TDA and Mainstream Topology

Geometric Realization of AATRN

IMSI, Chicago

Gunnar Carlsson, August 18, 2025

Congratulations to AATRN for its Achievements and Contributions!

Agenda

- Data science over a base (Brad Nelson)
- High dimensional α-complexes (E. and J. Carlsson)
- Hom-complexes and optimization (Speculation)
- Foundations of TDA (Speculation)

Data Science over a Base

Brad Nelson

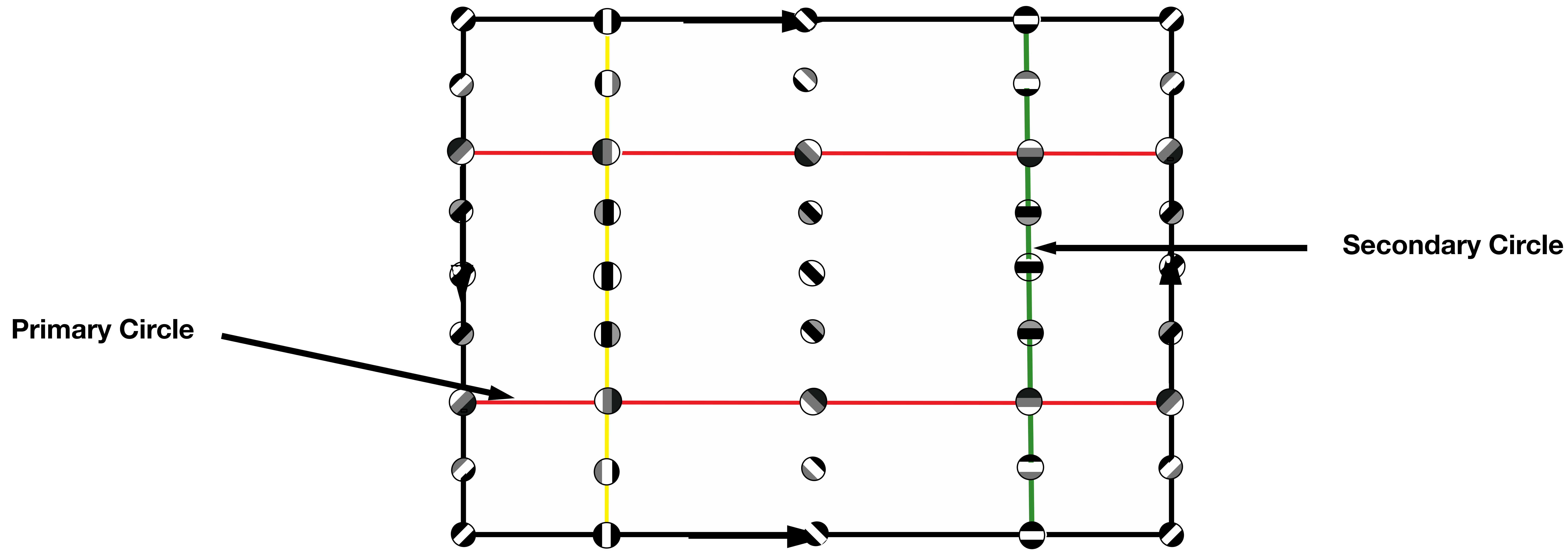
Topology Over a Base

- Homotopy theory over a reference base space
- Objects are spaces X with a reference map $f : X \to B$ to a fixed reference space
- Morphisms and homotopies required to respect the reference map
- Gives much additional structure: $H^*(X)$ is a $H^*(B)$-module, and maps respect that structure
- Many more low dimensional invariants.

Data Analysis Over a Base

- Early demonstration of value of persistent homology was analysis of Mumford image patch data set
- Ultimate finding was that frequently occurring image patches concentrated around a Klein bottle.
- Obtaining the finding was painful.
- Finding can be made much more direct with “data science over a base”.
- In particular, varying notions of density over fibers of map are important.

