Persistent Homology Approximations on Network Distances

Weiyu Huang and Alejandro Ribeiro

Department of Electrical and Systems Engineering

University of Pennsylvania

whuang@seas.upenn.edu

http://www.weiyuhuang.net

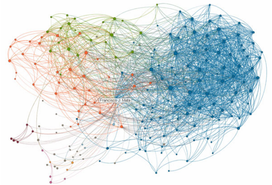
December 16, 2015

GlobalSIP'2015

Networked data

Internet

Online social media

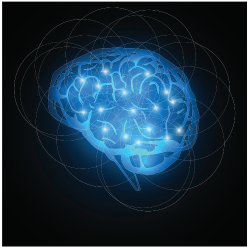


Clean energy & grid

- Networked data structures encode relationships between elements
- How to evaluate dissimilarities between networks remain unclear

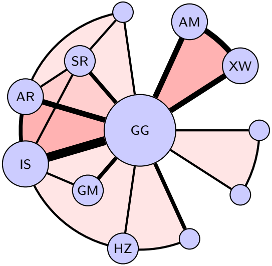
Network comparison

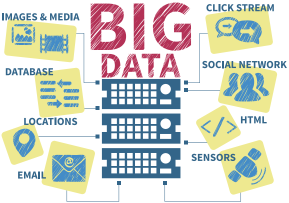
- Neurodegenerate
  - Association with brain network
  - Feature heuristics so far
  - Not specific to a region of the brain
  - But more about global properties
  - Need to be compared as unlabeled entities



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 $N_X^K = \left(X, r_X^0, r_X^1, \ldots, r_X^K\right)$

  $\Rightarrow$ r_X^k is a mapping $\overbrace{X \times X \cdots \times X}^{k+1} \rightarrow \mathbb{R}_+$

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

- Proximity network P_X^K if $r_X^k(x_{0:k}) \leq r_X^{k-1}(x_{0:k-1})$

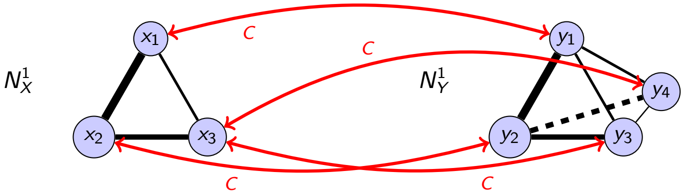
  $\Rightarrow$ Order decreasing

- Dissimilarity network D_X^K if $r_X^k(x_{0:k}) \geq r_X^{k-1}(x_{0:k-1})$

  $\Rightarrow$ Order increasing

Correspondence

- 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



Network metric

- Given proximity networks P_X^K and P_Y^K and a correspondence C
- Define the k-order network difference with respect to C as

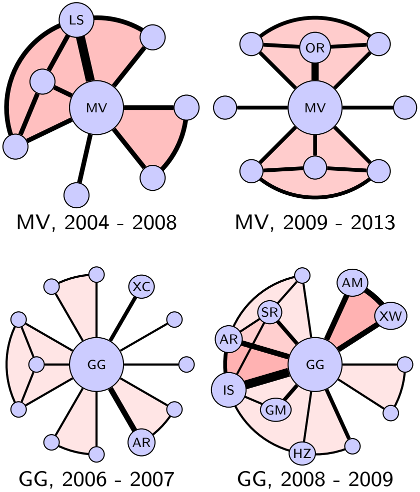
$$\Gamma_{X,Y}^k(C) \ := \max_{(x_{0:k}, y_{0:k}) \in C} \left| r_X^k(x_{0:k}) - r_Y^k(y_{0:k}) \right| .$$

- The k-order proximity network distance is defined as

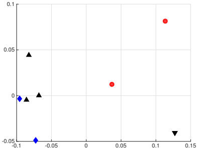
$$d_{\mathcal{P}}^k(P_X^K, P_Y^K) \ := \min_{C \in \mathcal{C}(X,Y)} \left\{ \Gamma_{X,Y}^k(C) \right\} .$$

Theorem

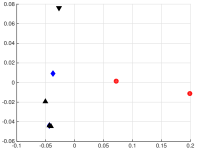
$d_{\mathcal{P}}^k : \mathcal{P}^K \times \mathcal{P}^K \to \mathbb{R}_+$ is a metric in space $\mathcal{P}^K \mod \cong_k$ for $k \geq 1$ and a pseudometric in $\mathcal{P}^K \mod \cong_0$.


