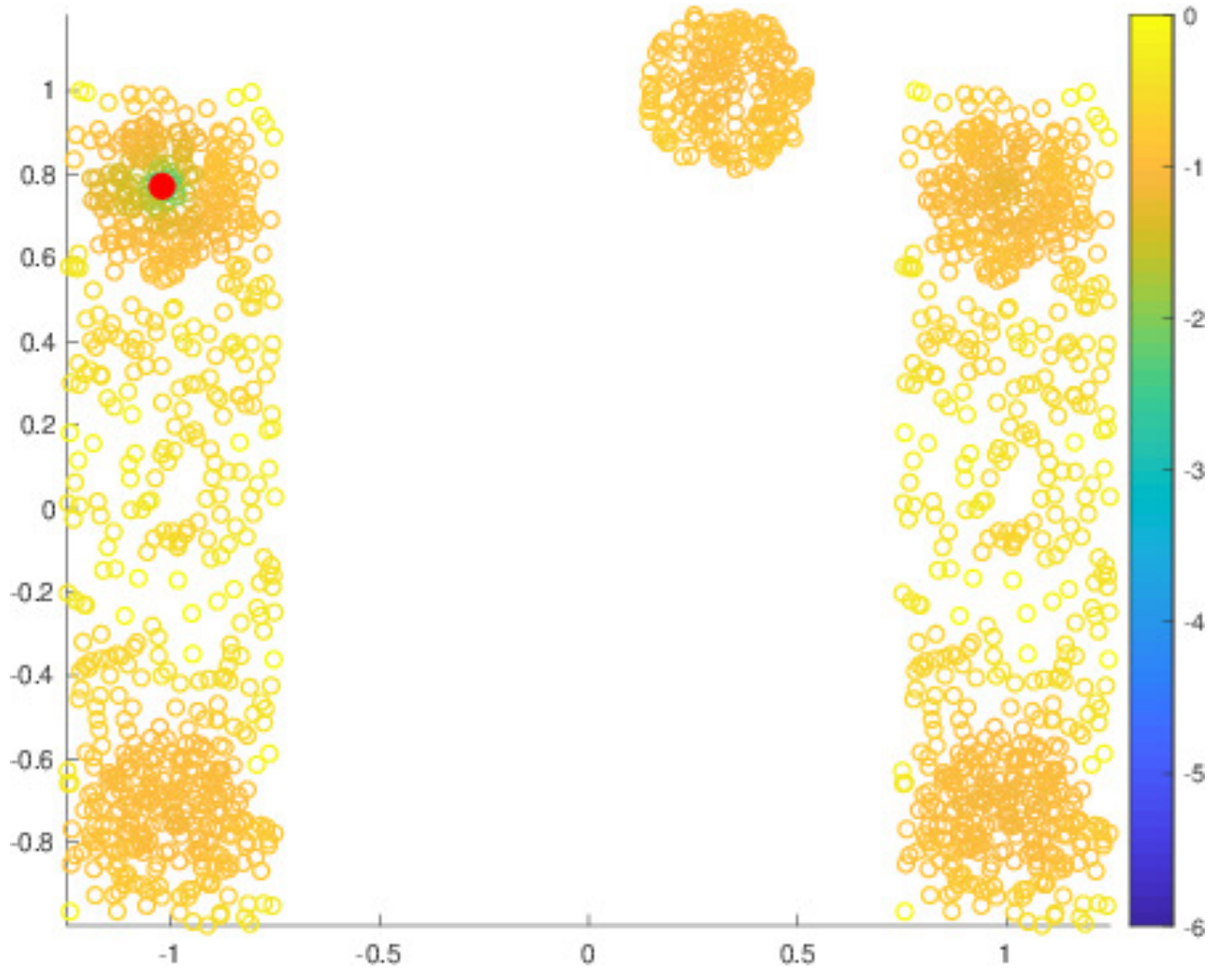
$$D_t^{in} = \max_k \max_{x,y \in X_k} d_t(x,y), \quad D_t^{btw} = \min_{k \neq k'} \min_{x \in X_k, y \in X_{k'}} d_t(x,y)$$

Theorem. *(Maggioni, **M.**) Let $X = \bigcup_{k=1}^{K} X_k$ and let P be a corresponding Markov transition matrix on X, inducing diffusion distances $\{d_t\}_{t \geq 0}$. Then there exist constants $\{C_i\}_{i=1}^{5} \geq 0$ such that the following holds: for any $\epsilon > 0$, and for any t satisfying $C_1 \ln\left(\frac{C_2}{\epsilon}\right) < t < C_3 \epsilon$, we have*

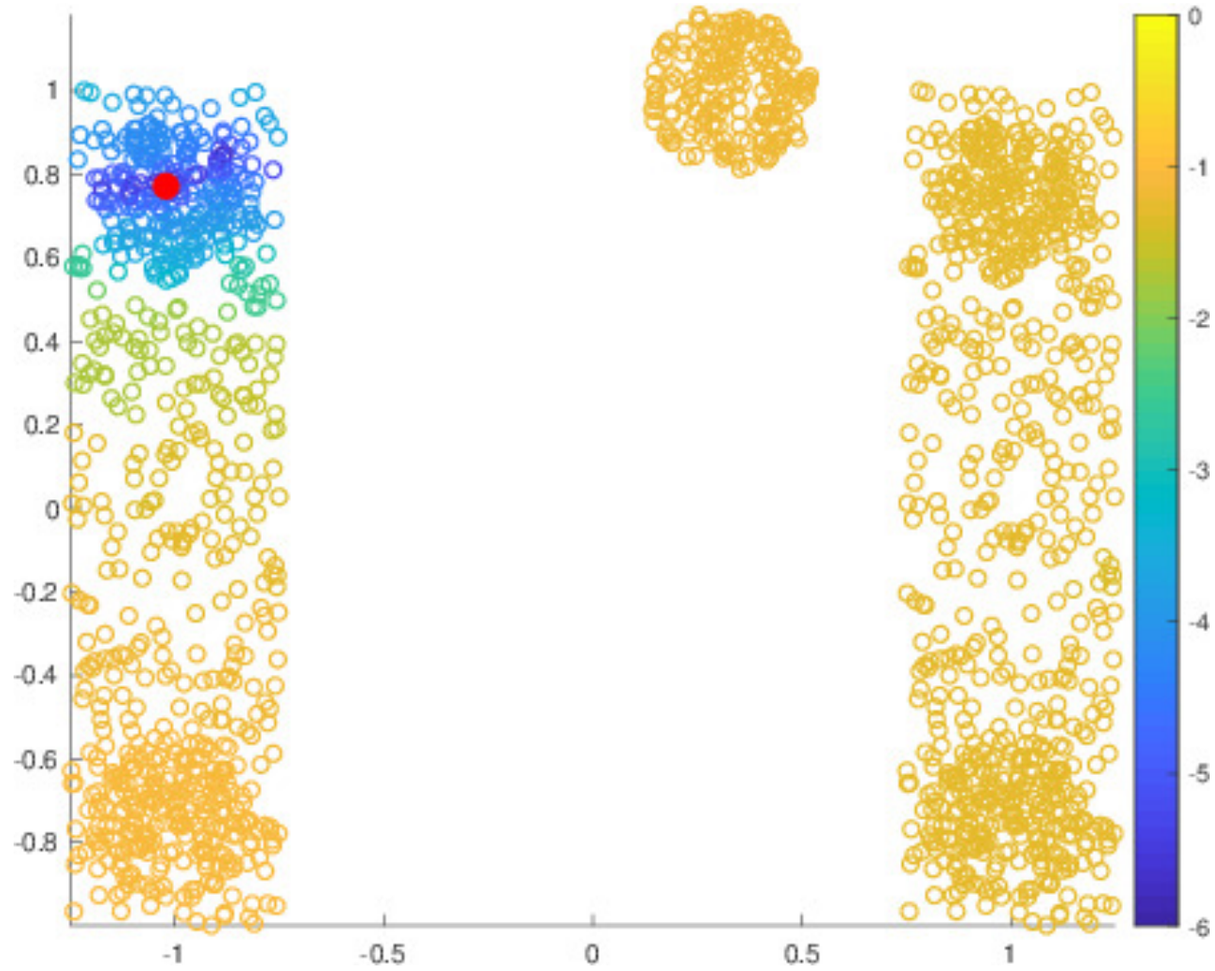
$$D_t^{in} \leq C_4 \epsilon, \quad D_t^{btw} \geq C_5 - C_4 \epsilon.$$

The constants $\{C_i\}_{i=1}^{5}$ depend on the data. More separation between clusters and cohesion within cluster lead to better constants.

