Stratification Learning through Homology Inference

ICIAM 2011

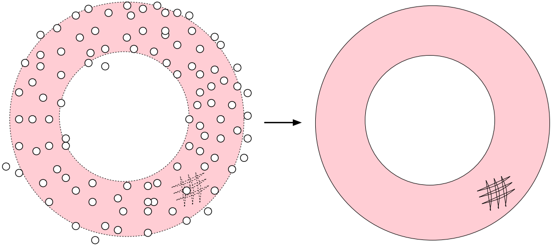
Sayan Mukherjee

Departments of Statistical Science, Computer Science, Mathematics
Institute for Genome Sciences & Policy, Duke University
Joint work with:
Bei Wang (University of Utah), Paul Bendich (Duke University)

July 18, 2011

Manifold learning

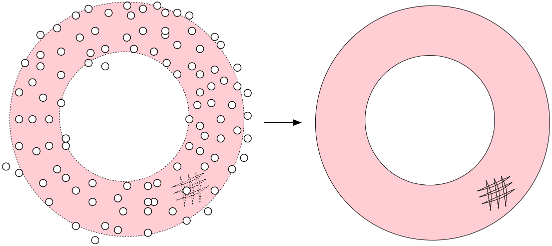
Manifold learning



1. Build better predictive models, dimension reduction.

Manifold learning

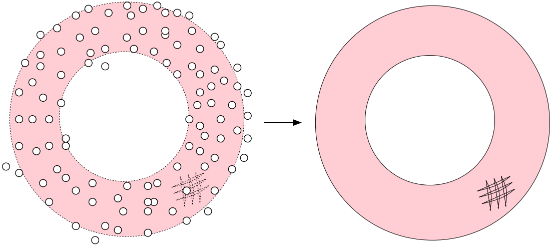
Manifold learning



1. Build better predictive models, dimension reduction.
2. Function estimation: fewer variables/more compact representation.

Manifold learning

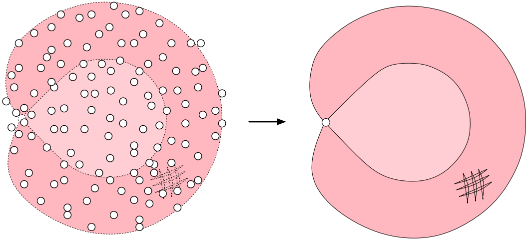
Manifold learning



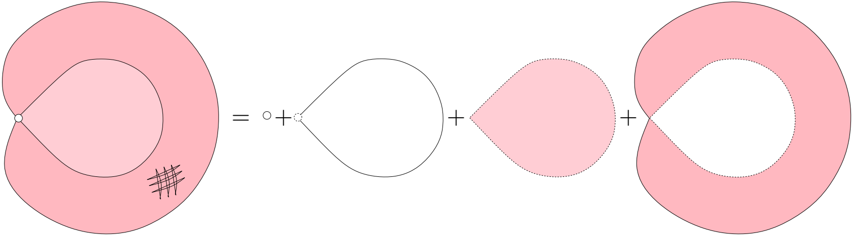
1. Build better predictive models, dimension reduction.
2. Function estimation: fewer variables/more compact representation.
3. Modeling parameter space: faster mixing in Markov chains.

Stratification learning

Stratification learning: singularities, mixed dimension



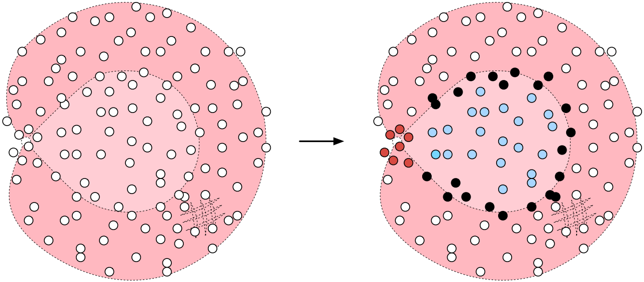
Stratification



1. Decompose into manifold pieces (strata).
2. Pieces fit “nicely” – Whitney conditions.

Stratification learning

Clustering: points whose local structure glue together nicely belong to the same cluster.



