TDA and Dimensionality Reduction

Nick Sale - Swansea TDA Seminar - 19/05/20

- Rahul Paul and Stephan K. Chalup, **A study on validating non-linear dimensionality reduction using persistent homology**, 2017, https://www.sciencedirect.com/science/article/pii/S0167865517303513
- Michael Moor, Max Horn, Bastian Rieck and Karsten Borgwardt, **Topological Autoencoders**, 2019, https://arxiv.org/abs/1906.00722
- Lin Yan, Yaodong Zhao, Paul Rosen, Carlos Scheidegger and Bei Wang, **Homology-Preserving Dimensionality Reduction via Manifold Landmarking and Tearing**, 2018

TDA to validate low-dimensional embeddings

Idea: A good dimensionality reduction / manifold learning technique should preserve the topology of the data

In ‘A study on validating non-linear dimensionality reduction using persistent homology’, Rahul Paul and Stephan K. Chalup compare the Betti numbers of data with its embedding using Isomap and Locally-Linear Embedding (LLE)

Betti numbers are estimated using Vietoris-Rips persistent homology

Isomap

1. Find the k-NN graph of the data (k chosen large enough to be connected)
2. Compute the graph / geodesic distance between each pair of points
3. Apply MDS using these distances (low dimensional embedding that most closely maintains the distances between points)

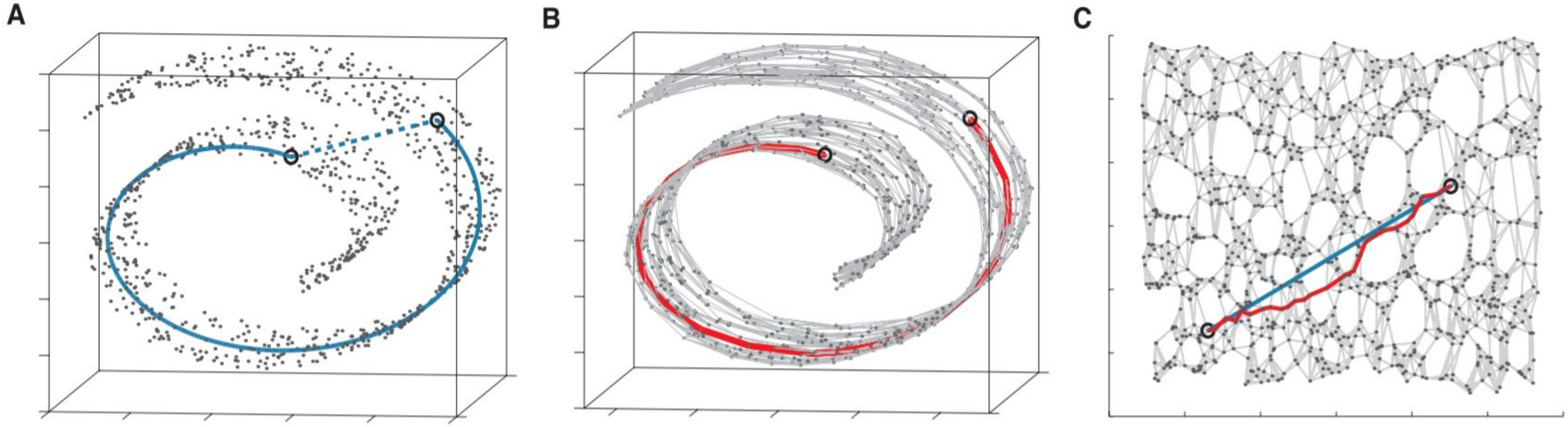


Image from: ‘A Global Geometric Framework for Nonlinear Dimensionality Reduction’ by Joshua B. Tenenbaum, Vin de Silva, John C. Langford

Locally-Linear Embedding (LLE)

1. Find the k-NN of each point

2. Express each point as a linear combination of its neighbours (i.e. minimise $\mathcal{E}(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2$)

3. Find the low dimensional embedding which best maintains these linear relations (i.e. minimise $\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2$)

