Novel Homology-Based Centrality Measures for Weighted Graphs

A Presentation for GTDAML 2021

John Rick Dolor Manzanares

University of the Philippines Baguio

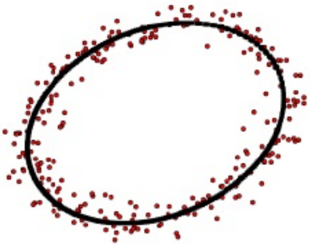
1 August 2021

Topological Data Analysis

- Properties from the "shape" of data sets. These data sets are usually embedded in metric spaces.

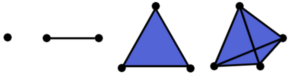
Algebraic Topology and Simplicial Homology

- Focus of our study will be holes.

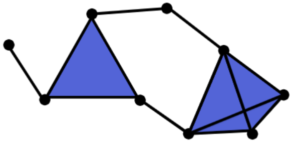


Vietoris-Rips Simplicial Complex

- Connect, using an edge, any two points ϵ units apart.
- A higher-dimensional simplex forms assuming its codimension-1 faces appear.



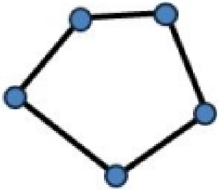
- The resulting structure is a **simplicial complex**, a generalization of graphs.

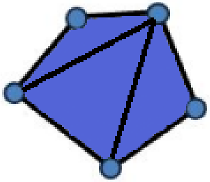


- Increasing the parameter ϵ, we obtain a sequence of simplicial complexes known as a **filtration**.

Persistence of Homology Classes

- Holes and voids, as equivalence classes, are called **homology classes**.





- For any homology class, we collect the ordered pairs (b, d) where b is the **birth** and d is the **death** of the homology class. The difference $d - b$ is called the **persistence**.

Quantifying Importance of a Homology Class

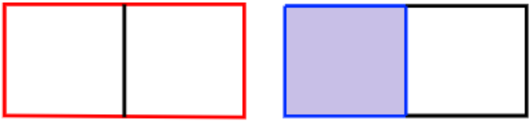
- More persistent homology classes are more important.
- Short-lived homology classes are important in estimation problems such as determining curvatures of surfaces [2].
- (Graph Theory) Loop Centrality is an importance based on the number of walks passing a cycle or loop.

Merging Homology Classes

- A homology class either trivializes or merges with another homology class.
- The elder rule states that an elder feature in the filtration survives while the other feature dies when two features merge.
- Two kth homology classes $[\sigma]$ and $[\delta]$ **merge** at ϵ if

$$\sigma + \delta = \rho$$

for some kth boundary ρ.



Proposed Centrality Measures

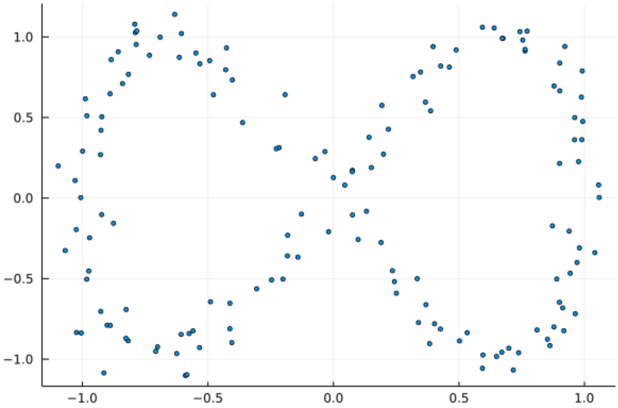
- The form of the proposed centrality measures of homology class $[\sigma]$ is given by

$$J(\sigma, \epsilon) = \begin{cases} 0 & \text{for } \epsilon \leq b(\sigma) \\ [\epsilon - b(\sigma)] + \sum_{\varsigma \in M_1[\sigma,\epsilon]} cP(\varsigma) & \text{for } b(\sigma) < \epsilon \leq d(\sigma) \\ P(\sigma) + \sum_{\varsigma \in M_1[\sigma,\epsilon]} cP(\varsigma) & \text{for } \epsilon > d(\sigma) \end{cases}$$

  where $c \in (0, 1]$.
  - Let $c = 1$ to denote full weight on the persistence of merging homology classes.
  - Consider the merging time and death to give preference to early and late merging.
  - Extend those three centrality measures by also considering merging classes of each homology class in the merging class.

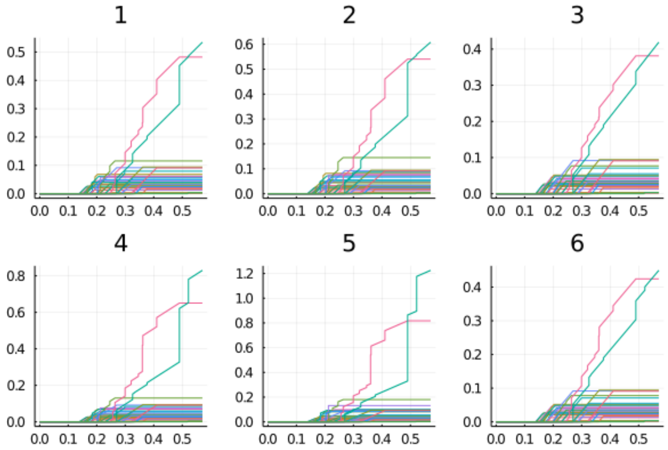
Numerical Simulations on Synthetic Data Sets

Consider the point cloud depicting a noisy infinity sign (or figure eight).