Sampling a stratified space

Remove the problems of singularities and varying dimension:

- M_1 : a mixture model. Lebesgue measure $\mu_i(\mathbb{S}_i)$ on the closure of each maximal strata, with corresponding density ν_i

$$f(x) = \sum_{i=1}^{K} \frac{1}{K} \nu_i(X = x).$$

Sampling a stratified space

Remove the problems of singularities and varying dimension:

- M_1 : a mixture model. Lebesgue measure $\mu_i(\mathbb{S}_i)$ on the closure of each maximal strata, with corresponding density ν_i

$$f(x) = \sum_{i=1}^{K} \frac{1}{K} \nu_i(X = x).$$

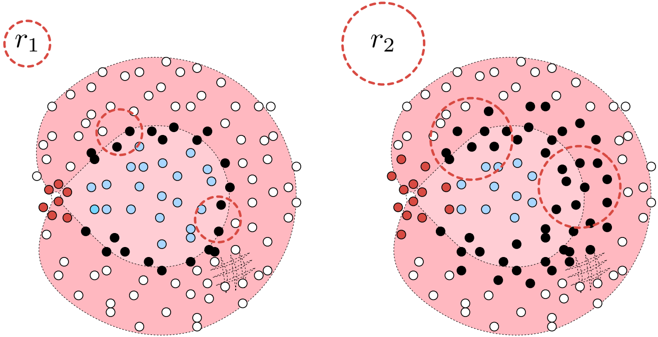
- M_2 : replace $\mathbb{X}$ by a slightly thickened version $\mathbb{X} \equiv \mathbb{X}_\delta$.
  Placing an appropriate measure on the highest dimensional strata to ensure that lower dimensional strata will be sampled from.

Informal learning statement

Given $U = \{x_1, x_2, ..., x_n\} \overset{iid}{\sim} f(x)$ for what n can we state with probability $1 - \delta$ that we correctly group points in the same strata together.

Stratification learning at multi-scale

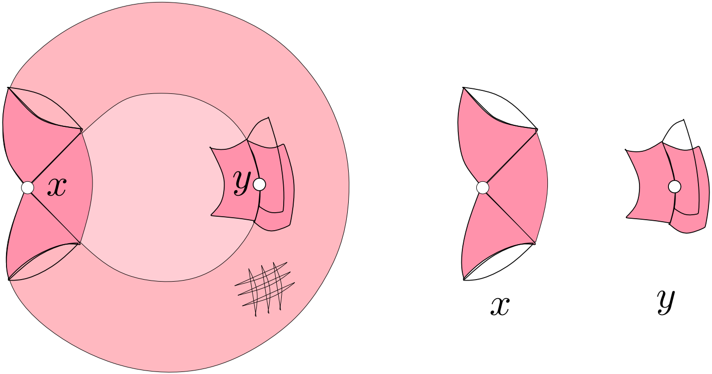
Our goal: clustering points, study multi-scale stratified structure.



Coming up next: a gentle introduction to local homology and persistence.

Local structure

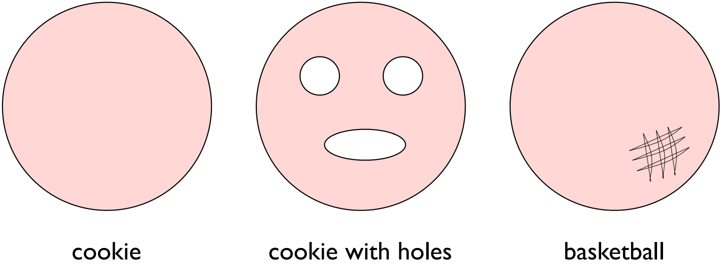
Points in the same strata have same local structure.



Local homology

Local homology is a tool to study local structure.

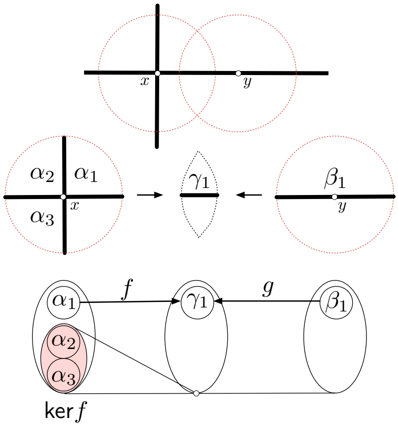
What is homology? Count "components" or "holes".



Local homology intersection map

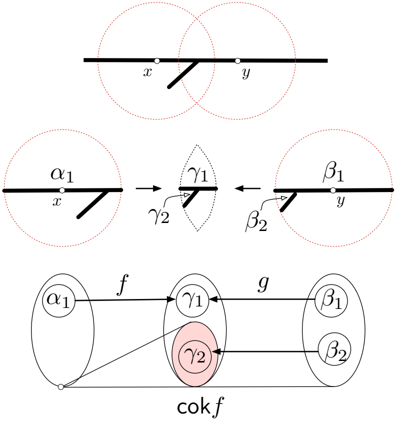
How are local structures of two nearby points "glued together"?

