A Local Homology Based Distance

Brittany Terese Fasy, Tulane University
joint work with Mahmuda Ahmed and Carola Wenk

27 March 2014
www.mapconstruction.org
www.mapconstruction.org
Examples
Examples

- Road networks: GPS trajectories of cars.
Examples

- Road networks: GPS trajectories of cars.
- Hiking paths: GPS paths of people.
Examples

- Road networks: GPS trajectories of cars.
- Hiking paths: GPS paths of people.
- Migration paths: GPS on animals.
Examples

- Road networks: GPS trajectories of cars.
- Hiking paths: GPS paths of people.
- Migration paths: GPS on animals.
- Fillaments of galaxies: point cloud data.
Examples

- Road networks: GPS trajectories of cars.
- Hiking paths: GPS paths of people.
- Migration paths: GPS on animals.
- Fillaments of galaxies: point cloud data.
- Hurricane paths: historical paths.
Examples

- Road networks: GPS trajectories of cars.
- Hiking paths: GPS paths of people.
- Migration paths: GPS on animals.
- Fillaments of galaxies: point cloud data.
- Hurricane paths: historical paths.
- **Networks:** path-constrained trajectories or point sets.
A road network is represented as an embedded graph $G = (V, E) \subset D$, where $D \subset \mathbb{R}^2$ is compact.
Motivating Questions
What is the Distance Between Embedded Graphs?

Can we compare different road construction techniques?
Motivating Questions

What is the Distance Between Embedded Graphs?

Can we compare different road construction techniques?

Can we detect if (and where) changes have occurred?
Different Approaches


Hausdorff Distance

Road Networks are Sets of Points

\[ G_1 = (V_1, E_1) \]

\[ G_2 = (V_2, E_2) \]

Question

What is \( d(G_1, G_2) \)?
Hausdorff Distance

Road Networks are Sets of Points

\[ G_1 = (V_1, E_1) \]
\[ G_2 = (V_2, E_2) \]

Question

What is \( d(G_1, G_2) \)?

Definition (Hausdorff Distance)

\[
\begin{align*}
    d^i_{h} \rightarrow j &= \sup_{x \in G_i} \inf_{y \in G_j} \| x - y \|_2 \\
    d_H(G_1, G_2) &= \max(d^1_{h} \rightarrow ^2, d^2_{h} \rightarrow ^1)
\end{align*}
\]
Hausdorff Distance

Road Networks are Sets of Points

\[ G_1 = (V_1, E_1) \]
\[ G_2 = (V_2, E_2) \]

Question

What is \( d(G_1, G_2) \)?

Definition (Hausdorff Distance)

\[
\begin{align*}
    d_{h}^{i \rightarrow j} &= \sup_{x \in G_i} \inf_{y \in G_j} ||x - y||_2 \\
    d_\mathcal{H}(G_1, G_2) &= \max(d_{h}^{1 \rightarrow 2}, d_{h}^{2 \rightarrow 1})
\end{align*}
\]
Different Approaches


Hansel and Gretel Distance

Paths Define Neighborhood

- \( p_{01} = \frac{k}{m} \)
- \( r_{01} = \frac{k}{n} \)
- \( d(B(G_0, G_1)) = 2p_{01}r_{01}p_{01} + r_{01} \)
Hansel and Gretel Distance

Paths Define Neighborhood

$r$ is the local radius

$\{x_i\} \subset D$ is a set of seeds
Hansel and Gretel Distance

Paths Define Neighborhood

\( r \) is the local radius
\( \{x_i\} \subset D \) is a set of seeds
\( m \) is number of marks in \( G_0 \)
\( n \) is number of marks in \( G_1 \)
Hansel and Gretel Distance

Paths Define Neighborhood

\( r \) is the local radius
\( \{x_i\} \subset D \) is a set of seeds
\( m \) is number of marks in \( G_0 \)
\( n \) is number of marks in \( G_1 \)
\( k \) is number of pairs
Hansel and Gretel Distance