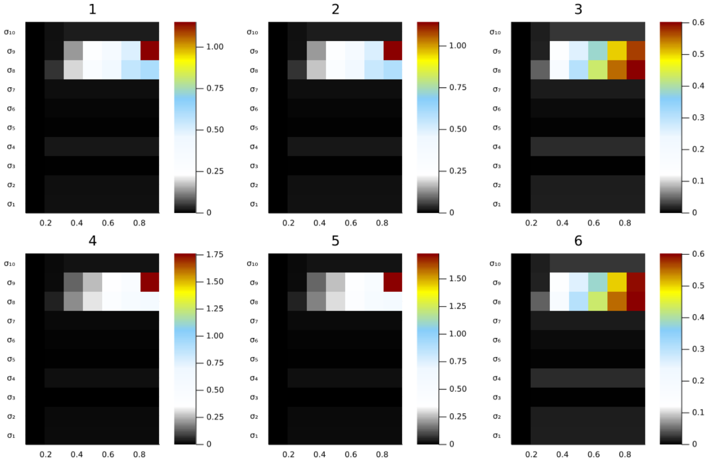


Representation of Centrality Measures

As expected, we see two relatively important holes. Those two holes are the largest holes on the left and right symmetrical parts of the infinity sign.

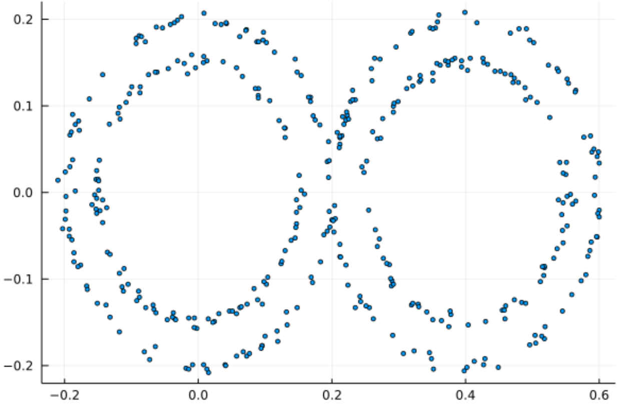


Alternative Representation



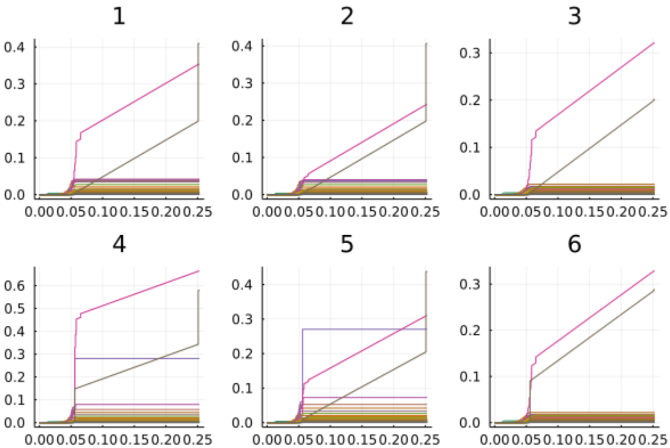
Another Numerical Simulation

Now, we consider a noisy infinity sign but added with an annular region. Two large holes are visible and expected to be captured by the centrality plot.

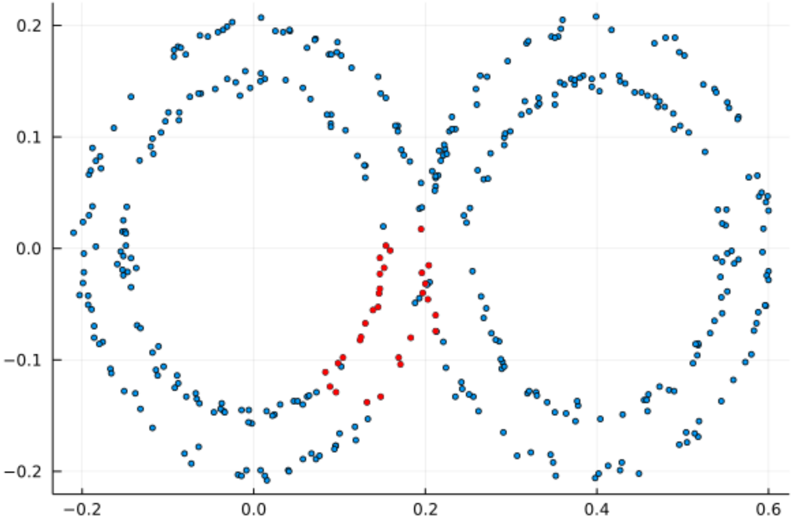


Another Interpretation

The extended centrality measures capture a third relatively important hole. Centrality 5 gives preference to early merging which means the third important hole merges early with other holes.