Image Patches Parametrized by Klein bottle

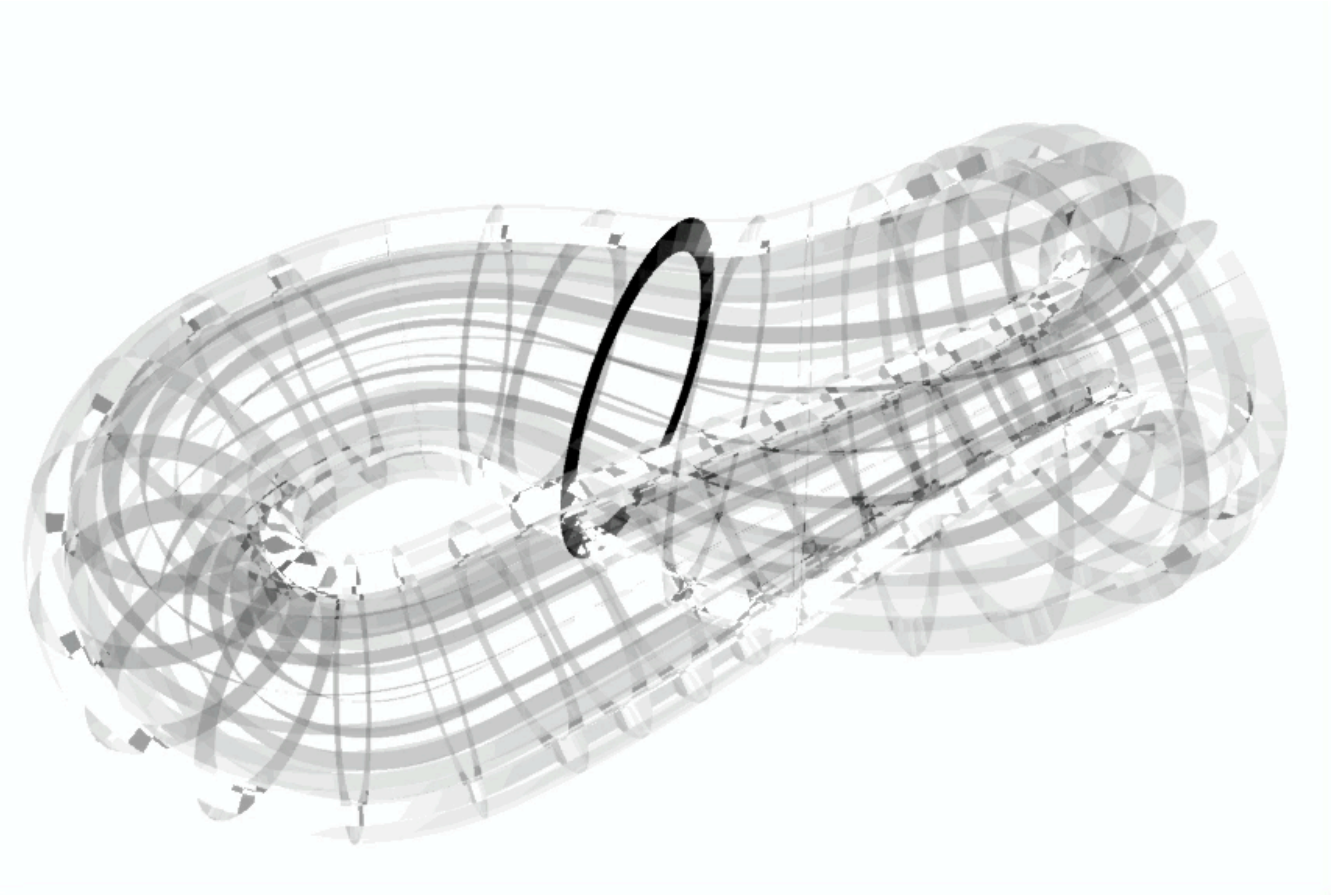


Density much greater in the vertical and horizontal angles

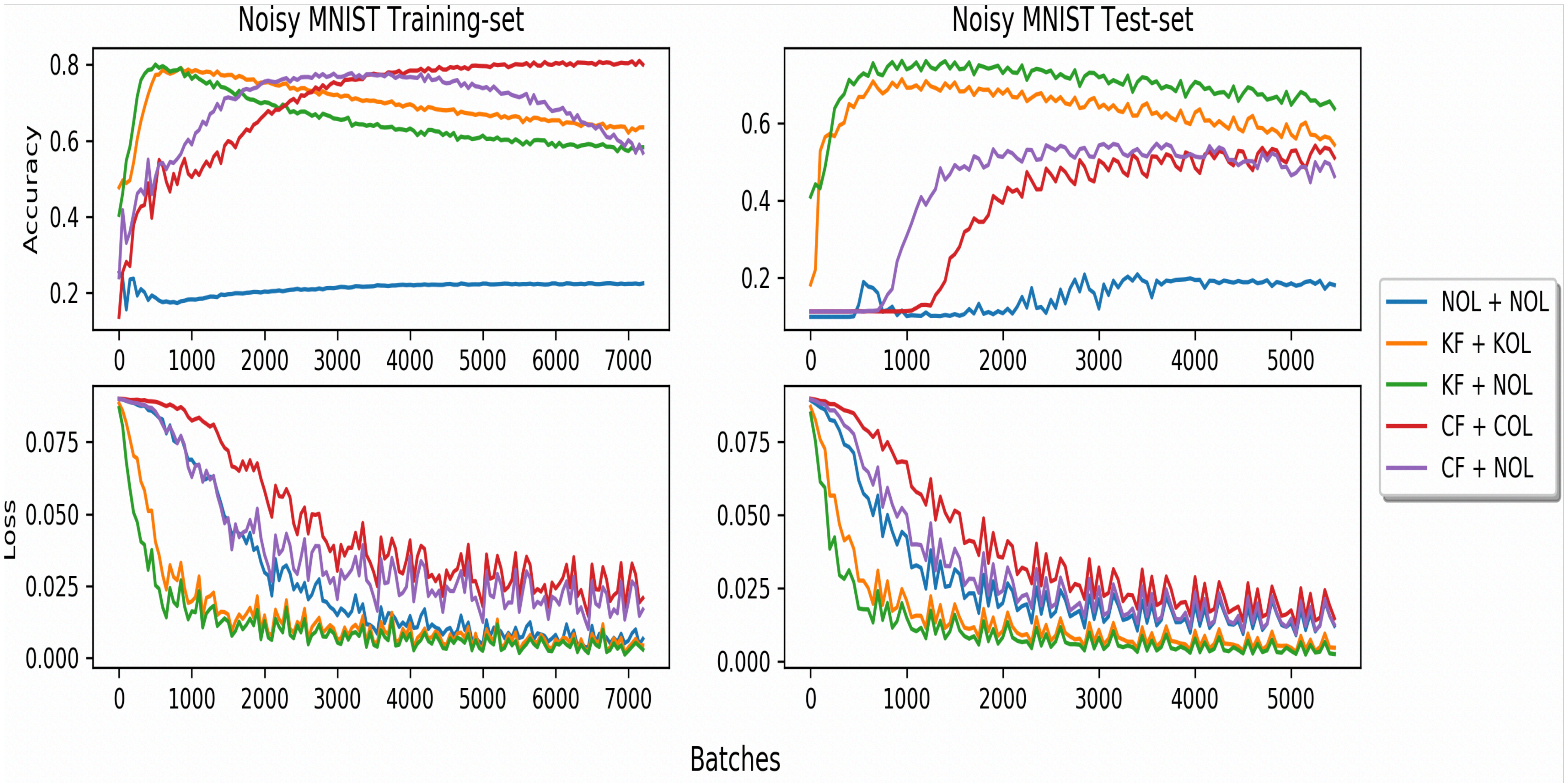
Data Analysis Over a Base

- First analysis (primary circle) showed that the "directionality" of a patch is important.
- View this as a new feature on the data set, i.e. a map from the data set to the circle.
- We have a data set "over the base circle S^1"
- This means compute the homology of the fibers over various points - they turn out to be secondary circles themselves
- Key point: density is estimated fiberwise, so in each fiber F_b select points based on a percentage density threshold *in* F_b

Summary: The data set is concentrated on a circle bundle, with varying density on the fibers



Improving CNNs



E. Love, B. Filippenko, V. Maroulas, G.Carlsson, *Topological Convolutional Layers for Deep Learning,* JMLR, 24(59)10ths -35, 2023

Summary

- After “fibrewise density filtration” has been applied, the correct computation comes out directly.
- Because of tools for “local to global” analysis, one reduces to one dimensional homology calculations.
- The correct structure is as a space with filtration over a base with a filtration.

