Edimburgh - July 3, 2012

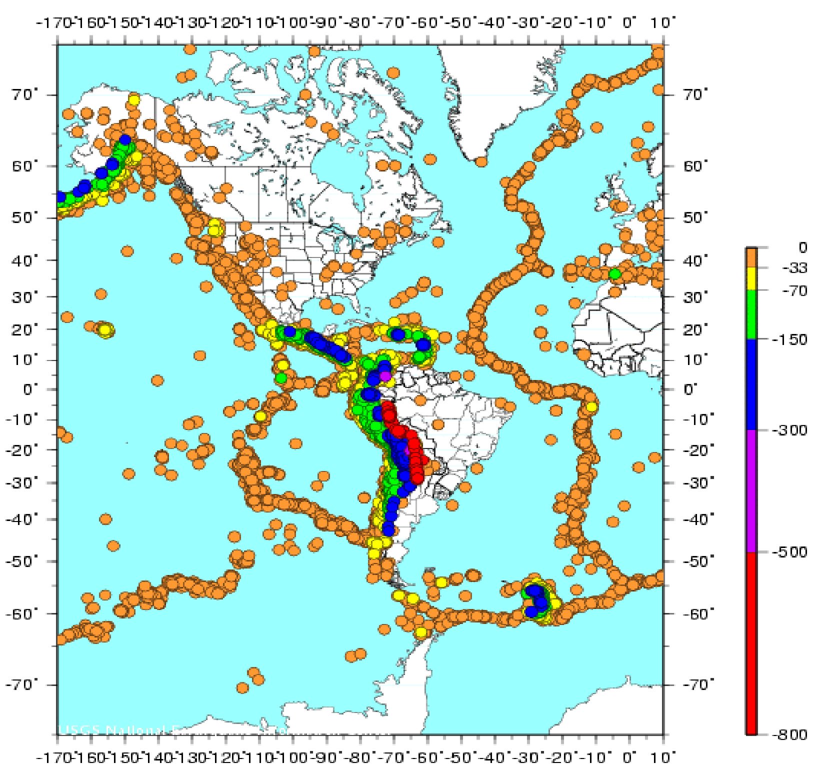
ATMCS 2012

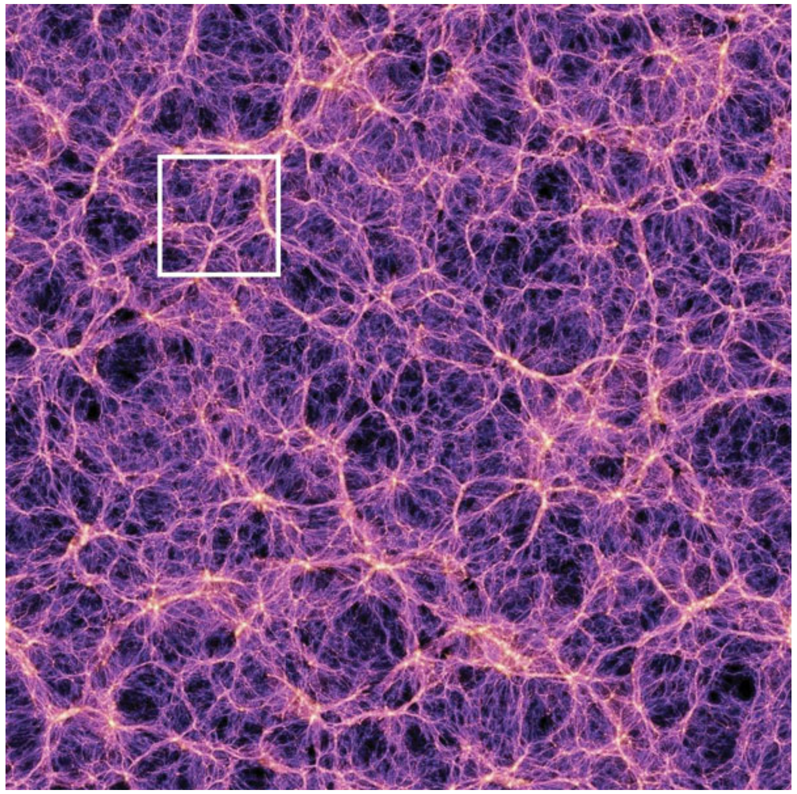
Detection and approximation of linear structures in metric spaces

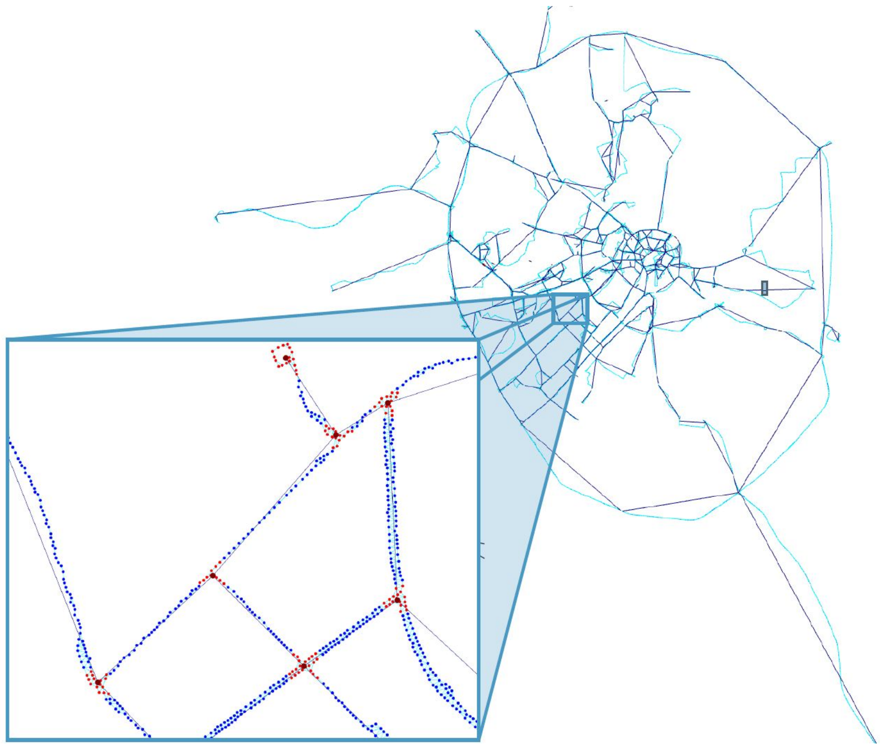
Frédéric Chazal
Geometrica group
INRIA Saclay

Joint work with M. Aanjaneya, D. Chen, M. Glisse, L. Guibas, D. Morozov and on-going work with Jian Sun

Introduction







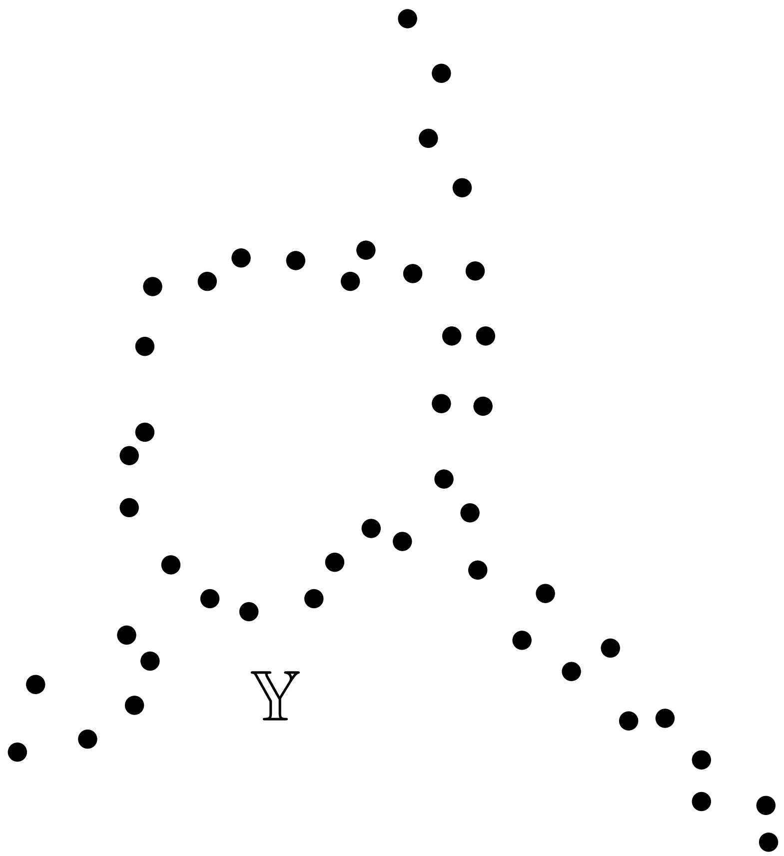
- Branching filamentary structures (*metric graphs*) appear in a wide of real world data sets.
- Data possibly not embedded in Euclidean space and only coming with (loca) pairwise distance information $\rightarrow$ metric spaces

Problem:

Can we recover the underlying metric graph structure from approximating data?

Problem statement

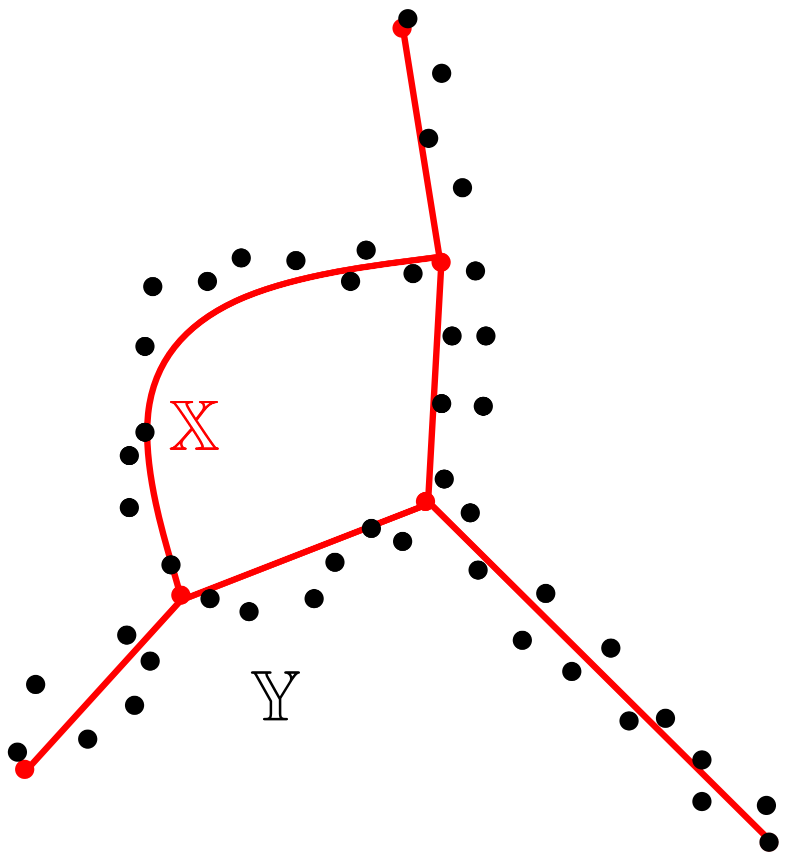
Input: a metric space $(\mathbb{Y}, d_{\mathbb{Y}})$ (e.g. a point cloud with distance matrix).



Problem statement

Input: a metric space $(\mathbb{Y}, d_{\mathbb{Y}})$ (e.g. a point cloud with distance matrix).

Promise: the input is close to some unknown *metric graph* $\mathbb{X}$



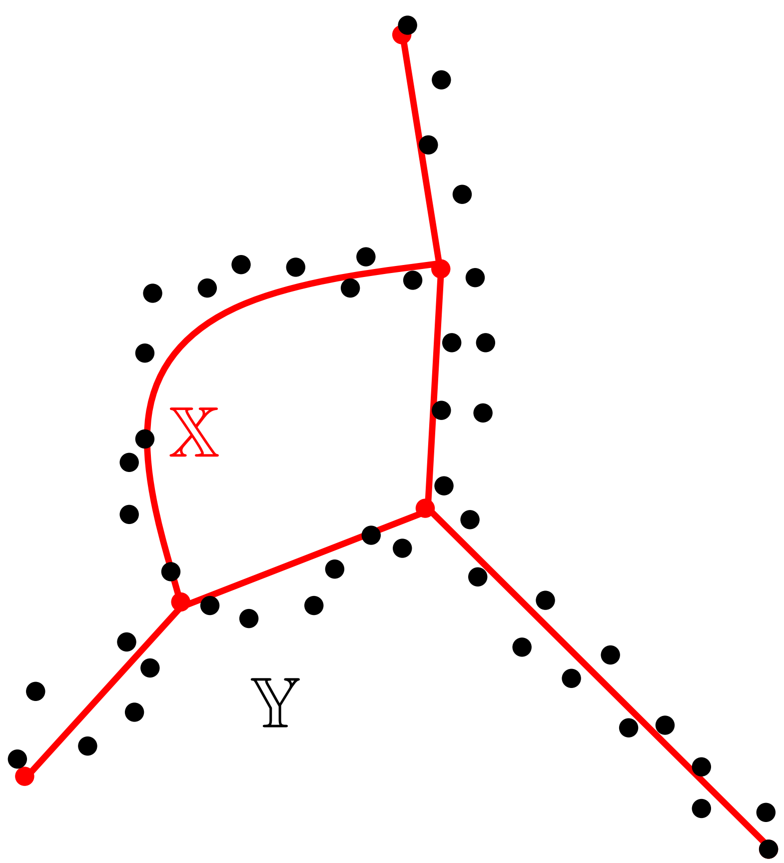
Problem statement

Input: a metric space $(\mathbb{Y}, d_{\mathbb{Y}})$ (e.g. a point cloud with distance matrix).

Promise: the input is close to some unknown *metric graph* $\mathbb{X}$

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



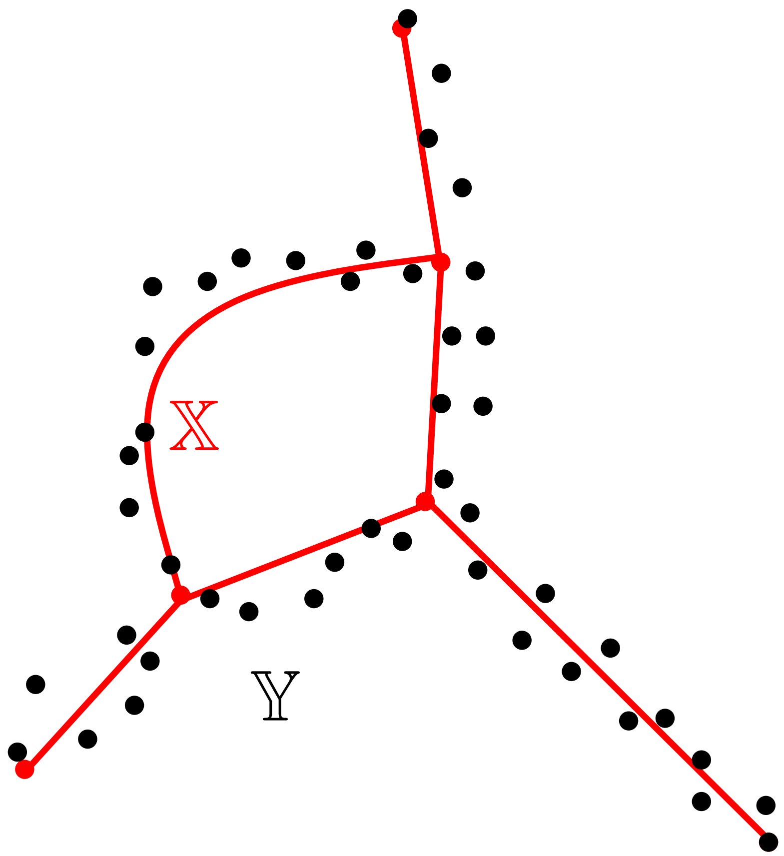
Problem statement

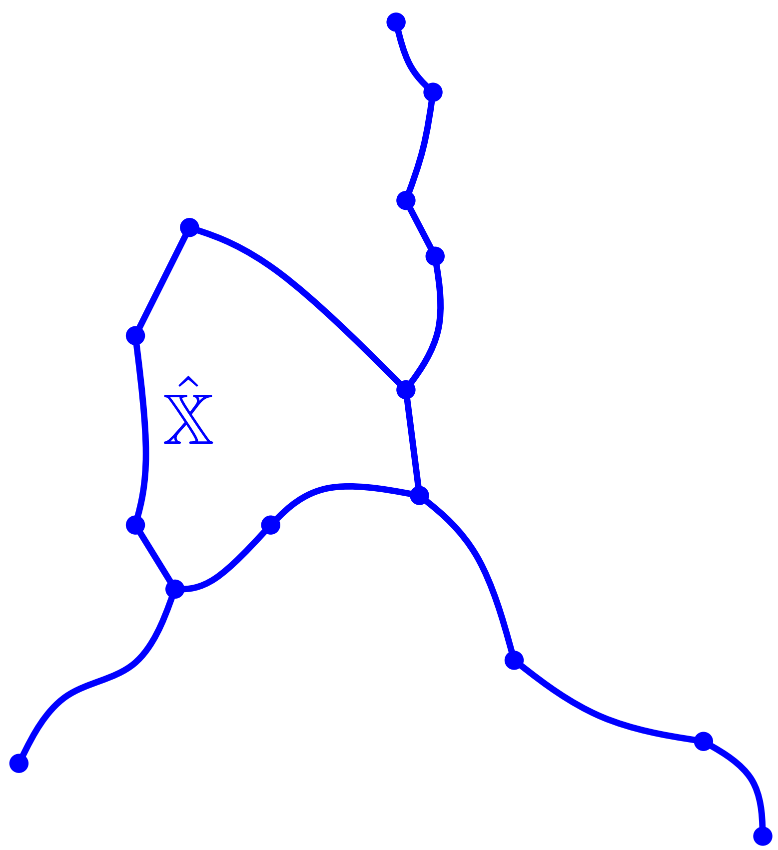
Input: a metric space $(\mathbb{Y}, d_\mathbb{Y})$ (e.g. a point cloud with distance matrix).

Promise: the input is close to some unknown *metric graph* $\mathbb{X}$

Output:

- a metric graph $(\hat{\mathbb{X}}, d_{\hat{\mathbb{X}}})$ that is close and if possible homeomorphic to $\mathbb{X}$





Problem statement

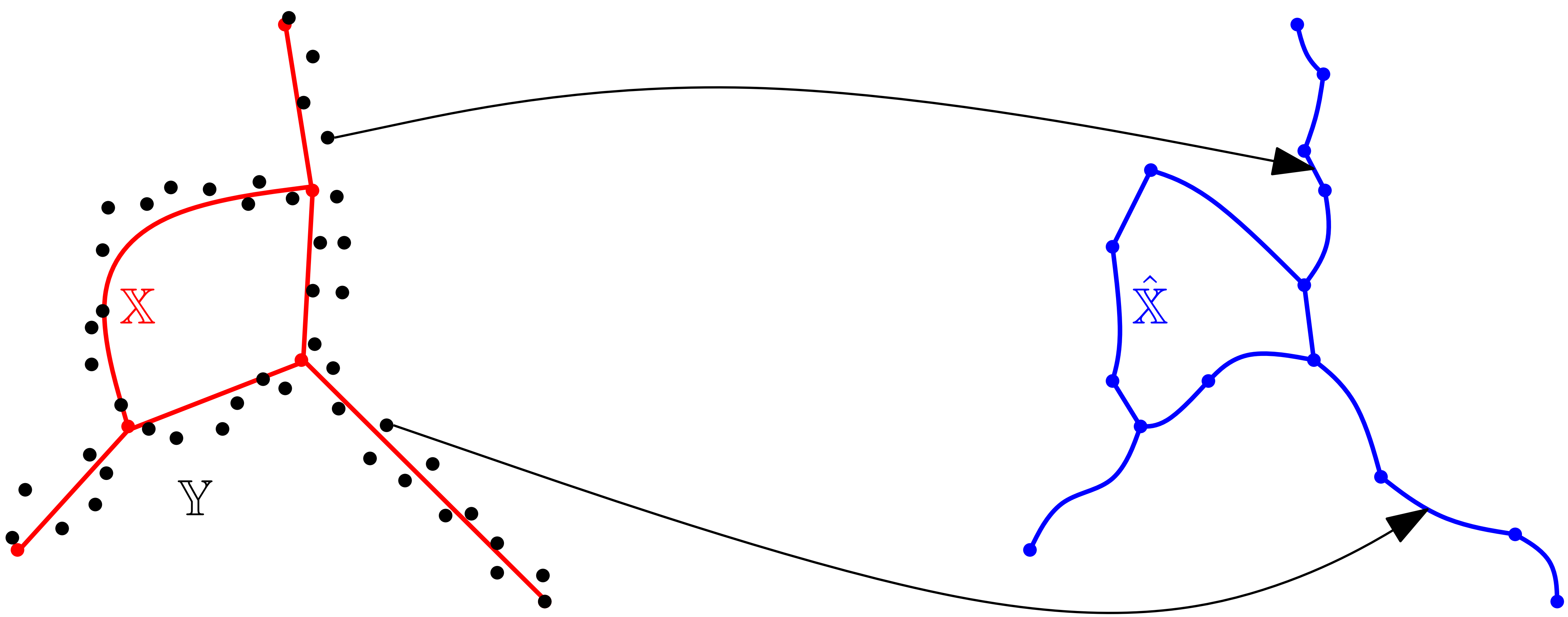
Input: a metric space $(\mathbb{Y}, d_{\mathbb{Y}})$ (e.g. a point cloud with distance matrix).