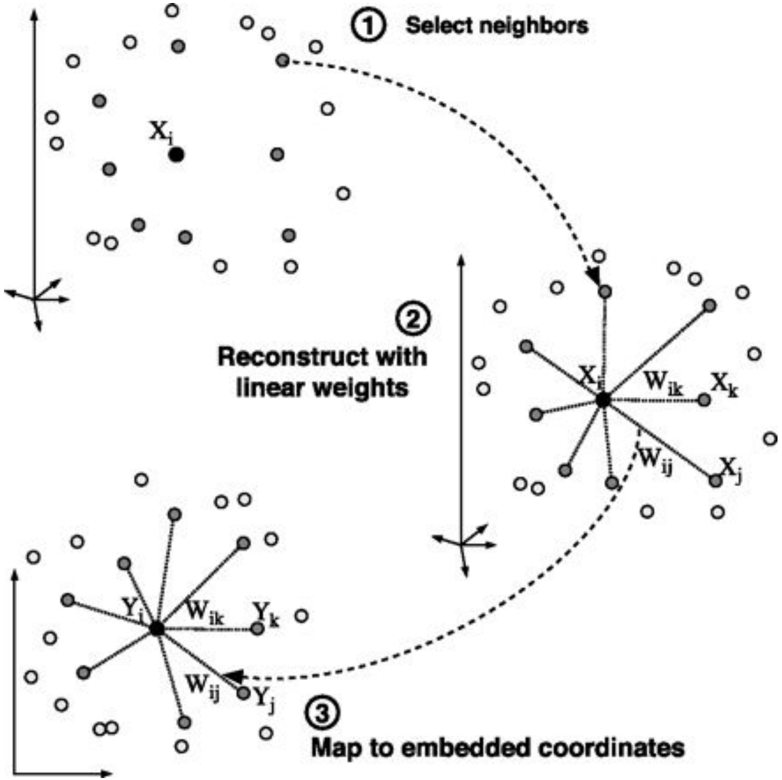
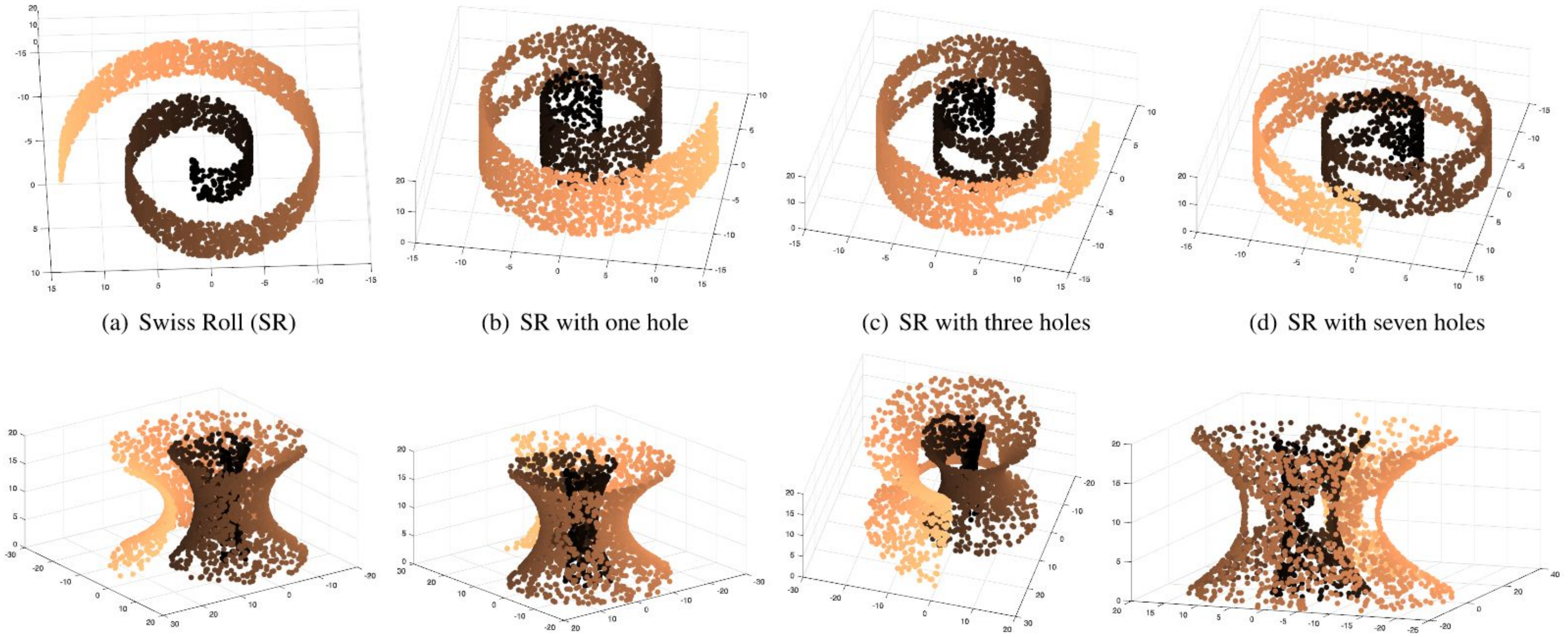