LS
MV
OR
MV
GG
ZT
SR
AR
IS
GG
GM
JB
MV, 2004 - 2008
MV, 2009 - 2013
GG, 2004 - 2008
GG, 2009 - 2013
XC
GG
AR
AM
SR
XW
AR
GG
IS
GM
HZ
GM
JB
AD
VK
GG
SF
GM
GG
PF
GG, 2006 - 2007
GG, 2008 - 2009
GG, 2010 - 2011
GG, 2012 - 2013

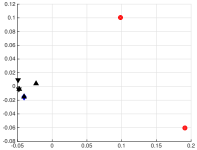
Applications



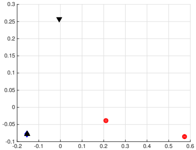
$d^0_{\mathcal{P}}$



$d^1_{\mathcal{P}}$



$d^2_{\mathcal{P}}$



$d_{\mathcal{P},1}$

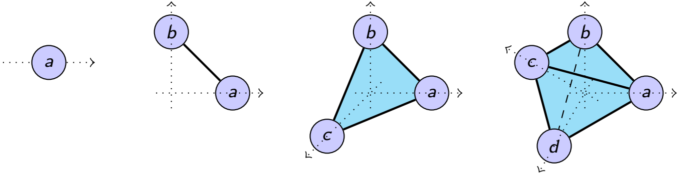
Motivation for lower bounds

- Computable for large networks
- Admittedly, lower bounds may suffer same problems as features
- Large lower bound entails a large distance
- Upper bound easy to establish

  $\Rightarrow$ Used to estimate distance intervals

Simplex

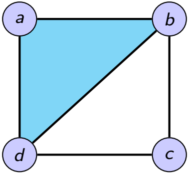
- k-simplex $[x_{0:k}]$ is the convex hull of the set of points $x_{0:k}$



vertex $[a]$	edge $[a, b]$	triangle $[a, b, c]$	tetrahedron $[a, b, c, d]$
0-simplex	1-simplex	2-simplex	3-simplex

Simplicial complex

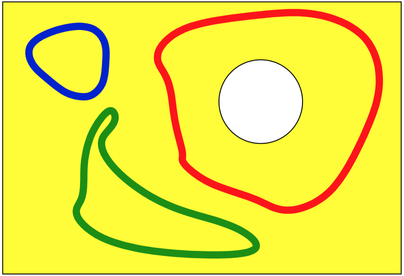
- Simplicial complex L is the collection of simplices glued together



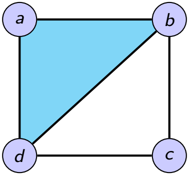
Homology (1 of 2)

- We want to describe holes that do not have interiors
- Why do we consider them $\Rightarrow$ think of rubber bands

  $\Rightarrow$ Rubber band enclosing them cannot be diminished



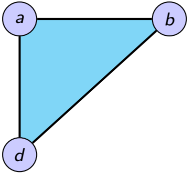
- We want to describe holes that do not have interiors
  - ⇒ Homological features are defined to formalize this
  - ⇒ Cycles without any interiors



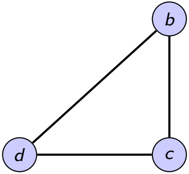
- Homological features describe hole with no interior
- $[a, b], [b, d], [d, a]$ forms a hole that has interior

  $\Rightarrow$ Not a Homological feature
- $[d, b], [b, c], [c, d]$ forms a hole with no interior

  $\Rightarrow$ Is homological feature

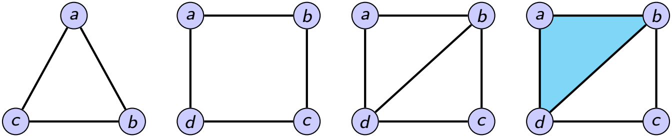


- Homological features describe hole with no interior
- $[a, b], [b, d], [d, a]$ forms a hole that has interior
  - $\Rightarrow$ Not a Homological feature
- $[d, b], [b, c], [c, d]$ forms a hole with no interior
  - $\Rightarrow$ Is homological feature



Homological features

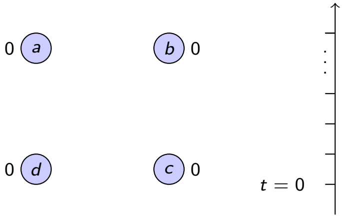
- Homological features describe hole with no interior