Promise: the input is close to some unknown *metric graph* $\mathbb{X}$

Output:

- a metric graph $(\hat{\mathbb{X}}, d_{\hat{\mathbb{X}}})$ that is close and if possible homeomorphic to $\mathbb{X}$

- a map $\mathbb{Y} \to \hat{\mathbb{X}}$ that roughly preserves distances



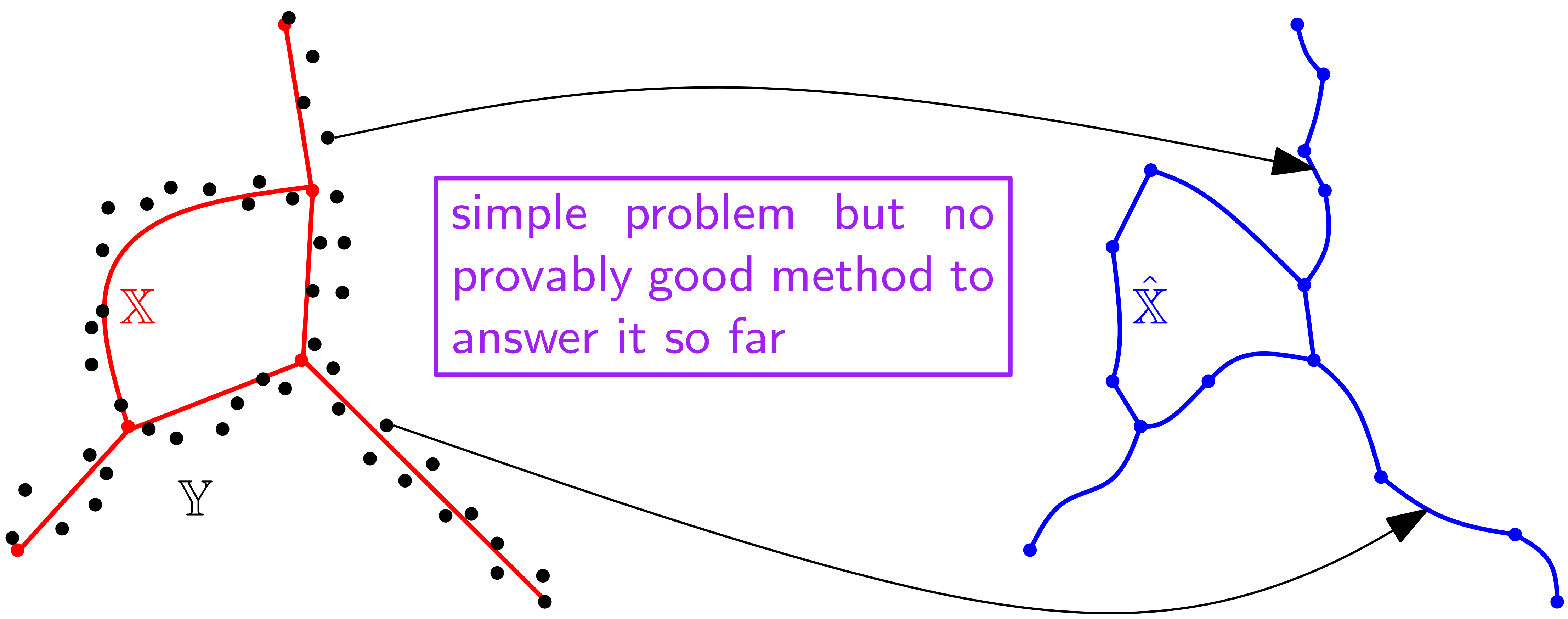
Problem statement

Input: a metric space $(\mathbb{Y}, d_{\mathbb{Y}})$ (e.g. a point cloud with distance matrix).

Promise: the input is close to some unknown *metric graph* $\mathbb{X}$

Output:

- a metric graph $(\hat{\mathbb{X}}, d_{\hat{\mathbb{X}}})$ that is close and if possible homeomorphic to $\mathbb{X}$
- a map $\mathbb{Y} \to \hat{\mathbb{X}}$ that roughly preserves distances



Problem statement

Input: a metric space $(\mathbb{Y}, d_{\mathbb{Y}})$ (e.g. a point cloud with distance matrix).

Promise: the input is close to some unknown *metric graph* $\mathbb{X}$

Output:

- a metric graph $(\hat{\mathbb{X}}, d_{\hat{\mathbb{X}}})$ that is close and if possible homeomorphic to $\mathbb{X}$

- a map $\mathbb{Y} \to \hat{\mathbb{X}}$ that roughly preserves distances

In this talk:

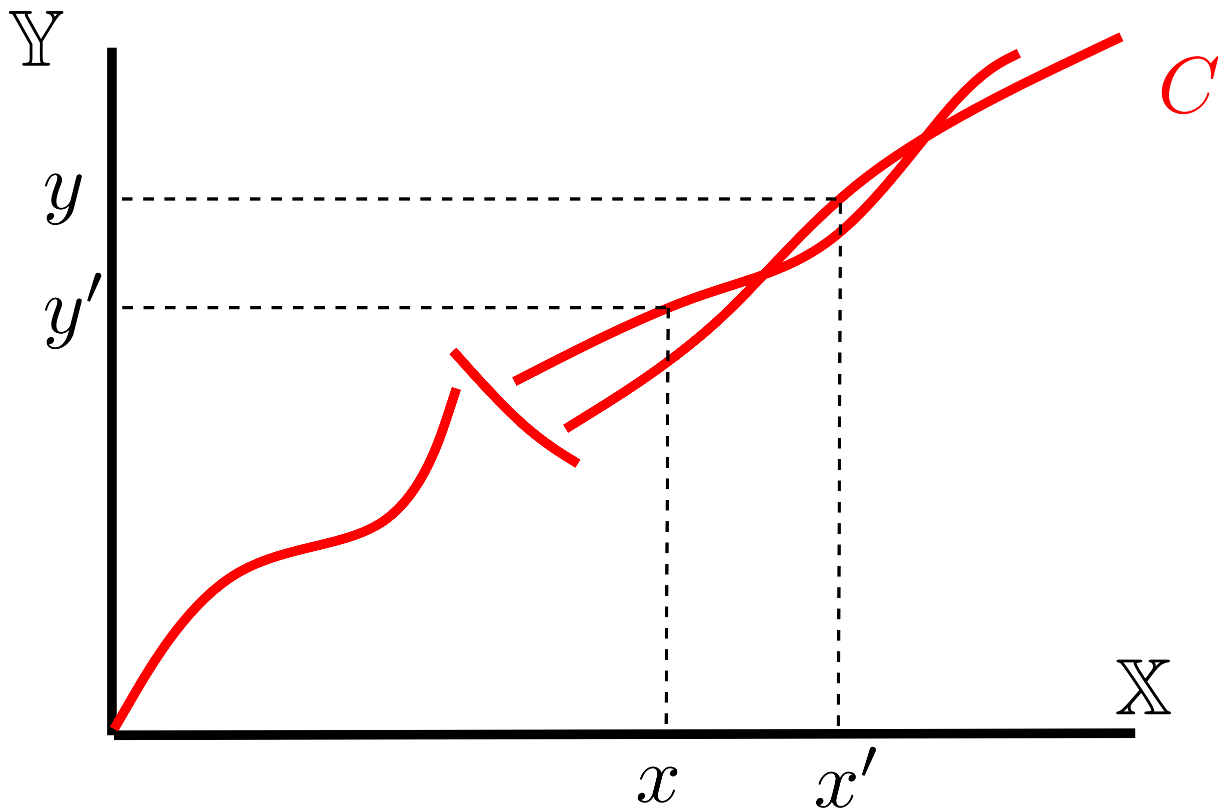
- A few basic topological and geometric ideas to address this problem.

- Topology guaranteed graph reconstruction based upon degree inference.

- "Linear structure" detection and tree reconstruction through metric hyperbolic geometry.

Metric spaces approximation

Let $(\mathbb{X}, d_{\mathbb{X}})$, $(\mathbb{Y}, d_{\mathbb{Y}})$ be two metric spaces.



An *ε-correspondence* between $(\mathbb{X}, d_{\mathbb{X}})$ and $(\mathbb{Y}, d_{\mathbb{Y}})$ is a set $C \subset \mathbb{X} \times \mathbb{Y}$ s. t.

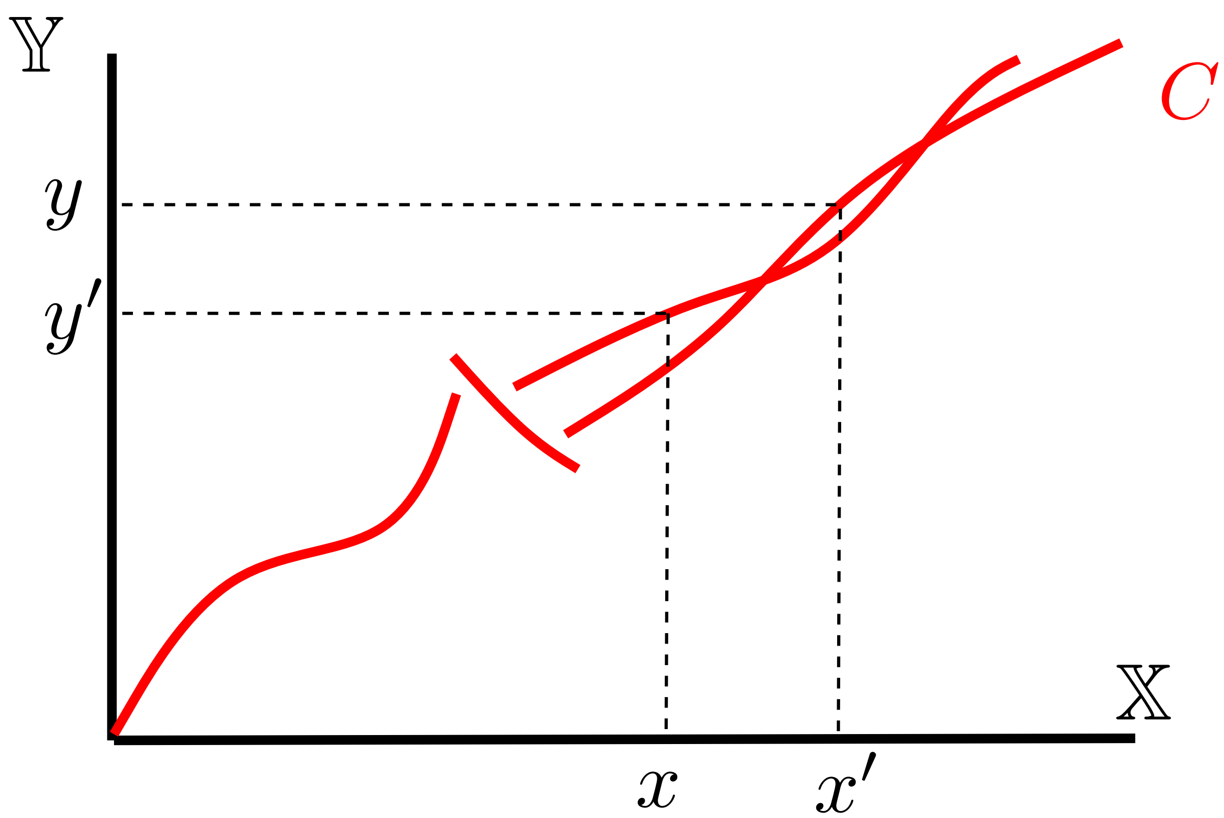
1. for any $x \in \mathbb{X}$ (resp. $y \in \mathbb{Y}$), there exists $y \in \mathbb{Y}$ (resp. $x \in \mathbb{X}$) s. t. $(x, y) \in C$.
2. For any $(x, y), (x', y') \in C$, $|d_{\mathbb{X}}(x, x') - d_{\mathbb{Y}}(y, y')| \leq \varepsilon$.

The *Gromov-Hausdorff distance*:

$$d_{GH}(\mathbb{X}, \mathbb{Y}) = \frac{1}{2} \inf\{\varepsilon > 0 : \text{there exists an } \varepsilon\text{-correspondence between} X \text{and} Y\}$$

Metric spaces approximation

Let $(\mathbb{X}, d_{\mathbb{X}})$, $(\mathbb{Y}, d_{\mathbb{Y}})$ be two metric spaces.



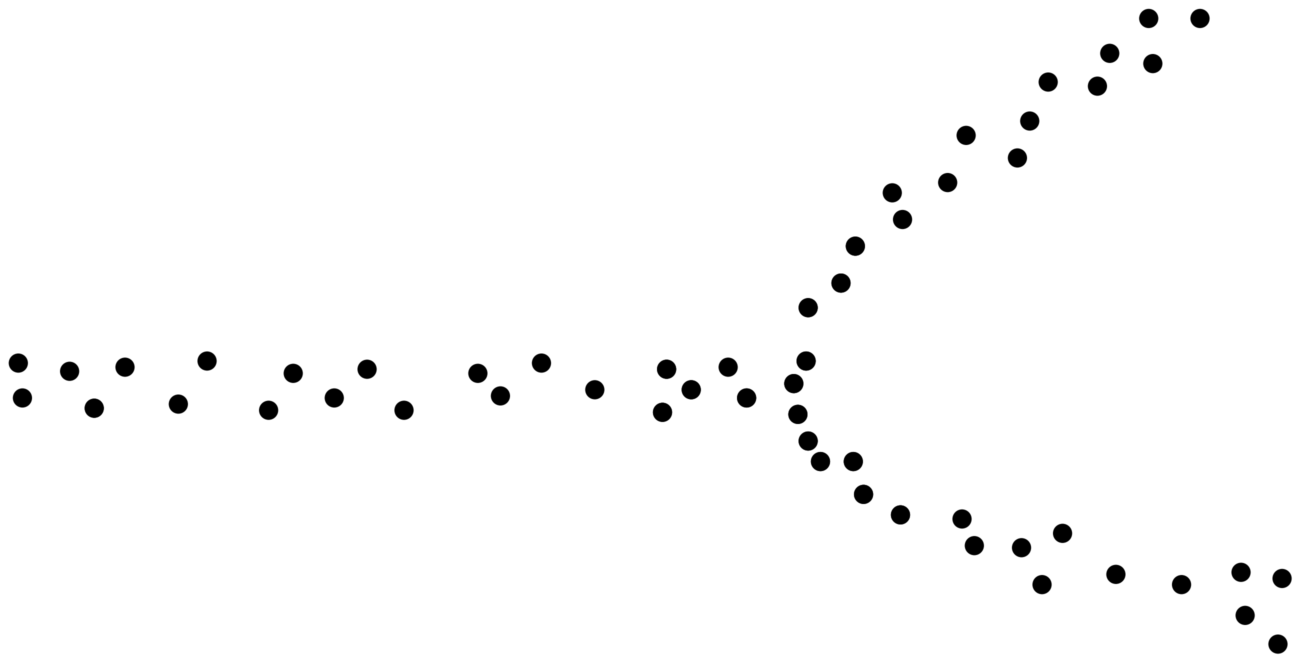
An *(ε, R)-correspondence* between $(\mathbb{X}, d_{\mathbb{X}})$ and $(\mathbb{Y}, d_{\mathbb{Y}})$ is a set $C \subset \mathbb{X} \times \mathbb{Y}$ s. t.

1. for any $x \in \mathbb{X}$ (resp. $y \in \mathbb{Y}$), there exists $y \in \mathbb{Y}$ (resp. $x \in \mathbb{X}$) s. t. $(x, y) \in C$.
2. For any $(x, y), (x', y') \in C$ s. t. $\min(d_{\mathbb{X}}(x, x'), d_{\mathbb{Y}}(y, y')) \leq R$, $|d_{\mathbb{X}}(x, x') - d_{\mathbb{Y}}(y, y')| \leq \varepsilon$.

$\rightarrow$ $(\mathbb{Y}, d_{\mathbb{Y}})$ is an (ε, R)-*approximation* of $(\mathbb{X}, d_{\mathbb{X}})$.

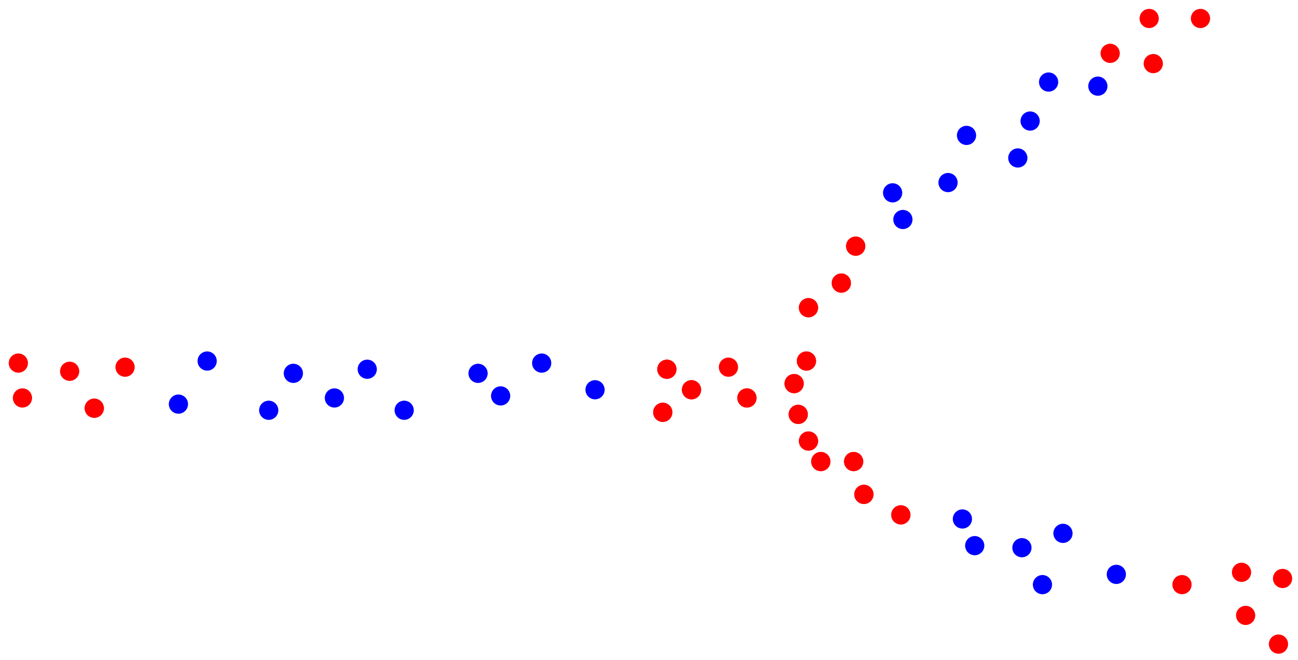
A first approach using the local topology of data

[Aanjaneya, C., Chen, Glisse, Guibas, Morozov, SoCG'11, IJCGA (2012)]