High Dimensional α-Complexes

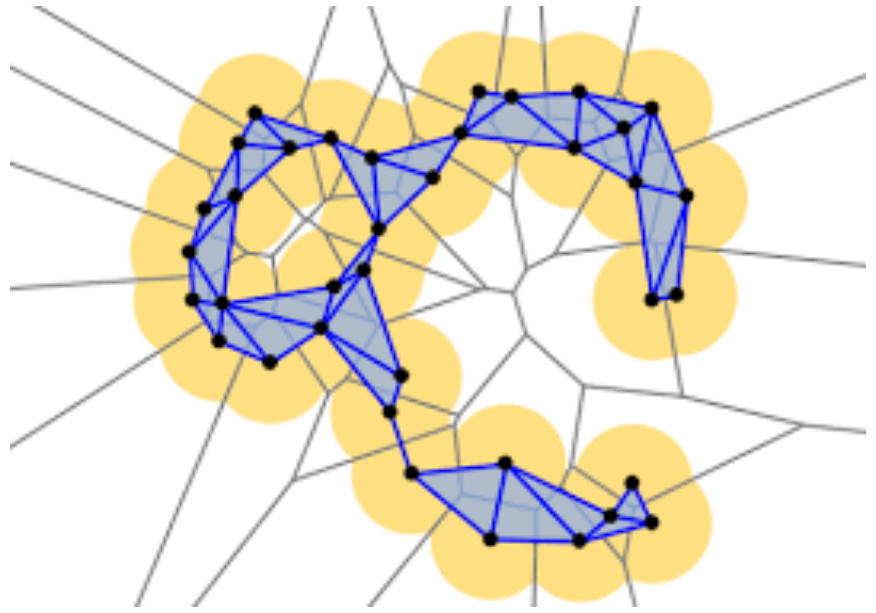
Work of E. and J. Carlsson on α-Complexes

- Much smaller than Čech or Rips constructions
- Produces coverings but is agnostic of the shape of the sets in the covering, is not restricted to balls as the Čech construction is
- Some sets in the coverings may be much larger than others (good thing)
- Direct constructions are computationally expensive and are impractical in even moderately high dimensions, e.g. 10
- E. and J. Carlsson circumvent these problems in

Carlsson E, Carlsson J. Computing the alpha complex using dual active set quadratic programming. Scientific Reports. 2024 Aug 27;14(1):19824.

α -Complexes

- We denote the Voronoi cells for a subset $S = \{x_0, \ldots, x_n\}$ of $\mathbb{R}^n$ by V_{x_i}
- For a given radius r and point x, the *radius restricted Voronoi cell* $V_x(r)$ is defined to be $V_x \cap B_r(x)$
- The nerve of the covering of $\bigcup_{x \in X} V_r(x)$ by the sets $V_r(x)$ is called the *r-restricted α-complex*



α-Complexes

- How to determine if $\sigma = \{x_0, \ldots, x_l\} \subseteq S$ spans a simplex in the r-restricted α -complex?
- Amounts to solving the constrained quadratic optimization problem

$$\underset{y \in \mathbb{R}^m}{\text{Minimize}} \ |y - x|^2$$

$$\text{Subject to } |y - x_i| \leq |y - x_j| \qquad x_i \in \sigma,\ x_j \in S - \sigma$$

$$|y - x_i| = |y - x_j| \qquad x_i, x_j \in \sigma,\ i \neq j$$

Where $x = x_i$ for some (any) i

and finding that the minimizer is less than r^2. Constraints can be made linear by squaring and cancelling quadratic terms.

Dual Linear Programming

- Typical linear programming problem is

$$\text{minimize } c^T x \text{ subject to } Ax \le b, x \ge 0$$

- Dual problem is

$$\text{maximize } b^T y \text{ subject to } A^T y = c, y \ge 0$$

- Solutions are identical. Note that if the primal has m variables and n constraints, the dual problem has n variables and m constraints. This also applies to quadratic programming.