Third Relatively Important Homology Class



Differences to Common Topological Summaries

- A barcode or a persistence diagram does not capture the third hole as relatively important.
- Less persistent holes may be important different to popular notion of importance for homology classes.

Properties of the Centrality Measures

- The centrality measure is a monotonic function.
- For each homology class $[\sigma]$ in the filtered simplicial complex, we define

$$\|J_\sigma(\epsilon)\|_p = \left\{ \int_{[0,d(\sigma)]} |J_\sigma(x)|^p \ dx \right\}^p$$

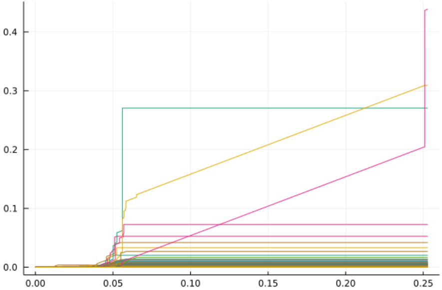
for $1 \leq p < \infty$. Furthermore, we form the set

$$\{\|J_\sigma(\epsilon)\|_p : \sigma \text{ is a homology class}\}.$$

We want to define a distance between two centrality plots.

Properties (Continuation)

The landscape distance [1] compares the kth order landscapes. The centrality plot cannot have such reasonable ordering. Hence, we consider all possible matchings between the centrality curves similar to the idea of the bottleneck distance.



Properties (Continuation)

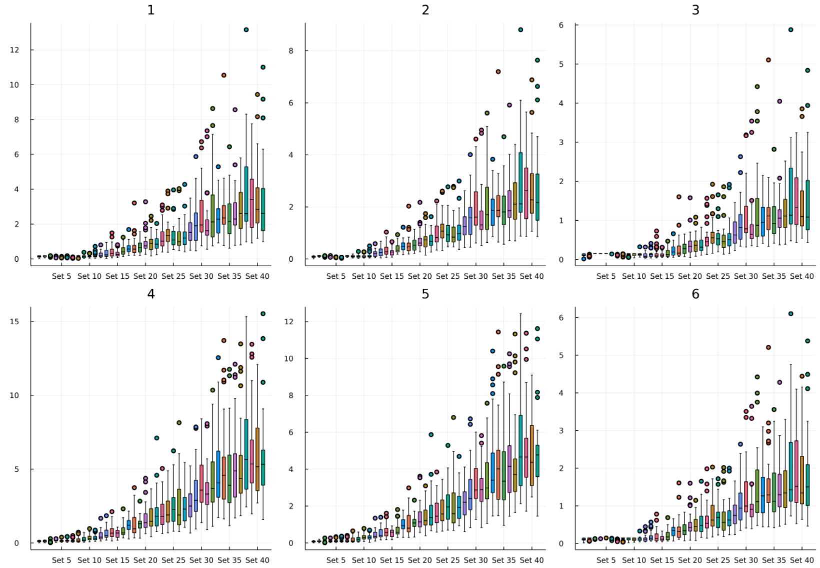
Assuming the birth times are equal, the implementation [4] of the bottleneck distance between two persistence diagrams depending only on the death times, δ_x and $\delta_{\phi(x)}$, is given by $\inf_\phi \sup_{x \in X} \|x - \phi(x)\|_\infty$ where

$$\|x - \phi(x)\|_\infty = \begin{cases} \frac{1}{2} \max\{\delta_x, \delta_{\phi(x)}\} & \text{if } \phi(x) \in \Delta \\ |\delta_x - \delta_{\phi(x)}| & \text{otherwise} \end{cases}.$$

The image $\phi(x)$ of x under a bijection ϕ is a matched point for x and Δ is the set of diagonal points. We replace the death times to $|J_\sigma(\epsilon)|_p$ which is a number that only depends on $d(\sigma)$.

- (Current Work) Stability of centrality plots under the defined distance.

Empirical Evidence



References I

P. Bubenik and P. Dłotko, *A persistence landscapes toolbox for topological statistics*, Journal of Symbolic Computation, 78 (2017), pp. 91 – 114.

P. Bubenik, M. Hull, D. Patel, and B. Whittle, *Persistence homology detects curvature*, Inverse Problems, 36 (2020).

G. E. Carlsson, *Topology and data*, Bulletin (New Series) of the American Mathematical Society, 46 (2009), pp. 255–308.

P. S. Ignacio, J.-A. Bulauan, and D. Uminsky, *Lumáwig: An efficient algorithm for dimension zero bottleneck distance computation in topological data analysis*, Algorithms, 13 (2020).

Thank you for listening! Any Questions?