Metric Graph Reconstruction from Noisy Data

M. Aanjaneya, F. Chazal, D. Chen, M. Glisse, L. Guibas, D. Morozov
• Branching filamentary structures (metric graphs) appear in a wide variety of real world data sets.

• Data possibly not embedded in Euclidean space and only coming with local metric information → metric spaces

Problem:
Can we recover the underlying metric graph structure from approximating data?
Metric space approximation

Let \((X, d_X), (Y, d_Y)\) be two metric spaces.

An \(\varepsilon\)-correspondence between \((X, d_X)\) and \((Y, d_Y)\) is a set \(C \subset X \times Y\) s. t. 

1. for any \(x \in X\) (resp. \(y \in Y\)), there exists \(y \in Y\) (resp. \(x \in X\)) s. t. \((x, y) \in C\).

2. For any \((x, y), (x', y') \in C\), \(|d_X(x, x') - d_Y(y, y')| \leq \varepsilon\).

The Gromov-Hausdorff distance \(d_{GH}(X, Y)\) is the infimum of the \(\varepsilon \geq 0\) such that there exists an \(\varepsilon\)-correspondence between \((X, d_X)\) and \((Y, d_Y)\).
Metric space approximation

Let \((X, d_X), (Y, d_Y)\) be two metric spaces.

\[ d_{GH}(X, Y) \geq \text{diam}(Y) - \text{diam}(X) \]
An \((\varepsilon, R)\)-correspondence between \((X, d_X)\) and \((Y, d_Y)\) is a set \(C \subset X \times Y\) s. t.

1. for any \(x \in X\) (resp. \(y \in Y\)), there exists \(y \in Y\) (resp. \(x \in X\)) s. t. \((x, y) \in C\).

2. For any \((x, y), (x', y') \in C\) s. t. \(\min(d_X(x, x'), d_Y(y, y')) \leq R\),

\[
|d_X(x, x') - d_Y(y, y')| \leq \varepsilon.
\]

\(\rightarrow (Y, d_Y)\) is an \((\varepsilon, R)\)-approximation of \((X, d_X)\).
Metric space approximation

Let \((X, d_X), (Y, d_Y)\) be two metric spaces.

- \(d_X\): geodesic distance
- \(d_Y\): restriction of the Euclidean distance

For any \(\varepsilon > 0\), the diagonal \(C = \{(x, x) : x \in X\} \subset X \times Y\) is an \((\varepsilon, O(\varepsilon^{1/3}))\)-correspondence.

\[
\text{diam}(X) = 2\pi, \text{diam}(Y) = 2 \Rightarrow d_{GH}(X, Y) \geq \pi - 2 > 0
\]

\[\Rightarrow (\varepsilon, R)\text{-correspondences do not bound Gromov-Hausdorff distance.}\]

But if \((X, d_X)\) is a length space (i.e. the distance between any pair of points is equal to the infimum of the lengths of the continuous curves joining them) and for any \(y, y' \in Y\) there exists a sequence \(y_0 = y, y_1, \ldots, y_{n-1}, y_n = y'\) such that for all \(i = 0, \ldots, n - 1\), \(d_Y(y_i, y_{i+1}) \leq R\) and \(d_Y(y, y') = \sum_{i=0}^{n-1} d_Y(y_i, y_{i+1})\) then

\[
|d_Y(y, y') - d_X(x, x')| \leq \left(\frac{\min(d_X(x, x'), d_Y(y, y'))}{R/2} + 1\right) \varepsilon.
\]
The metric graph reconstruction problem

A metric graph is a path metric space \((\mathbb{X}, d_\mathbb{X})\) that is homeomorphic to a 1-dimensional stratified space. A vertex of \(\mathbb{X}\) is a 0-dimensional stratum of \(\mathbb{X}\) and an edge of \(\mathbb{X}\) is a 1-dimensional stratum of \(\mathbb{X}\).

The distance between any pair of points is equal to the infimum of the lengths of the continuous curves joining them.
The metric graph reconstruction problem

A metric graph is a path metric space \((X, d_X)\) that is homeomorphic to a 1-dimensional stratified space. A vertex of \(X\) is a 0-dimensional stratum of \(X\) and an edge of \(X\) is a 1-dimensional stratum of \(X\).

**Input:** \((Y, d_Y)\) an \((\varepsilon, R)\)-approximation of a metric graph \((X, d_X)\).

**Goal:** build a metric graph \((\hat{X}, d_{\hat{X}})\) that is homeomorphic to \((X, d_X)\) and return a map \(\phi : Y \rightarrow \hat{X}\) that approximately preserves distances.
The metric graph reconstruction problem

A **metric graph** is a path metric space \((X, d_X)\) that is homeomorphic to a 1-dimensional stratified space. A **vertex** of \(X\) is a 0-dimensional stratum of \(X\) and an **edge** of \(X\) is a 1-dimensional stratum of \(X\).

\[
\text{length( } \bigcirc \text{ )} \approx d_Y(a, b)
\]
The main idea: degree inference

- The degree of a point $x$ on $X$ is the number of connected components of a sufficiently small (intrinsic) sphere centered at $x$.
- Vertices of $X$ are the points with degree 1 or larger than 2.
- The degree of (most) points can be inferred from $Y$ by looking at (intrinsic) spherical shells.