A first approach using the local topology of data

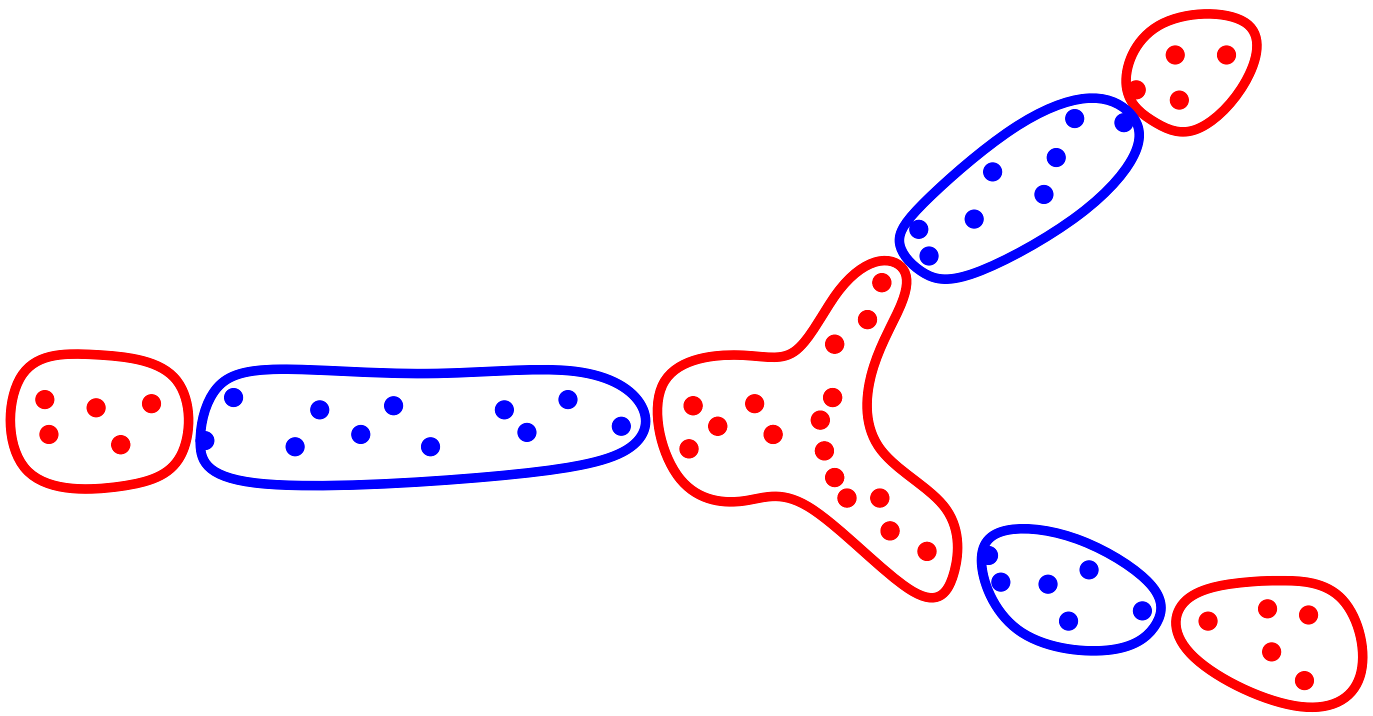
[Aanjaneya, C., Chen, Glisse, Guibas, Morozov, SoCG'11, IJCGA (2012)]



1. label each data point as edge point or vertex point

A first approach using the local topology of data

[Aanjaneya, C., Chen, Glisse, Guibas, Morozov, SoCG'11, IJCGA (2012)]

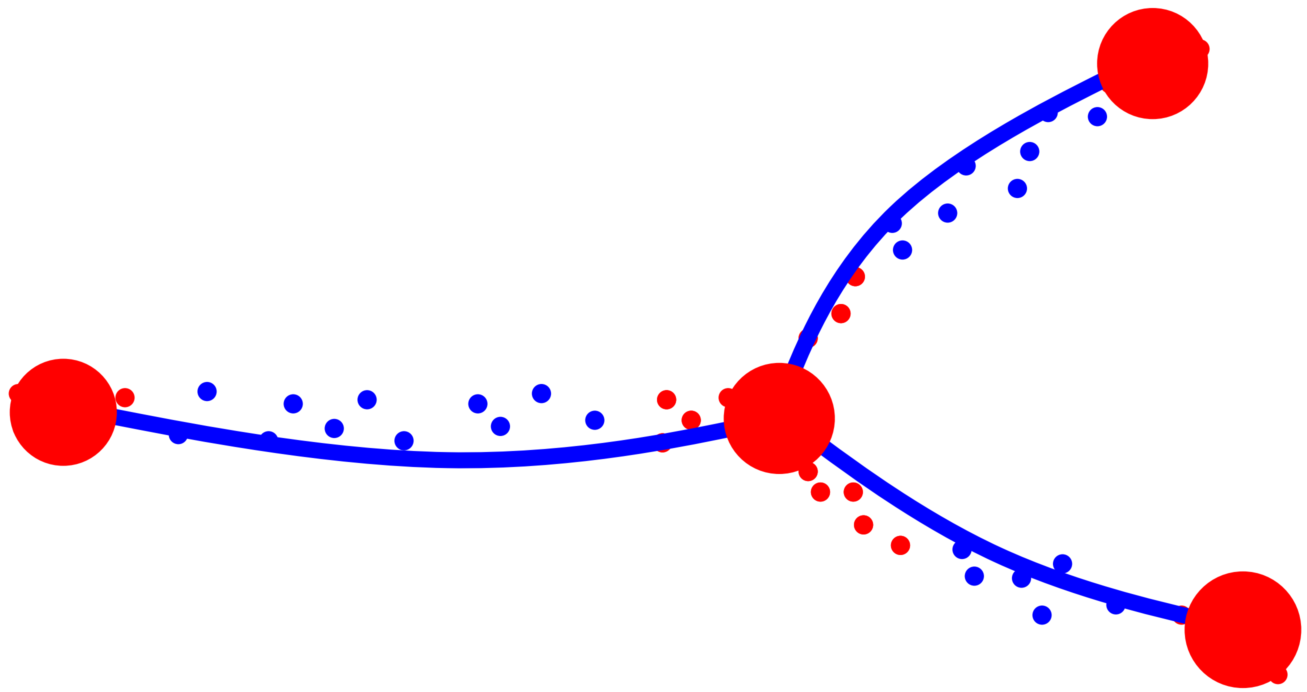


1. label each data point as edge point or vertex point

2. partition data points into edge clusters and vertex clusters

A first approach using the local topology of data

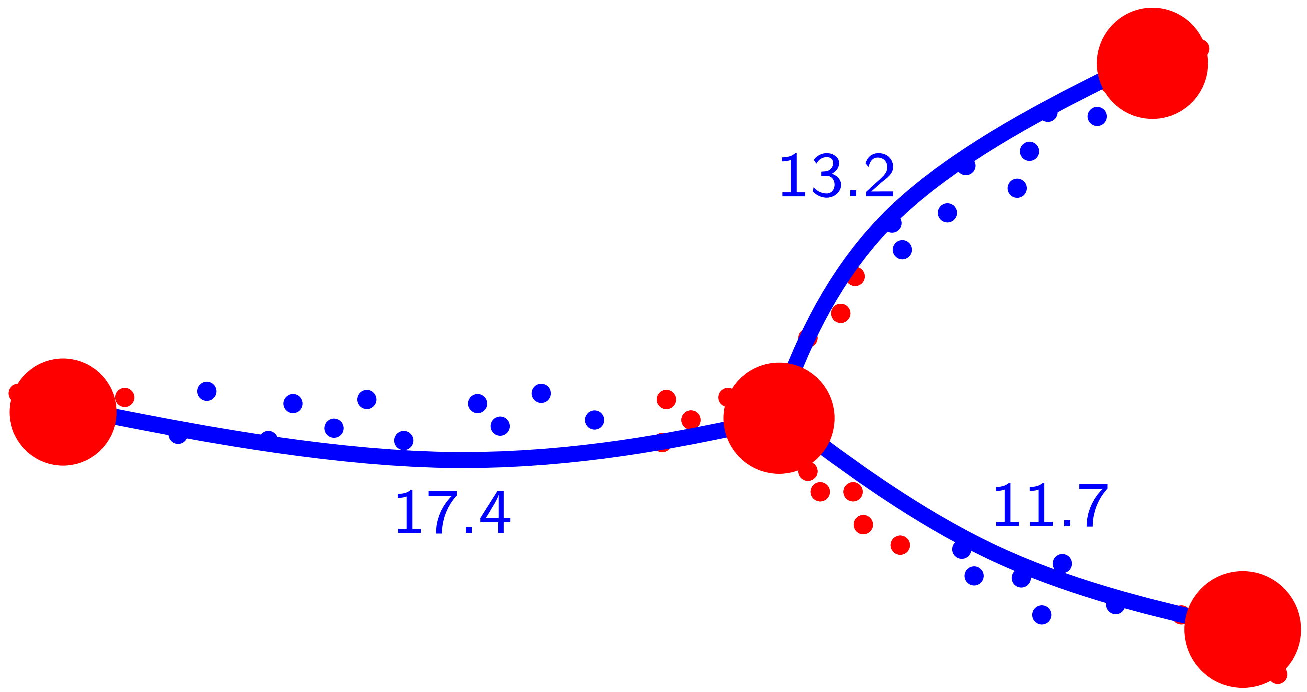
[Aanjaneya, C., Chen, Glisse, Guibas, Morozov, SoCG'11, IJCGA (2012)]



1. label each data point as edge point or vertex point
2. partition data points into edge clusters and vertex clusters
3. reconstruct graph structure

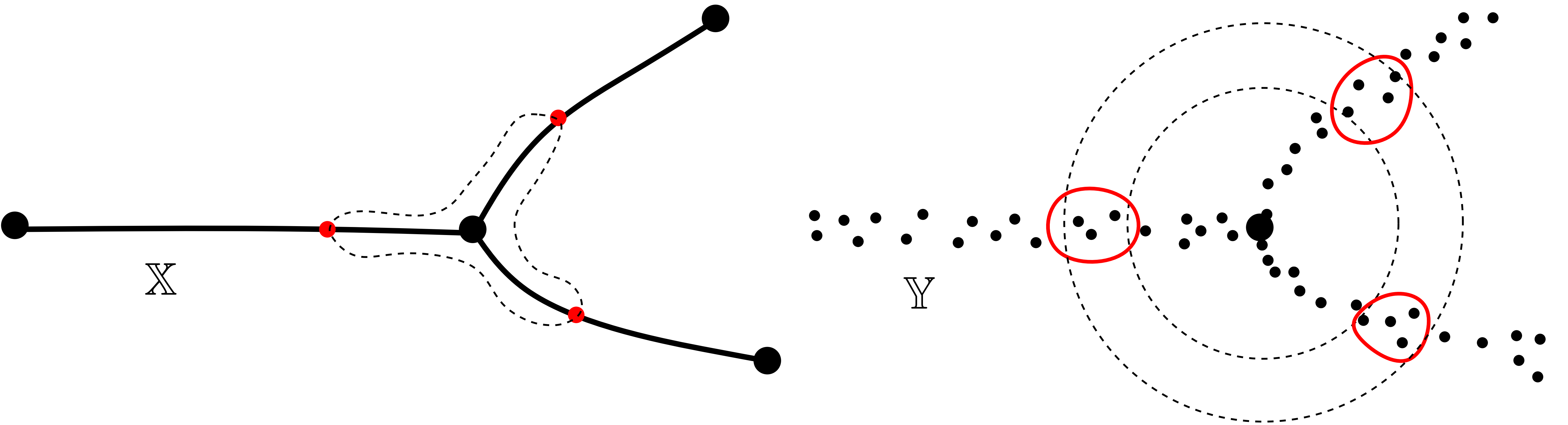
A first approach using the local topology of data

[Aanjaneya, C., Chen, Glisse, Guibas, Morozov, SoCG'11, IJCGA (2012)]



1. label each data point as edge point or vertex point
2. partition data points into edge clusters and vertex clusters
3. reconstruct graph structure
4. reconstruct metric (and mapping)

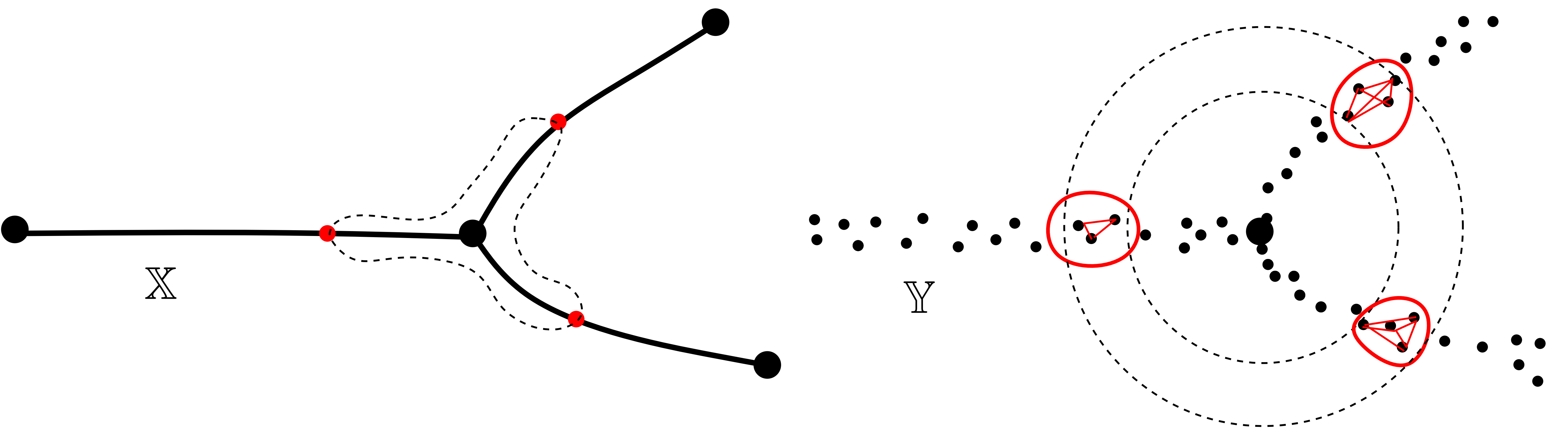
The main idea: degree inference



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

Use inferred degree to identify vertices of $\mathbb{X}$ and reconstruct its edges.

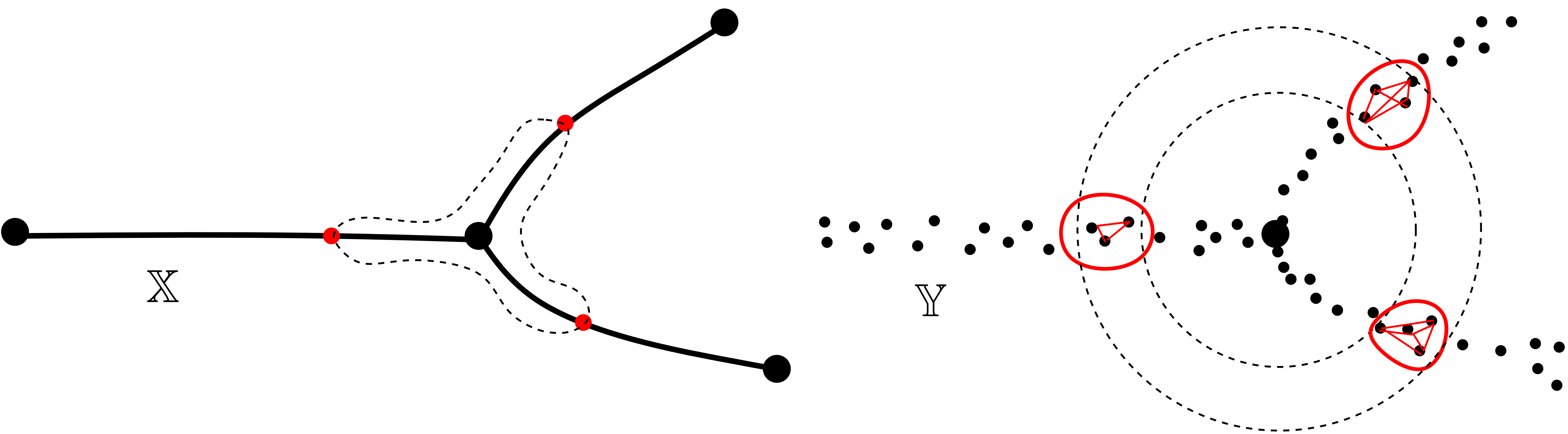
The main idea: degree inference



Given a metric space $\mathbb{M}$ and a real number $r > 0$, the Rips-Vietoris graph $\mathcal{R}_r(\mathbb{M})$ 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 (ε, 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 $\mathcal{R}_{4r/3}(\mathrm{B}_{\mathbb{Y}}(y, 5r/3) \setminus \mathrm{B}_{\mathbb{Y}}(y, r))$

The main idea: degree inference



Degree Inference Theorem: Let $(\mathbb{Y}, d_{\mathbb{Y}})$ be an (ε, R)-approximation of $\mathbb{X}$. Let $C \subset \mathbb{X} \times \mathbb{Y}$ be an (ε, R)-correspondence between $\mathbb{X}$ and $\mathbb{Y}$, let $(x, y) \in C$.

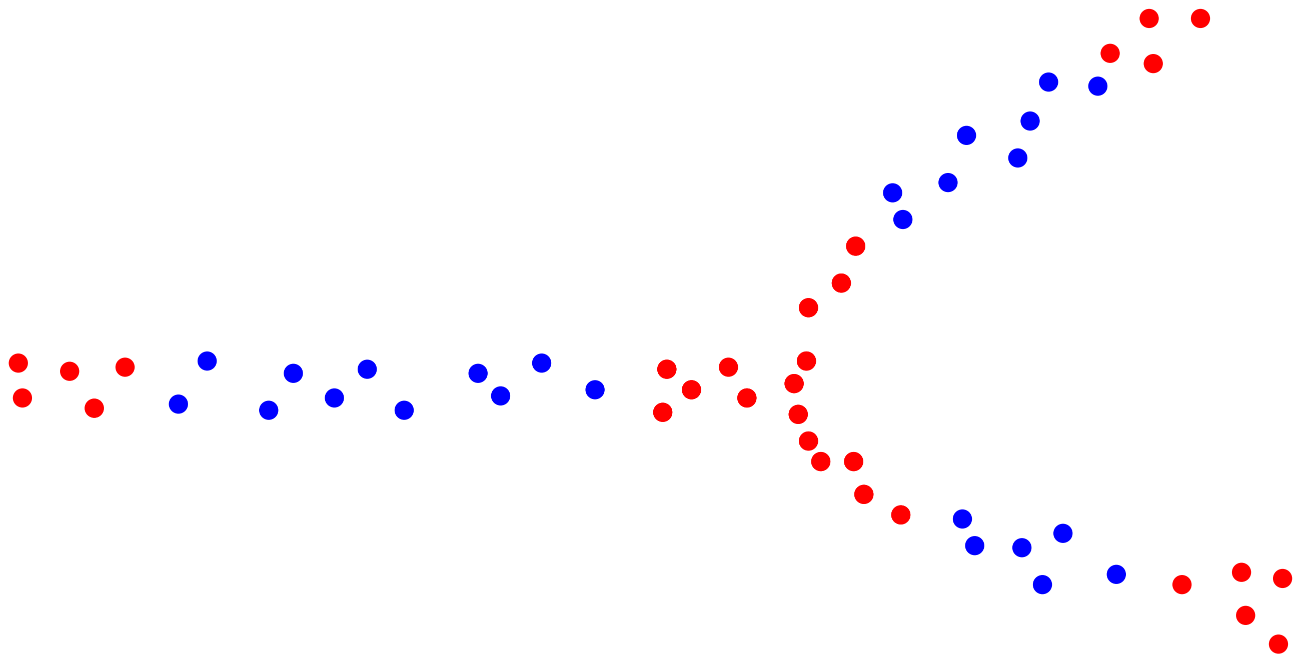
i) If the distance d_0 from x to any vertex of $\mathbb{X}$ is larger than $\frac{17}{2}\varepsilon$, then for $\frac{9}{2}\varepsilon < r < \min(\frac{R}{2}, \frac{3(d_0-\varepsilon)}{5})$, $deg_r(y)$ is equal to the degree of x in $\mathbb{X}$ (i.e. 2).

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

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 $(\hat{\mathbb{X}}, d_{\hat{\mathbb{X}}})$.



