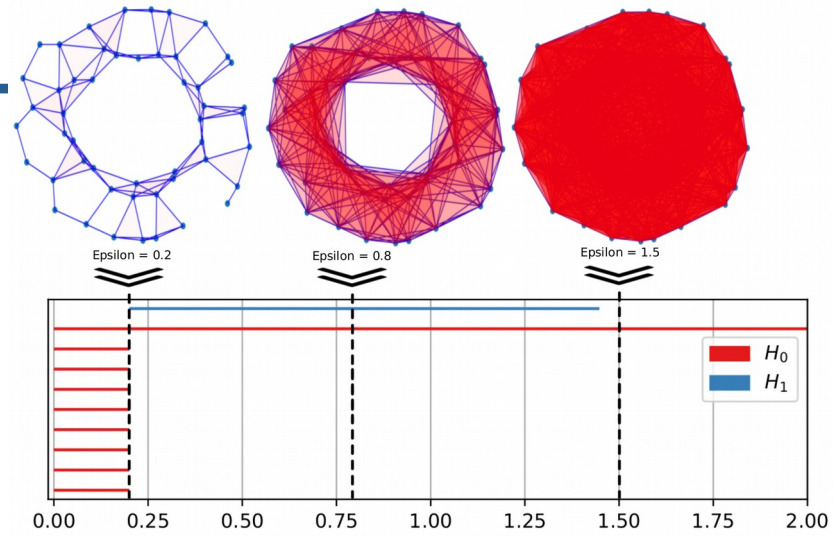
III:Small: Partitioning Big Data for High Performance Computation of Persistent Homology

PI: Philip A. Wilsey
Institutions: University of Cincinnati

Award #: IIS-1909096



Persistent Homology

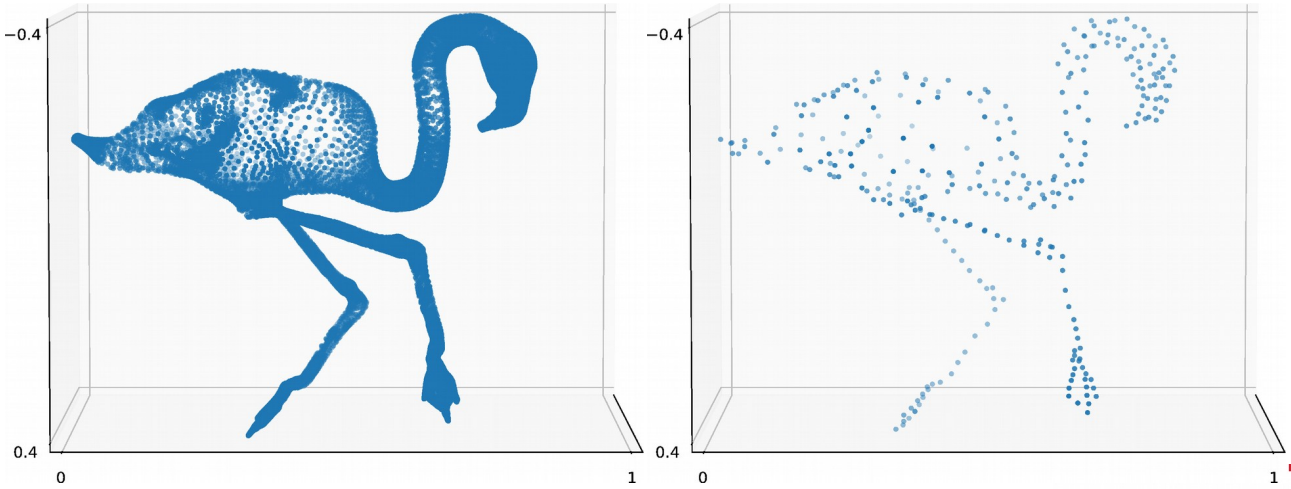
- Characterizes data from *Topological Features*
- Exponential complexity (time & space)
- Important Applications: brain artery analysis, viral evolution, COPD lung patient analysis

Data Reduction & Partitioning

- Data reduction w/ clustering
  - Centroids → Reduced Data (large features)
    - upscaling: improve feature boundary
  - Clusters → Partitions (small features)
  - Possible feature loss

Performance Implications

- Embarassingly parallel
- 3-4 orders of magnitude performance
- Minimal loss in persistence interval loss



Data reduction: 27K → 300 points

Utilizing k-means++ centroids

Contact information:
Philip A. Wilsey
High Performance Computing Lab
Dept of EECS, PO Box 210030
Cincinnati, OH 45221-0030

University of CINCINNATI