Image from: https://cs.nyu.edu/~roweis/lle/algorithm.html

The Data



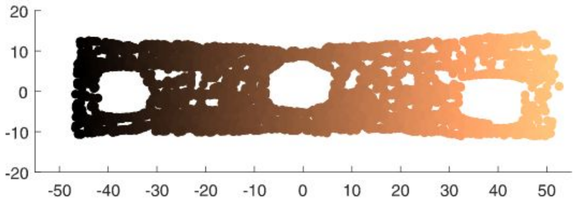
(a) Swiss Roll (SR)
(b) SR with one hole
(c) SR with three holes
(d) SR with seven holes
(e) Heated Swiss Roll (HR)
(f) HR with one hole
(g) HR with three holes
(h) HR with seven holes

Detecting when things go wrong

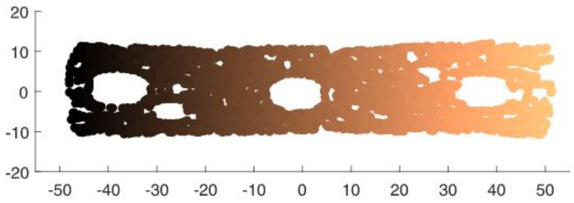
A standard measure to check an embedding

$$1 - R^2(D_M, D_L)$$

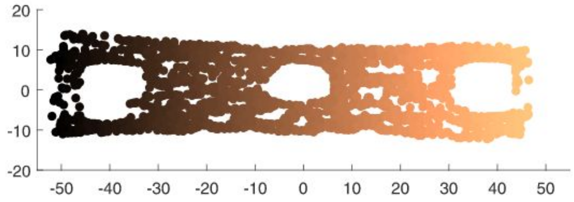
fails to notice the topological change but the Betti numbers estimated using PH do



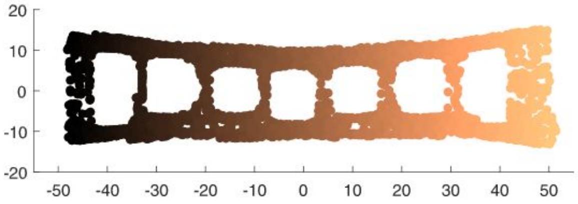
(a) Successful embedding



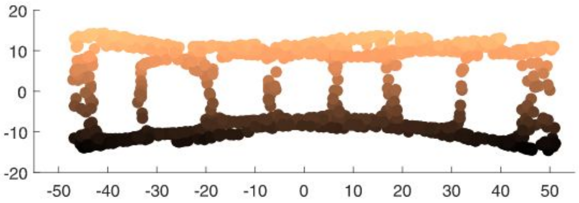
(b) A significant 4th hole occured on the left



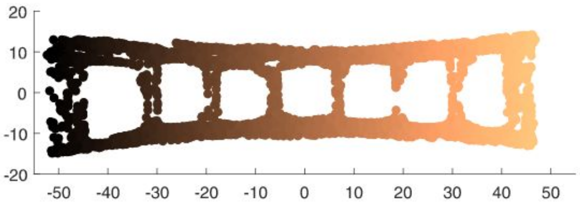
(c) The hole on the right was ripped open.