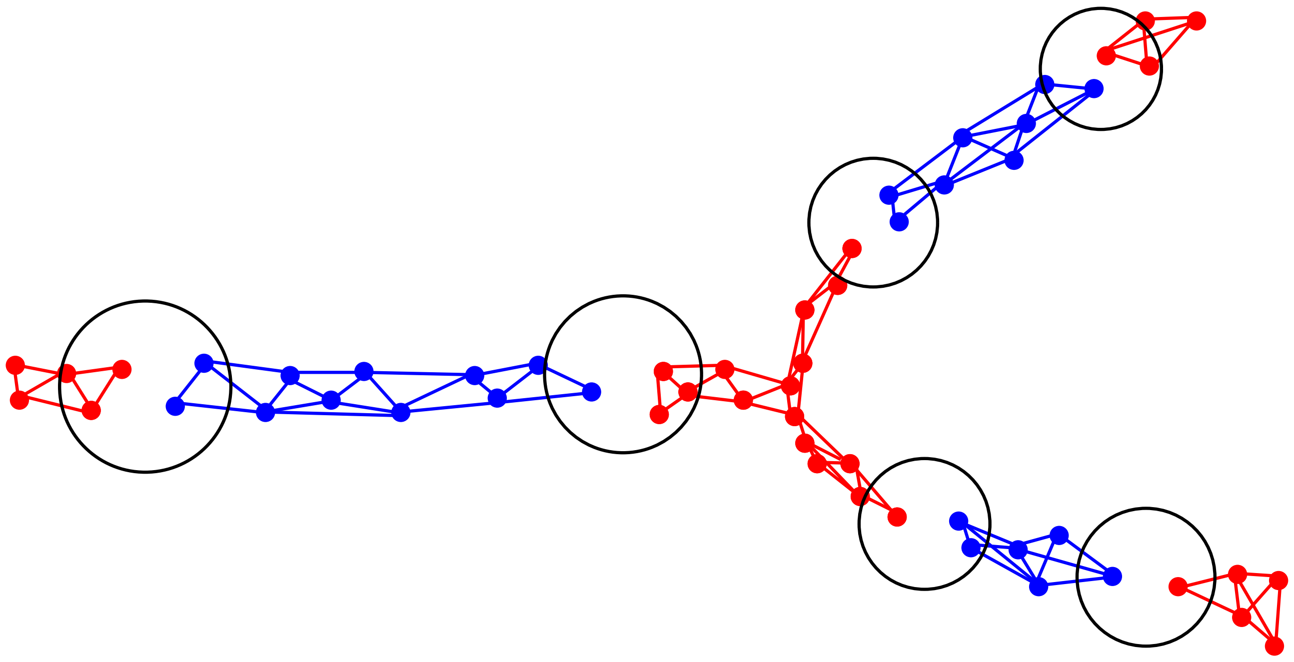
1. Labelling points as edge or branch

- **for all** $y \in \mathbb{Y}$ do
- **if** $deg_r(y) = 2$ **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: $(\mathbb{Y}, d_{\mathbb{Y}})$ approximating a metric graph $(\mathbb{X}, d_{\mathbb{X}})$ and parameter $r > 0$.

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



Graph structure reconstruction

• $\mathbb{E} \leftarrow$ points of $\mathbb{Y}$ labeled as edge points; $\mathbb{V} \leftarrow$ points of $\mathbb{Y}$ labeled as branch points.

• Compute the connected components of the Rips-Vietoris graphs $\mathcal{R}_{2r}(\mathbb{E})$ and $\mathcal{R}_{2r}(\mathbb{V})$.

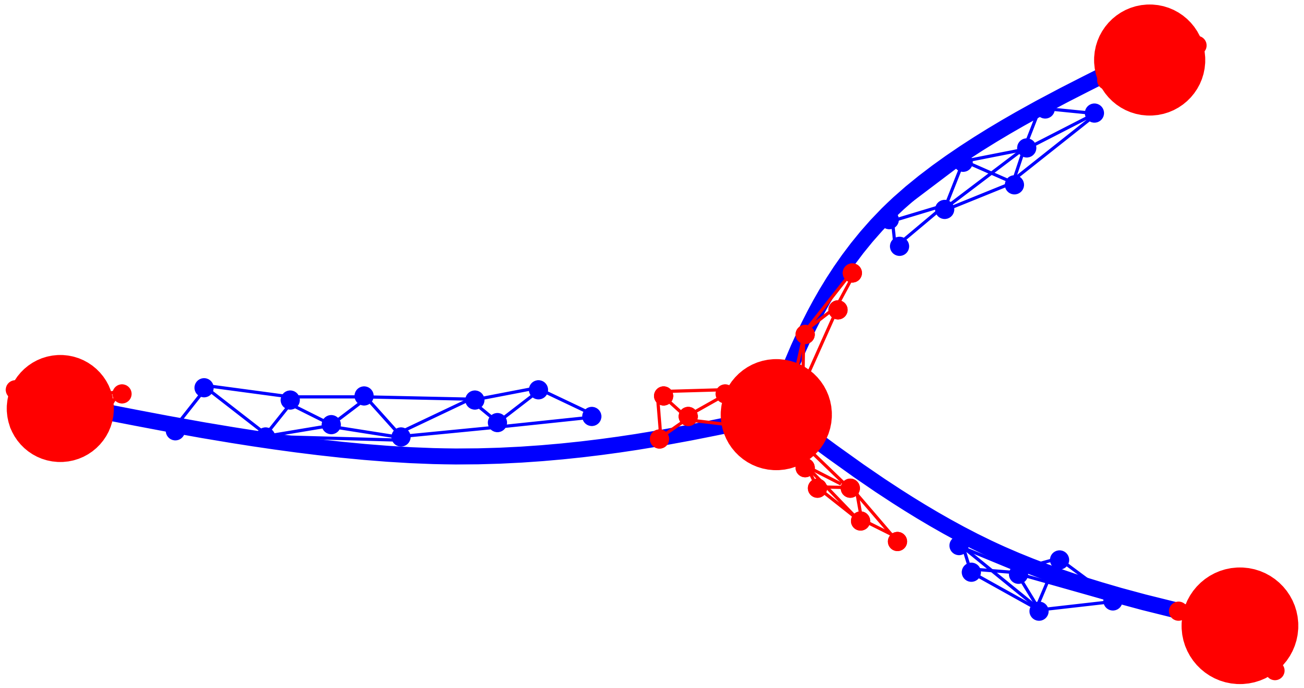
• Vertices of $\hat{\mathbb{X}} \leftarrow$ connected components of $\mathcal{R}_{2r}(\mathbb{V})$.

• Put an edge between vertices of $\hat{\mathbb{X}}$ if their corresponding components in $\mathcal{R}_{2r}(\mathbb{V})$ contain points at distance less than $2r$ from the same component of $\mathcal{R}_{2r}(\mathbb{E})$.

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 $(\hat{\mathbb{X}}, d_{\hat{\mathbb{X}}})$.



Reconstructing the metric

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

Theoretical guarantees

Let $(\mathbb{Y}, d_{\mathbb{Y}})$ be an (ε, R)-approximation of a metric graph $(\mathbb{X}, d_{\mathbb{X}})$ for some $\varepsilon, R > 0$.

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

Theoretical guarantees

Let $(\mathbb{Y}, d_{\mathbb{Y}})$ be an (ε, R)-approximation of a metric graph $(\mathbb{X}, d_{\mathbb{X}})$ for some $\varepsilon, R > 0$.

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

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

Theoretical guarantees

Let $(\mathbb{Y}, d_{\mathbb{Y}})$ be an (ε, R)-approximation of a metric graph $(\mathbb{X}, d_{\mathbb{X}})$ for some $\varepsilon, R > 0$.

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

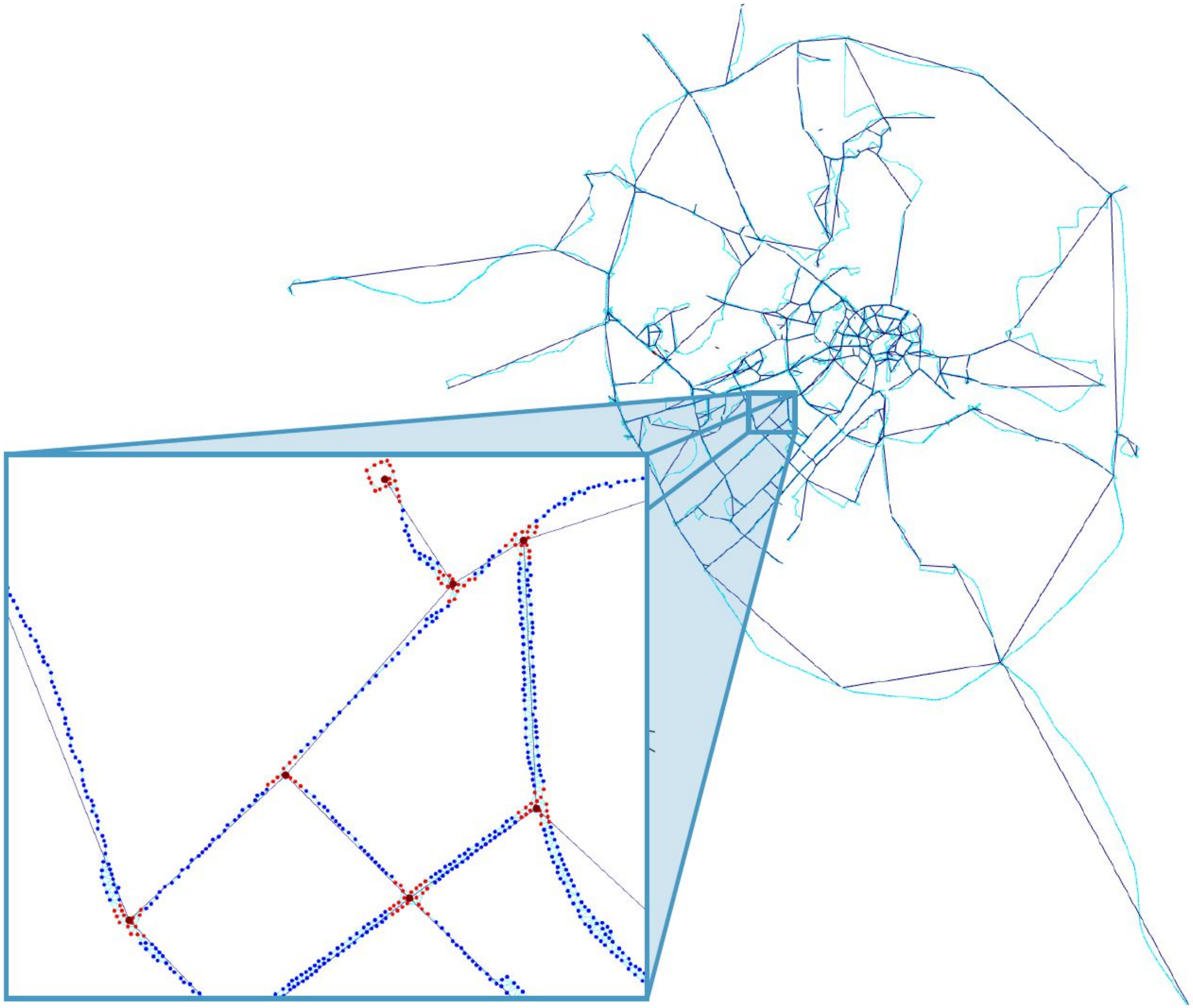
Metric Reconstruction theorem: Under the assumptions of the previous Theorem there exists a homeomorphism $\phi : \mathbb{X} \to \hat{\mathbb{X}}$ such that for any $x, x' \in \mathbb{X}$, $(1-\kappa)d_{\mathbb{X}}(x,x') \leq d_{\hat{\mathbb{X}}}(\phi(x), \phi(x')) \leq (1+\kappa')d_{\mathbb{X}}(x,x')$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

Theorem: There exists a map $\psi : \mathbb{Y} \to \hat{\mathbb{X}}$ such that for any $y, y' \in \mathbb{Y}$

$$(1-\kappa)\left((1-\frac{2\varepsilon}{R})d_{\mathbb{Y}}(y,y') - \varepsilon\right) \leq d_{\hat{\mathbb{X}}}(\psi(y), \psi(y')) \leq (1+\kappa')\left((1+\frac{2\varepsilon}{R})d_{\mathbb{Y}}(y,y') + \varepsilon\right)$$

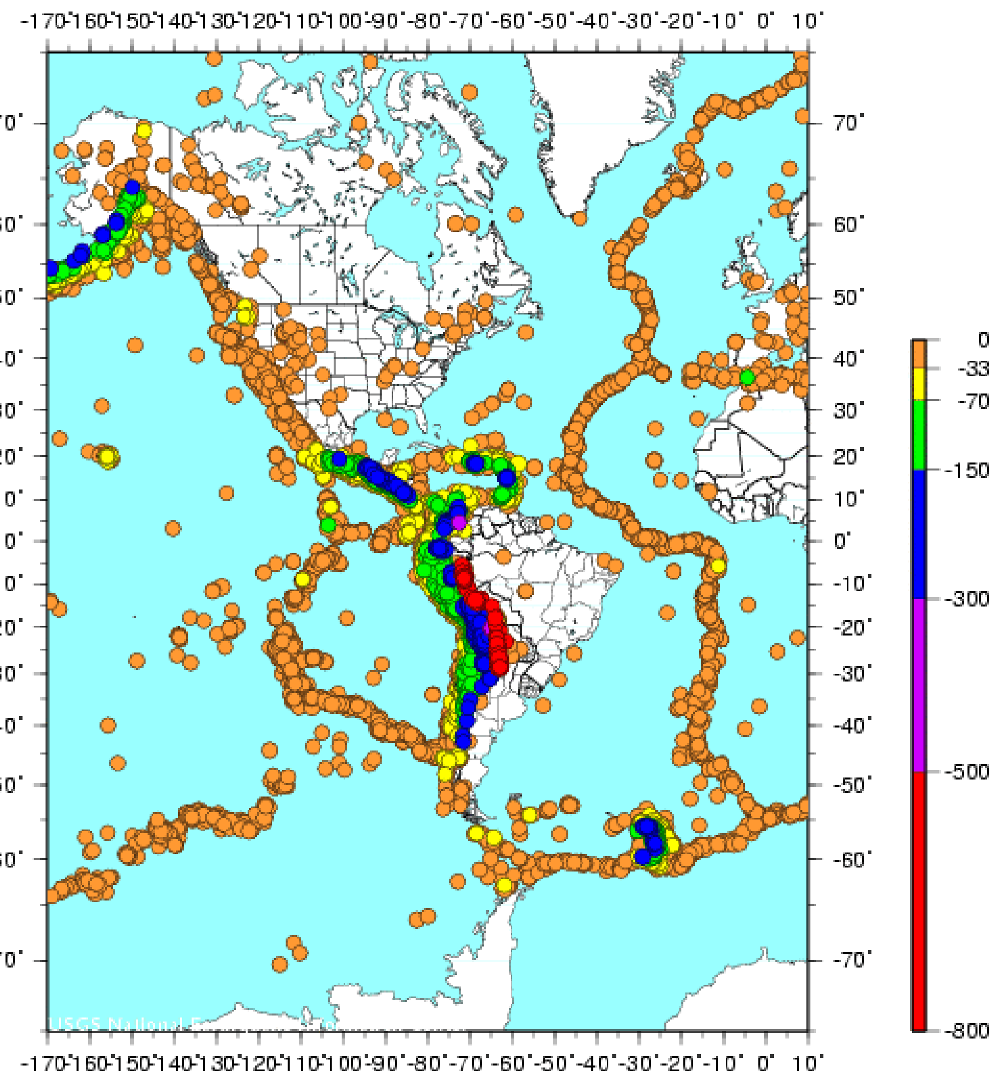
with κ and κ' as in the Metric Reconstruction Theorem.

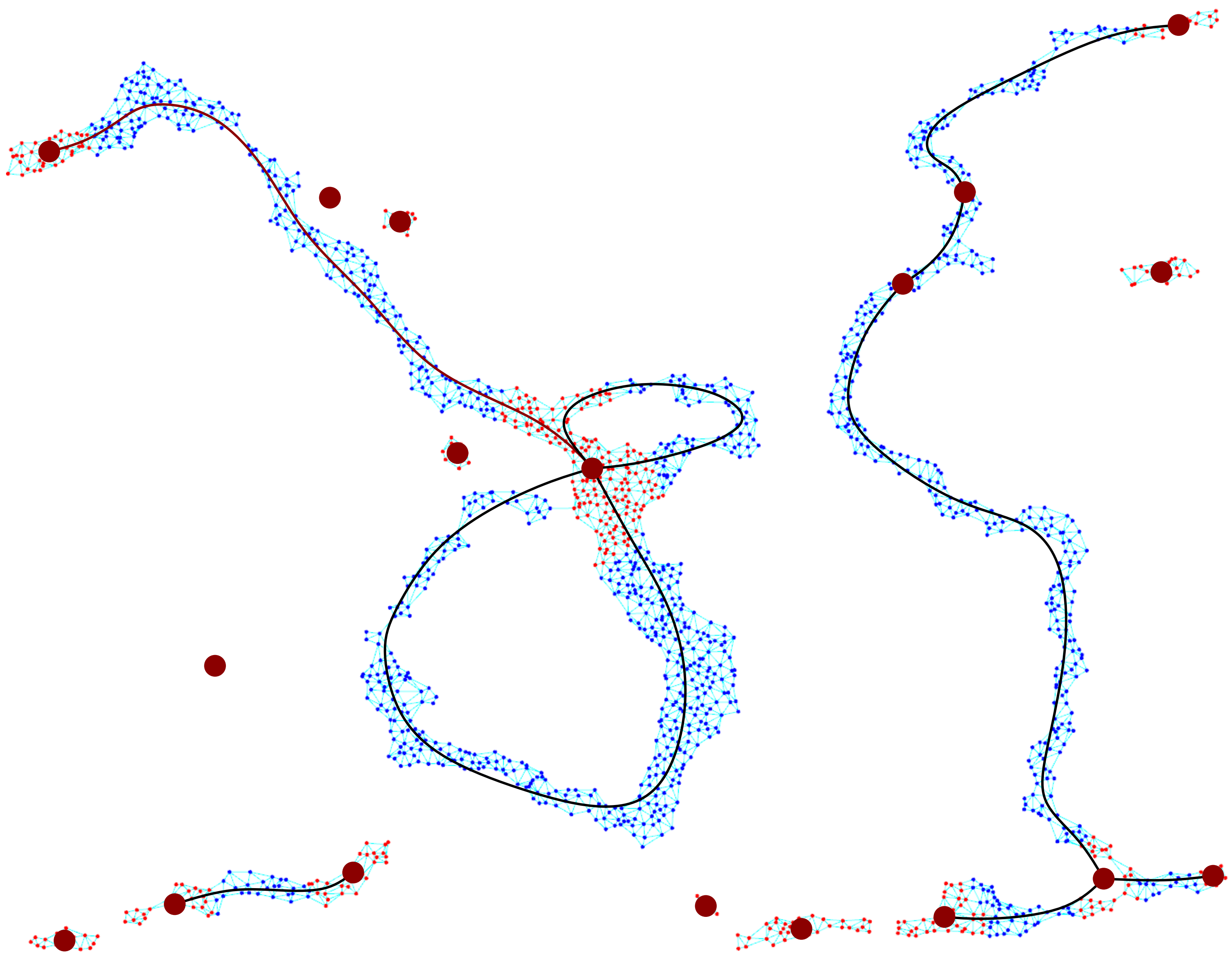
Experimental results



- Data: GPS traces (sampled curves along a road network).
- $(\mathbb{Y}, d_{\mathbb{Y}})$: a neighborhood graph (Rips) built on the data set with its intrinsic metric.

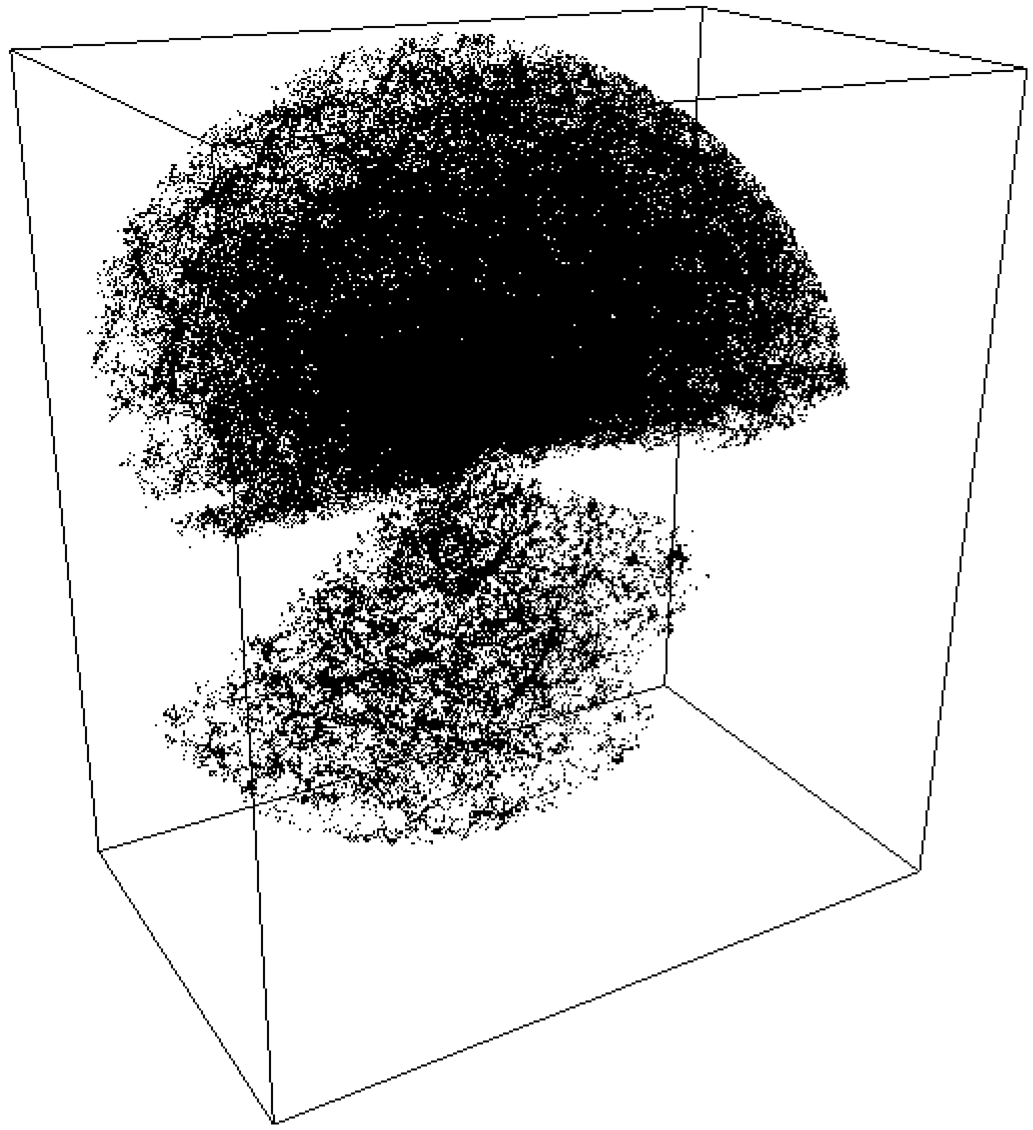
Experimental results





- Data: earthquakes epicenters (preprocessed to remove "outliers/noise").
- $(\mathbb{Y}, d_{\mathbb{Y}})$: a neighborhood graph (Rips) with its intrinsic metric.