Multiscale Equilibria I



$t = 0$

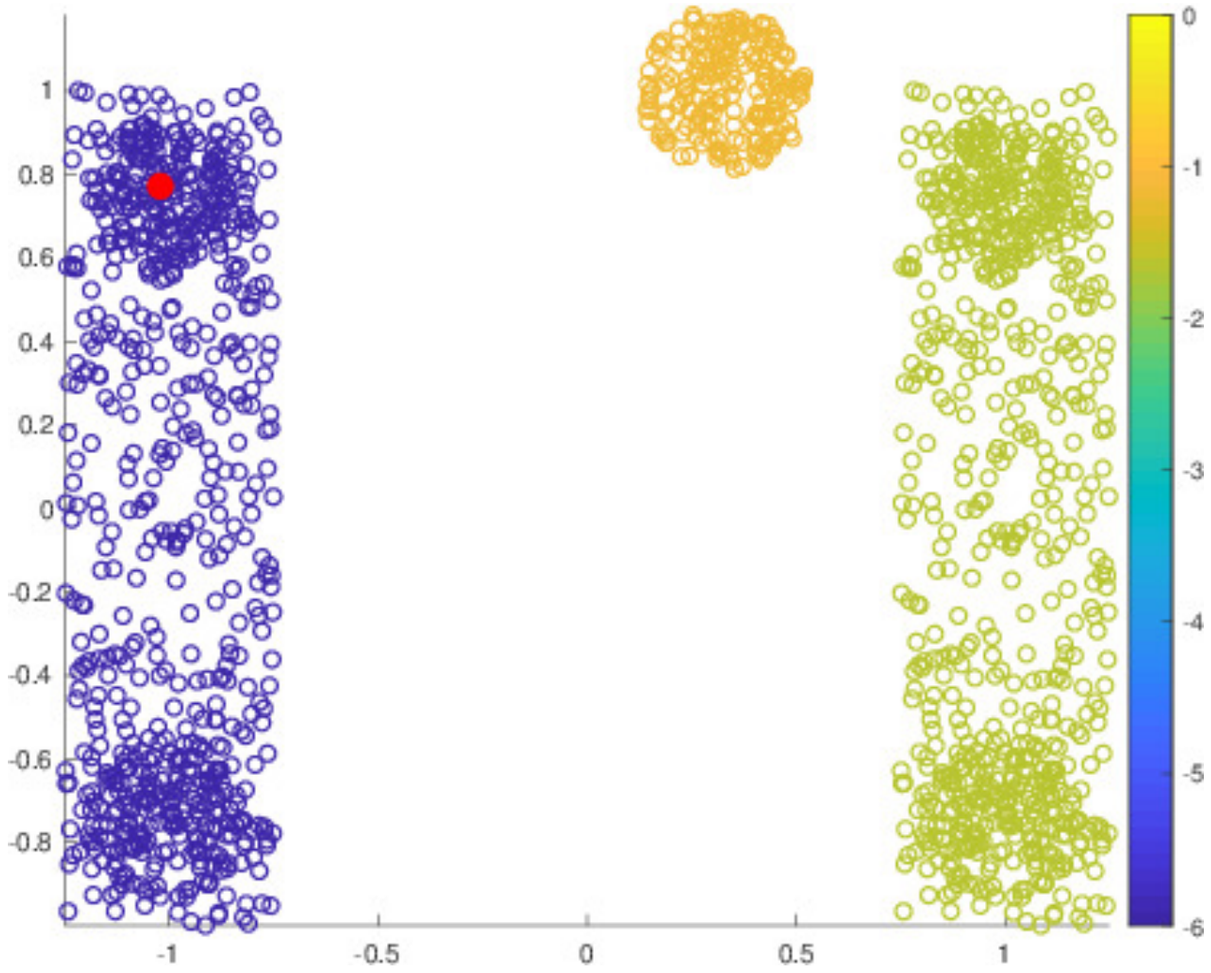


$t = 10^2$

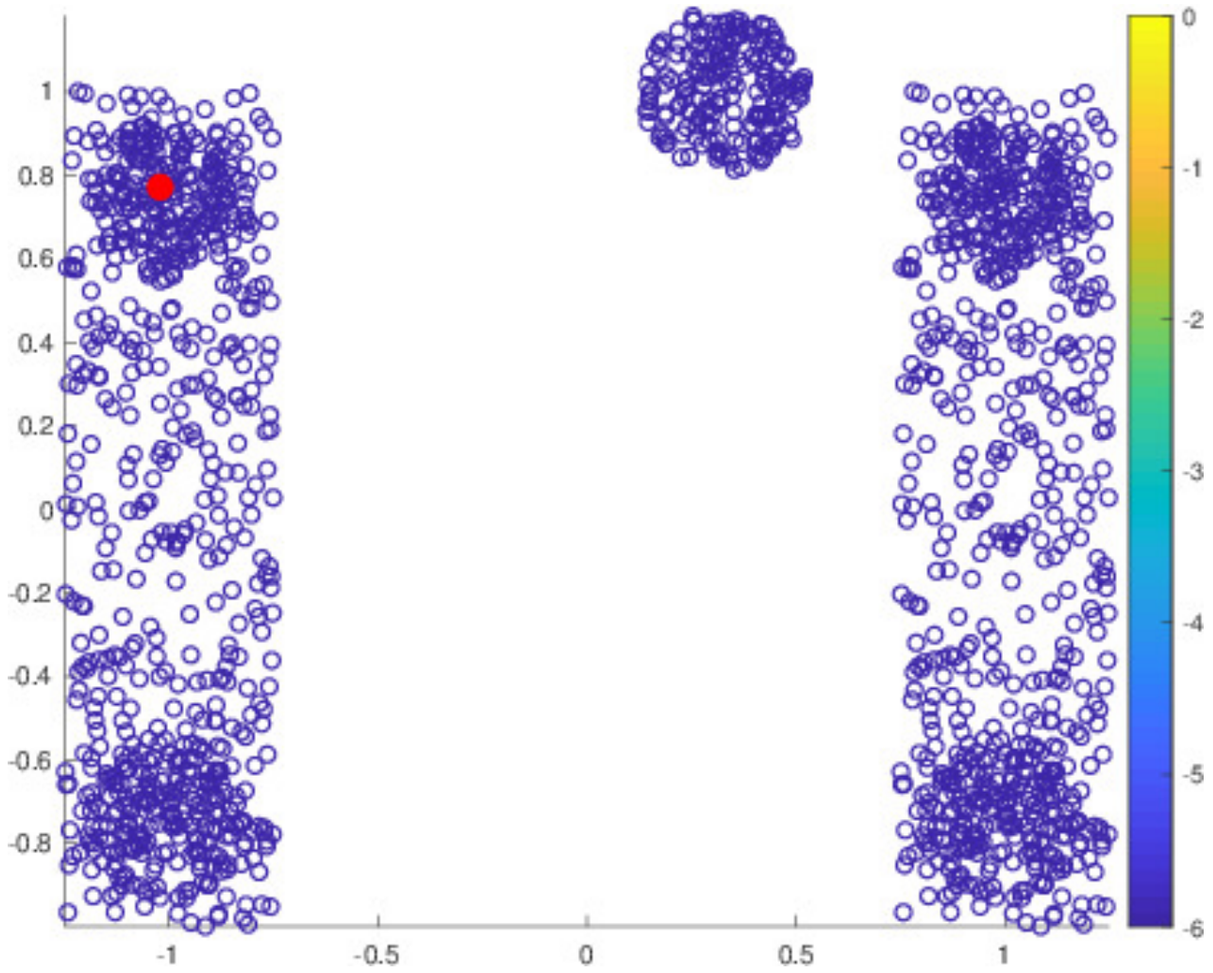
Diffusion distances from red point in log scale:
small times lead to local mixing.

Tufts UNIVERSITY

Multiscale Equilibria II



$t = 10^8$



$t = 10^{16}$

Diffusion distances from red point in log scale: as time increases, mesoscopic equilibria, then global equilibrium is reached.

Tufts UNIVERSITY

Mathematical Guarantees

$$M = \{p(x) \mid \exists k \text{ such that } x = \text{argmax}_{y \in X_k} p(y)\}$$

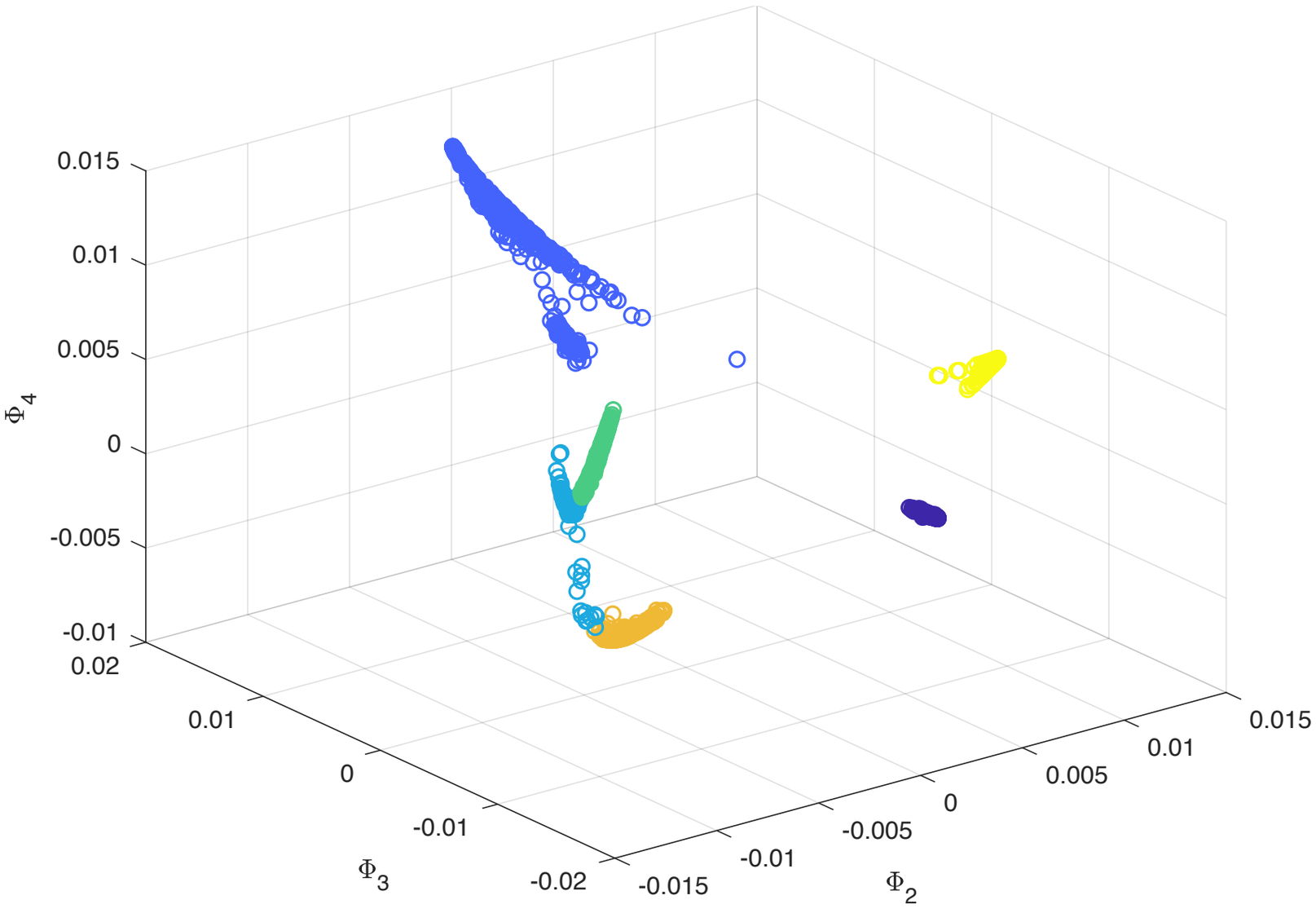
Theorem *(Maggioni, **M.**) Let $X = \bigcup_{k=1}^{K} X_k$ be data to cluster. If K is known a priori, then LUND achieves perfect accuracy if*

$$\frac{D_t^{in}}{D_t^{btw}} < \frac{\min(M)}{\max(M)}.$$

- The more well-separated and internally cohesive the clusters are, the greater time range in which accuracy is assured.
- Similar result available when K is unknown a priori.
- Proofs based on analysis of Markov matrices in relationship to near reducibility and mixing times.