(a) Successful embedding of a 3100 point sample

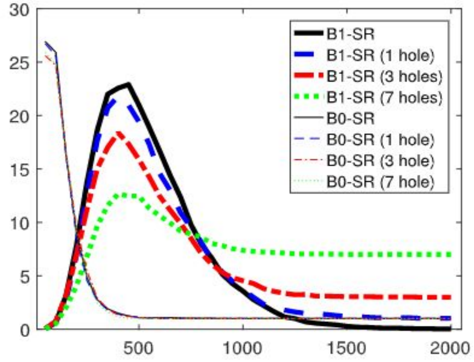


(b) Successful embedding of a 1000 point sample

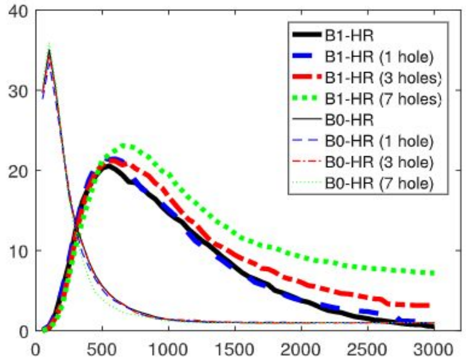


(c) A 3100 point sample where hole 5 and 6 merged

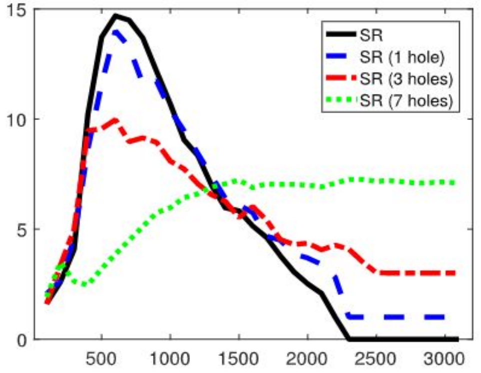
Finding a good enough sample size



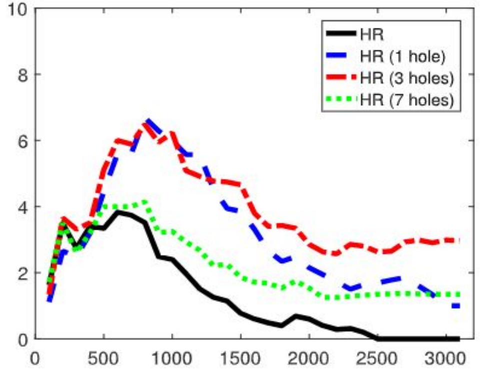
(a) Betti numbers of the Swiss Roll data in 3-dimensions in dependency of the number of sample points



(b) Betti numbers of the Heated Swiss Roll data in 3-dimensions in dependency of the number of sample points

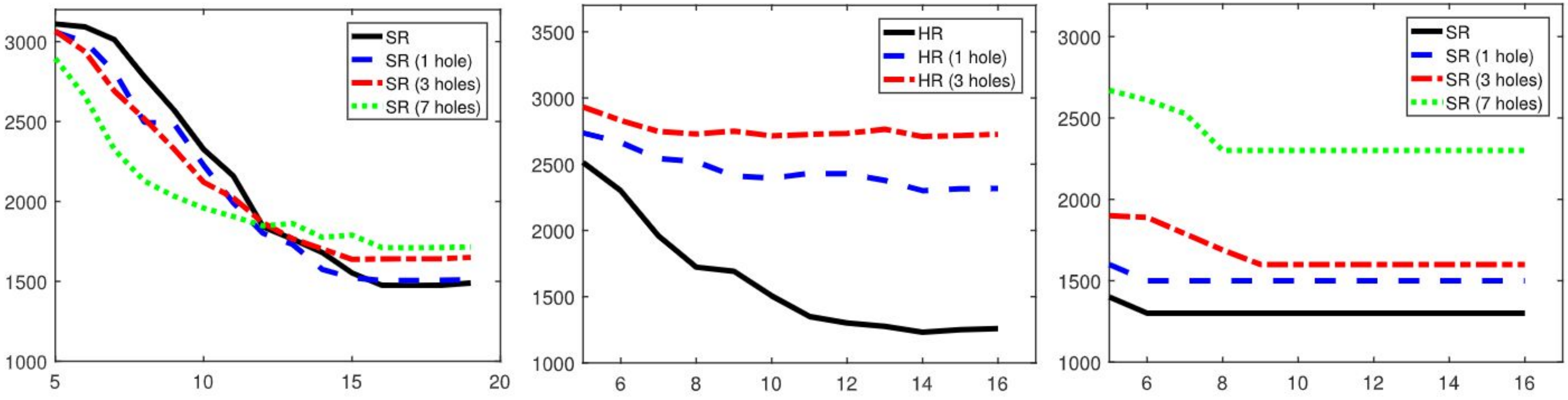


(a) B_1 for the embedded SR with different number of holes.



(b) B_1 for the embedded HR with different number of holes.