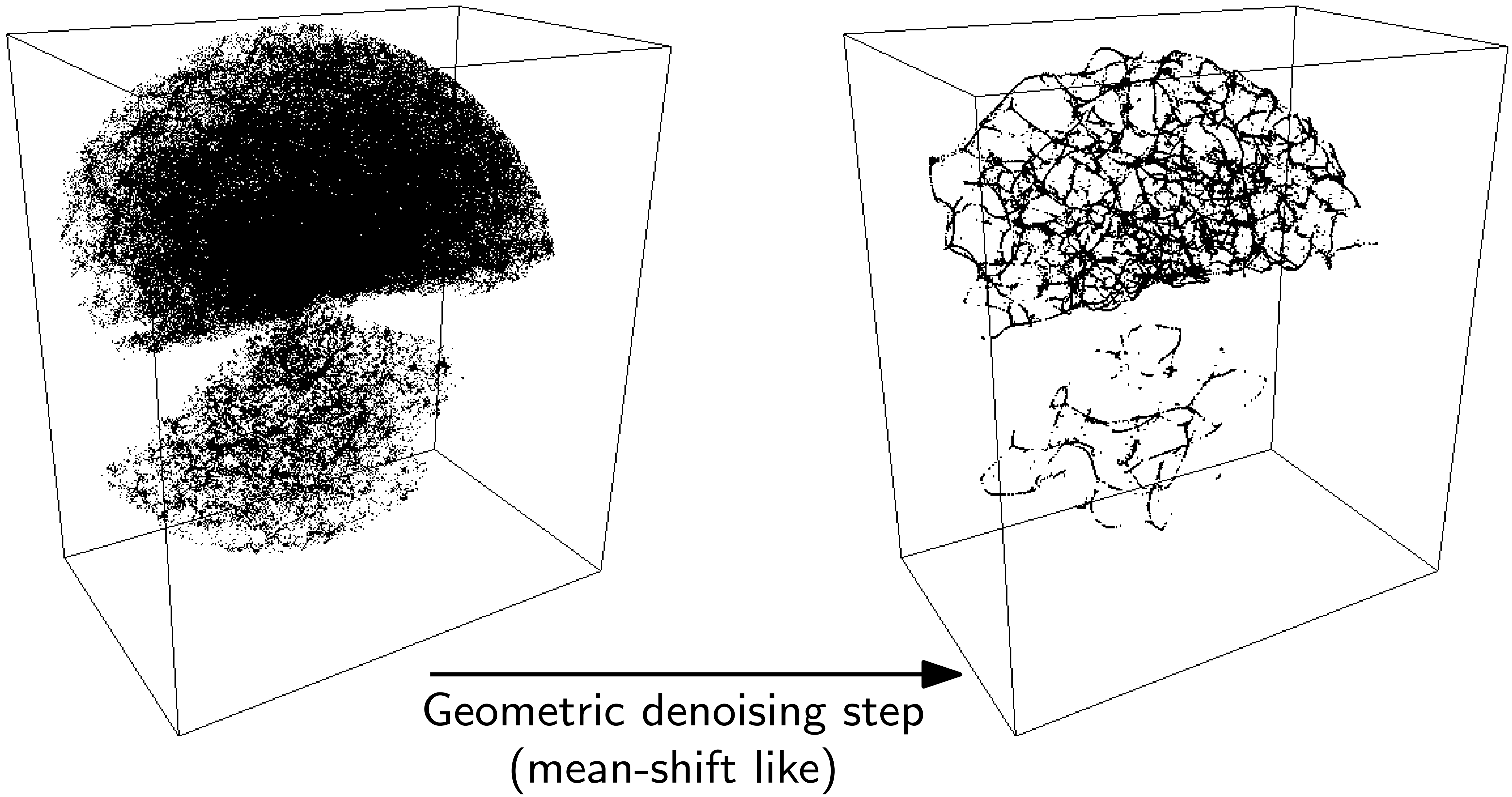
Experimental results



Data: galaxies positions

Data courtesy of the Sloan Digital Sky Survey and R. van de Weygaert

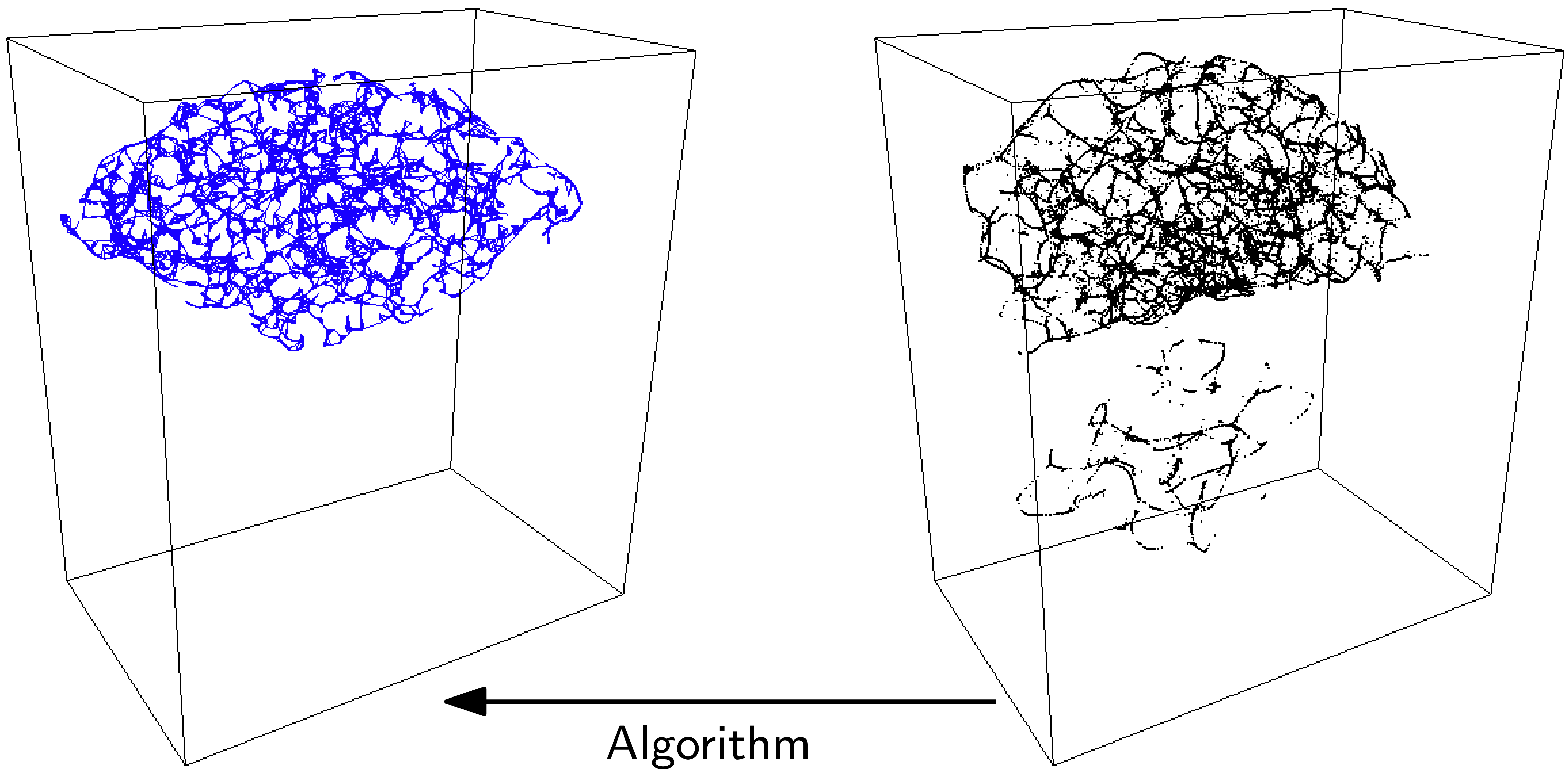
Experimental results



Data: galaxies positions

Data courtesy of the Sloan Digital Sky Survey and R. van de Weygaert

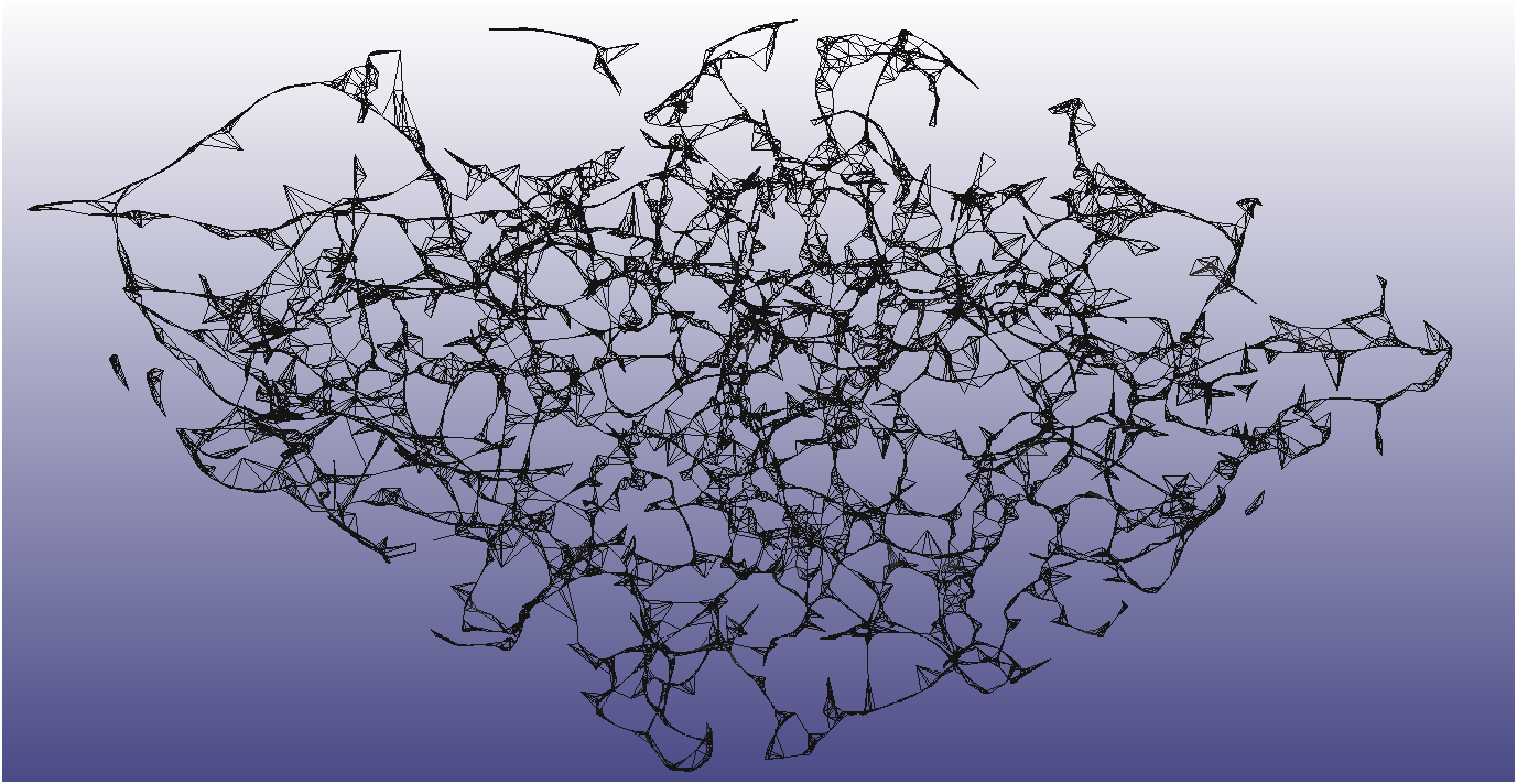
Experimental results



Data: galaxies positions

Data courtesy of the Sloan Digital Sky Survey and R. van de Weygaert

Experimental results



How close to the real cosmic web is this metric graph ?

→ the presence of some remaining noise prevent to get a very nice result.

Advantages and drawbacks of the previous algorithm

Good points:

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

Advantages and drawbacks of the previous algorithm

Good points:

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

Bad points and open questions:

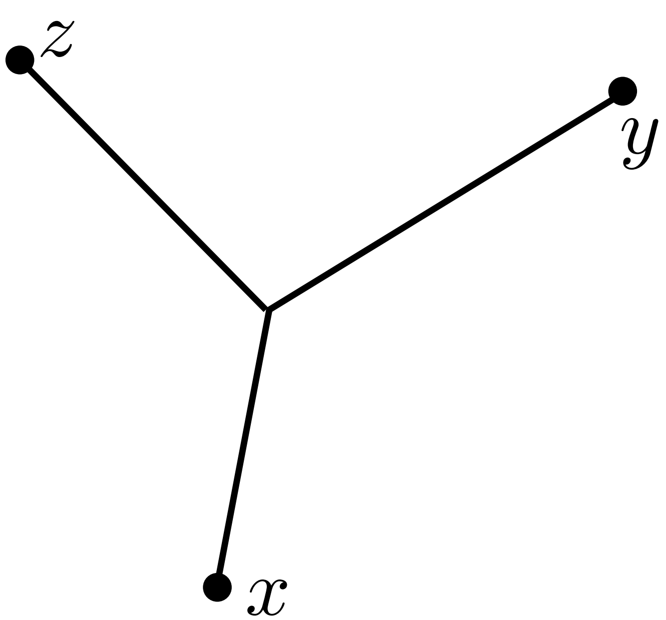
- Choice of the parameter r (that relies on ε, R and b): how to do it in practice? How to make r dependent of the local quality of the approximation?
- Interpretation of the output of the algorithm when the sampling conditions are not fullfilled: the algorithm might not output a graph (edges adjacent to more than 2 vertices).

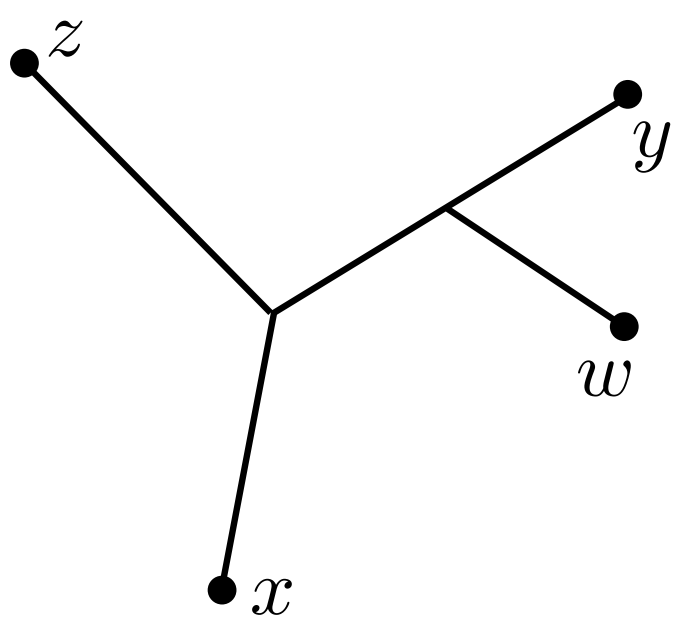
Another approach based upon basic metric geometry

[F. C., Jian Sun, work in progress]

- Metric graphs locally are trees...

- In metric trees the geodesic triangles/tetrahedra have a specific shape : metric trees are 0-*hyperbolic*.



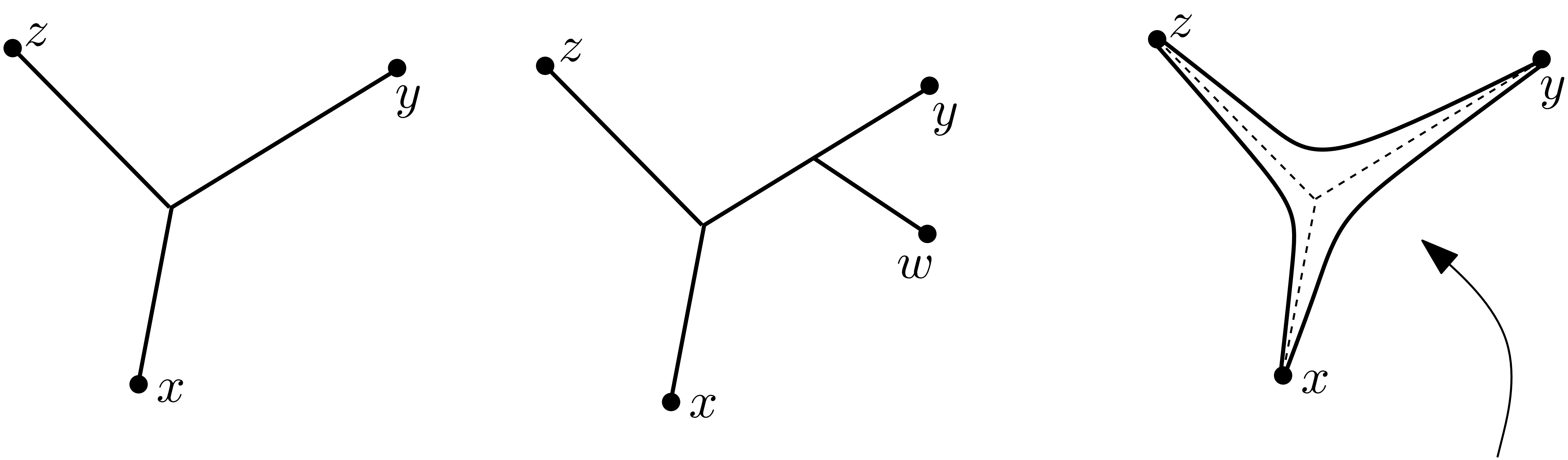


Another approach based upon basic metric geometry

[F. C., Jian Sun, work in progress]

- Metric graphs locally are trees...

- In metric trees the geodesic triangles/tetrahedra have a specific shape : metric trees are 0-*hyperbolic*.