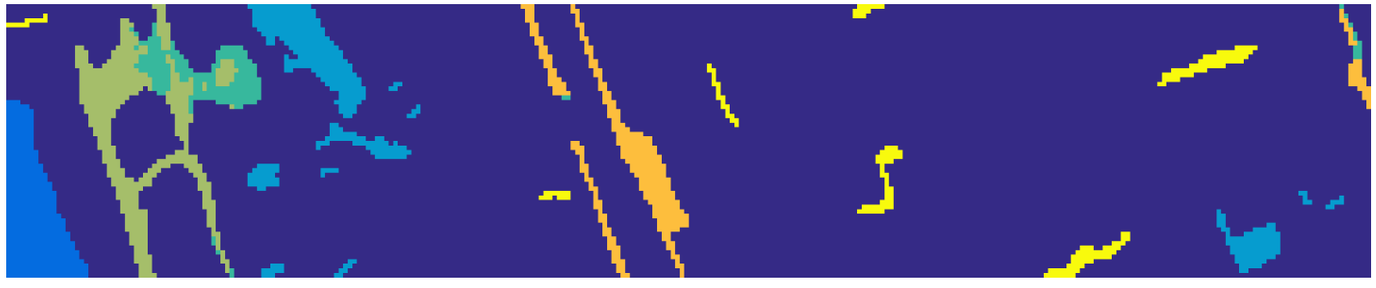
Tufts UNIVERSITY

The value of spatial information for HSI



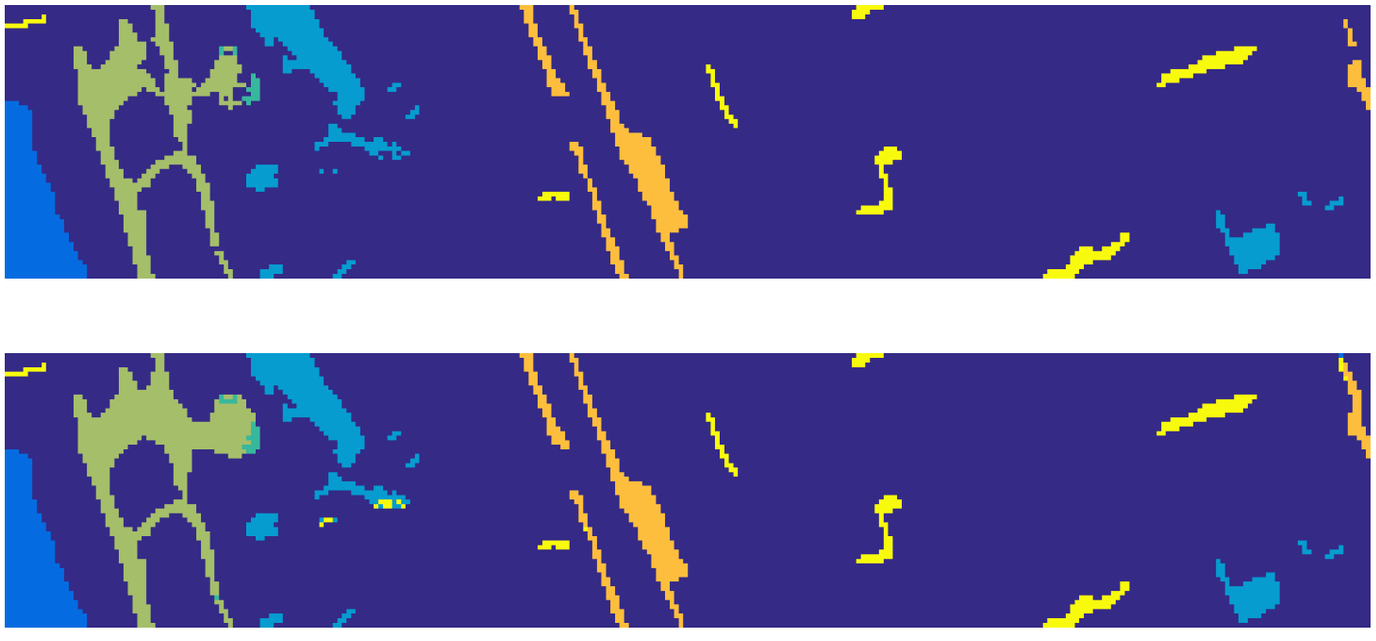
Spectral information only…some problems near class boundaries.

Spectral-only labels

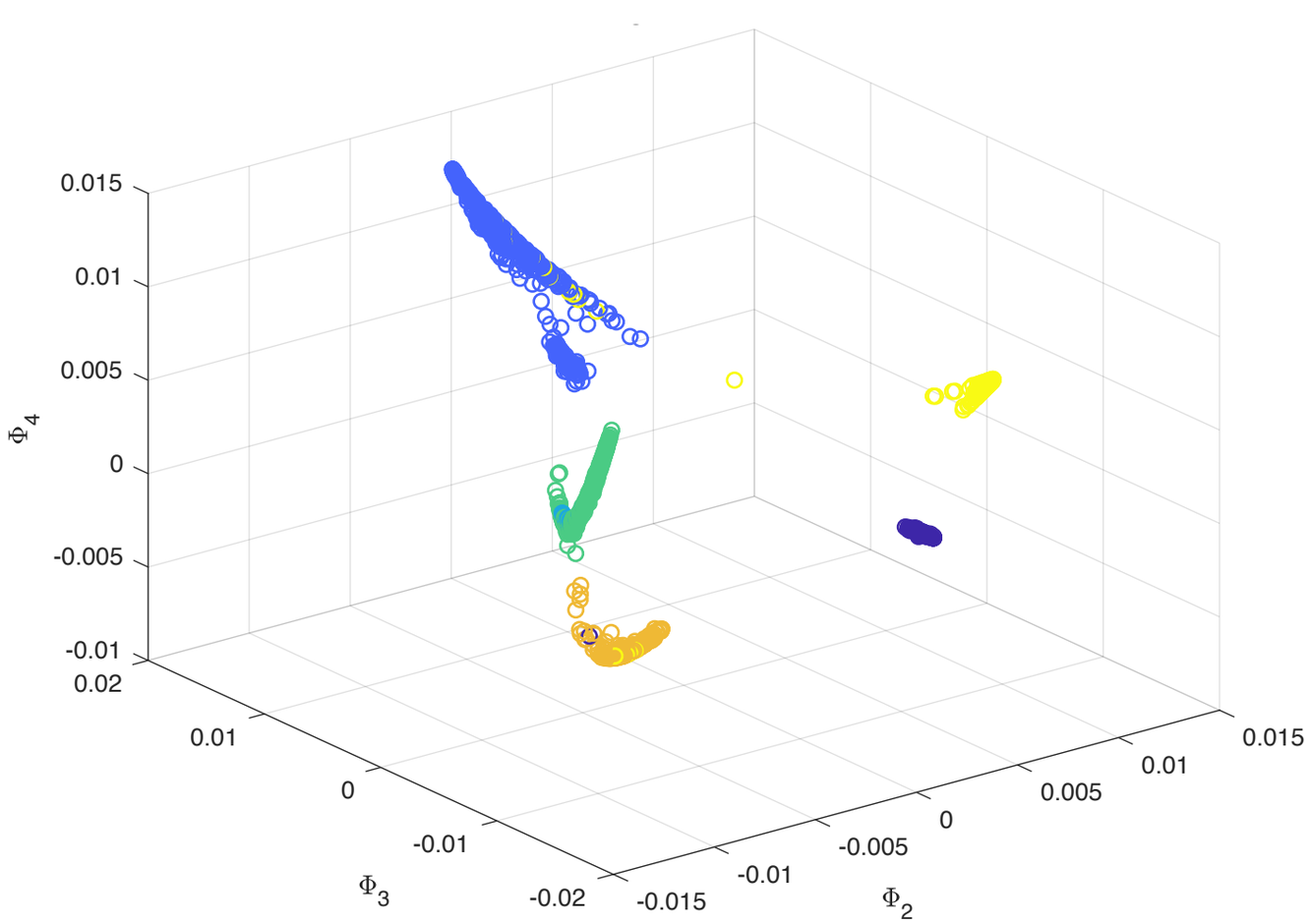


Tufts UNIVERSITY

Two Stage Labeling



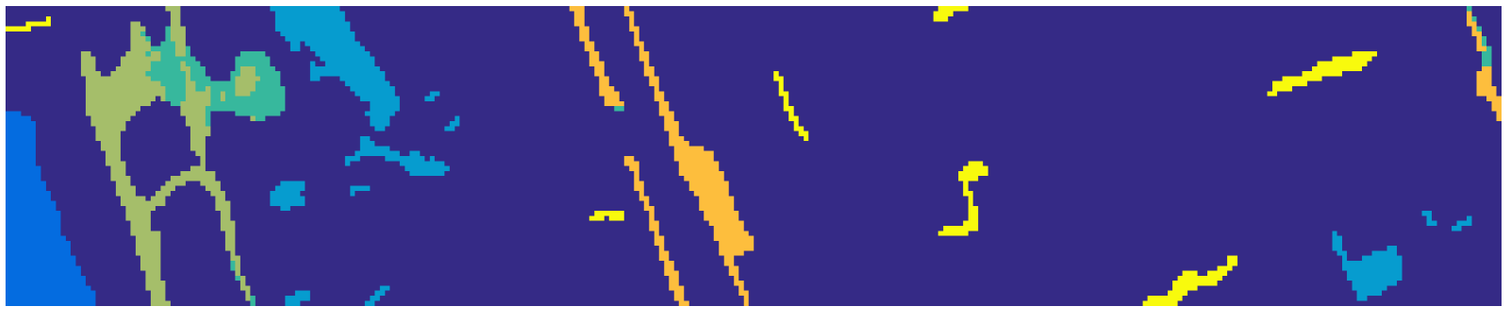
Stage 1: Only label points that can confidently be labeled with spectral information.



Stage 2: Then, use spatial information to help fill in the remaining, spectrally ambiguous points.

Tufts UNIVERSITY

Empirical Clustering Results



Spectral-Only Overall Accuracy:

.8494

Spatial-Spectral Overall Accuracy:

.9461

Ground Truth

Compares very well to state-of-the-art….and fast!

Tufts UNIVERSITY

Generalization to Active Learning

Active learning: Given O(1) labels, which points to query?

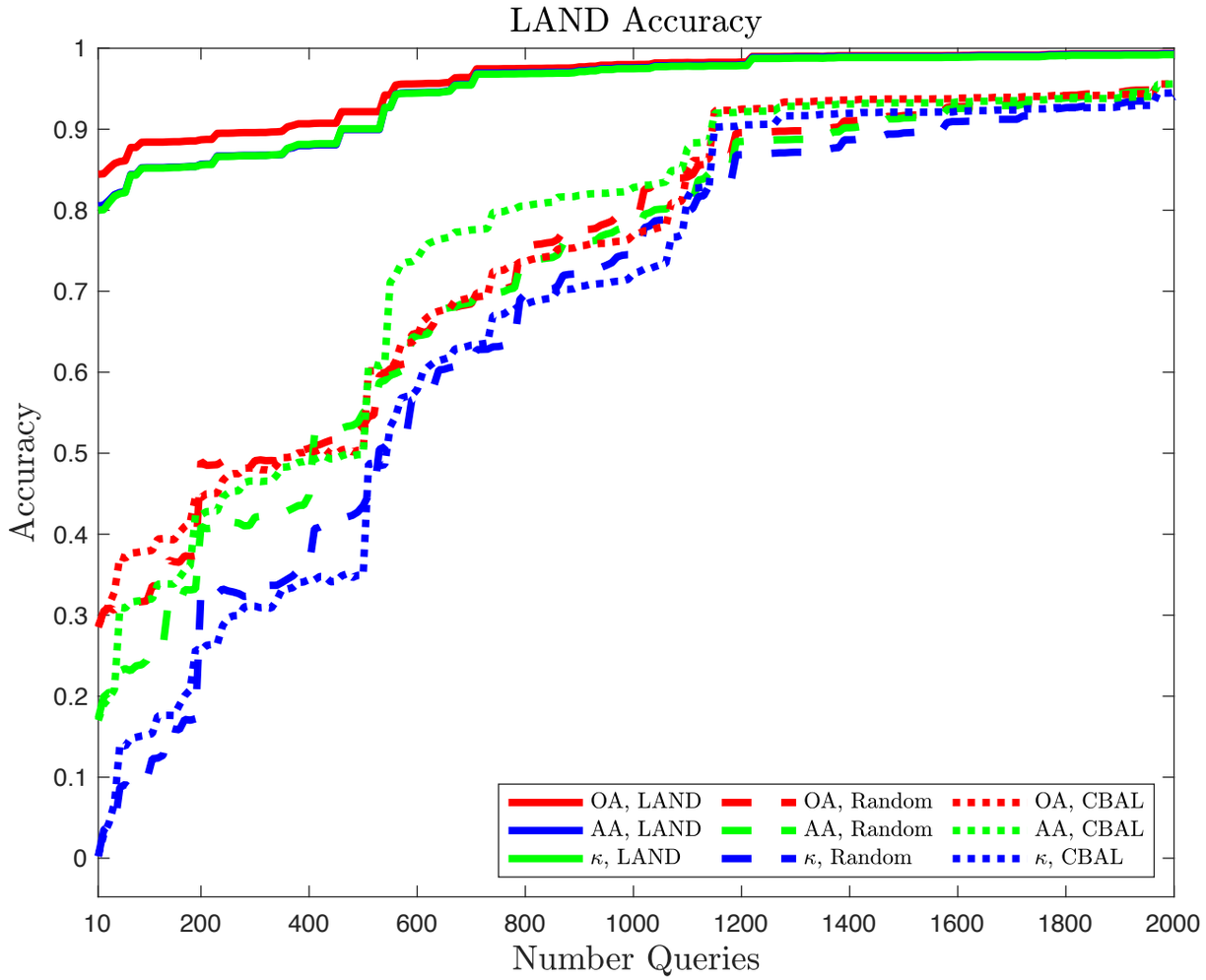
Two major paradigms for active learning:

1. **Margin-based**: set the boundaries between classes as quickly as possible.
2. **Cluster-based**: use latent cluster structure to sample ambiguous regions in the data.