Use inferred degree to identify vertices of $X$ and reconstruct its edges.

Identify the points $y \in Y$ that are paired to vertices (resp. edge points) in $X$ for any $(\varepsilon, R)$-correspondence $C$ between $X$ and $Y$:

- The degree of a point $x$ on $X$ is the number of connected components of a sufficiently small (intrinsic) sphere centered at $x$.
- Vertices of $X$ are the points with degree 1 or larger than 2.
- The degree of (most) points can be inferred from $Y$ by looking at (intrinsic) spherical shells.
The main idea: degree inference

Given a metric space $\mathbb{M}$ and a real number $r > 0$, the Rips-Vietoris graph $R_r(\mathbb{M})$ is the graph with vertex set $\mathbb{M}$ and edges connecting all pairs of vertices at distance at most $r$.

Let $(\mathbb{Y}, d_\mathbb{Y})$ be an $(\varepsilon, R)$-approximation of $\mathbb{X}$. Given $0 < r < R/2$, the $r$-degree $\deg_r(y)$ of $y \in \mathbb{Y}$ is the number of connected components of the Rips-Vietoris graph $R_{4r/3}(B_\mathbb{Y}(y, 5r/3) \setminus B_\mathbb{Y}(y, r))$. 

Given $(\mathbb{Y}, d_\mathbb{Y}, \varepsilon, R)$ be an $(\varepsilon, R)$-approximation of $\mathbb{X}$. Given $0 < r < R/2$, the $r$-degree $\deg_r(y)$ of $y \in \mathbb{Y}$ is the number of connected components of the Rips-Vietoris graph $R_{4r/3}(B_\mathbb{Y}(y, 5r/3) \setminus B_\mathbb{Y}(y, r))$.
The main idea: degree inference

Degree Inference Theorem: Let \((Y, d_Y)\) be an \((\varepsilon, R)\)-approximation of \(X\). Let \(C \subset X \times Y\) be an \((\varepsilon, R)\)-correspondence between \(X\) and \(Y\), let \((x, y) \in C\).

i) If the distance \(d_0\) from \(x\) to any vertex of \(X\) is larger than \(\frac{17}{2}\varepsilon\), then for \(\frac{9}{2}\varepsilon < r < \min\left(\frac{R}{2}, \frac{3(d_0-\varepsilon)}{5}\right)\), \(\deg_r(y)\) is equal to the degree of \(x\) in \(X\) (i.e. 2).

ii) If \(x\) is at distance less than \(\varepsilon\) from a vertex \(x_0\) of \(X\) and if the length \(l_0\) of the shortest edge adjacent to \(x_0\) is larger than \(\frac{27}{2}\varepsilon\) then for \(\frac{15}{2}\varepsilon < r < \min\left(\frac{R}{2}, \frac{3(l_0-2\varepsilon)}{5}\right)\), \(\deg_r(y)\) is equal to the degree of \(x_0\) in \(X\).
The algorithm

Input: \((\mathbb{Y}, d_{\mathbb{Y}})\) approximating a metric graph \((\mathbb{X}, d_{\mathbb{X}})\) and parameter \(r > 0\).

Output: A metric graph \((\hat{\mathbb{X}}, d_{\hat{\mathbb{X}}})\).

1. Labelling points as edge or branch
   • for all \(y \in \mathbb{Y}\) do
   • if \(\text{deg}_r(y) = 2\) then label \(y\) as an edge point.
   • else label \(y\) as a branch point.
   • Label all points within distance \(2r\) from a preliminary branch point as branch points.
The algorithm

**Input:** \((Y, d_Y)\) approximating a metric graph \((X, d_X)\) and parameter \(r > 0\).

**Output:** A metric graph \((\hat{X}, d_{\hat{X}})\).

**Graph structure reconstruction**

- \(E \leftarrow\) points of \(Y\) labeled as edge points; \(V \leftarrow\) points of \(Y\) labeled as branch points.
- Compute the connected components of the Rips-Vietoris graphs \(\mathcal{R}_{2r}(E)\) and \(\mathcal{R}_{2r}(V)\).
- Vertices of \(\hat{X} \leftarrow\) connected components of \(\mathcal{R}_{2r}(V)\).
- Put an edge between vertices of \(\hat{X}\) if their corresponding components in \(\mathcal{R}_{2r}(V)\) contain points at distance less than \(2r\) from the same component of \(\mathcal{R}_{2r}(E)\).
The algorithm

Input: \((Y, d_Y)\) approximating a metric graph \((X, d_X)\) and parameter \(r > 0\).