Choosing k



(a) Isomap (b) Isomap (c) LLE

Persistent homology to do dimensionality reduction

Idea: Make preservation of topological signature part of the objective for a low-dimensional representation

In ‘Topological Autoencoders’, Michael Moor, Max Horn, Bastian Rieck and Karsten Borgwardt develop a regularisation term to be used for autoencoders which compares the critical values of Vietoris-Rips filtrations on the original data and low-dimensional representation

Autoencoders

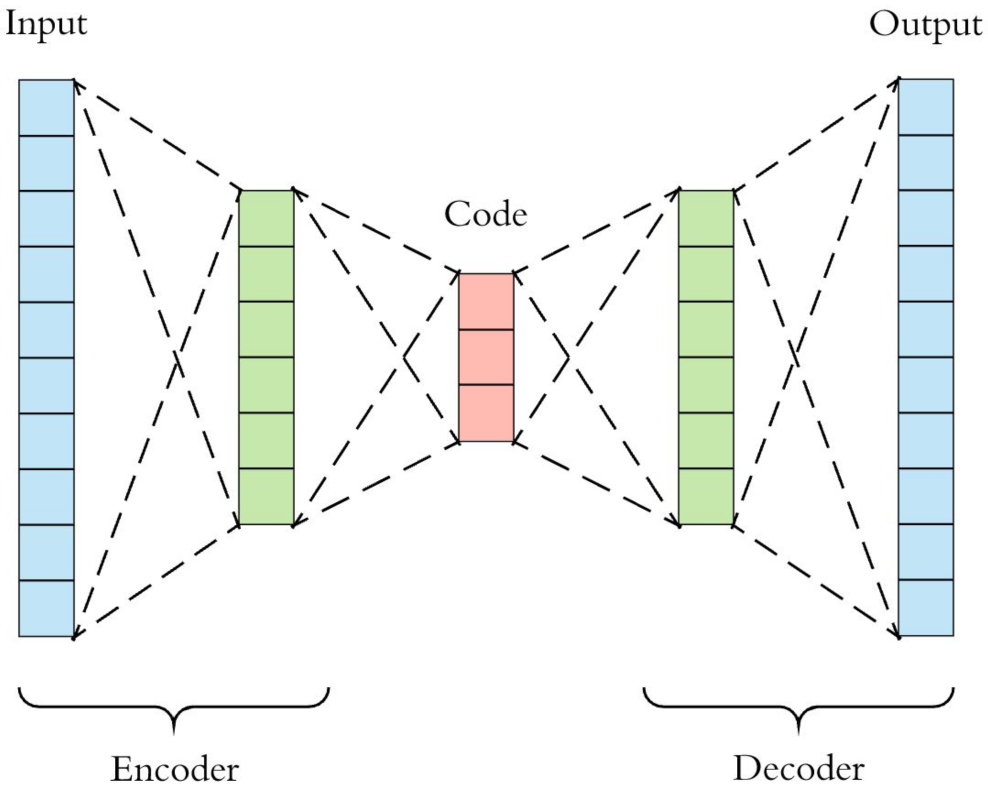


Image from: https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

Compare: *Connectivity-Optimized Representation Learning via Persistent Homology*, Hofer et al

Their approach

- Identify the critical edges from 0-dim persistence
- Try to preserve the lengths of these edges in the low-dim representation
- Since they use mini-batch to train their network, ensure that subsamples of the data are likely to give a similar set of critical values

$$\mathcal{L} := \mathcal{L}_r(X, h(g(X))) + \lambda\, \mathcal{L}_t$$

$$\mathcal{L}_{\mathcal{X}\to\mathcal{Z}} := \frac{1}{2}\left\|\mathbf{A}^X\left[\pi^X\right] - \mathbf{A}^Z\left[\pi^X\right]\right\|^2$$

$$\mathcal{L}_{\mathcal{Z}\to\mathcal{X}} := \frac{1}{2}\left\|\mathbf{A}^Z\left[\pi^Z\right] - \mathbf{A}^X\left[\pi^Z\right]\right\|^2$$