LUND lends itself to the second paradigm, where cluster modes are queried for labels. We call this *Learning by Active Nonlinear Diffusion (LAND).*

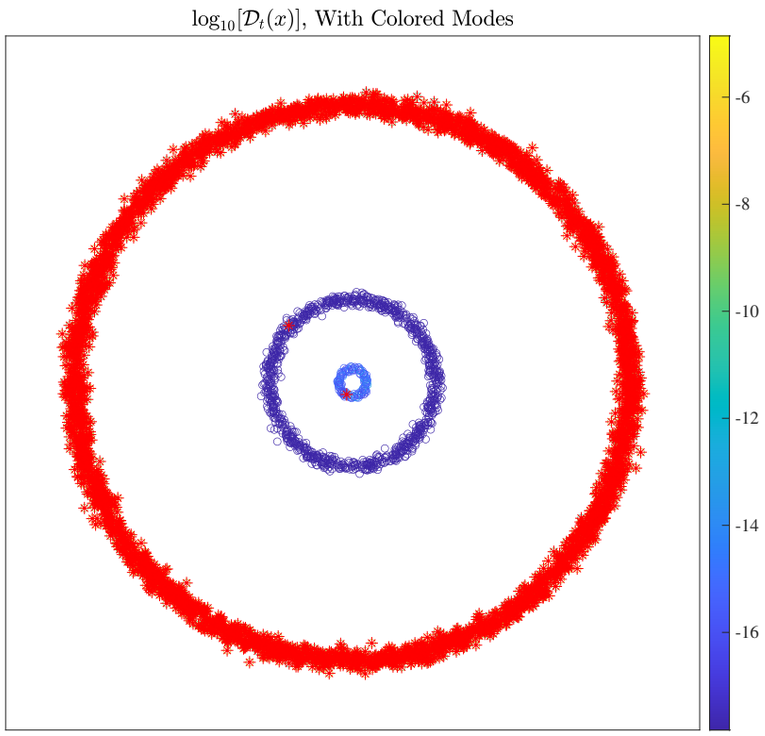
Learning by Active Nonlinear Diffusion (LAND)

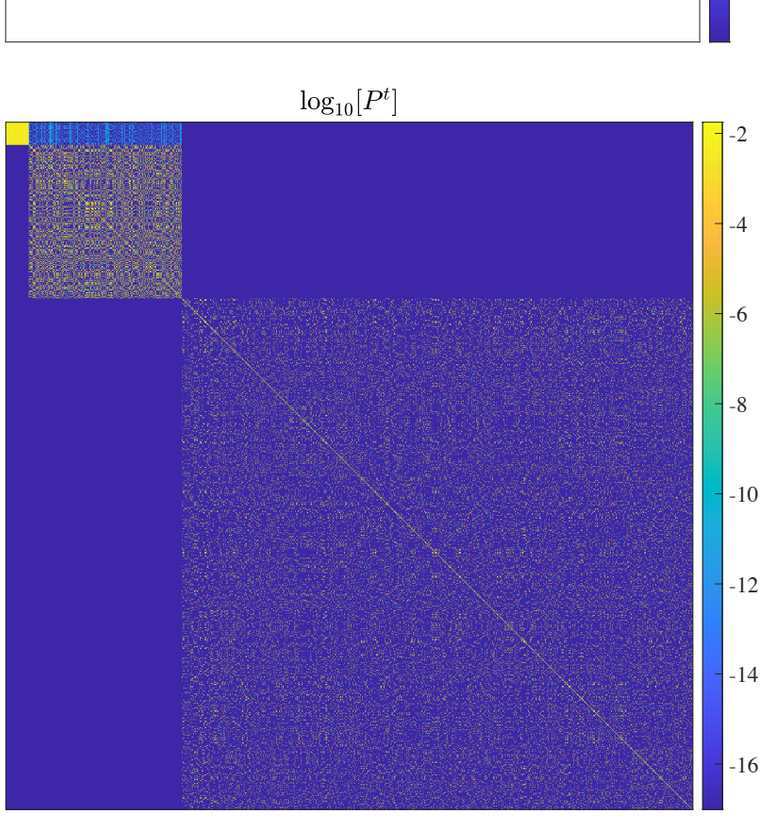
Theorem *(Maggioni, **M.**) Let $X = \bigcup_{k=1}^{K} X_k$ be data to classify. Suppose that $D_t^{in} < D_t^{btw}$, and that the B maximizers of $\mathcal{D}_t$ include the elements of $\mathcal{M}$. Then LAND with a budget of size B achieves perfect classification accuracy.*



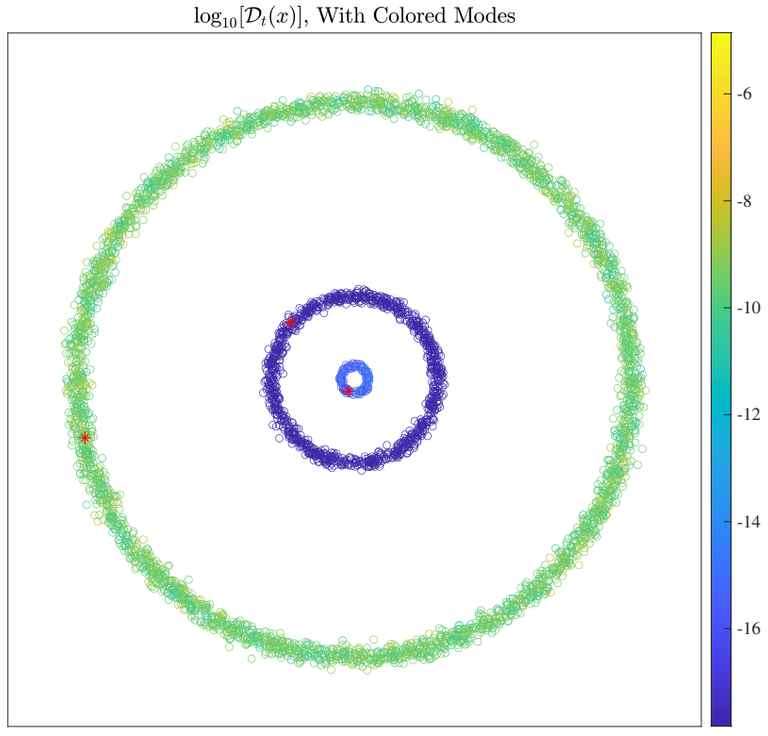
On Salinas A, LAND achieves improvements in accuracy rapidly, compared to Euclidean-based active learning methods and random sampling.