Map local structure to the neighborhood intersection.



kernel not empty $\equiv$ local structures disappear during mapping $\Rightarrow$ not the same local structure.

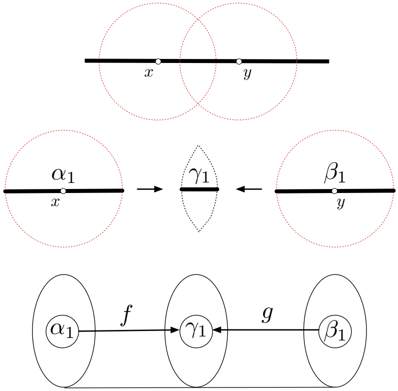
Local homology intersection map



Cokernel not empty $\equiv$ extra local structures exist in the intersection $\Rightarrow$ not the same local structure.

Local homology intersection map

Kernel/cokernel both empty ≡ local structures have one-to-one correspondance ⇒ same local structure.



Persistent homology philosophy

Persistent homology studies multi-scale features (“holes”) of spaces:

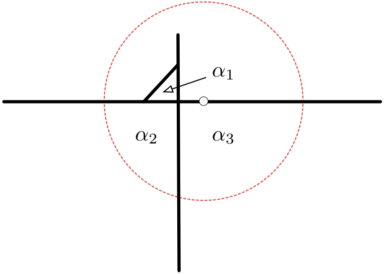
1. If the space is known, gives multi-scale representation of its features.
2. Given a point cloud sample, it describes features at different resolution. It separates features from noise.
3. Here, we explain the theories assuming ideal spaces, later on replacing the spaces with point cloud samples.

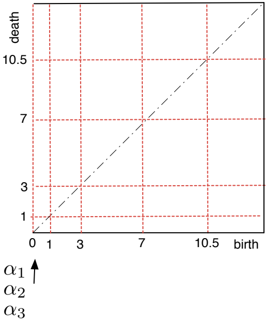
Persistent homology

A tool to study multiscale features ("holes") of space.

Some holes are larger (more persistent) than others.

We simulate the scale by "thickening" the space.



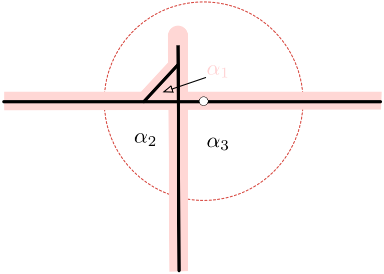


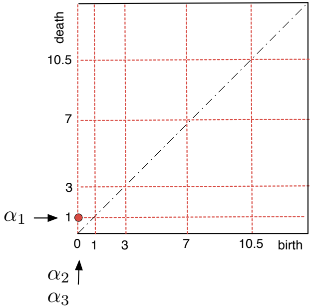
Persistent homology

A tool to study multiscale features ("holes") of space.

Some holes are larger (more persistent) than others.

We simulate the scale by "thickening" the space.



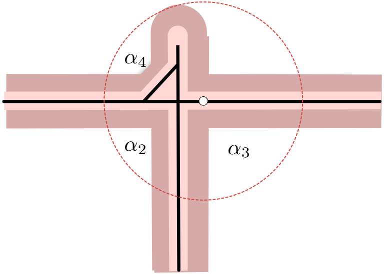


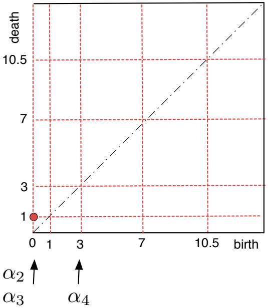
Persistent homology

A tool to study multiscale features ("holes") of space.

Some holes are larger (more persistent) than others.

We simulate the scale by "thickening" the space.



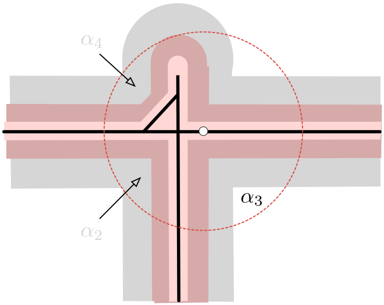


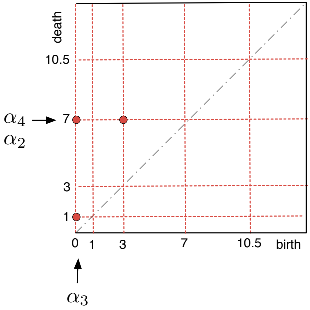
Persistent homology

A tool to study multiscale features ("holes") of space.

Some holes are larger (more persistent) than others.