1 feature   1 feature   2 features   1 feature

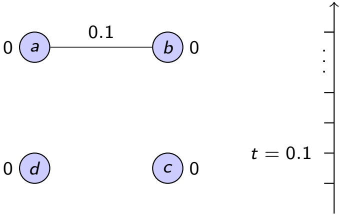
Filtrations

- 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



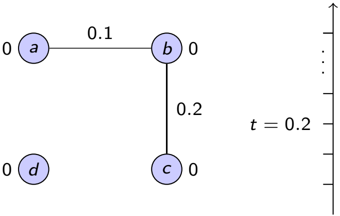
Filtrations

- 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



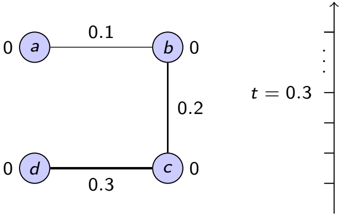
Filtrations

- 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

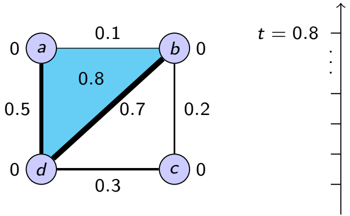


Filtrations

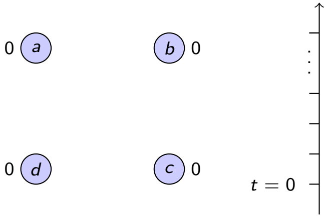
- Simplicial complexes form unweighted networks
  - No way to incorporate weights
  - To solve this problem, assign each simplex a value
  - The time when this simplex appears

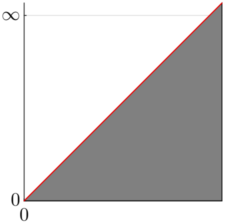


- Simplicial complexes form unweighted networks
  - No way to incorporate weights
  - To solve this problem, assign each simplex a value
  - The time when this simplex appears



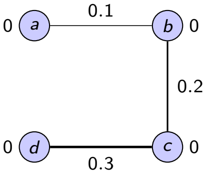
- Quantify when do holes and interiors appear

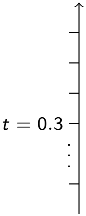


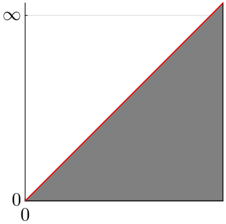


Persistent Homology (1 of 2)

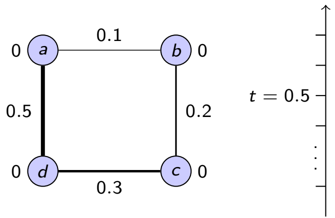
- Quantify when do holes and interiors appear

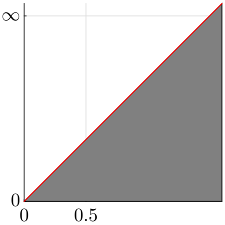




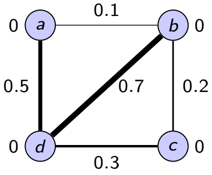


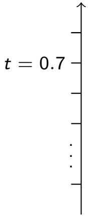
- Quantify when do holes and interiors appear

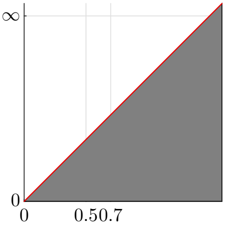




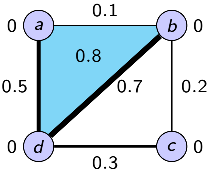
- Quantify when do holes and interiors appear

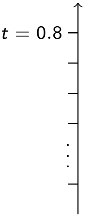


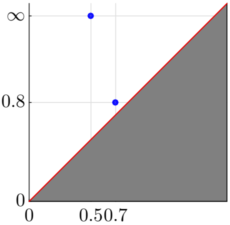




- Quantify when do holes and interiors appear

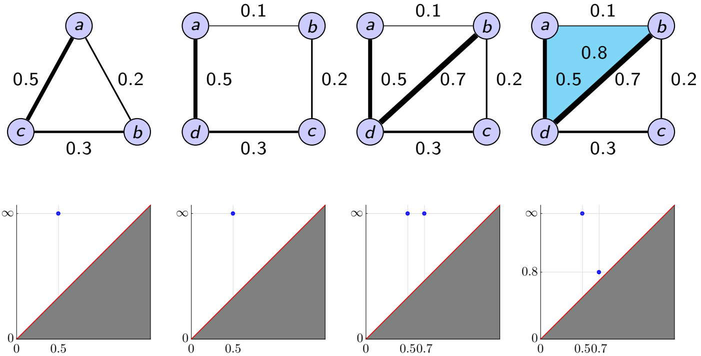






Persistent Homology (2 of 2)