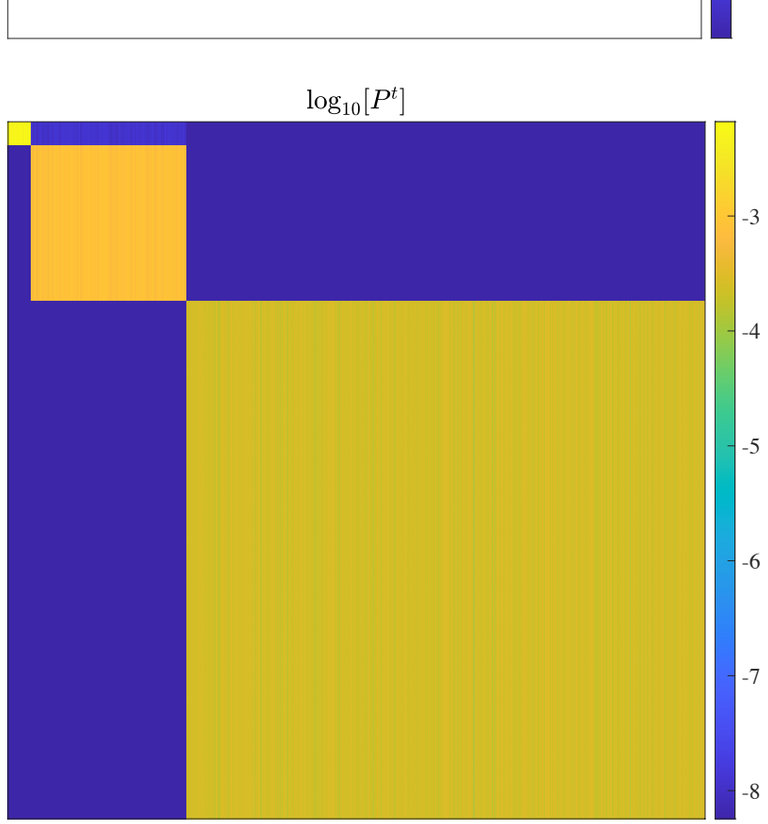
M-LUND: Multiscale Cluster Models



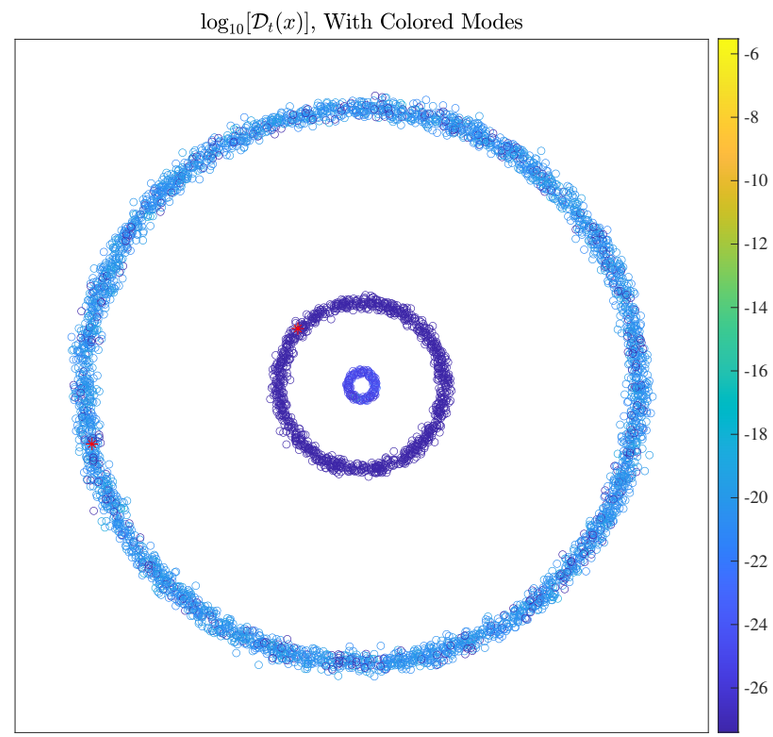


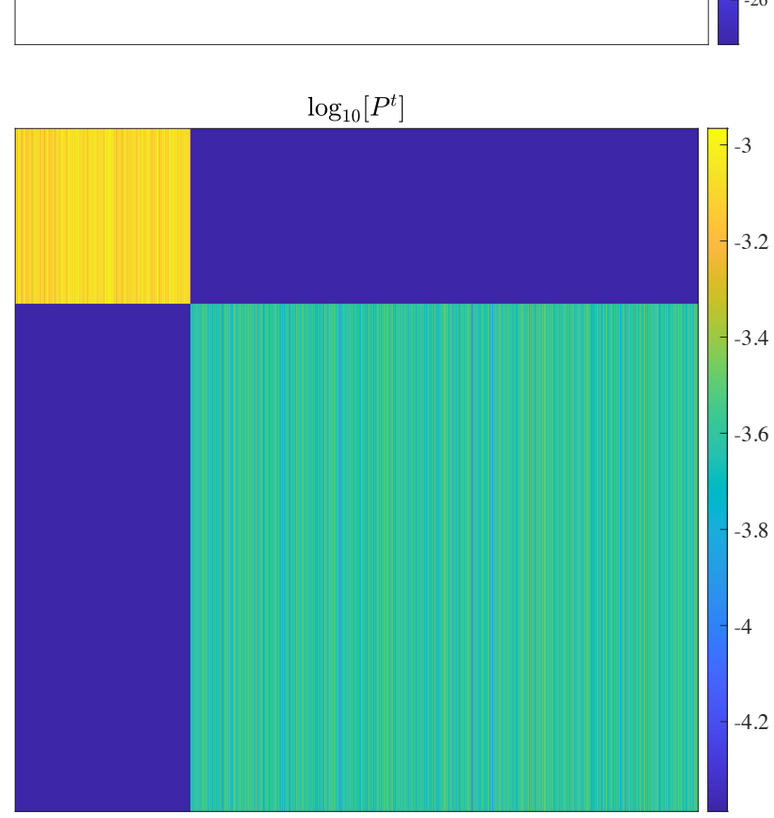
t small





t medium





t large

Tufts UNIVERSITY

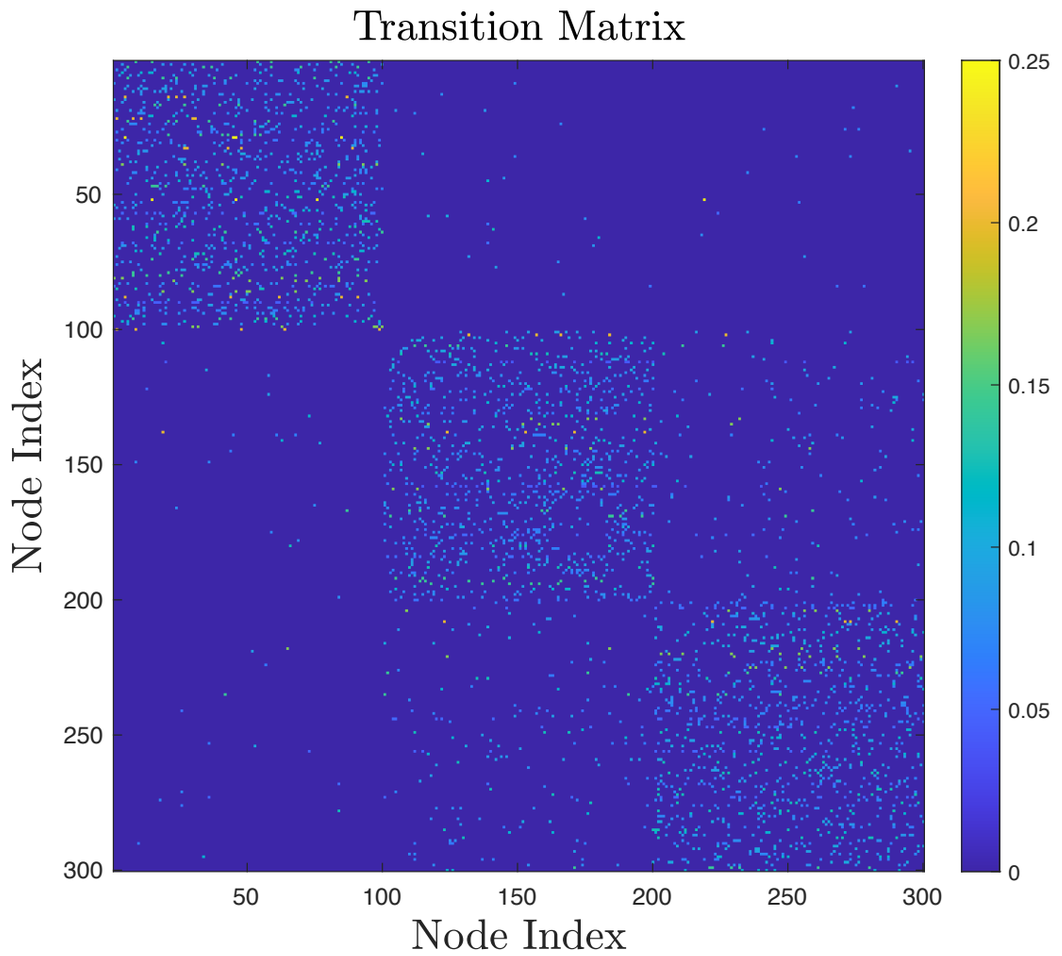
Analyzing Data with Multiscale Structure

What if there is more than one time scale that makes sense for the data?

Diffusion State Distances:

$$\mathcal{D}_w^p(x_i, x_j) = \left\| \sum_{t=0}^{\infty} (e_i - e_j) P^t \right\|_{\ell^p(w)}$$

$$= \left(\sum_{\ell=1}^{n} \left| \sum_{t=0}^{\infty} P_{i\ell}^t - P_{j\ell}^t \right|^p w(\ell) \right)^{\frac{1}{p}}$$