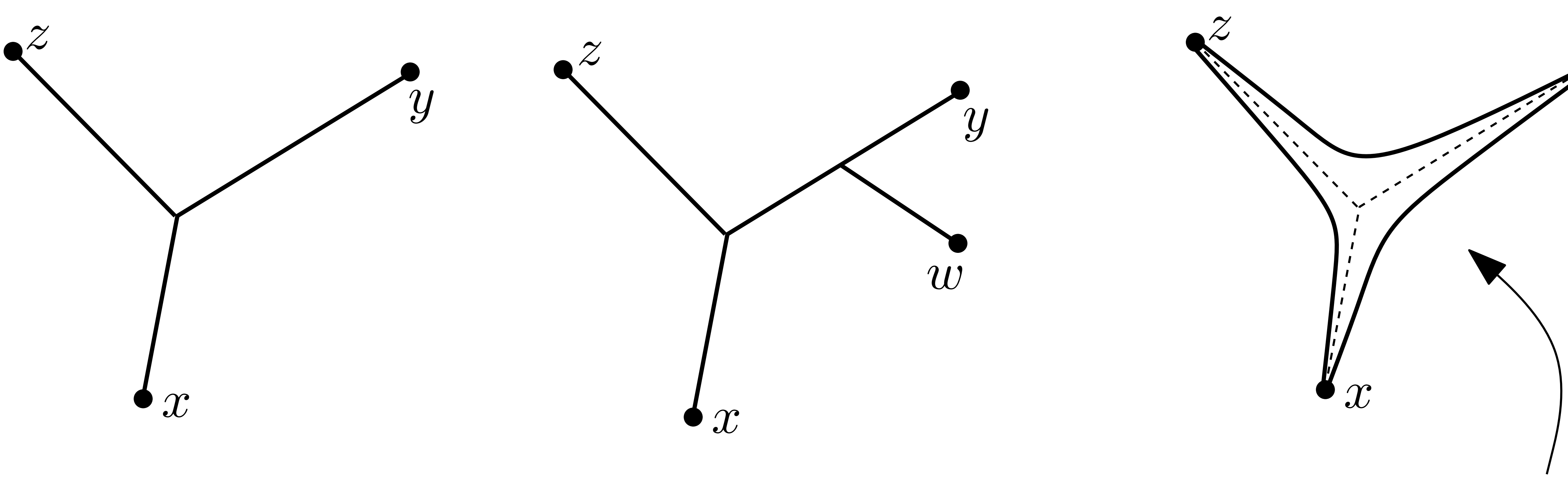
- Metric hyperbolicity is a robust notion w.r.t. Gromov-Hausdorff distance.

Another approach based upon basic metric geometry

[F. C., Jian Sun, work in progress]

- Metric graphs locally are trees...

- In metric trees the geodesic triangles/tetrahedra have a specific shape : metric trees are 0-*hyperbolic*.



- Metric hyperbolicity is a robust notion w.r.t. Gromov-Hausdorff distance.

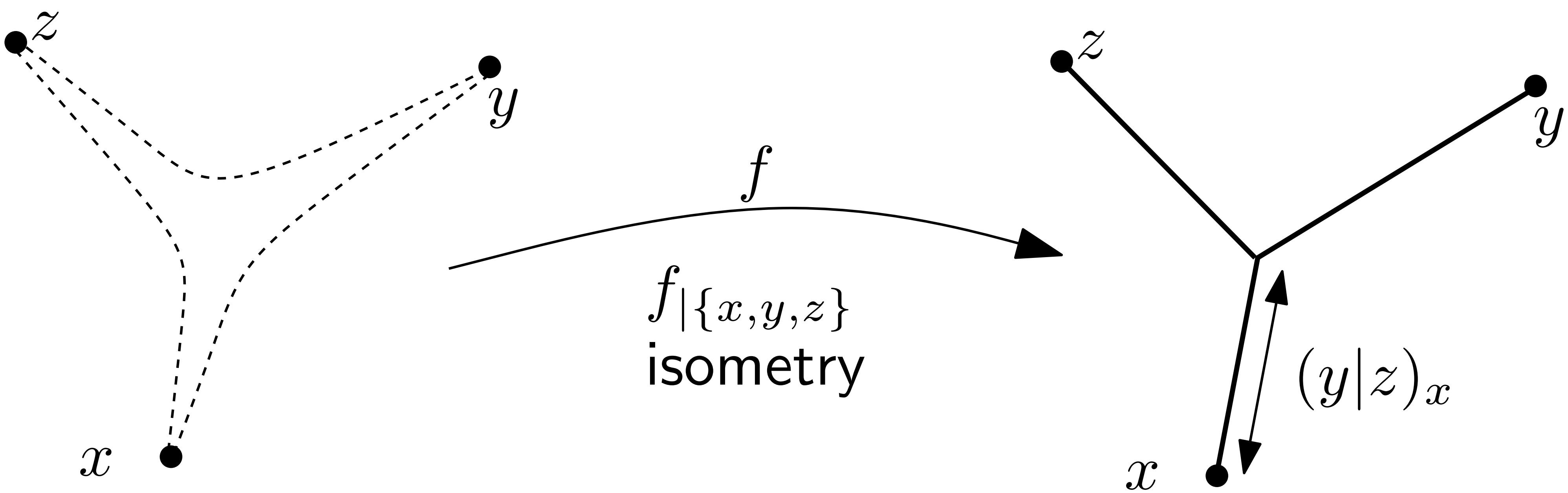
Goal: Use these remarks/properties to reconstruct metric trees and graphs from approximations.

Gromov product and δ-hyperbolic spaces

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a metric space.

Definition (Gromov product):

$$(y|z)_x = \frac{1}{2}\left(d_{\mathbb{X}}(x,y) + d_{\mathbb{X}}(x,z) - d_{\mathbb{X}}(y,z)\right)$$



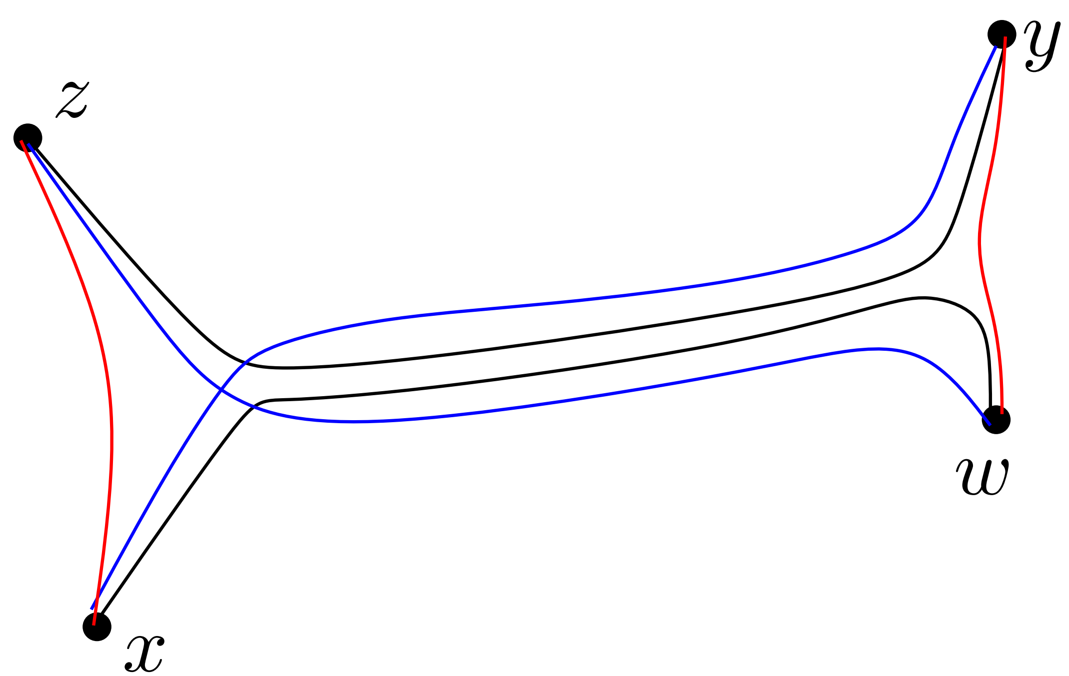
The Gromov product $(y|z)_x$ quantifies the “default” of triangle equality.

Gromov product and δ-hyperbolic spaces

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a metric space.

Definition (Gromov product):

$$(y|z)_x = \frac{1}{2}\left(d_{\mathbb{X}}(x,y) + d_{\mathbb{X}}(x,z) - d_{\mathbb{X}}(y,z)\right)$$



$d(x,y) + d(z,w)$
$d(x,w) + d(z,y)$
$d(x,z) + d(y,w)$

Given $\delta > 0$, $\mathbb{X}$ is δ-**hyperbolic** if for any $x, y, z, w \in \mathbb{X}$

$$(x|z)_w \geq \min\{(x|y)_w, (y|z)_w\} - \delta$$

Equivalently, $\mathbb{X}$ is δ-hyperbolic if the two larger of the distance sums $d_{\mathbb{X}}(x,w) + d_{\mathbb{X}}(y,z)$, $d_{\mathbb{X}}(x,y) + d_{\mathbb{X}}(z,w)$ and $d_{\mathbb{X}}(x,z) + d_{\mathbb{X}}(y,w)$ differ by at most 2δ.

Gromov product and δ-hyperbolic spaces

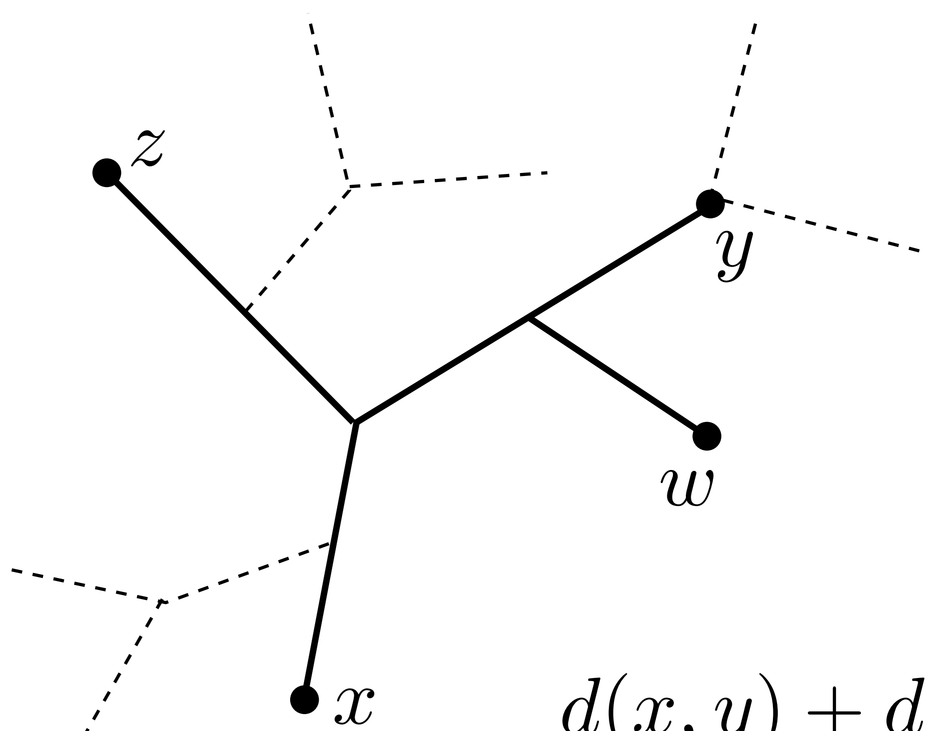
Given $\delta > 0$, $\mathbb{X}$ is δ-**hyperbolic** if for any $x, y, z, w \in \mathbb{X}$

$$(x|z)_w \geq \min\{(x|y)_w, (y|z)_w\} - \delta$$

Equivalently, $\mathbb{X}$ is δ-hyperbolic if the two larger of the distance sums $d_{\mathbb{X}}(x,w)+d_{\mathbb{X}}(y,z)$, $d_{\mathbb{X}}(x,y)+d_{\mathbb{X}}(z,w)$ and $d_{\mathbb{X}}(x,z)+d_{\mathbb{X}}(y,w)$ differ by at most 2δ.

Basic properties:

- Metric trees are 0-hyperbolic. This is indeed a characterization of (real) trees.



$$d(x,y) + d(z,w) = d(x,w) + d(z,y)$$

Gromov product and δ-hyperbolic spaces

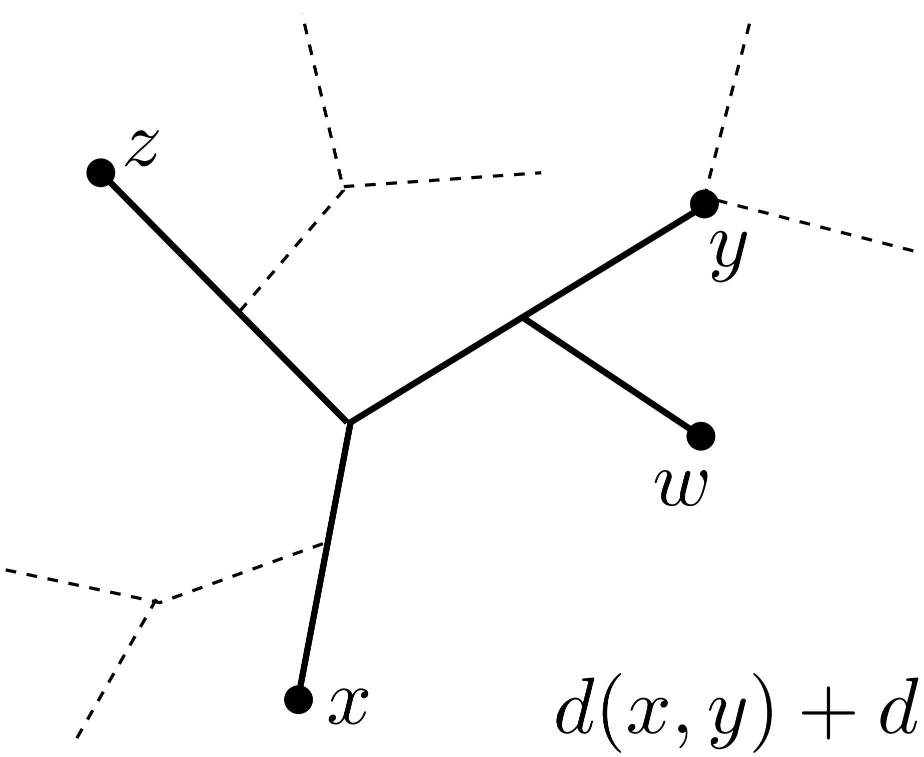
Given $\delta > 0$, $\mathbb{X}$ is δ-**hyperbolic** if for any $x, y, z, w \in \mathbb{X}$

$$(x|z)_w \geq \min\{(x|y)_w, (y|z)_w\} - \delta$$

Equivalently, $\mathbb{X}$ is δ-hyperbolic if the two larger of the distance sums $d_{\mathbb{X}}(x,w) + d_{\mathbb{X}}(y,z)$, $d_{\mathbb{X}}(x,y) + d_{\mathbb{X}}(z,w)$ and $d_{\mathbb{X}}(x,z) + d_{\mathbb{X}}(y,w)$ differ by at most 2δ.

Basic properties:

- Metric trees are 0-hyperbolic. This is indeed a characterization of (real) trees.
- if $d_{GH}(\mathbb{X}, \mathbb{Y}) < \varepsilon$ then ($\mathbb{X}$ is δ-hyperbolic) $\Rightarrow$ ($\mathbb{Y}$ is $(\delta + 2\varepsilon)$-hyperbolic).



$$d(x,y) + d(z,w) = d(x,w) + d(z,y)$$

Gromov product and δ-hyperbolic spaces

Given $\delta > 0$, $\mathbb{X}$ is δ-**hyperbolic** if for any $x, y, z, w \in \mathbb{X}$

$$(x|z)_w \geq \min\{(x|y)_w, (y|z)_w\} - \delta$$

Equivalently, $\mathbb{X}$ is δ-hyperbolic if the two larger of the distance sums $d_{\mathbb{X}}(x,w) + d_{\mathbb{X}}(y,z)$, $d_{\mathbb{X}}(x,y) + d_{\mathbb{X}}(z,w)$ and $d_{\mathbb{X}}(x,z) + d_{\mathbb{X}}(y,w)$ differ by at most 2δ.

Basic properties:

- Metric trees are 0-hyperbolic. This is indeed a characterization of (real) trees.
- if $d_{GH}(\mathbb{X}, \mathbb{Y}) < \varepsilon$ then ($\mathbb{X}$ is δ-hyperbolic) $\Rightarrow$ ($\mathbb{Y}$ is $(\delta + 2\varepsilon)$-hyperbolic).

Question: Can we use δ-hyperbolicity to characterize (metric) data that are "close" (w.r.t. d_{GH}) to a metric tree?

Gromov product and δ-hyperbolic spaces

Given $\delta > 0$, $\mathbb{X}$ is δ-**hyperbolic** if for any $x, y, z, w \in \mathbb{X}$

$$(x|z)_w \geq \min\{(x|y)_w, (y|z)_w\} - \delta$$

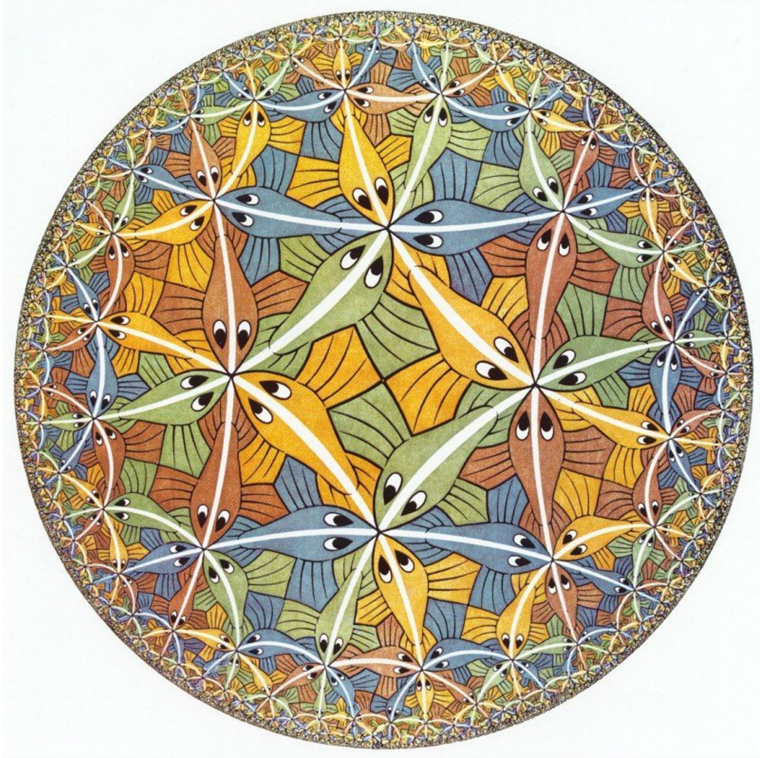
Equivalently, $\mathbb{X}$ is δ-hyperbolic if the two larger of the distance sums $d_{\mathbb{X}}(x,w) + d_{\mathbb{X}}(y,z)$, $d_{\mathbb{X}}(x,y) + d_{\mathbb{X}}(z,w)$ and $d_{\mathbb{X}}(x,z) + d_{\mathbb{X}}(y,w)$ differ by at most 2δ.

Basic properties:

- Metric trees are 0-hyperbolic. This is indeed a characterization of (real) trees.
- if $d_{GH}(\mathbb{X}, \mathbb{Y}) < \varepsilon$ then ($\mathbb{X}$ is δ-hyperbolic) $\Rightarrow$ ($\mathbb{Y}$ is $(\delta + 2\varepsilon)$-hyperbolic).

Question: Can we use δ-hyperbolicity to characterize (metric) data that are "close" (w.r.t. d_{GH}) to a metric tree?

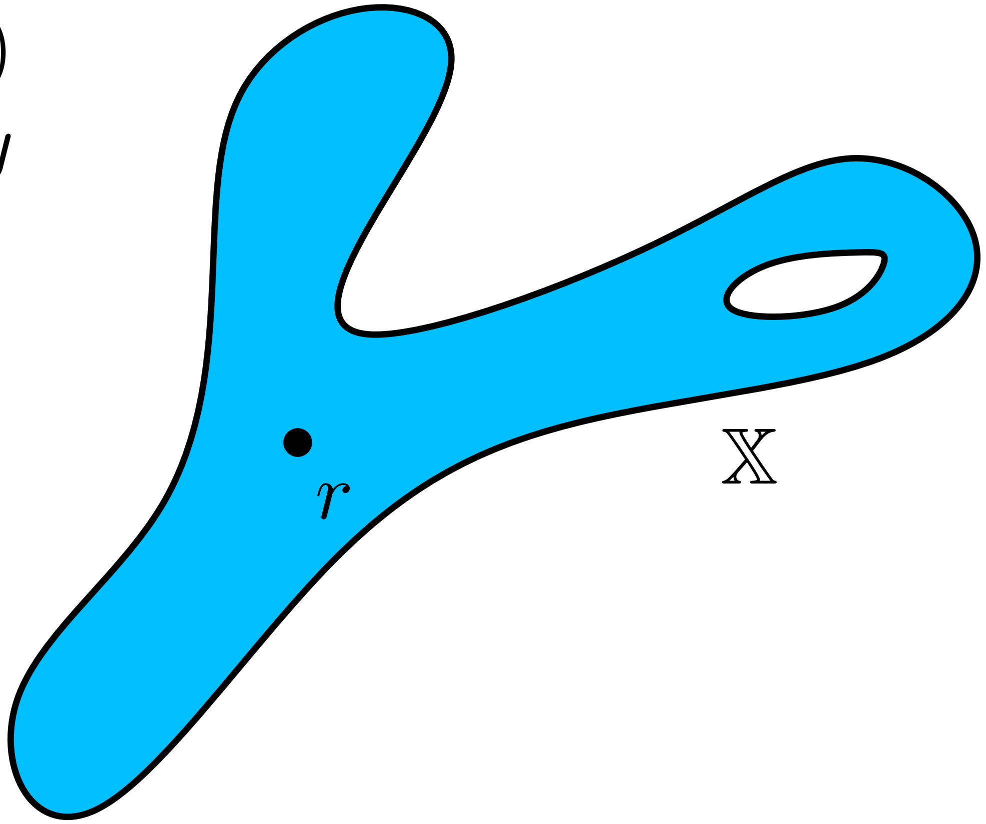
Not so easy:
the Poincaré disc is $\log(3)$-hyperbolic!