We simulate the scale by "thickening" the space.



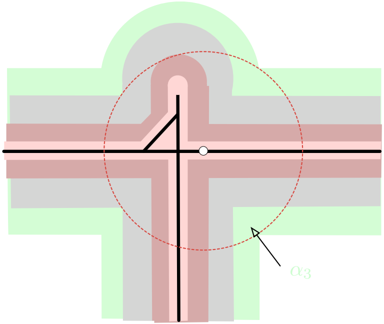


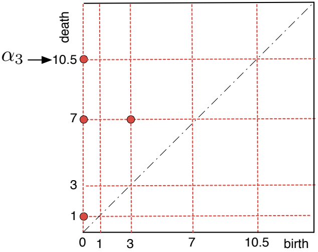
Persistent homology

A tool to study multiscale features ("holes") of space.

Some holes are larger (more persistent) than others.

We simulate the scale by "thickening" the space.



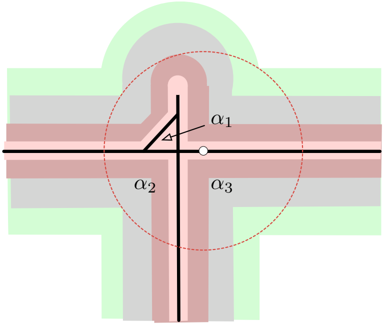


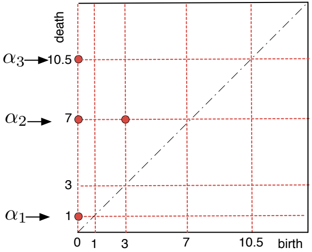
Persistent homology

A tool to study multiscale features ("holes") of space.

Some holes are larger (more persistent) than others.

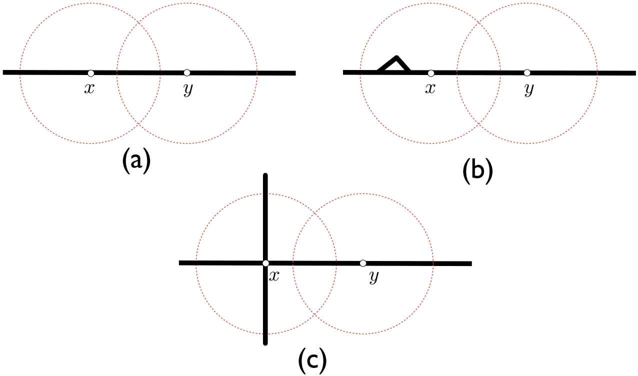
We simulate the scale by "thickening" the space.



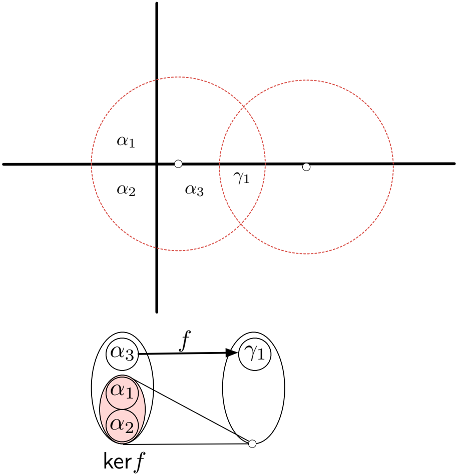


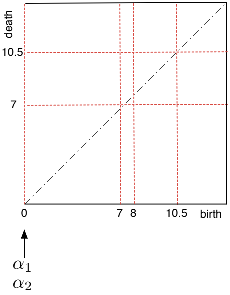
Kernel persistent homology

Study extra local structure in the kernel with high persistence.