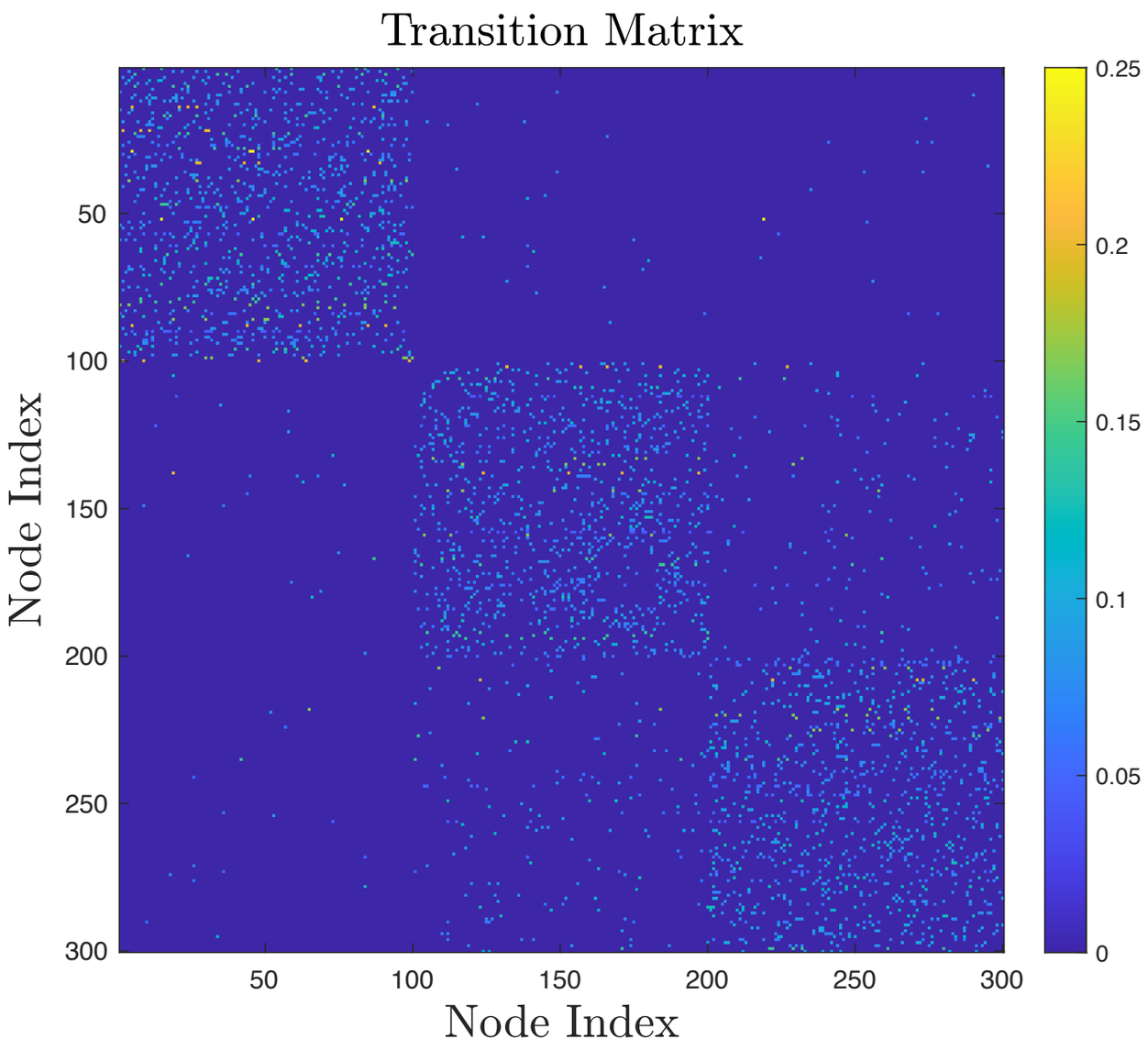


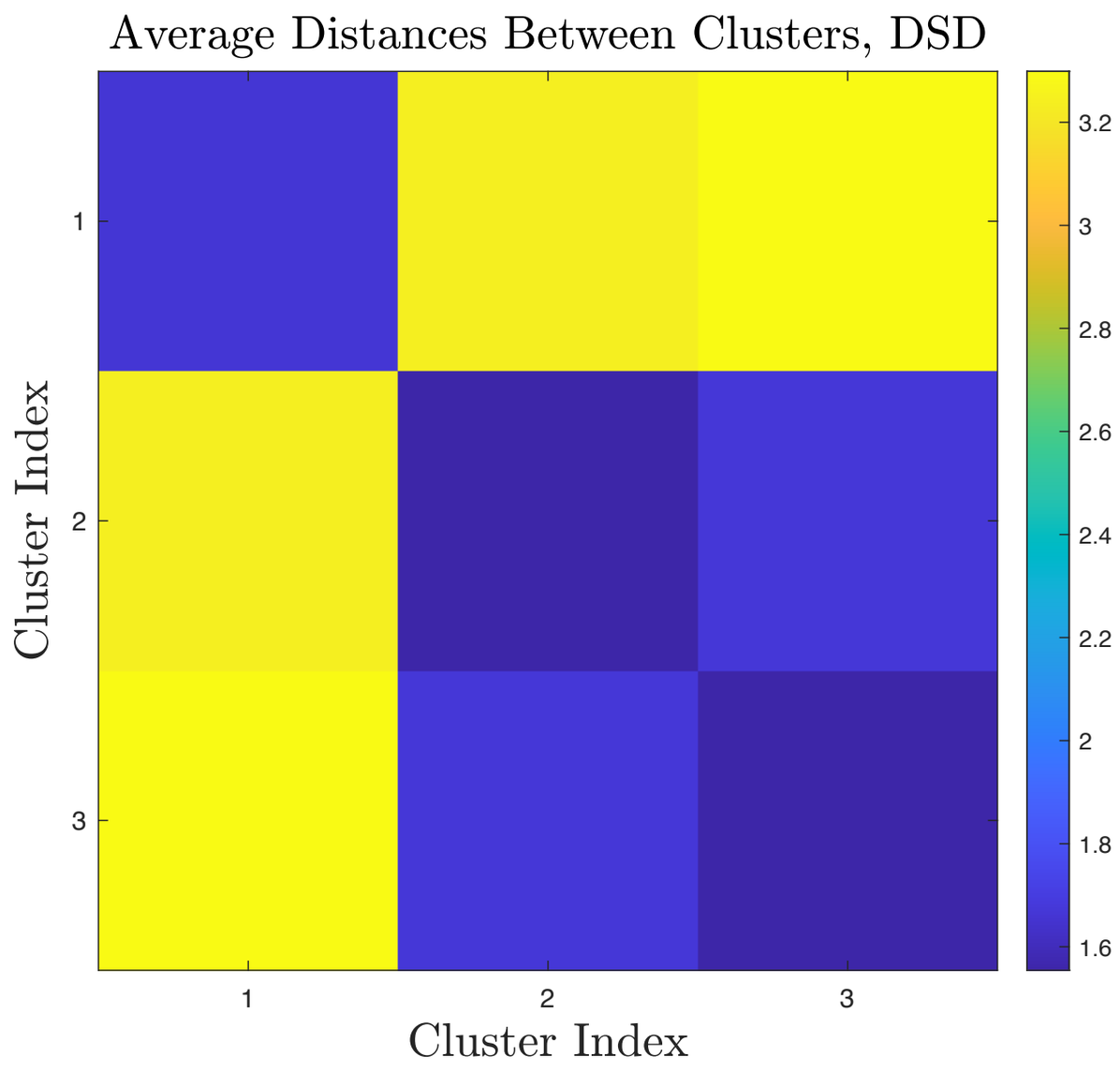
$p = 2, w = \frac{1}{\pi}$ natural choices.

Accounts for *multiscale structure* in the data.

Tufts UNIVERSITY

Hierarchical SBM and DSD





DSD captures the hierarchical structure in the connectivity structure of the HSBM.

Efficient, Low-Dimensional Embedding

Theorem *(Cowen, Devkota, Hu, **M.**, Wu) The diffusion state distance with weight $w = 1/\pi$ and $p = 2$ admits the decomposition*

$$\mathcal{D}^2_{1/\pi}(x_i, x_j) = \| \sum_{t=0}^{\infty} (e_i - e_j) P^t \|_{\ell^2(1/\pi)} = \sqrt{\sum_{\ell=1}^{n} (1-\lambda_\ell)^{-2} (\psi_\ell(i) - \psi_\ell(j))^2},$$

where $\{(\lambda_\ell, \psi_\ell)\}_{\ell=1}^n$ are the eigenvalues and right eigenvectors of P.

Low-dimensional embedding:

$$x_i \mapsto (\tfrac{1}{1-\lambda_2}\psi_2(i), \ldots, \tfrac{1}{1-\lambda_m}\psi_m(i))$$

Fast solvers (e.g. AMG) allow this to scale on big networks.

Tufts UNIVERSITY

Protein-Protein Interaction Networks

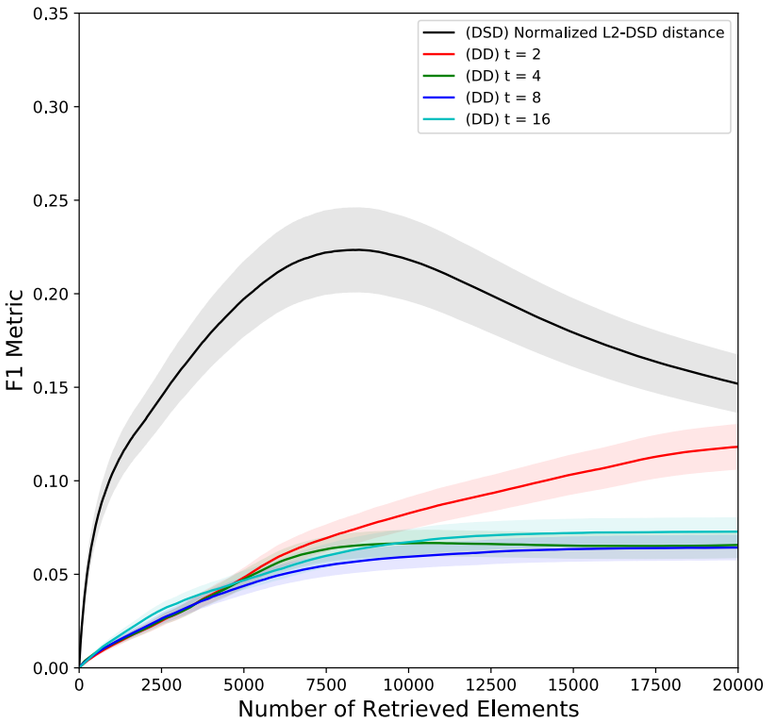
- Protein-protein interaction (PPI) networks: vertices represent proteins, edge connections between them.

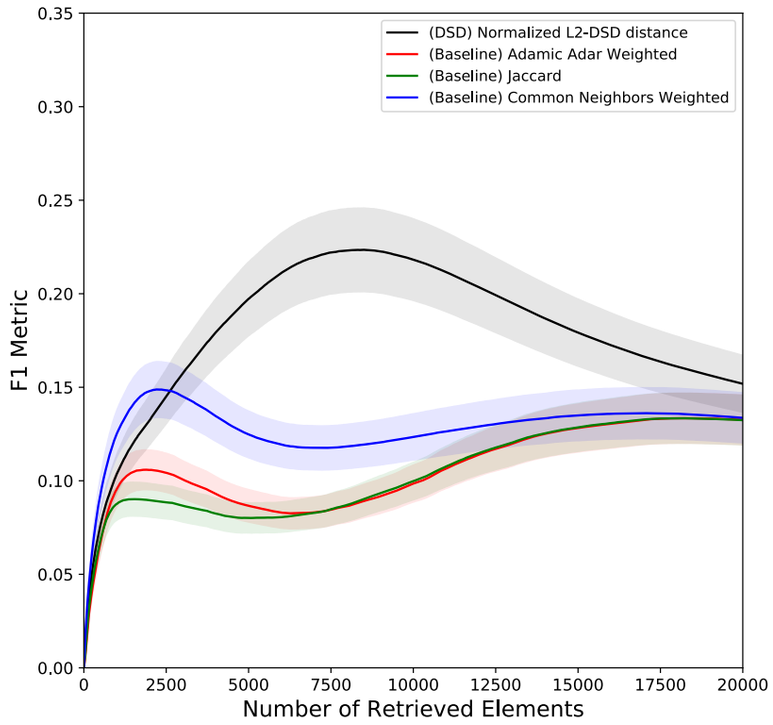
- Two proteins are connected by an edge for reasons of, e.g. experimental evidence that they bind in the cell; expressed in the same human tissues; similar function.

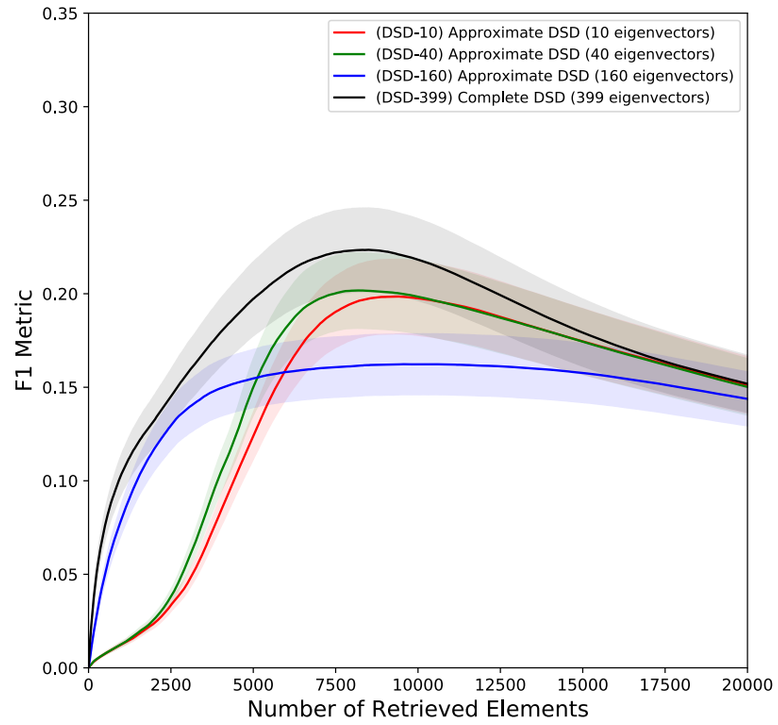
- The benchmark DREAM networks are fairly large and somewhat sparse (~20000 nodes, average degree 100).

- Want to predict protein functions and classes based on a *very small number* of labels and the network properties.

DSD Predicts Missing Links In PPI Networks







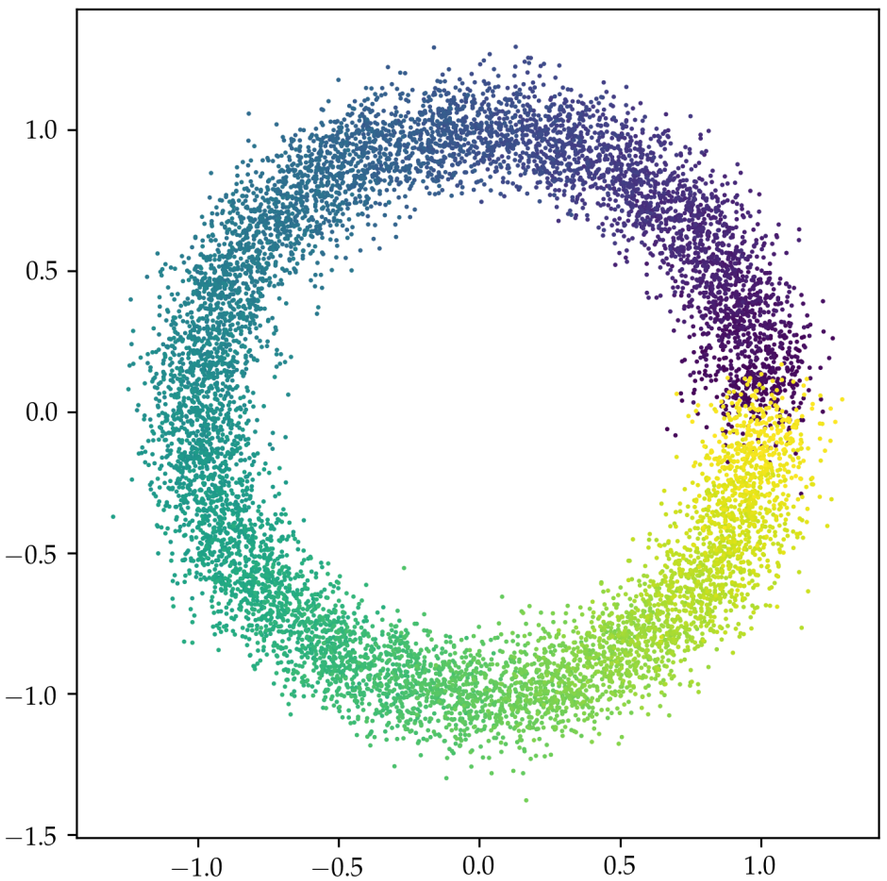
Using DSD to predict missing links in benchmark PPI yields state-of-the-art empirical results, and computational speedup can be achieved with truncated spectral decompositions.