Paths Define Neighborhood

\( r \) is the local radius
\( \{x_i\} \subset D \) is a set of seeds
\( m \) is number of marks in \( G_0 \)
\( n \) is number of marks in \( G_1 \)
\( k \) is number of pairs

Definition (HG Distance)

\[
p_{01} = \frac{k}{m}, \quad r_{01} = \frac{k}{n}
\]

\[
d_B(G_0, G_1) = \frac{2p_{01}r_{01}}{p_{01} + r_{01}}
\]
Hausdorff Distance

Road Networks are Sets of Points

\[ G_1 = (V_1, E_1) \]
\[ G_2 = (V_2, E_2) \]

**Question**

What is \( d(G_1, G_2) \)?

**Definition (Hausdorff Distance)**

\[ d_h^{i \rightarrow j} = \sup_{x \in G_i} \inf_{y \in G_j} \| x - y \|_2 \]

\[ d_H(G_1, G_2) = \max(d_h^{1 \rightarrow 2}, d_h^{2 \rightarrow 1}) \]
Hausdorff Distance

Road Networks are Sets of Points

\[ G_1 = (V_1, E_1) \]
\[ G_2 = (V_2, E_2) \]

Question

What is \( d(G_1, G_2) \)?

Definition (Hausdorff Distance)

\[
    d_{i \rightarrow j}^h = \sup_{x \in G_i} \inf_{y \in G_j} \| x - y \|_2
\]

\[
    d_H(G_1, G_2) = \max(d_{1 \rightarrow 2}^h, d_{2 \rightarrow 1}^h)
\]
Different Approaches


The Fréchet Distance

Definition (Fréchet Distance)

\[ d_F(\alpha, \beta) = \inf_{\rho} \max_t \{\alpha(t), \beta(\rho(t))\} \]
Path-Based Method

Road Networks are Sets of Paths

\[ \Pi_1 \] is the set of paths in \( G_1 \).
\[ \Pi_2 \] is the set of paths in \( G_2 \).

Question

What is \( d(G_1, G_2) \)?

Definition (Path-Based Distance)

\[
d_p(G_1, G_2) = \sup_{\alpha \in \Pi_1} \inf_{\beta \in \Pi_2} d_F(\alpha, \beta)
\]
Path-Based Method

Road Networks are Sets of Paths

\[ \Pi_1 \] is the set of paths in \( G_1 \).
\[ \Pi_2 \] is the set of paths in \( G_2 \).

Question

What is \( d(G_1, G_2) \)?

Definition (Path-Based Distance)

\[
d_p(G_1, G_2) = \sup_{\alpha \in \Pi_1} \inf_{\beta \in \Pi_2} d_F(\alpha, \beta)
\]
Path-Based Method

Road Networks are Sets of Paths

\[ \Pi_1 \text{ is the set of paths in } G_1. \]
\[ \Pi_2 \text{ is the set of paths in } G_2. \]

Question
What is \( d(G_1, G_2) \)?

Definition (Path-Based Distance)

\[
d_p(G_1, G_2) = \sup_{\alpha \in \Pi_1} \inf_{\beta \in \Pi_2} d_F(\alpha, \beta)
\]
Path-Based Method
Road Networks are Sets of Paths

\[ \Pi_1 \text{ is the set of paths in } G_1. \]
\[ \Pi_2 \text{ is the set of paths in } G_2. \]

Question
What is \( d(G_1, G_2) \)?

Definition (Path-Based Distance)

\[
d_p(G_1, G_2) = \sup_{\alpha \in \Pi_1} \inf_{\beta \in \Pi_2} d_F(\alpha, \beta)
\]
**Path-Based Method**

Road Networks are Sets of Paths

\[ \Pi_1 \] is the set of paths in \( G_1 \).

\[ \Pi_2 \] is the set of paths in \( G_2 \).