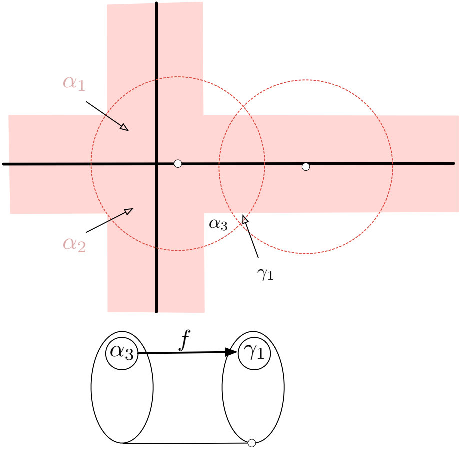


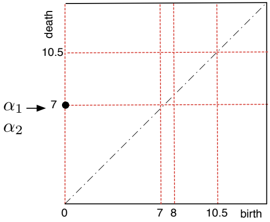
Kernel persistent homology: example



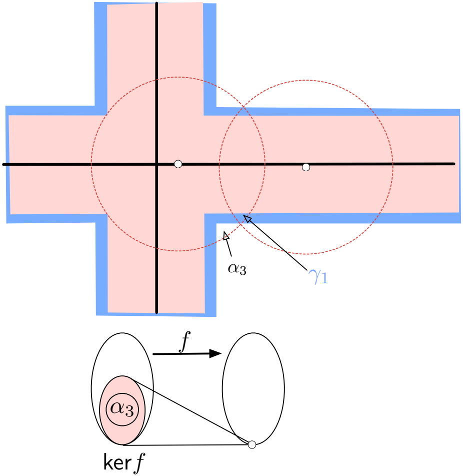


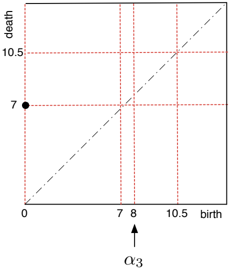
Kernel persistent homology: example



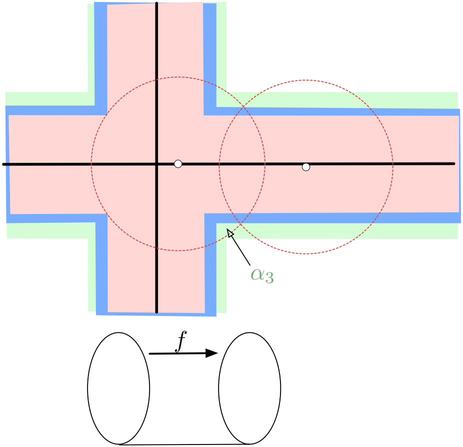


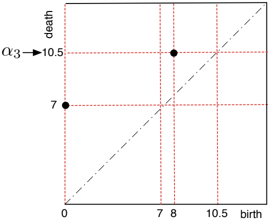
Kernel persistent homology: example



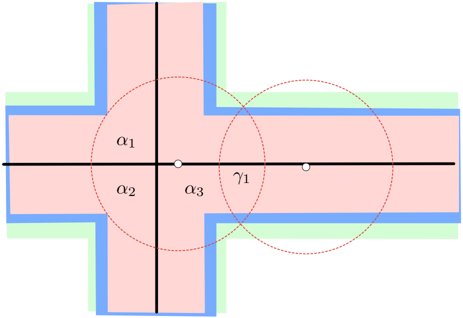


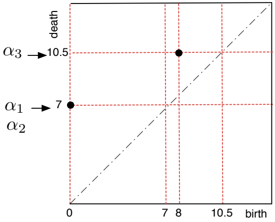
Kernel persistent homology: example



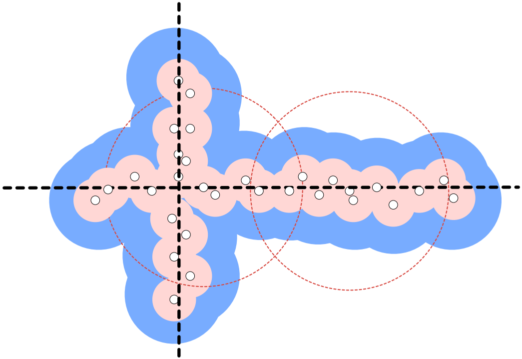


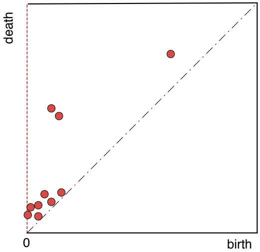
Kernel persistent homology: example



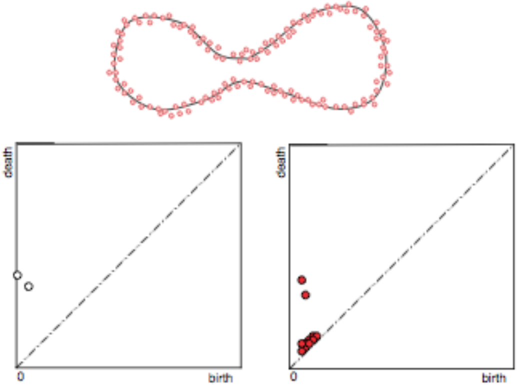


Kernel persistent homology for point cloud



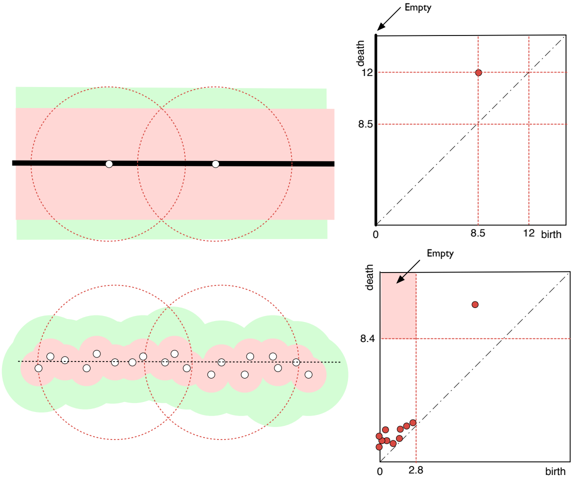


Persistence diagram stability

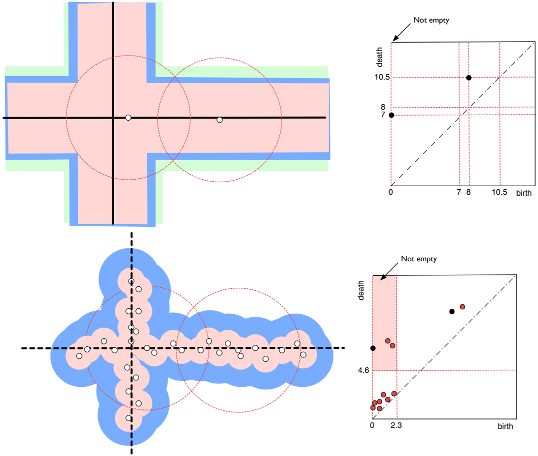


If two spaces are ϵ close (Haussdorff) the diagrams are ϵ close (Wasserstein).

Kernel persistence diagram stability



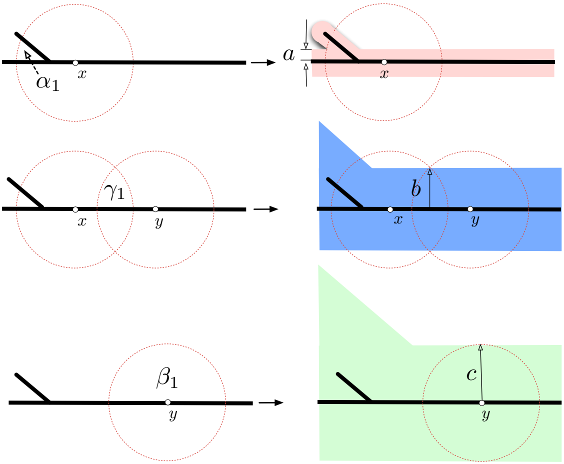
Kernel persistence diagram stability



What ϵ ?