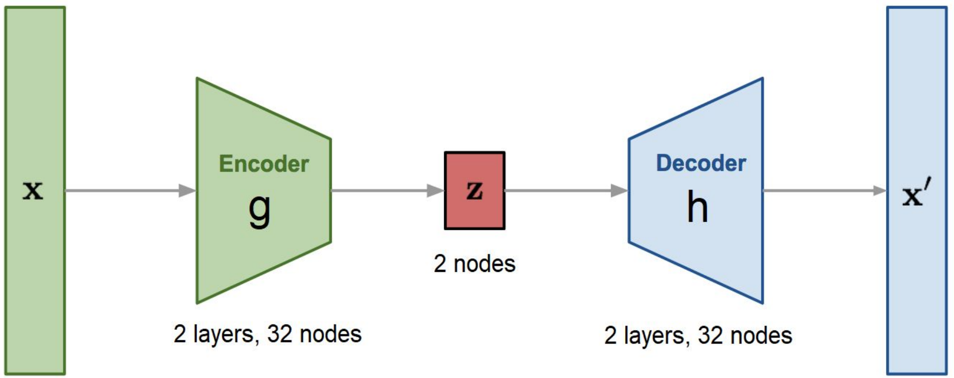
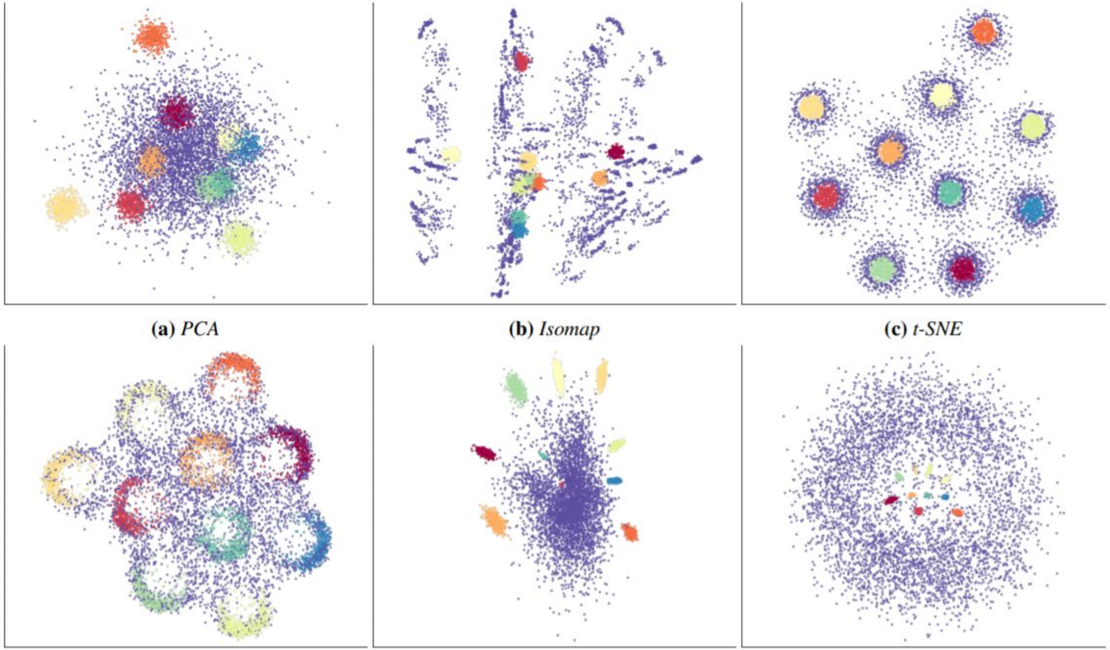


Image adapted from: https://lilianweng.github.io/lil-log/2018/08/12/from-autoencoder-to-beta-vae.html

SPHERES dataset

10 spheres surrounded by 1 larger sphere



(a) *PCA* (b) *Isomap* (c) *t-SNE*

(d) *UMAP* (e) *AE* (f) *TopoAE*

TDA to inform dimensionality reduction

Idea: Use the Reeb graph (or an approximation) to choose meaningful landmarks for landmark-based dimensionality reduction