Duality for α-complex optimization problem

- For a data set with N points and n variables, the α-complex optimization problem (for one simplex) has n variables and $\sim N$ constraints.
- The dual problem has $\sim N$ variables and n constraints
- For this problem, the number of constraints is the main driver of the computational complexity
- Using the dual problem permits calculations in higher dimensions than has been accomplished before.

Example

- Constructed a data set of 20232 points for C_4, a subspace of the configuration space $C_4(\mathbb{R}^2)$ of 4-tuples of distinct points in $\mathbb{R}^2$ related to the space of “hard discs”. Subset is embedded in $\mathbb{R}^8$.
- Ripser was able to compute $\beta_0, \beta_1,$ and β_2 in 40 minutes, was unable to compute β_3.
- The α-complex was able to compute $\beta_0, \beta_1, \beta_2,$ and β_3 in 4 minutes.

Hom-Complexes and Feature Creation

Non-Linear Feature Creation

- Suppose we are asked to create features with values in a manifold rather than in a linear space.
- Vejdemo-Johansson and de Silva have defined and worked with circular coordinates.
- Jose Perea has done a lot of great work for projective spaces and other non-linear manifolds
- We'll propose a way of looking at the feature generation problem for general manifolds.
- Likely expensive and complex.

Minimal Cycles

- One thread in the use of homology, within and outside mathematics, is the construction of minimal cycles under some appropriate appropriate notion of minimality
- Harmonic forms in de Rham theory context
- Cycles with minimal number of non-zero coefficients in the case of simplicial complexes
- Minimal cycles expected to be more interpretable

Hom Complex

- Given two chain complexes C_* and D_*, there is a complex $Hom(C_*, D_*)$ defined as $C^* \otimes D_*$, where the grading on C^* is negative
- $H_0(Hom(C_*, D_*))$ is the group of chain homotopy classes of chain maps from C_* to D_*
- If $C_* = C_*(X)$ and $D_* = C_*(Y)$, we have the maps induced by actual simplicial maps $X \longrightarrow Y$
- Can one do anything with the Hom-complex to get at the actual maps, and construct maps?
- Is there a notion of minimal cycles in the Hom-complex?

Minimal Cycles in the Hom Complex

- Here is a typical element in the degree zero part of the Hom complex. $c^*_{i_k}$ denotes the dual basis element to c_{i_k}, and c_{i_k} and d_{j_k} have the same degree. Suppose we are dealing with $\mathbb{F}_2$ coefficients, and consider a typical element .

$$\xi = \sum_k a_{i_k j_k} c^*_{i_k} \otimes d_{j_k}$$

- For each simplex c of the complex X, let $\lambda(c)$ denote the number of simplices d in Y for which the coefficient of $c^* \otimes d$ is non-zero.
- Assuming there is a metric on the sets of simplices of Y, let $\lambda_\mu(c)$ denote the diameter of the set of simplices d for which the coefficient of $c^* \otimes d$ in ξ is non-zero.
- The element ξ is assigned a value $\theta = \sum_{c \in S_k(X)} \lambda(c)$ or $\theta_\mu = \sum_{c \in S_k(X)} \lambda_\mu(c)$