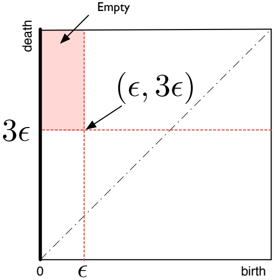
Minimum feature size

Minimum feature size (mfs) : minimum non-zero thickening parameter where local structures changes.



Local topological homology inference

Given ϵ-approximation where $\epsilon < \text{mfs}/4$, if (co)kernel persistence diagrams contain no points in $[0, \epsilon] \times [3\epsilon, \infty]$ then x, y are locally equivalent, $x \sim_r y$.



Local homology inference theorem

Theorem (Local homology theorem)

Given an ϵ-sample U from $\mathbb{X}$. For a pair of points $p, q \in \mathbb{R}^d$ with mfs$(p, q, r) \geq 4\epsilon$, $p \sim_r q$ iff

$$\mathrm{Dgm}(\ker \phi^U_{p,q})[\epsilon, 3\epsilon] \cup \mathrm{Dgm}(\mathrm{cok}\, \phi^U_{p,q})[\epsilon, 3\epsilon] = \emptyset.$$

Strata inference theorem

Theorem (Strata clustering theorem)

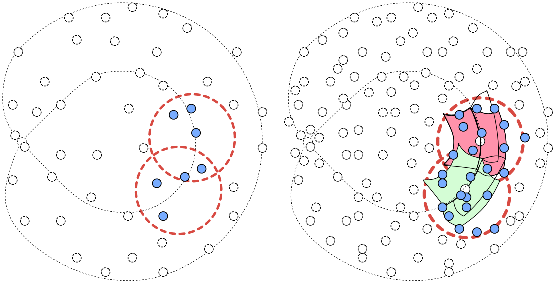
Given an ϵ-sample U from $\mathbb{X}$ with mfs$(p, q, r) \geq 4\epsilon$ $\forall p, q \in U$, each cluster C_i is the transitive closure of $p, q \in U$ with $p \sim_r q$.