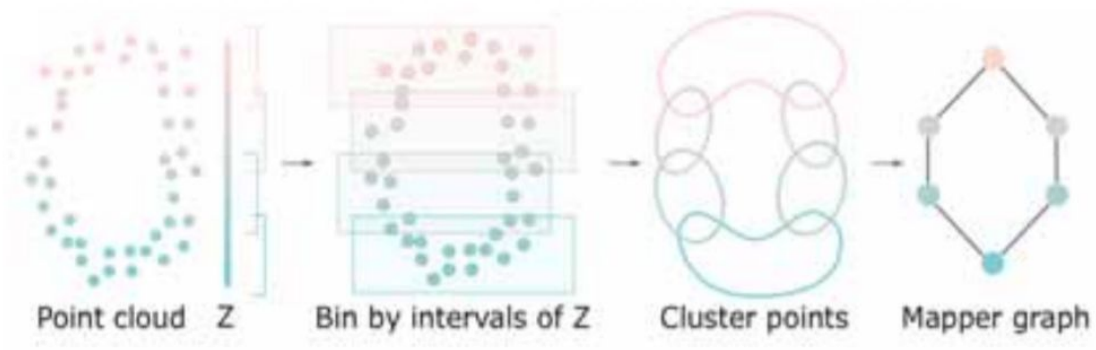
In 'Homology-Preserving Dimensionality Reduction via Manifold Landmarking and Tearing', Lin Yan, Yaodong Zhao, Paul Rosen, Carlos Scheidegger and Bei Wang use Mapper to obtain landmarks for use with Landmark Isomap (L-Isomap)

Landmark Isomap (L-Isomap)

1. Find the k-NN graph of the N data points
2. Somehow choose a set of n landmarks among the data points
3. Compute the N x n matrix of graph distances from each point to each of the landmark points
4. Embed just the landmarks based on their graph distances to one another (Eigendecomposition of n x n matrix)
5. Embed the rest of the points based on their graph distances to the landmarks
6. (Optional) Apply PCA to rescale according to the distribution of all the points rather than just the landmarks

Their approach

- Use Mapper (Distance-to-Basepoint filter) to approximate the Reeb graph
- Select the centroids of the clusters represented by the vertices as landmarks
- Apply L-Isomap



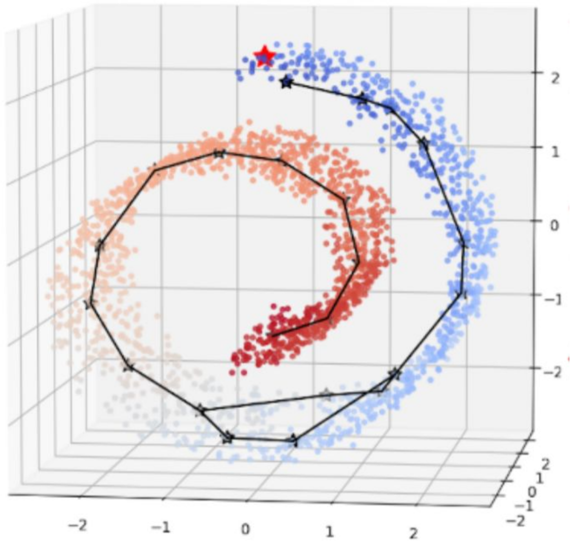
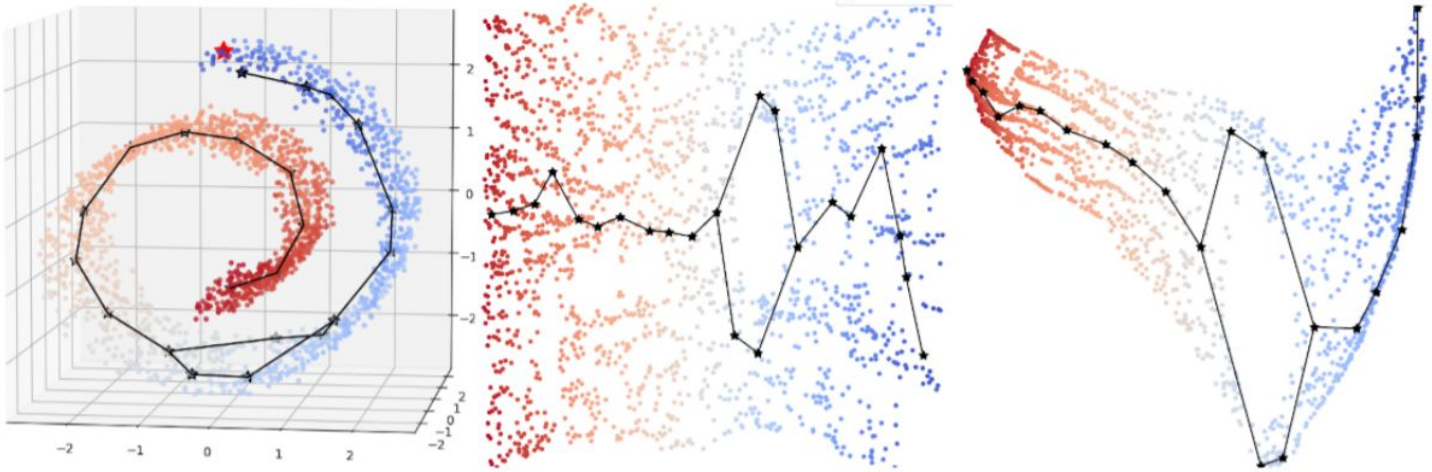