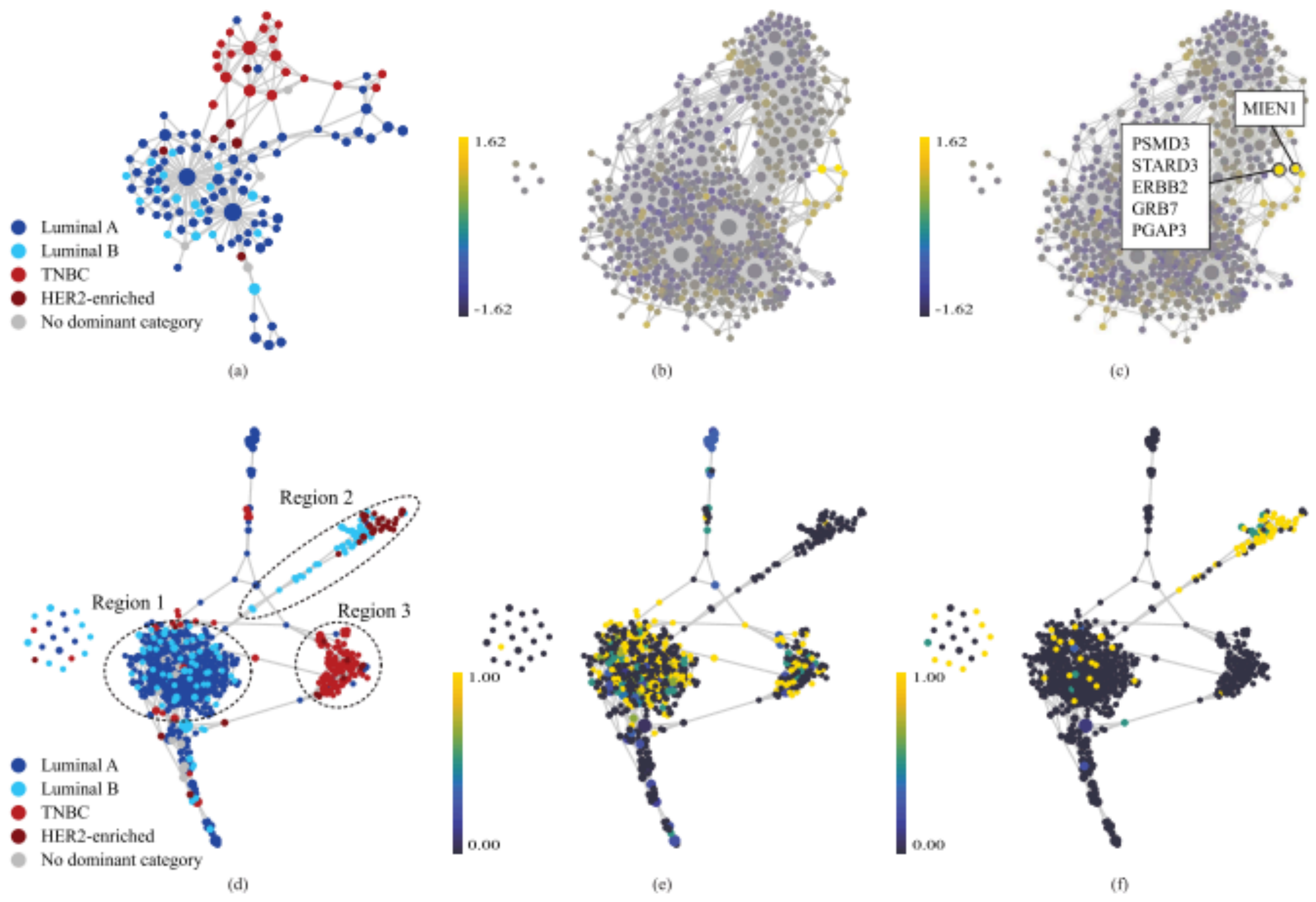
Minimal Cycles in the Hom Complex

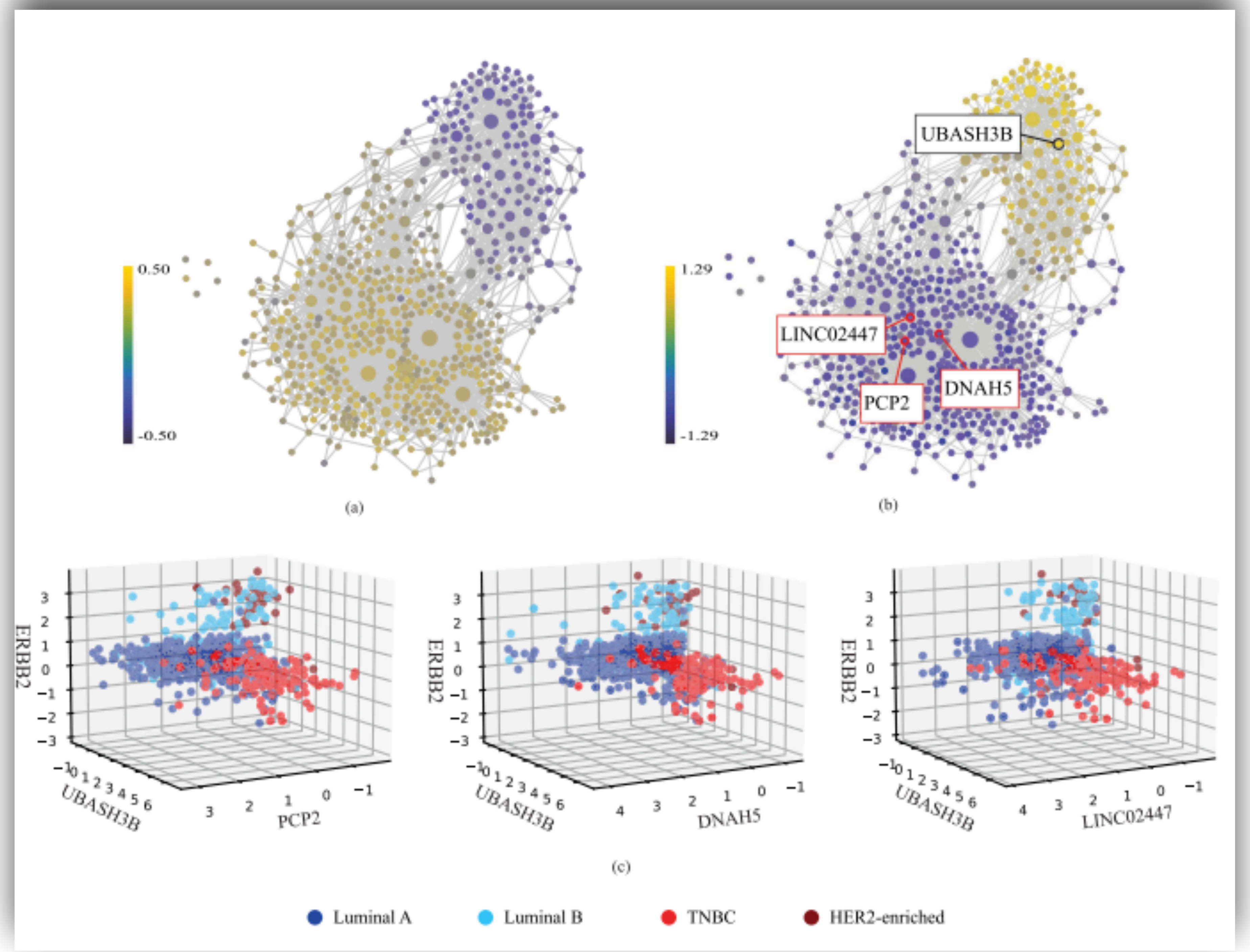
- After selection of a chain homotopy class of chain maps from $C_*(X)$ to $C_*(Y)$, i.e. a coset in $\sum_i C^i(X) \otimes C_i(Y)$, minimize one of the values θ or θ_μ over that coset.
- If $\alpha =$ Min θ_μ is sufficiently small, we can hope that the map comes from a map of simplicial complexes from X to the nerve of a covering of Y, namely the covering by balls of radius α
- If we minimize using θ, we hope that it will also create a small value of α
- So "minimal cycles" in the Hom-complex correspond to actual geometric maps
- Topological feature creation for constrained feature sets.

Foundations of TDA

Example: Splitting Luminal B

Z. Rostani, D. Fooshee, S. Subramaniam, GC





Rostami, Z., Fooshee, D., Carlsson, G., & Subramaniam, S. (2025). Topological Data Analysis Reveals a Subgroup of Luminal B Breast Cancer. *IEEE Open Journal of Engineering in Medicine and Biology*, *6*, 465-471.

Mapper

- Mapper is often used to identify “meaningful” sets within a data set.
- Data-base theorists differentiate between *Intentional* and *Extensional* sets
- Intentional sets are characterized by inequalities and equations
- Extensional sets are defined by enumeration
- Mapper allows us to define and work with extensional sets, using point and click methods as well as graph clustering techniques.

Topology

- The insight that Hausdorff had was that one can capture geometry by the properties of an algebra of subsets, the closed sets, rather than by the properties of equations and inequalities.

- This observation recognized the power of extensional sets.

- For data science, can one make formal a notion of *meaningful* sets analogous to the notion of closed sets?

- For example, can one build a “soft” version of closed sets?

Mathematics of Data, Learning, and Intelligence

https://mdlijournal.org/index.php/mdli

Thank You!