- Quantify when do holes and interiors appear

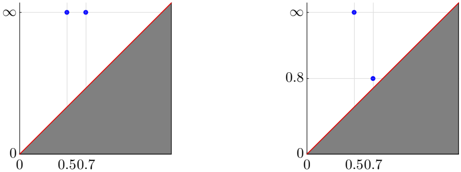


Difference between persistence diagrams

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

$$\color{red}{d_B^\infty}(Q, \tilde{Q}) = \min_{\color{blue}{\pi}} \max_{q \in Q} \|q - \color{blue}{\pi}(q)\|_\infty ,$$

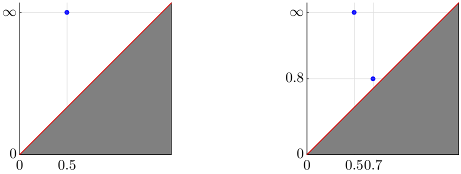
  - where $\color{blue}{\pi}$ ranges over all bijections from Q to $\tilde{Q}$
  - $|Q| = |\tilde{Q}|$ are point sets in two dimensional space



$$d_B^\infty(Q, \tilde{Q}) = \max\{|\infty - 0.8|, |0.7 - 0.7|\} = \infty$$

- d_{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 : $\min_{\pi} \max_i c(q_i, \tilde{q}_{\pi(i)})$

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



$$d_{\mathrm{B}}^{\infty}(Q, \tilde{Q}) = \tfrac{1}{2}|0.8 - 0.7| = 0.05$$

Network metric lower bounds

Theorem

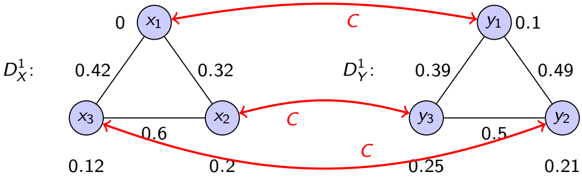
d_B^∞ *between the k-persistence diagrams of the filtrations* $\mathcal{L}(D_X^K)$ *and* $\mathcal{L}(D_Y^K)$ *is at most* $d_{\mathcal{D},\infty}(D_X^K, D_Y^K)$ *for any* $0 \leq k \leq K$*, i.e.*

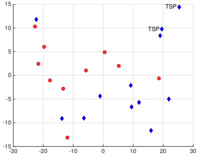
$$d_B^\infty(\mathcal{P}_k\mathcal{L}(D_X^K), \mathcal{P}_k\mathcal{L}(D_Y^K)) \leq d_{\mathcal{D},\infty}(D_X^K, D_Y^K).$$

Bound is tight

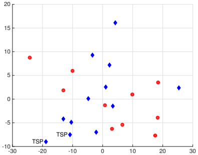
Theorem

Any k-order relationships between full rank tuples of D_X^K 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.

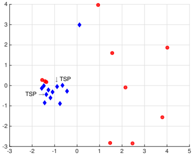




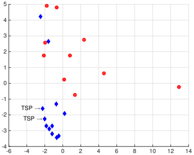
$d_{\mathrm{B}}^{\infty}(\mathcal{P}_0\mathcal{L})$ removed



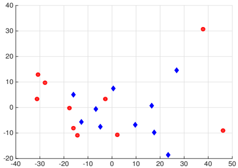
$d_{\mathrm{B}}^{\infty}(\mathcal{P}_0\mathcal{L})$ replaced



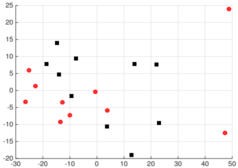
$d_{\mathrm{B}}^{\infty}(\mathcal{P}_1\mathcal{L})$ replaced



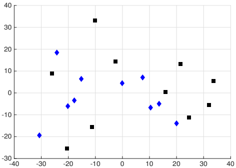
$d_{\mathrm{B}}^{\infty}(\mathcal{P}_2\mathcal{L})$ replaced



TAC - TSP



TAC - TWC



TSP - TWC

Future directions

- Applications
  - Brain networks at finer scale
  - Pattern recognition from time series of observations
- Theory
  - Clustering based on distance intervals
  - Graph structure inference