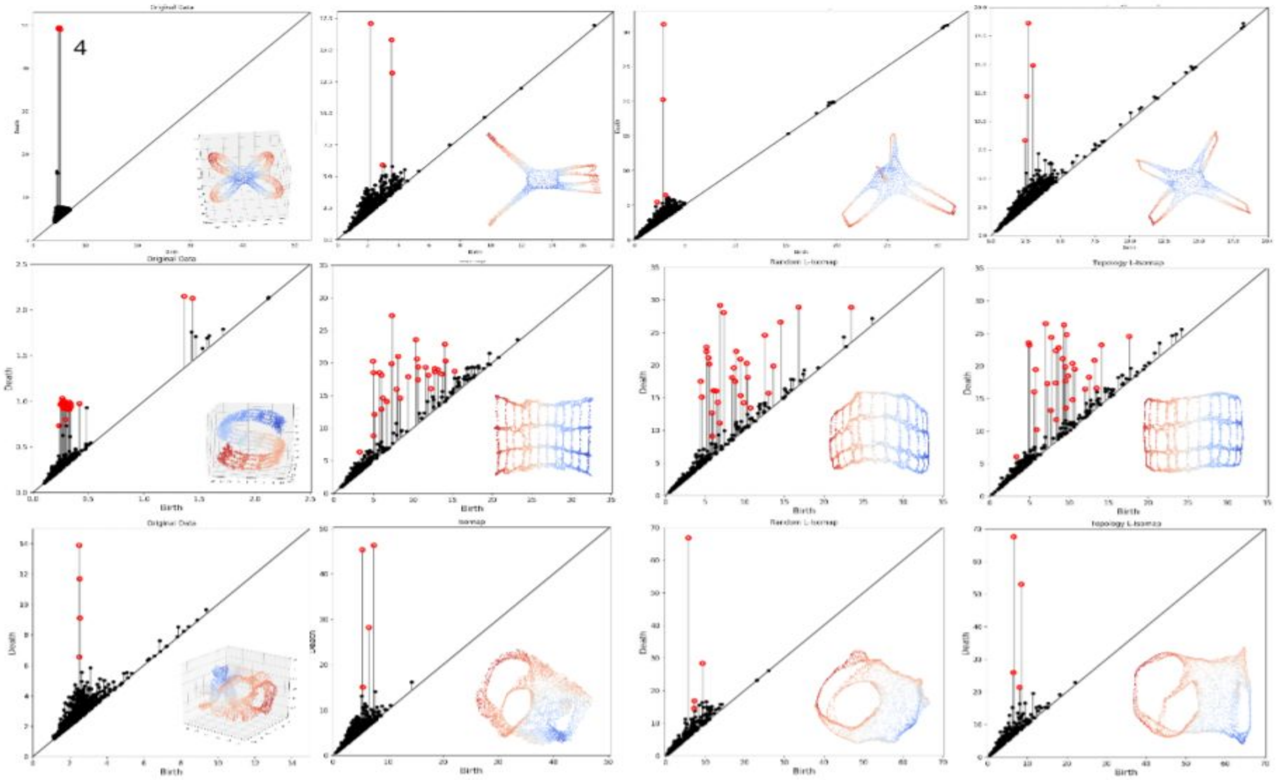
Image from: Modelling and Interpreting Mesoscale Network Dynamics, Khambhati et al

Example

- Swiss roll with hole
- Can obtain a decent embedding that captures the hole using only 21 landmarks out of 1983 points



Persistent homology to compare



Recap

Three perspectives:

- Persistent homology (persistent Betti numbers or Wasserstein distance) to validate how well data has been embedded
- Loss functions that incorporate 0-dim persistent homology to preserve cluster structure
- Mapper to identify small sets of landmarks which still capture topological features