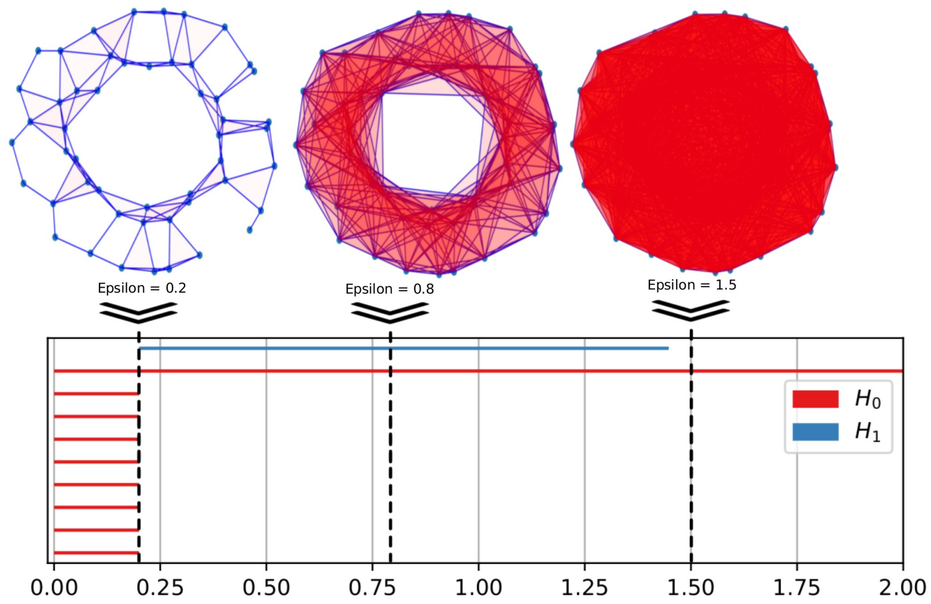
IIS--1909096
Award #: IIS-1909096

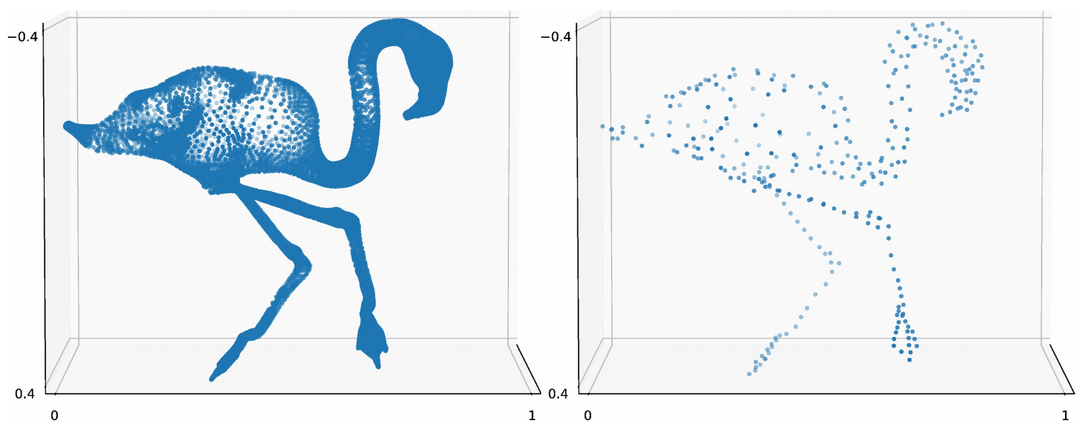
III:Small: Partitioning Big Data for High Performance Computation of Persistent Homology

PI: Philip A. Wilsey
Institution: University of Cincinnati

Topological Data Analysis/ Persistent Homology

- **Exponential Complexity** (time & space)
  - Limited to ~10K points in R^3 (64GB RAM)
- **Data reduction & Partitioning**
  - Use cluster centroids (samples)
  - Use clusters +δ as partitions
- **Parallelism and concurrency**