**Question**

What is \( d(G_1, G_2) \)?

**Definition (Path-Based Distance)**

\[
d_p(G_1, G_2) = \sup_{\alpha \in \Pi_1} \inf_{\beta \in \Pi_2} d_F(\alpha, \beta)
\]
Different Approaches


Homology and Relative Homology
Definitions by Pictures

\[ H_1(X) \]
Homology and Relative Homology
Definitions by Pictures

\[ H_1(X) \]
Homology and Relative Homology
Definitions by Pictures

\[ H_1(X) \]
Homology and Relative Homology
Definitions by Pictures

\[ H_1(X, A) \]
Homology and Relative Homology
Definitions by Pictures

\[ H_1(X, A) \]
Homology and Relative Homology

Definitions by Pictures

\[ H_1(X, A) \]
Local Homology

A Definition

Definition (Local Homology)

\[ H_1(\mathbb{X}, \mathbb{X} - x) = \lim_{r \to 0} H_1(\mathbb{X} \cap B_r(x), \mathbb{X} \cap \partial B_r(x)) \]
Local Homology

A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]
Local Homology
A Definition

Definition (Local Homology)

$$H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x))$$
Local Homology

A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]
Local Homology
A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]
Local Homology
A Definition

Definition (Local Homology)

$$H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x))$$
Local Homology
A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]
Local Homology
A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]
Local Homology

A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]
Local Homology

A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]
Local Homology

A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]
Local Homology
A Definition

Definition (Local Homology)

\[ H_1(X, X - x) = \lim_{r \to 0} H_1(X \cap B_r(x), X \cap \partial B_r(x)) \]

Theorem

\[ H_1(X \cap B_r(x), X \cap \partial B_r(x)) = H_1(X / X - B_r(x)) \]
Local Topology Approach [AFW]

We will use the embedding and the local homology to create a local distance signature.
Local Distance Signature
Finding the Local Persistence Diagram

$G_i \subset D$ is the road network.
$r > 0$ is the scale.
$x \in D$.

$L G_i(x, 0) = (G_i \cap B_r(x))/(G_1 \cap \partial B_r(x))$
Local Distance Signature
Finding the Local Persistence Diagram

\[ LG_i(x, \varepsilon) = \frac{(G_i^\varepsilon \cap B_r(x))}{(G_1^\varepsilon \cap \partial B_r(x))} \]

\( G_i \subset D \) is the road network.
\( r > 0 \) is the scale.
\( x \in D \).
Local Distance Signature
Finding the Local Persistence Diagram

\[ LG_i(x, 2\varepsilon) = \frac{G_i^{2\varepsilon} \cap B_r(x)}{G_1^{2\varepsilon} \cap \partial B_r(x)} \]

- \(G_i \subset D\) is the road network.
- \(r > 0\) is the scale.
- \(x \in D\).
Local Distance Signature

Finding the Local Persistence Diagram

\[ f_{1,x,r} : B_r(x) / \partial B_r(x) \to \mathbb{R} \]

\[ f_{1,x,r}(y) = d(y, G_i) \]
Local Distance Signature

Computing the Local Distance

\[ s(x, r) := W_\infty(Dgm_1(f_{1,x,r}), Dgm_1(f_{2,x,r})) \]
Local Distance Signature

Computing the Local Distance

\[ s(x, r) := W_\infty(\text{Dgm}_1(f_1, x, r), \text{Dgm}_1(f_2, x, r)) \]
Local Topology Based Distance

Definition (Local Homology Distance)

\[ d_{LH}(G_1, G_2) = \frac{1}{|D|} \int_D s(x, r) \, dx \]
Definition (Local Homology Distance)

\[ d_{LH}(G_1, G_2) = \frac{1}{|D|} \int_{r_0}^{r_1} \int_D s(x, r) \, dx \, dr \]
Local Topology Based Distance