Distance functions and persistence tree

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact path metric space. Given $r \in \mathbb{X}$, let $d(.) = d_{\mathbb{X}}(r, .)$ the distance function to r in $\mathbb{X}$ (assume d has a finite number of local maxima).

Equivalence relation: $x \sim y$ iff $d(x) = d(y)$ and there exists a continuous path from x to y contained in $d^{-1}([d(x), +\infty))$.

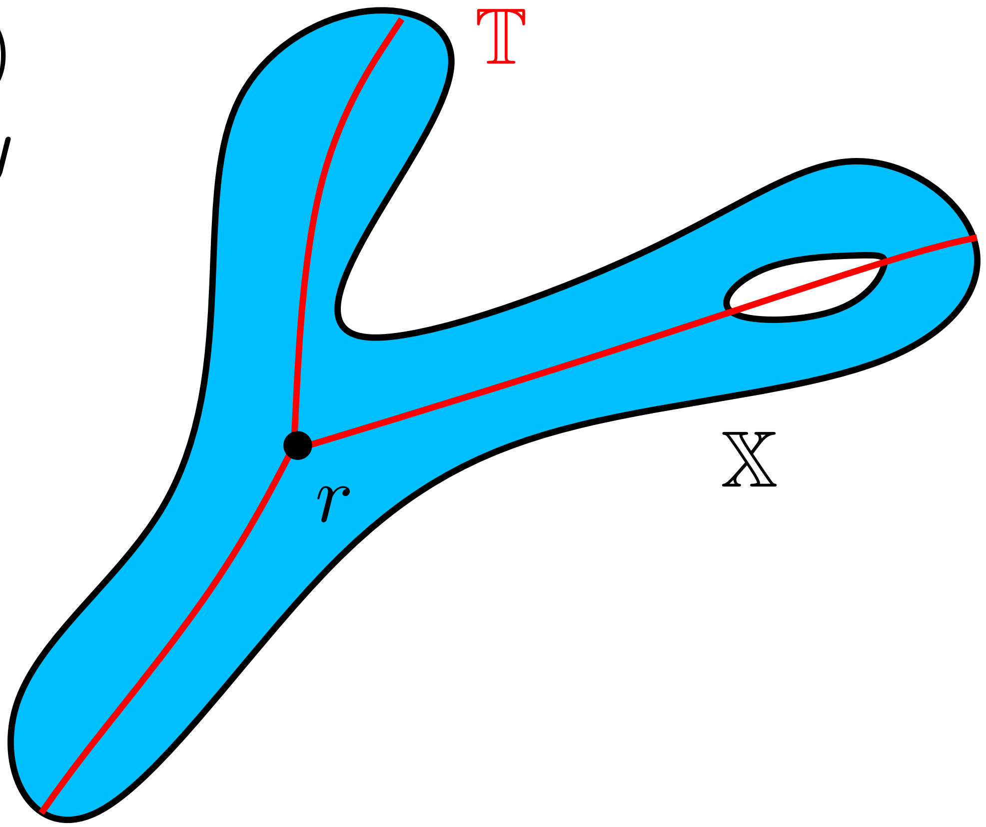


Distance functions and persistence tree

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact path metric space. Given $r \in \mathbb{X}$, let $d(.) = d_{\mathbb{X}}(r, .)$ the distance function to r in $\mathbb{X}$ (assume d has a finite number of local maxima).

Equivalence relation: $x \sim y$ iff $d(x) = d(y)$ and there exists a continuous path from x to y contained in $d^{-1}([d(x), +\infty))$.

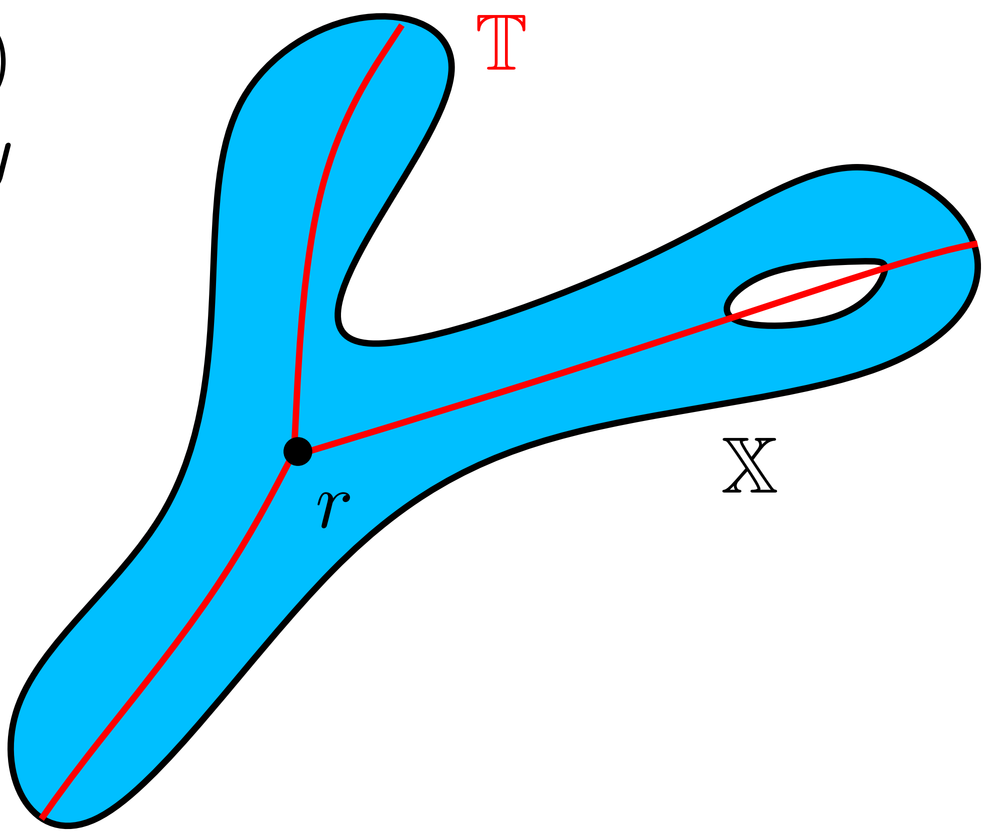
Persistence tree: $\mathbb{T} = \mathbb{X}/\sim$



Distance functions and persistence tree

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact path metric space. Given $r \in \mathbb{X}$, let $d(.) = d_{\mathbb{X}}(r, .)$ the distance function to r in $\mathbb{X}$ (assume d has a finite number of local maxima).

Equivalence relation: $x \sim y$ iff $d(x) = d(y)$ and there exists a continuous path from x to y contained in $d^{-1}([d(x), +\infty))$.



Persistence tree: $\mathbb{T} = \mathbb{X}/\sim$

Let $\Pi : \mathbb{X} \to \mathbb{T}$ be the quotient map.

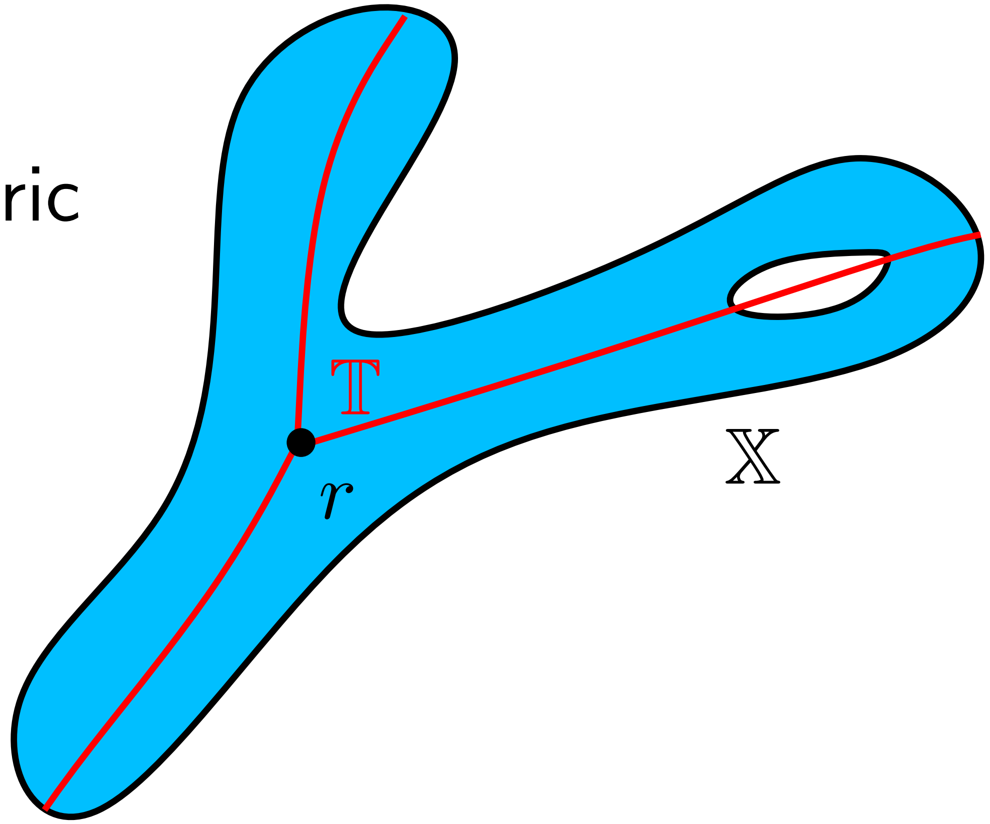
Lemma:
(i) $\mathbb{T}$ is a tree; its leaves are in one-to-one correspondence with the local maxima of d.
(ii) $\mathbb{T}$ inherits a metric structure from $d_{\mathbb{T}}(\Pi r, \Pi x) := d_{\mathbb{X}}(r, x)$ for all $x \in \mathbb{X}$.

In practice: $\mathbb{T}$ can be efficiently constructed using the 0-persistence algorithm (union find data structure) on the input data endowed with a neighboring graph.

Tree reconstruction

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact path metric space. Given $r \in \mathbb{X}$, let $d(.) = d_{\mathbb{X}}(r, .)$ the distance function to r in $\mathbb{X}$ (assume d has a finite number of local maxima).

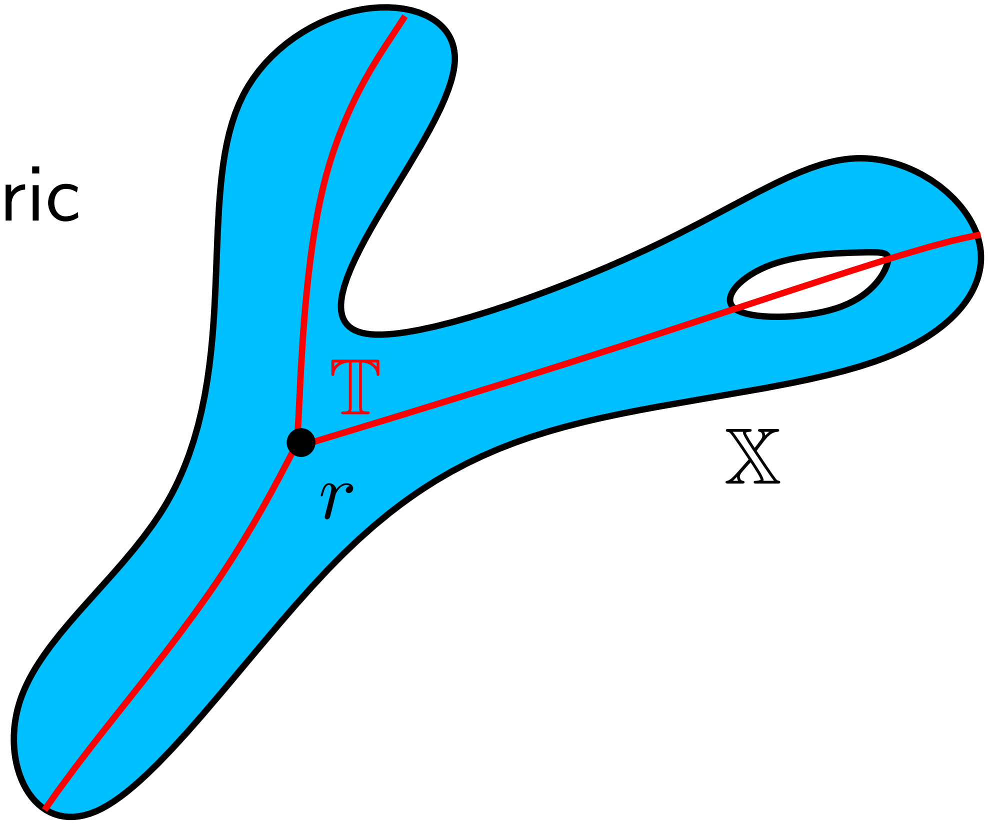
Lemma: if $\mathbb{X}$ is a metric tree then $\mathbb{T}$ is isometric to $\mathbb{X}$.



Tree reconstruction

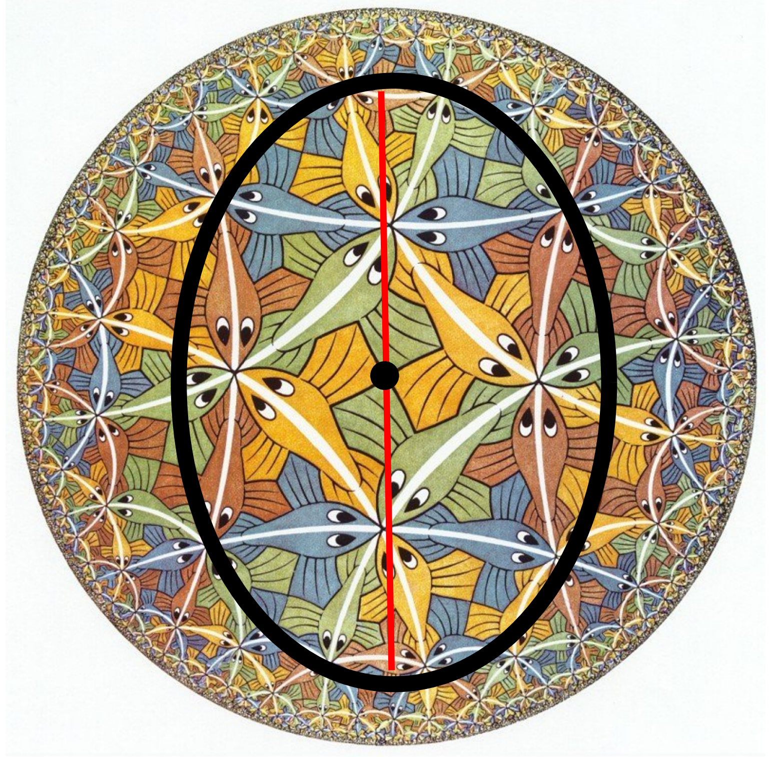
Let $(\mathbb{X}, d_\mathbb{X})$ be a compact path metric space. Given $r \in \mathbb{X}$, let $d(.) = d_\mathbb{X}(r, .)$ the distance function to r in $\mathbb{X}$ (assume d has a finite number of local maxima).

Lemma: if $\mathbb{X}$ is a metric tree then $\mathbb{T}$ is isometric to $\mathbb{X}$.



What can we say about $d_{GH}(\mathbb{X}, \mathbb{T})$ when $\mathbb{X}$ is δ-hyperbolic?

$\rightarrow$ In general nothing... But....

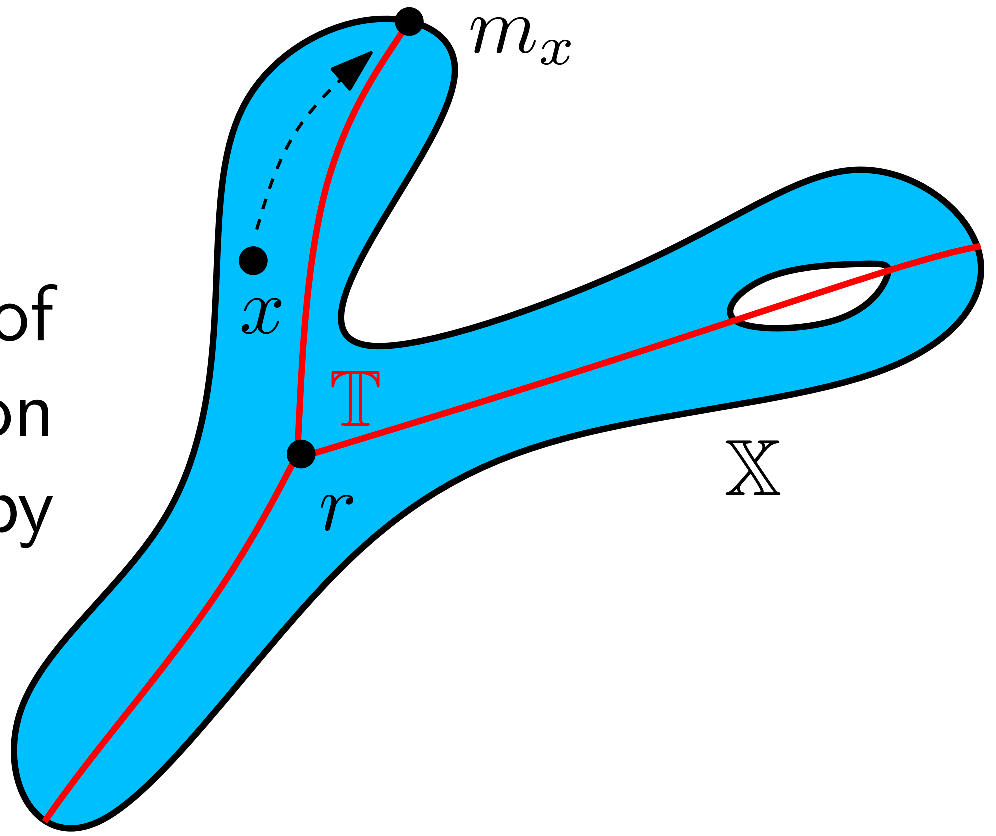


Tree reconstruction

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact path metric space. Given $r \in \mathbb{X}$, let $d(.) = d_{\mathbb{X}}(r, .)$ the distance function to r in $\mathbb{X}$ (assume d has a finite number of local maxima).

Not unique in general

For $x \in \mathbb{X}$ let $m_x \in X$ be in the c.c. of $d^{-1}([d(x), +\infty))$ containing x that maximizes d on this c.c. and let $f : \mathbb{X} \to \mathbb{R}_+$ be defined by $f(x) = (r|m_x)_x$.



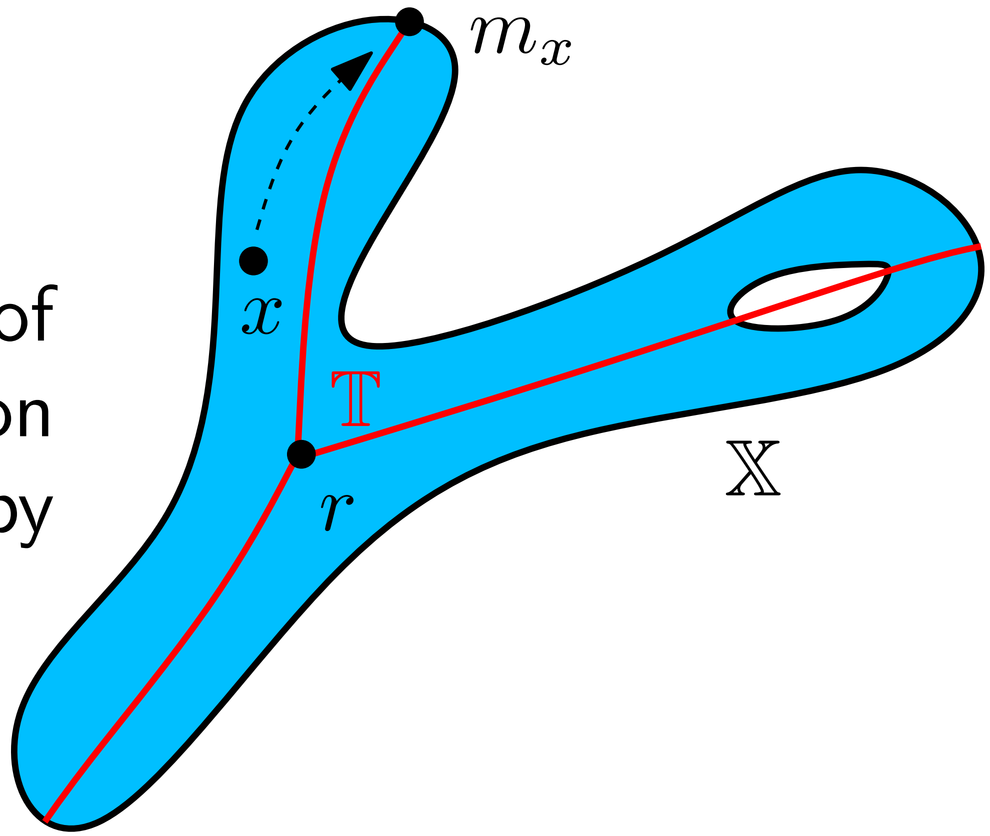
In practice: f can be easily computed as $\mathbb{T}$ is constructed using the 0-persistence algorithm.