Data reduction: 27K → 300 points

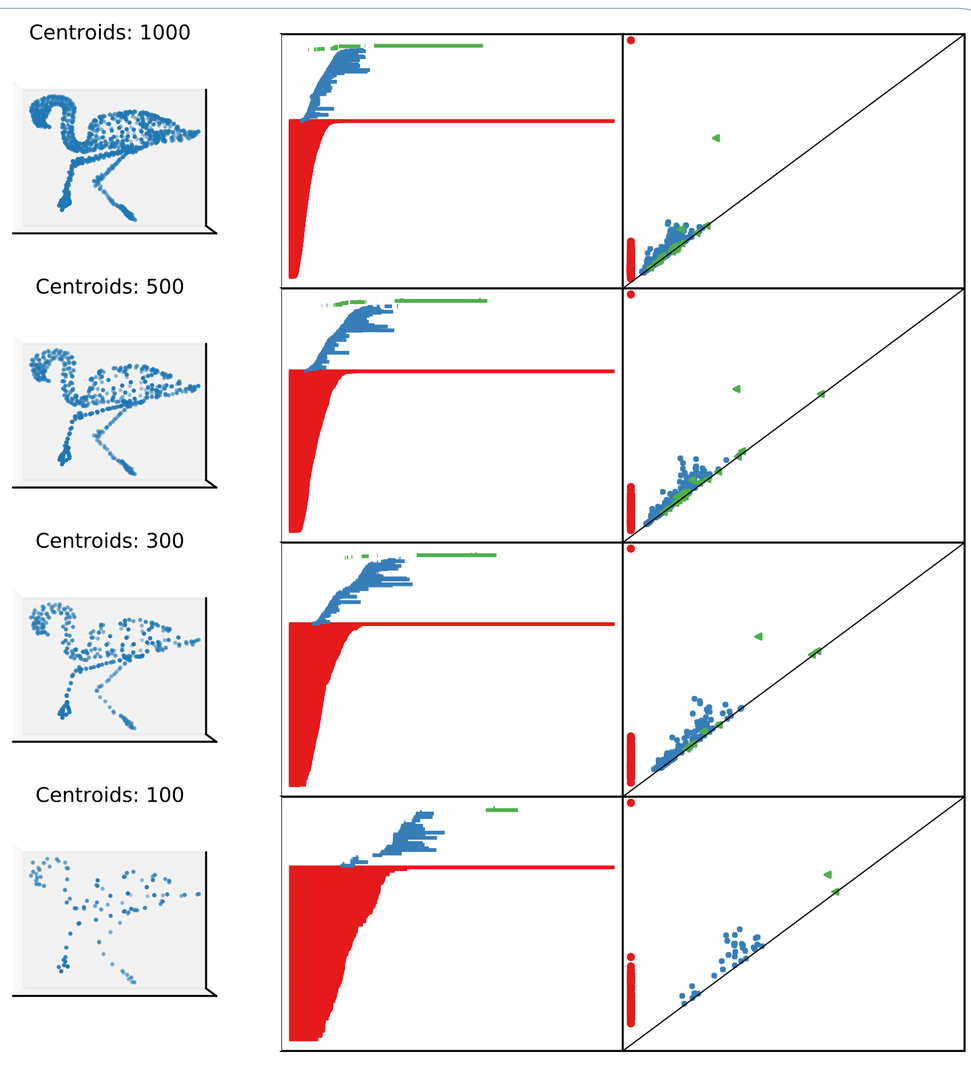
Utilizing k-means++

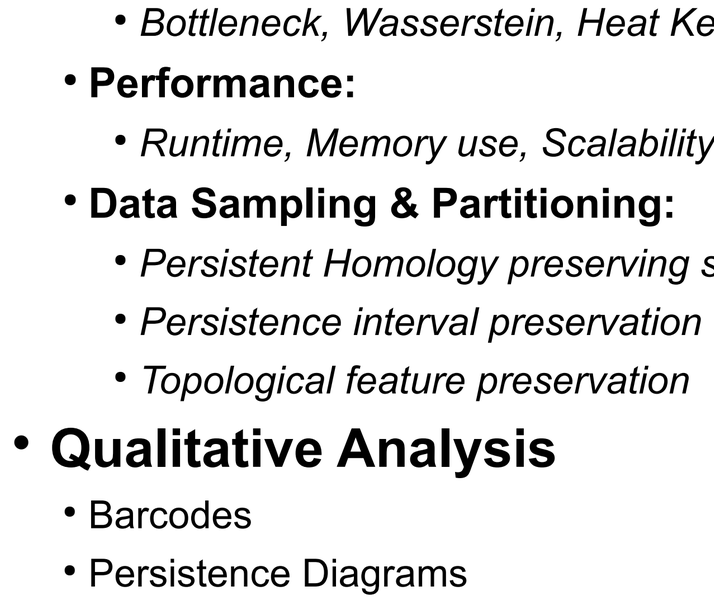
Data Sampling and Partitioning

- **Witness complex**
- **Random Sampling**
- **Clustering (enables upscaling)**
  - Density-based: DBScan
  - Grid-based (distance-independent)
  - Partition-based: k-Means++
  - Hierarchical-based: Agglomerative
- **3-4 orders of magnitude perf gains**

Output Analysis

- **Quantitative Analysis**
  - **Persistence Interval Comparison:**
    - *Bottleneck, Wasserstein, Heat Kernel Distances*
  - **Performance:**
    - *Runtime, Memory use, Scalability*
  - **Data Sampling & Partitioning:**
    - *Persistent Homology preserving sampling*
    - *Persistence interval preservation*
    - *Topological feature preservation*
- **Qualitative Analysis**
  - Barcodes
  - Persistence Diagrams
  - Landscape Diagrams
  - Persistence Images
  - Feature Boundary Extraction





Data Reduction performance Improvements

- **Reduced input points results in:**
  - **Reduced Memory Footprint**
  - **Faster CPU and Wall Time**
  - **Approximations of large features with bounded error**
  - **Upscaling to improve boundary identification of topological feature**

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

University of CINCINNATI