Points in each C_i belong to the same stratum (at resolution r).

Probabilistic local homology inference

$$U = \{x_1, x_2, ..., x_n\} \overset{iid}{\sim} f(x).$$

For $n > n_0$ with prob $\geq 1 - \xi$ we can infer local homology where $n_0(\xi, r, \text{mfs}, \text{vol}\,(\mathbb{X}))$.



If we do not sample enough points, locally the homology inference fails.

Probabilistic local homology inference

Theorem (Probabilistic local homology theorem)

Let $U = \{x_1, x_2, ..., x_n\} \overset{iid}{\sim} f(x)$ *For a pair of points* $p, q \in U$ *with* $\rho =$ *mfs*(p, q, r) *and*

$$v(\rho) = \inf_{x \in \mathbb{X}} \frac{\mathrm{vol}\left(B_{\rho/24}(x) \cap \mathbb{X}\right)}{\mathrm{vol}\left(\mathbb{X}\right)}.$$

If

$$n \geq n_0 = \frac{1}{v(\rho)}\left(\log \frac{1}{v(\rho)} + \log \frac{1}{\xi}\right),$$

then $p \sim_r q$ *with prob* $\geq 1 - \xi$.

Probabilistic homology inference

Theorem (Probabilistic homology theorem)

Let $U = \{x_1, x_2, ..., x_n\} \overset{iid}{\sim} f(x)$, *set* $\rho_{\min} = \min_{p,q \in U}$ *mfs*(p, q, r) *and*

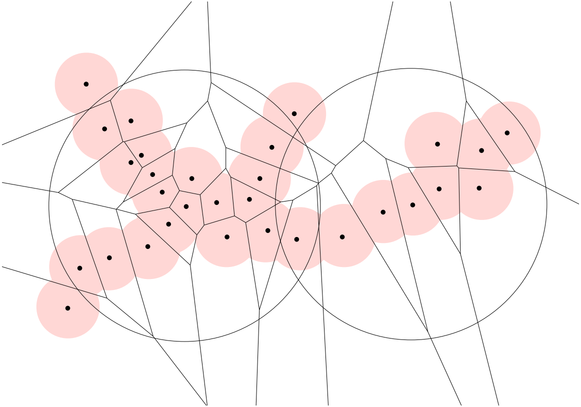
$$v(\rho_{\min}) = \inf_{x \in \mathbb{X}} \frac{\mathrm{vol}\left(B_{\rho_{\min}/24}(x) \cap \mathbb{X}\right)}{\mathrm{vol}\left(\mathbb{X}\right)}.$$

Each cluster C_i *is the transitive closure of* $p, q \in U$ *with* $p \sim_r q$.

Points in each C_i *belong to the same stratum (at resolution* r*).*

Compute simplicial complexes

Compute local structure through simplicial complexes.



Compute simplicial complexes

Compute local structure through simplicial complexes.

Graph embedding

- Weight matrix: $W(p,q) = h(\mathrm{Dgm}(\ker \phi^U_{p,q}), \mathrm{Dgm}(\mathrm{cok}\, \phi^U_{p,q}))$.

Graph embedding

- Weight matrix: $W(p,q) = h(\mathrm{Dgm}(\ker \phi^U_{p,q}), \mathrm{Dgm}(\mathrm{cok}\, \phi^U_{p,q}))$.
- $D(p,p) = \sum_q W(p,q)$

Graph embedding

- Weight matrix: $W(p,q) = h(\mathrm{Dgm}(\ker \phi^U_{p,q}), \mathrm{Dgm}(\mathrm{cok}\, \phi^U_{p,q}))$.
- $D(p,p) = \sum_q W(p,q)$
- Graph Laplacian: $L = D - W$

Graph embedding

- Weight matrix: $W(p,q) = h(\mathrm{Dgm}(\ker \phi^U_{p,q}), \mathrm{Dgm}(\mathrm{cok}\, \phi^U_{p,q}))$.
- $D(p,p) = \sum_q W(p,q)$
- Graph Laplacian: $L = D - W$
- Eigen-decomposition: $Lv = \lambda v$