Output: A metric graph \((\hat{X}, d_{\hat{X}})\).

Reconstructing the metric

- To each edge \(\hat{e}\) of \(\hat{X}\) assign a length equal to the diameter of the corresponding connected component of \(\mathcal{R}_{2r}(E)\) plus \(4r\).
Theoretical guarantees

Let $(Y, d_Y)$ be an $(\varepsilon, R)$-approximation of a metric graph $(X, d_X)$ for some $\varepsilon, R > 0$.

**Topological Reconstruction theorem:** If the length $b$ of the shortest edge of $X$ is larger than $16r$ and $15\varepsilon/2 < r < \min(R/4, 3(b - 2\varepsilon)/5)$ then the reconstructed graph $\hat{X}$ is homeomorphic to $X$. 
Theoretical guarantees

Let $(Y, d_Y)$ be an $(\varepsilon, R)$-approximation of a metric graph $(X, d_X)$ for some $\varepsilon, R > 0$.

**Topological Reconstruction theorem:** If the length $b$ of the shortest edge of $X$ is larger than $16r$ and $15\varepsilon/2 < r < \min(R/4, 3(b - 2\varepsilon)/5)$ then the reconstructed graph $\hat{X}$ is homeomorphic to $X$.

**Metric Reconstruction theorem:** Under the assumptions of the previous Theorem there exists a homeomorphism $\phi : X \to \hat{X}$ such that for any $x, x' \in X$, $(1 - \kappa)d_X(x, x') \leq d_{\hat{X}}(\phi(x), \phi(x')) \leq (1 + \kappa')d_X(x, x')$ with $\kappa = \frac{10r}{3b} + \left(\frac{5}{b} + \frac{2}{R}\right)\varepsilon$ and $\kappa' = \left(\frac{3}{b} + \frac{2}{R}\right)\varepsilon$. 
Theoretical guarantees

Let \((Y, d_Y)\) be an \((\varepsilon, R)\)-approximation of a metric graph \((X, d_X)\) for some \(\varepsilon, R > 0\).

**Topological Reconstruction theorem:** If the length \(b\) of the shortest edge of \(X\) is larger than \(16r\) and \(15\varepsilon/2 < r < \min(R/4, 3(b - 2\varepsilon)/5)\) then the reconstructed graph \(\hat{X}\) is homeomorphic to \(X\).

**Metric Reconstruction theorem:** Under the assumptions of the previous Theorem there exists a homeomorphism \(\phi : X \to \hat{X}\) such that for any \(x, x' \in X\), 
\[
(1 - \kappa)d_X(x, x') \leq d_{\hat{X}}(\phi(x), \phi(x')) \leq (1 + \kappa')d_X(x, x')
\]
with \(\kappa = \frac{10r}{3b} + \left(\frac{5}{b} + \frac{2}{R}\right)\varepsilon\) and \(\kappa' = \left(\frac{3}{b} + \frac{2}{R}\right)\varepsilon\).

**Theorem:** There exists a map \(\psi : Y \to \hat{X}\) such that for any \(y, y' \in Y\)
\[
(1 - \kappa)\left(1 - \frac{2\varepsilon}{R}\right)d_Y(y, y') - \varepsilon \leq d_{\hat{X}}(\psi(y), \psi(y')) \leq (1 + \kappa') \left(1 + \frac{2\varepsilon}{R}\right)d_Y(y, y') + \varepsilon
\]
with \(\kappa\) and \(\kappa'\) as in the Metric Reconstruction Theorem.
Experimental results

- Data: GPS traces (sampled curves along a road network).
- $\mathcal{Y}, d_\mathcal{Y}$: a neighborhood graph built on the sampled points with its intrinsic metric.
Experimental results

- Data: earthquakes epicenters (preprocessed to remove “outliers/noise”).
- \((\mathcal{Y}, d_\mathcal{Y})\): a neighborhood graph (Rips) with its intrinsic metric.
Experimental results

- Data: images connected according to their similarity.
Experimental results

- Data: images connected according to their similarity.
Conclusion:  

- A simple algorithm for metric graph reconstruction:  
  - coming with topological and metric guarantees,  
  - relying on intrinsic metric information (no need of coordinates).
Conclusion and open questions

Conclusion:

- A simple algorithm for metric graph reconstruction:
  - coming with topological and metric guarantees,
  - relying on intrinsic metric information (no need of coordinates).

Open questions / Future works:

- Choice of the parameter $r$: how to do it in practice? How to make $r$ dependent of the local metric structure of the data?
- Interpretation of the output of the algorithm when the sampling conditions are not fulfilled: scale dependent graph simplification?
- Extension to higher dimensions: local homology? (see a recent work by P. Bendich, S. Mukherjee and B. Wang) - no canonical stratification.