Persistent Homology Approximations of Network Distances

Weiyu Huang and Alejandro Ribeiro

Department of Electrical and Systems Engineering
University of Pennsylvania
whuang@seas.upenn.edu
http://www.weiyuhuang.net

December 16, 2015
GlobalSIP’2015
Networked data structures encode relationships between elements.

How to evaluate dissimilarities between networks remain unclear.
Network comparison

- Neurodegenerate
  - Association with brain network
  - Feature heuristics so far
  - Not specific to a region of the brain
  - But more about global properties
  - Need to be compared as unlabeled entities
My approach

- Define and estimate **distances** between unlabeled networks
  - More generalizable
  - Universal and avoid conflicting statement
- High order networks
  - Relationships between three or four nodes, or even singleton
High order networks

- Network $N^K_X = (X, r^0_X, r^1_X, \ldots, r^K_X)$

  $\Rightarrow r^k_X$ is a mapping $\prod_{i=0}^{k+1} X \rightarrow \mathbb{R}_+$

  $\Rightarrow r^2_X(x_1, x_2, x_2) = r^1_X(x_1, x_2)$

- Proximity network $P^K_X$ if $r^k_X(x_{0:k}) \leq r^{k-1}_X(x_{0:k-1})$

  $\Rightarrow$ Order decreasing

- Dissimilarity network $D^K_X$ if $r^k_X(x_{0:k}) \geq r^{k-1}_X(x_{0:k-1})$

  $\Rightarrow$ Order increasing
A correspondence $C$ between $X$ and $Y$ is $C \subseteq X \times Y$ s.t.

- $\forall x \in X$, there exists $y \in Y$ such that $(x, y) \in C$
- $\forall y \in Y$, there exists $x \in X$ such that $(x, y) \in C$
- Generalizes permutations
- $\mathcal{C}(X, Y)$ the set of all correspondences
Given proximity networks $P^K_X$ and $P^K_Y$ and a correspondence $C$,

Define the $k$-order network difference with respect to $C$ as

$$\Gamma^k_{X,Y}(C) := \max_{(x_0:k,y_0:k) \in C} |r^k_X(x_0:k) - r^k_Y(y_0:k)|.$$

The $k$-order proximity network distance is defined as

$$d^k_P(P^K_X, P^K_Y) := \min_{C \in \mathcal{C}(X,Y)} \{ \Gamma^k_{X,Y}(C) \}.$$

Theorem

$d^k_P : \mathcal{P}^K \times \mathcal{P}^K \rightarrow \mathbb{R}_+$ is a metric in space $\mathcal{P}^K \mod \cong_k$ for $k \geq 1$ and a pseudometric in $\mathcal{P}^K \mod \cong_0$. 
- **$k$-simplex** $[x_{0:k}]$ is the convex hull of the set of points $x_{0:k}$

- **0-simplex**: vertex $[a]$  
- **1-simplex**: edge $[a, b]$  
- **2-simplex**: triangle $[a, b, c]$  
- **3-simplex**: tetrahedron $[a, b, c, d]$
Simplicial complex $L$ is the collection of simplices glued together.
We want to describe holes that do not have interiors.

Why do we consider them? Think of rubber bands.

⇒ Rubber band enclosing them cannot be diminished.
We want to describe holes that do not have interiors.

⇒ Homological features are defined to formalize this.

⇒ Cycles without any interiors.
Homological features describe hole with no interior

- $[a, b], [b, d], [d, a]$ forms a hole that has interior
  ⇒ Not a Homological feature

- $[d, b], [b, c], [c, d]$ forms a hole with no interior
  ⇒ Is homological feature
Homological features describe hole with no interior

- $[a, b], [b, d], [d, a]$ forms a hole that has interior
  $\Rightarrow$ Not a Homological feature

- $[d, b], [b, c], [c, d]$ forms a hole with no interior
  $\Rightarrow$ Is homological feature
Homological features describe hole with no interior.
Filtrations

- Simplicial complexes form *unweighted* networks
  - No way to incorporate *weights*
  - To solve this problem, assign each simplex a *value*
  - The time when this simplex *appears*
Filtrations

- Simplicial complexes form unweighted networks
  - No way to incorporate weights
  - To solve this problem, assign each simplex a value
  - The time when this simplex appears

\[ t = 0.1 \]

Diagram:

- Vertices labeled as \( a, b, c, d \)
- Edges labeled as 0.1
- Time \( t = 0.1 \)
- Simplicial complexes form **unweighted** networks

- No way to incorporate **weights**
- To solve this problem, assign each simplex a **value**
- The time when this simplex **appears**

![Diagram of a simple network with edges labeled 0.1 and 0.2 and a vertical time axis labeled t = 0.2]
Filtrations

- Simplicial complexes form unweighted networks
  - No way to incorporate weights
  - To solve this problem, assign each simplex a value
  - The time when this simplex appears

\[
\begin{align*}
0 & \quad a \quad 0.1 \quad b \quad 0 \\
0 & \quad d \quad 0.3 \quad c \quad 0 \\
0 & \quad c \quad 0.2 \quad b \\
\end{align*}
\]

\[t = 0.3\]
Filtrations

- Simplicial complexes form **unweighted** networks
  - No way to incorporate **weights**
  - To solve this problem, assign each simplex a **value**
  - The time when this simplex **appears**

![Diagram of simplicial complexes with weights at different times](attachment:image.png)
Quantify when do **holes** and **interiors** appear.