Definition (Local Homology Distance)

\[
d_{LH}(G_1, G_2) = \frac{1}{|D|} \int_{r_0}^{r_1} \omega_r \int_D \eta_x s(x, r) \, dx \, dr
\]
Local Topology Based Distance

Definition (Local Homology Distance)

\[ d_{LH}(G_1, G_2) = \frac{1}{|D|} \int_{r_0}^{r_1} \omega_r \int_D \eta_x s(x, r) \, dx \, dr \]

Properties

1. \( r_0 = 0 \implies \) metric
2. \( d_{LH}(G_1, G_2) \leq \|f_1 - f_2\|_\infty \)
Properties

- $s(\cdot, r)$ is 1-Lipschitz
Properties

- \( s(\cdot, r) \) is 1-Lipschitz \( \Rightarrow \) can estimate with a finite cover
Properties

- $s(\cdot, r)$ is 1-Lipschitz $\implies$ can estimate with a finite cover
- Error bounded by (a function of) the min distance between ball centers.
Computation

Using the Nerve

\[
d(e_i, e_j)
\]
boundary-boundary:
\[
d(b_i, b_j) = 0
\]
boundary-edge:
\[
d(e_i, b_j)
\]
2-Skeleton
\[
e-e-e:
\]
\[
b-b-b:
\]
\[
b-e-e:
\]
\[
b-b-e:
\]

B. Fasy (Tulane)
Nerve
Nerve
Nerve
Nerve Lemma

If $U$ is a collection of sets such that all intersections are either empty or contractible, then $H(U) = H(N(U))$. 
Computation
Using the Nerve
Local Homology Distance

Computation
Using the Nerve
Computation

Using the Nerve
Computation
Using the Nerve
Computation

Using the Nerve
Local Homology Distance

Computation
Using the Nerve
Computation
Using the Nerve
Computation
Using the Nerve

1-Skeleton

edge-edge: \( d(e_i, e_j) \)
boundary-boundary: \( d(b_i, b_j) = 0 \)
boundary-edge: \( d(e_i, b_j) \)
Computation
Using the Nerve

1-Skeleton
edge-edge: $d(e_i, e_j)$
boundary-boundary: $d(b_i, b_j) = 0$
boundary-edge: $d(e_i, b_j)$

2-Skeleton
e-e-e: $d(e_i, e_j, e_k)$
b-b-b: $d(b_i, b_j, b_k) = 0$
b-e-e: $d(b_i, e_j, e_k)$
b-b-e: $d(b_i, b_j, e_k)$
Dataset 1: Athens

www.mapconstruction.org

- 129 GPS trajectories from school buses.
- \( D = 2.6 \text{ km} \times 6 \text{ km} \)
- Two reconstruction algorithms: [BE-12a] and [KP-12].
- Trajectories: 13-47 samples
Dataset 1: Athens

www.mapconstruction.org
Dataset 1: Athens

www.mapconstruction.org
Dataset 1: Athens

www.mapconstruction.org
Results

Athens: Comparing Two Different Reconstructions
Results

Athens: Comparing Two Different Reconstructions
Ongoing

- Apply this to different construction algorithms versus the ground truth to rank algorithms.
- Fast Implementation of distance measure (currently, estimating the distance).
- Improve theoretical guarantees.
- Input Model: other noise models?
- Output Model: road category, direction, intersection regions, ...
Summary

Map Comparison

[AFW-14] Local Homology Based Distance Between Maps.

- Provide a local distance signature.
- ... that is computable.
- ... and extends to a distance metric between graphs.
On Another Note ...

Joint Work with F. Chazal, F. Lecci, A. Rinaldo, L. Wasserman

- Confidence sets for persistence diagrams: $(\hat{P}, \delta)$
- Confidence bands for functional summaries of persistence diagrams.
- Hypothesis testing in TDA.
Thank You!

Contact me: brittany.fasy@alumni.duke.edu
www.mapconstruction.org, www.fasy.us