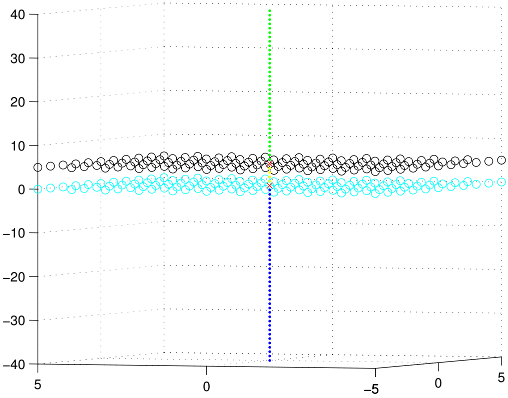
Graph embedding

- Weight matrix: $W(p,q) = h(\mathrm{Dgm}(\ker \phi^U_{p,q}), \mathrm{Dgm}(\mathrm{cok}\, \phi^U_{p,q}))$.
- $D(p,p) = \sum_q W(p,q)$
- Graph Laplacian: $L = D - W$
- Eigen-decomposition: $Lv = \lambda v$
- Embed: $\Phi(p) : p \to (v_1(p), ..., v_m(p)), \quad \forall p \in U$

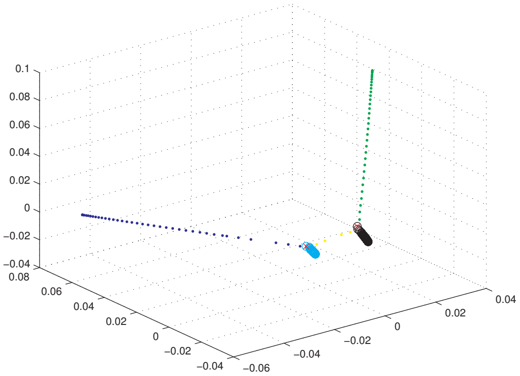
Graph embedding

- Weight matrix: $W(p,q) = h(\mathrm{Dgm}(\ker \phi^U_{p,q}), \mathrm{Dgm}(\mathrm{cok}\, \phi^U_{p,q}))$.
- $D(p,p) = \sum_q W(p,q)$
- Graph Laplacian: $L = D - W$
- Eigen-decomposition: $Lv = \lambda v$
- Embed: $\Phi(p) : p \to (v_1(p), ..., v_m(p)), \quad \forall p \in U$
- Cluster: n points in $\mathbb{R}^{m-1}$.

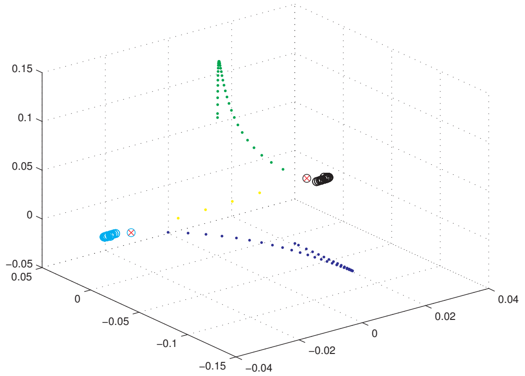
Data



Distance based weight matrix 3D embedding



Ker/Cok weight matrix 3D embedding



Open problems

- Faster algorithms in practice: Rips/Witness complexes, dimension reduction, random projection.

Open problems

- Faster algorithms in practice: Rips/Witness complexes, dimension reduction, random projection.
- Scaling with dimension.

Open problems

- Faster algorithms in practice: Rips/Witness complexes, dimension reduction, random projection.
- Scaling with dimension.
- Robustness of clustering, combinatorial Laplacian.

Open problems

- Faster algorithms in practice: Rips/Witness complexes, dimension reduction, random projection.
- Scaling with dimension.
- Robustness of clustering, combinatorial Laplacian.
- Fractional weights between pairs of points, probabilistic inference.

Open problems

- Faster algorithms in practice: Rips/Witness complexes, dimension reduction, random projection.
- Scaling with dimension.
- Robustness of clustering, combinatorial Laplacian.
- Fractional weights between pairs of points, probabilistic inference.
- Estimation of dimension of strata.

Paper for talk

Towards Stratification Learning through Homology Inference
http://ftp.stat.duke.edu/WorkingPapers/10-18.html

Acknowledgments

Herbert Edelsbrunner, John Harer, Shmuel Weinberger, Partha Niyogi, Mikael Vejdemo-Johansson, David Cohen-Steiner, Dmitriy Morozov, Yuriy Mileyko, Chao Chen, Brittany Terese Fasy, Michael Kerber

Acknowledgments

Herbert Edelsbrunner, John Harer, Shmuel Weinberger, Partha Niyogi, Mikael Vejdemo-Johansson, David Cohen-Steiner, Dmitriy Morozov, Yuriy Mileyko, Chao Chen, Brittany Terese Fasy, Michael Kerber

Funding:

- Center for Systems Biology at Duke
- NSF DMS and CCF
- AFOSR
- NIH