![Diagram showing persistent homology](attachment:image.png)
Quantify when do holes and interiors appear
Quantify when do holes and interiors appear
Quantify when do holes and interiors appear
Quantify when do holes and interiors appear.
Quantify when do holes and interiors appear.
Difference between persistence diagrams

- Bottleneck distance $d_B^\infty(Q, \tilde{Q})$ between two point sets $Q$ and $\tilde{Q}$

\[
d_B^\infty(Q, \tilde{Q}) = \min_{\pi} \max_{q \in Q} \| q - \pi(q) \|_\infty,
\]

- where $\pi$ ranges over all bijections from $Q$ to $\tilde{Q}$
- $|Q| = |\tilde{Q}|$ are point sets in two dimensional space

\[
d_B^\infty(Q, \tilde{Q}) = \max\{|\infty - 0.8|, |0.7 - 0.7|\} = \infty
\]
Diagrams with different cardinalities

- $d^\infty_B$ ill-defined if for diagrams with different cardinalities
- Homological features trivialized at the same time they appear
- Add diagonal points to the persistence diagram with fewer nodes
- Linear Bottleneck Assignment Problem: \( \min_\pi \max_i c(q_i, \tilde{q}_\pi(i)) \)

\[
c(q, \tilde{q}) = \min \left\{ \|q - \tilde{q}\|_\infty, \frac{1}{2} \max \left\{|q_x - q_y|, |\tilde{q}_x - \tilde{q}_y|\right\} \right\}.
\]

\[
d^\infty_B(Q, \tilde{Q}) = \frac{1}{2} |0.8 - 0.7| = 0.05
\]
Theorem

\(d_B^\infty \) between the \(k\)-persistence diagrams of the filtrations \(\mathcal{L}(D_X^K)\) and \(\mathcal{L}(D_Y^K)\) is at most \(d_{D,\infty}(D_X^K, D_Y^K)\) for any \(0 \leq k \leq K\), i.e.

\[d_B^\infty (\mathcal{P}_k \mathcal{L}(D_X^K), \mathcal{P}_k \mathcal{L}(D_Y^K)) \leq d_{D,\infty}(D_X^K, D_Y^K).\]
Theorem

Any $k$-order relationships between full rank tuples of $D^K_X$ appear either in the death time of the $(k - 1)$-th dimensional homological features or the birth time of the $k$-th dimensional homological features.

\begin{figure}
\centering
\begin{tikzpicture}
\node[shape=circle,draw=blue] (X1) at (1,3) {$x_1$};
\node[shape=circle,draw=blue] (X2) at (1,1) {$x_2$};
\node[shape=circle,draw=blue] (X3) at (1,-1) {$x_3$};
\node[shape=circle,draw=blue] (Y1) at (5,3) {$y_1$};
\node[shape=circle,draw=blue] (Y2) at (5,1) {$y_2$};
\node[shape=circle,draw=blue] (Y3) at (5,-1) {$y_3$};

\path[->,draw=red] (X1) edge node[above] {0} (X2);
\path[->,draw=red] (X2) edge node[above] {0.32} (Y1);
\path[->,draw=red] (X3) edge node[above] {0.42} (X1);
\path[->,draw=red] (X3) edge node[above] {0.6} (X2);
\path[->,draw=red] (Y1) edge node[above] {C} (Y3);
\path[->,draw=red] (Y2) edge node[above] {0.5} (Y3);
\path[->,draw=red] (Y3) edge node[above] {C} (Y2);
\path[->,draw=red] (X1) edge node[above] {C} (Y1);
\path[->,draw=red] (X2) edge node[above] {C} (Y2);
\path[->,draw=red] (X3) edge node[above] {C} (Y3);

\node at (2.5,3) {$D^1_X$: 0.42};
\node at (3.5,1) {$D^1_Y$: 0.39};
\node at (5.5,3) {$D^1_Y$: 0.49};
\node at (5.5,1) {$D^1_X$: 0.12};
\node at (5.5,-1) {$D^1_X$: 0.6};
\node at (2.5,-1) {$D^1_Y$: 0.25};
\node at (3.5,0) {$C$};
\node at (4.5,0) {$C$};
\node at (5.5,0) {$C$};
\end{tikzpicture}
\end{figure}
Application

$d_B^\infty(P_0L)$ removed  $d_B^\infty(P_0L)$ replaced  $d_B^\infty(P_1L)$ replaced  $d_B^\infty(P_2L)$ replaced
Future directions

Applications
- Brain networks at finer scale
- Pattern recognition from time series of observations

Theory
- Clustering based on distance intervals
- Graph structure inference