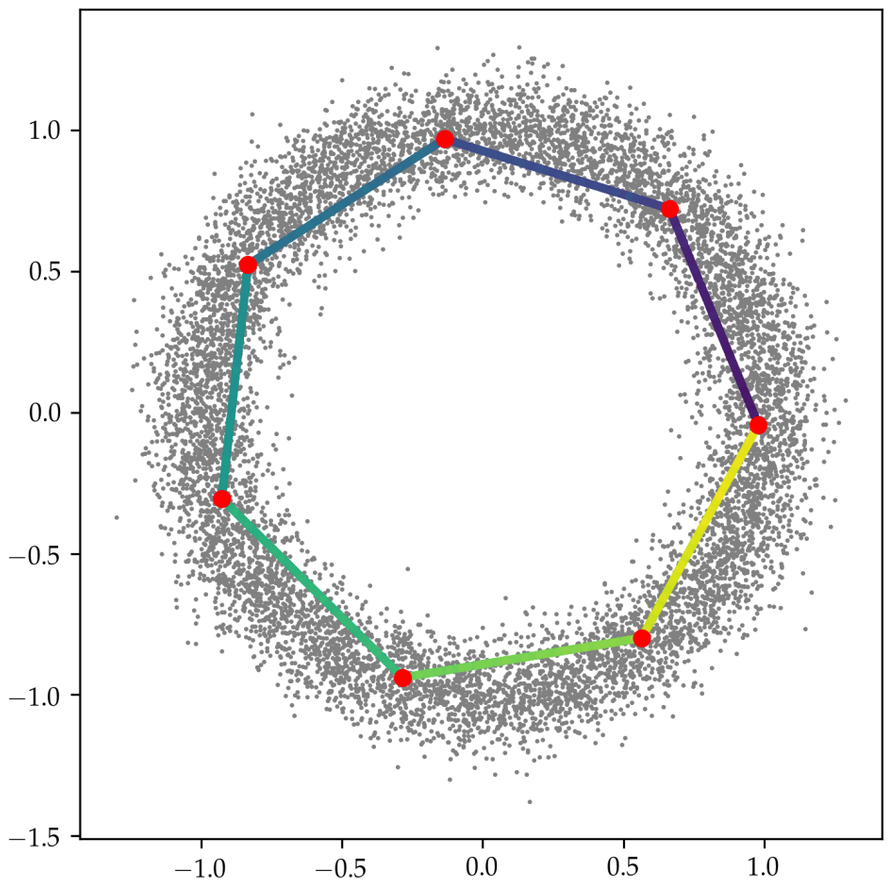
Dictionary Learning for Clustering

Given observations $\mathbf{Y} \in \mathbb{R}^{d \times n}$, find atoms $\mathbf{A} \in \mathbb{R}^{d \times m}$ and coefficieints $\mathbf{X} \in \mathbb{R}^{m \times n}$ such that

$$\|\mathbf{Y} - \mathbf{A}\mathbf{X}\|_2 + \mathcal{R}(\mathbf{A}, \mathbf{X}, \mathbf{Y})$$

is small for some regularizer $\mathcal{R}$.





What meaning do the coefficients have for clustering?

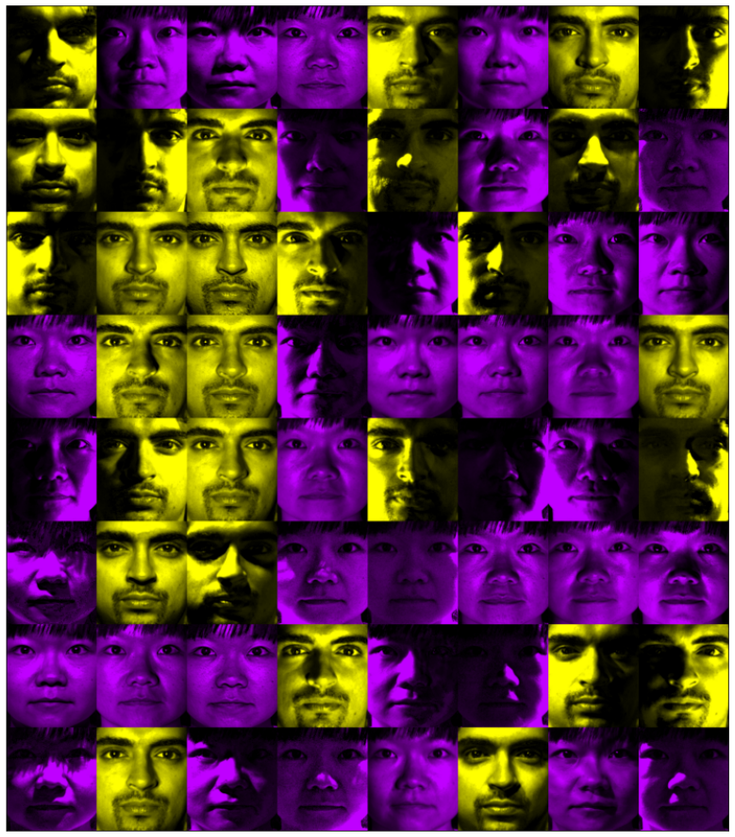
K-Deep Simplex (KDS)

$$\min_{\substack{\mathbf{A}\in\mathbb{R}^{d\times m}\\ \mathbf{X}\in\mathbb{R}^{m\times n}}} \sum_{i,j} x_{ij}\|\mathbf{y}_i - \mathbf{a}_j\|^2$$

$$\text{s.t.} \quad \mathbf{Y} = \mathbf{A}\mathbf{X},$$

$$\mathbf{X}^\top \mathbf{1} = \mathbf{1},$$

$$x_{ij} \geqslant 0, \quad \text{for all } i \text{ and } j.$$

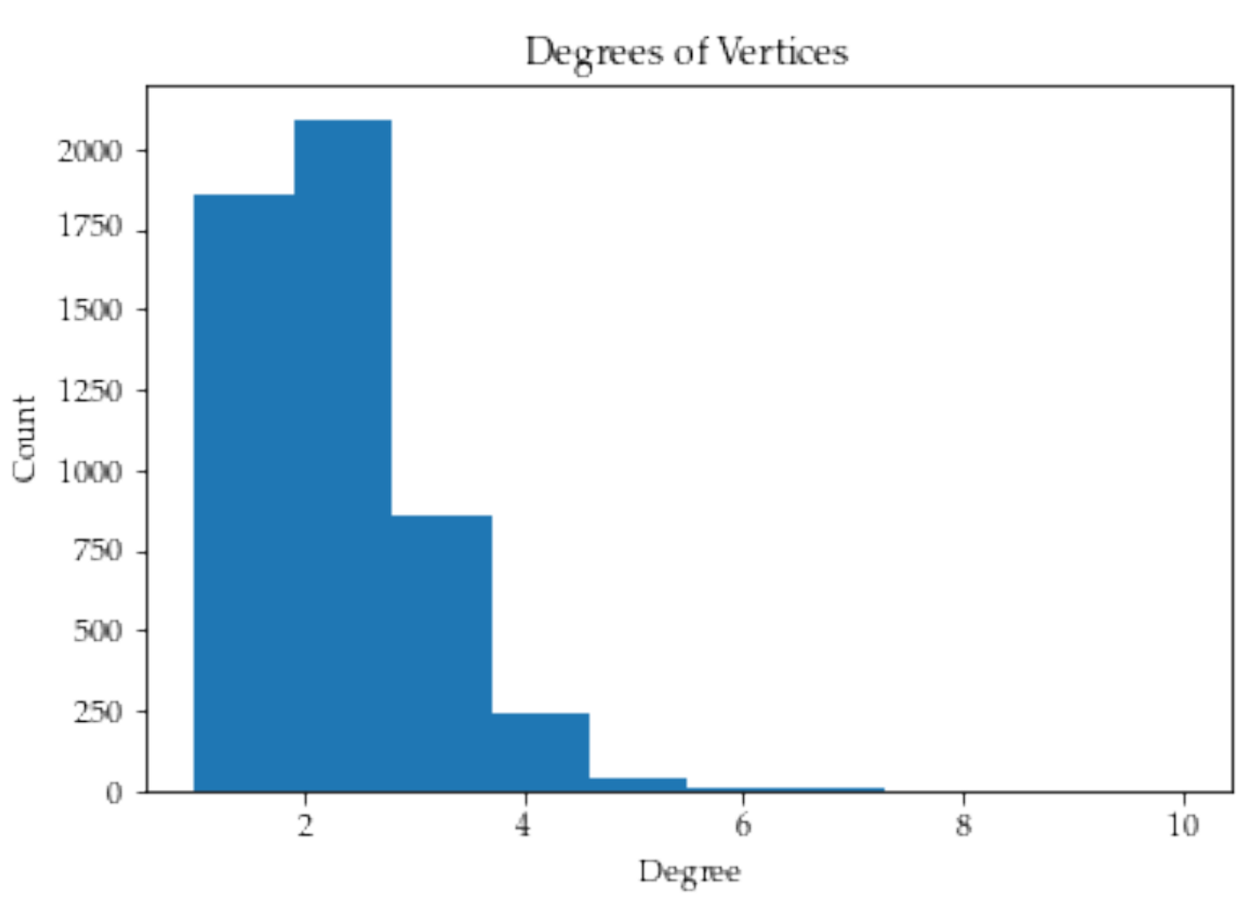


- The $\{x_{ij}\}$ are coefficients in the learned dictionary $\mathbf{A}$.
- The functional we minimize stitches the atoms together by enforcing *locality of reconstructions*.
- One can immediately use $\{x_{ij}\}$ as coefficients to cluster (e.g. as graph nodes in classical spectral clustering).

Geometric Sparsity?

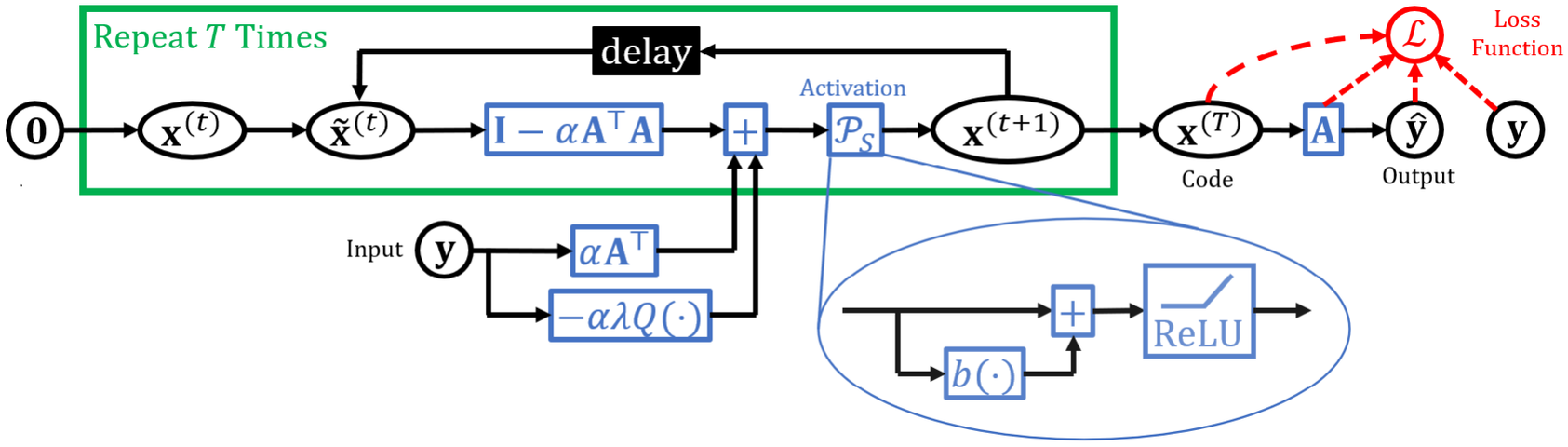
- One does not need a large number of atoms to represent well and sparsely in this framework (e.g. $m = 500, n = 35000$).
- Indeed, the optimization program generates (in highly idealized cases) Delauney triangulations.
- This suggests a good representations can be achieved with a number of atoms m depending only on data geometry.

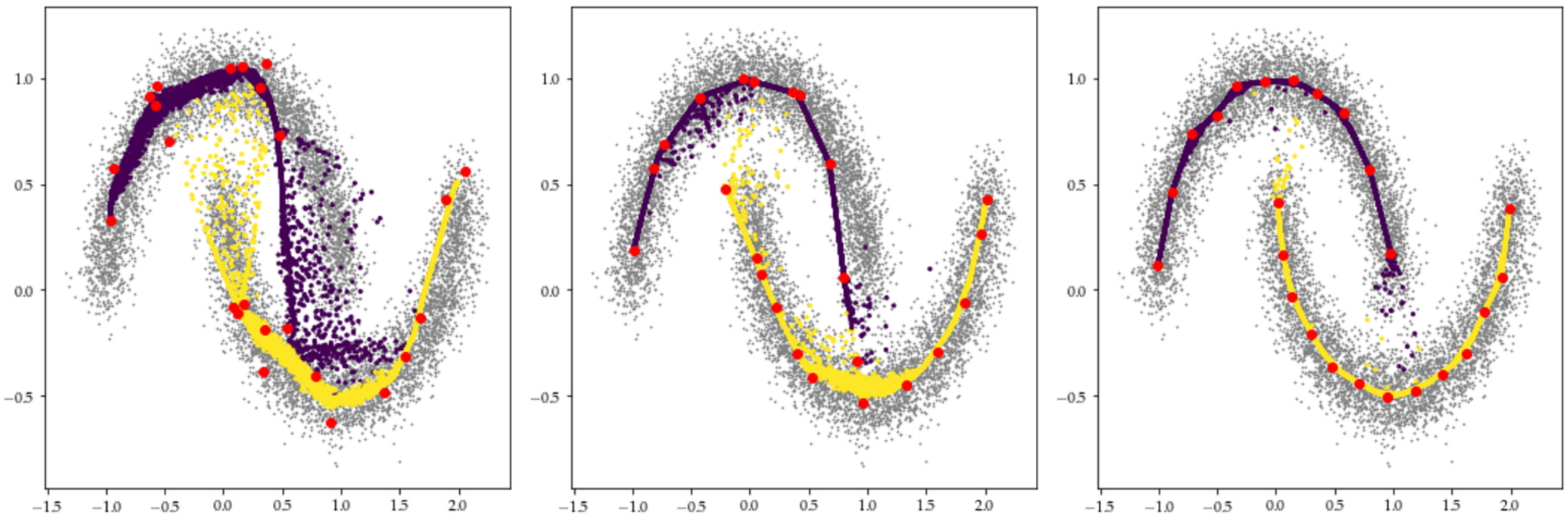
Left: Original MNIST digits.
Right: Reconstruction colored by learned label.



Deep Algorithm Unrolling

- Solving this via ADMM is possible but terribly slow.
- *Unrolling* with a structured deep network gives fast approximations that work extremely well in practice.





0 Epochs

250 Epochs

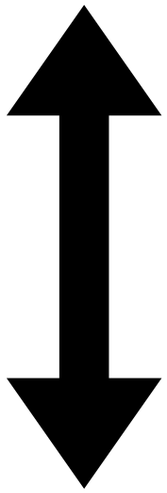
1000 Epochs

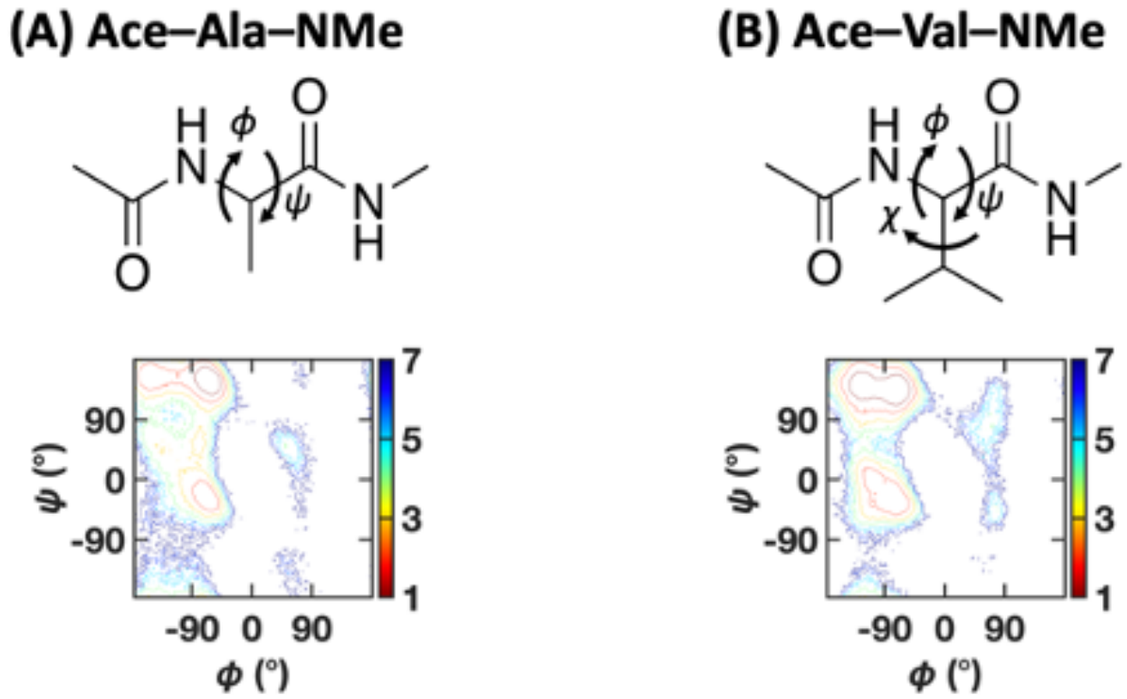
Tufts UNIVERSITY

Some Ongoing Work

- Intersecting manifolds: curvature-based graph constructions and flat geodesics.

- Molecular dynamics compression:





- Wasserstein clustering for data consisting of probability measures.

- Connections between clustering and segregation on geography networks—a quantitative framework for political science.

Collaborators


D. Ba, Harvard


L.Cowen, Tufts


K. Devkota, Tufts


X. Hu, Tufts


A. Little, Utah

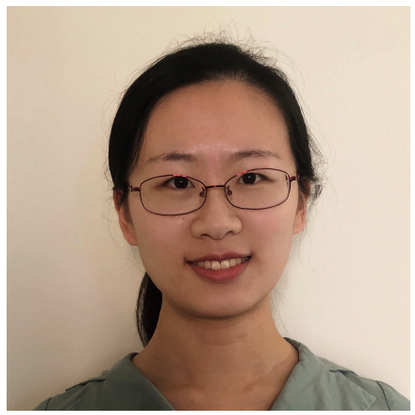

M. Maggioni, JHU


D. McKenzie, UCLA


S. Polk, Tufts


P. Tankala, Harvard


A. Tasissa, Tufts


K. Wu, Tufts

Tufts UNIVERSITY

Support

THE CAMILLE & HENRY DREYFUS FOUNDATION

DMS 1912737, DMS 1924513, CCF-1934553

Tufts
UNIVERSITY

References

- Cowen, Devokota, Hu, Murphy, and Wu. "Diffusion State Distances: Multitemporal Analysis, Fast Algorithms, and Applications to Biological Networks". *SIAM Journal on the Mathematics of Data Science*. 2021.

- Devkota, Murphy, and Cowen. "GLIDE: Combining Local Methods and Diffusion State Embeddings to Predict Missing Interactions in Biological Networks". *Bioinformatics*. 2020.

- Little, Maggioni, and Murphy. "Path-Based Spectral Clustering: Guarantees, Robustness to Outliers, and Fast Algorithms." *Journal of Machine Learning Research*. 2020.

- Little, McKenzie, and Murphy. "Balancing Geometry and Density: Path Distances on High-Dimensional Data." *SIAM Journal on the Mathematics of Data Science*. 2021+.

- Maggioni and Murphy. "Learning by Active Nonlinear Diffusion." *Foundations of Data Science*. 2019.

- Maggioni and Murphy. "Learning by Unsupervised Nonlinear Diffusion." *Journal of Machine Learning Research*. 2019.

- Murphy. "Spatially regularized active diffusion learning for high-dimensional images." *Pattern Recognition Letters*. 2020.

- Murphy and Maggioni. "Spectral-Spatial Diffusion Geometry for Hyperspectral Image Clustering." *IEEE Geoscience and Remote Sensing Letters*. 2020.

- Murphy, and Maggioni. "Unsupervised Clustering and Active Learning of Hyperspectral Images with Nonlinear Diffusion." *IEEE Transactions on Geoscience and Remote Sensing*. 2019.

- Murphy and Polk. "A Multiscale Environment for Learning by Diffusion." *arXiv*. 2021.

- Tankala, Tasissa, Murphy, and Ba. Manifold Learning and Deep Clustering with Local Dictionaries. *arXiv*. 2020.

Tufts UNIVERSITY

Code and Contact Information

Code: https://jmurphy.math.tufts.edu/Code/

Contact: jm.murphy@tufts.edu

Thanks for Your Attention!

Tufts UNIVERSITY