Tree reconstruction

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact path metric space. Given $r \in \mathbb{X}$, let $d(.) = d_{\mathbb{X}}(r, .)$ the distance function to r in $\mathbb{X}$ (assume d has a finite number of local maxima).

Not unique in general

For $x \in \mathbb{X}$ let $m_x \in X$ be in the c.c. of $d^{-1}([d(x), +\infty))$ containing x that maximizes d on this c.c. and let $f : \mathbb{X} \to \mathbb{R}_+$ be defined by $f(x) = (r|m_x)_x$.



Theorem: Let $M = \|f\|_\infty = \max_{x \in X} f(x)$. If $\mathbb{X}$ is δ-hyperbolic then

$$d_{GH}(\mathbb{X}, \mathbb{T}) < M + 9\delta$$

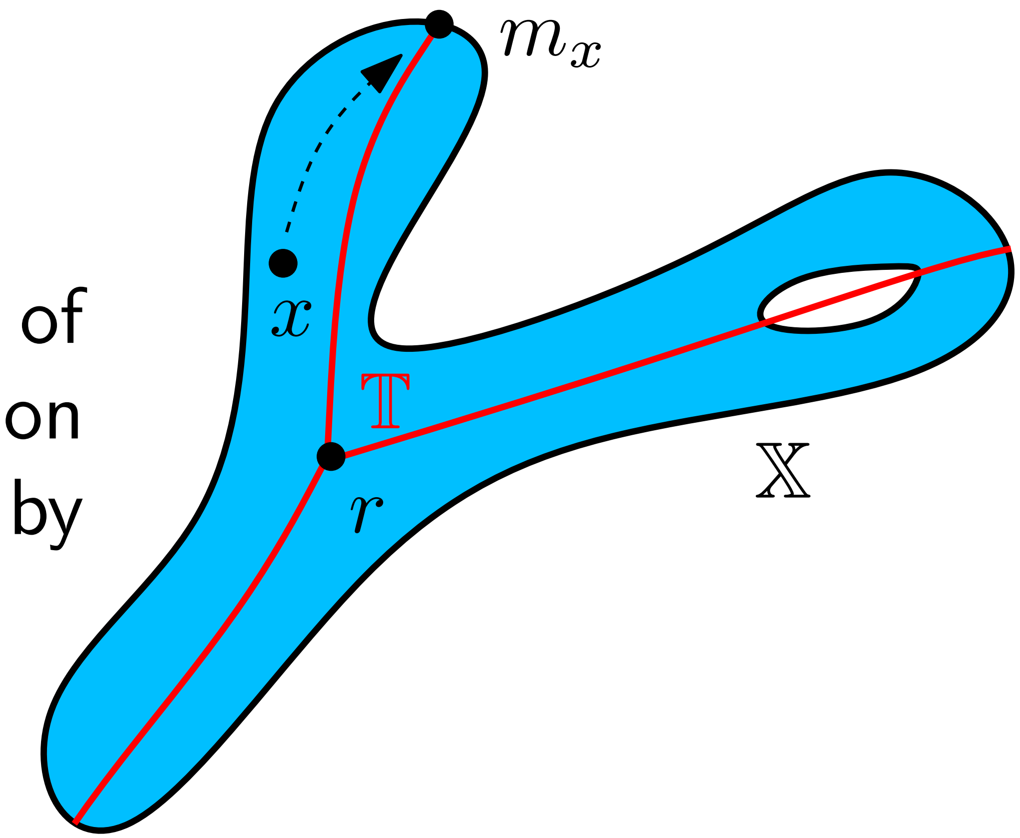
and for any $x, y \in X$, $|d_T(\Pi x, \Pi y) - d_X(x, y)| \leq 2(M + 9\delta)$.

Tree reconstruction

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a compact path metric space. Given $r \in \mathbb{X}$, let $d(.) = d_{\mathbb{X}}(r, .)$ the distance function to r in $\mathbb{X}$ (assume d has a finite number of local maxima).

Not unique in general

For $x \in \mathbb{X}$ let $m_x \in X$ be in the c.c. of $d^{-1}([d(x), +\infty))$ containing x that maximizes d on this c.c. and let $f : \mathbb{X} \to \mathbb{R}_+$ be defined by $f(x) = (r|m_x)_x$.



Theorem: Let $M = \|f\|_\infty = \max_{x \in X} f(x)$. If $\mathbb{X}$ is δ-hyperbolic then

$$d_{GH}(\mathbb{X}, \mathbb{T}) < M + 9\delta$$

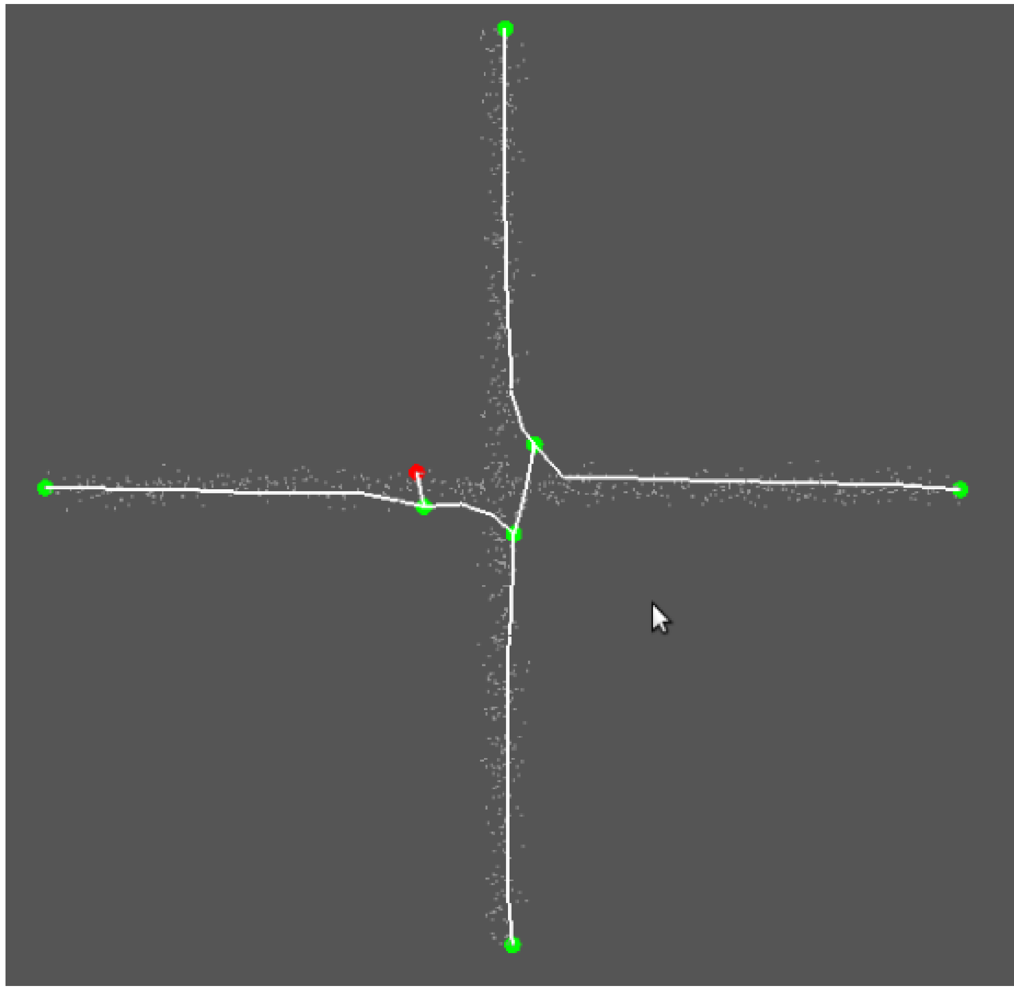
and for any $x, y \in X$, $|d_T(\Pi x, \Pi y) - d_X(x, y)| \leq 2(M + 9\delta)$.

Corollary: there exists a constant $C < 100$ s.t. if $\mathbb{T}_0$ is a tree and if $d_{GH}(\mathbb{X}, \mathbb{T}_0) < \varepsilon$ then $d_{GH}(\mathbb{T}, \mathbb{T}_0) < C\varepsilon$.

Some practical considerations

When $\mathbb{X}$ is a neighboring graph (i.e. Rips-Vietoris) built on top of a finite metric data set $(\mathbb{Y}, d_{\mathbb{Y}})$...

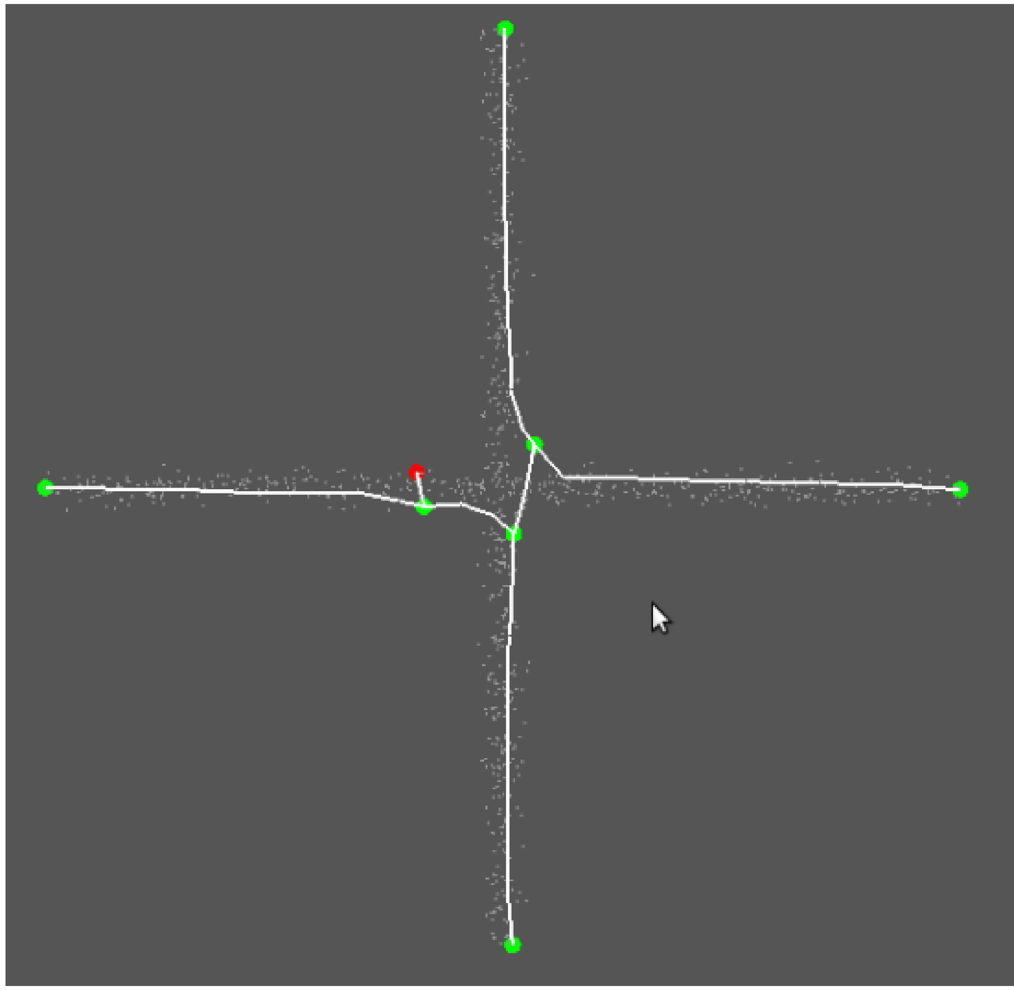
- $\mathbb{T}$, $\Pi : \mathbb{Y} \to \mathbb{T}$ and M can be computed efficiently using the 0-persistence algorithm in almost linear time (in the number of edges of $\mathbb{X}$)!



Some practical considerations

When $\mathbb{X}$ is a neighboring graph (i.e. Rips-Vietoris) built on top of a finite metric data set $(\mathbb{Y}, d_{\mathbb{Y}})$...

- $\mathbb{T}$, $\Pi : \mathbb{Y} \to \mathbb{T}$ and M can be computed efficiently using the 0-persistence algorithm in almost linear time (in the number of edges of $\mathbb{X}$)!
- δ can also be computed from the data but in $O(|Y|^4)$
  $\to$ can be improved to $O(|Y|^3)$; estimation from random samples of tetrahedra?

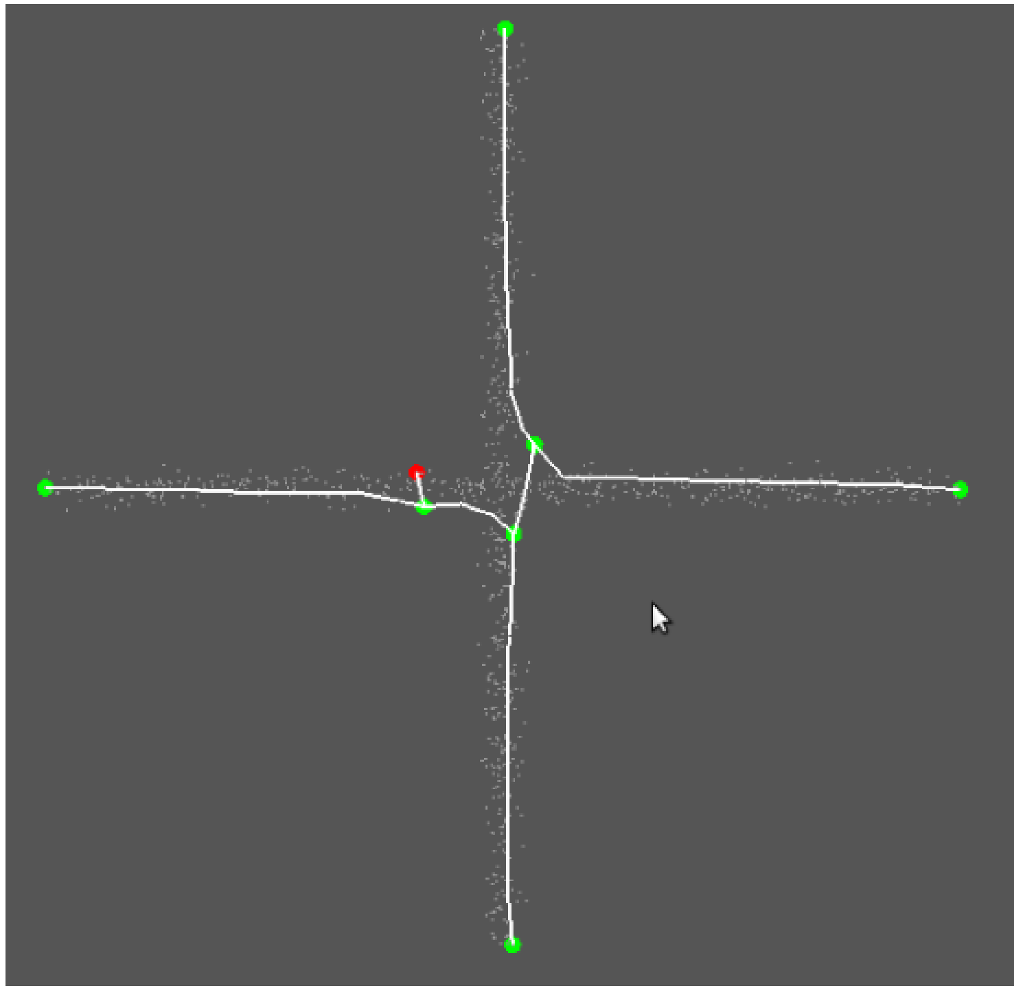


Some practical considerations

When $\mathbb{X}$ is a neighboring graph (i.e. Rips-Vietoris) built on top of a finite metric data set $(\mathbb{Y}, d_{\mathbb{Y}})$...

- $\mathbb{T}$, $\Pi : \mathbb{Y} \to \mathbb{T}$ and M can be computed efficiently using the 0-persistence algorithm in almost linear time (in the number of edges of $\mathbb{X}$)!

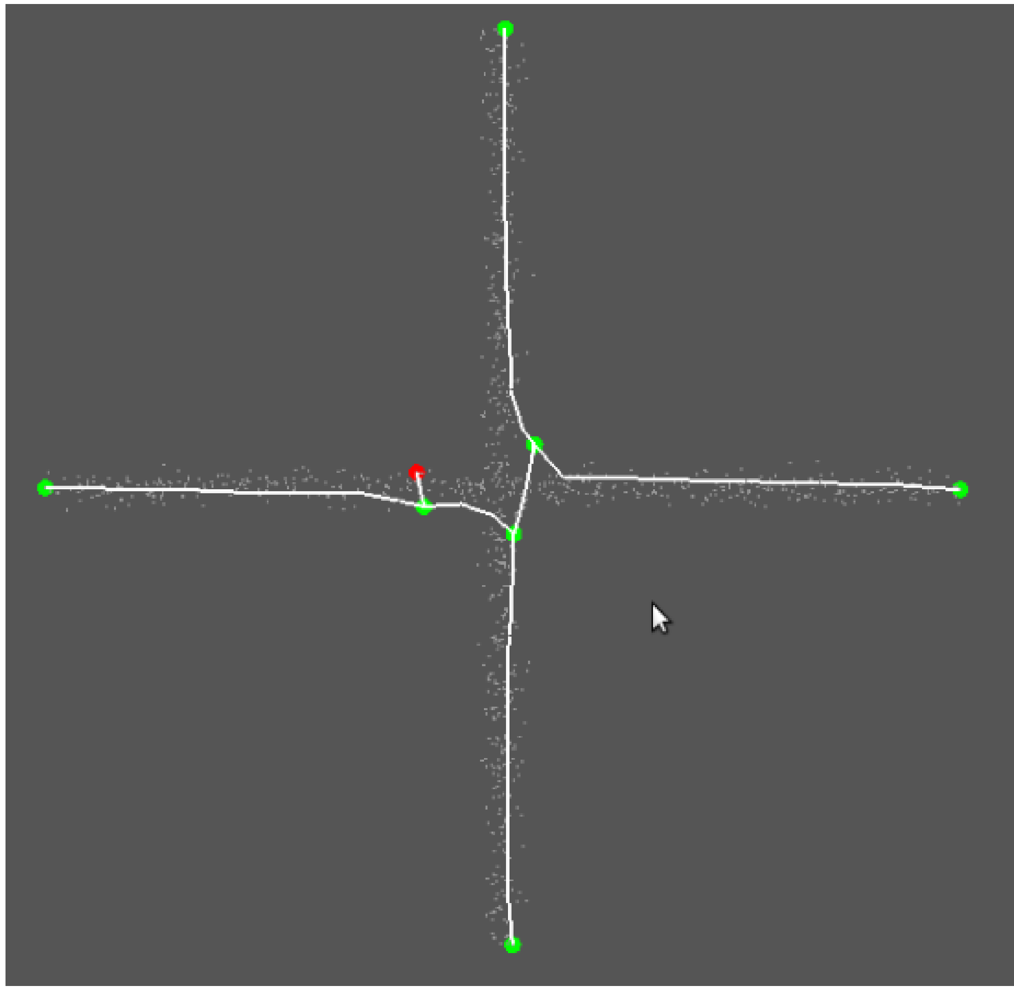
- δ can also be computed from the data but in $O(|Y|^4)$
  $\to$ can be improved to $O(|Y|^3)$; estimation from random samples of tetrahedra?



- Persistence can also be used to remove branches of the reconstructed tree.

Some practical considerations

When $\mathbb{X}$ is a neighboring graph (i.e. Rips-Vietoris) built on top of a finite metric data set $(\mathbb{Y}, d_{\mathbb{Y}})$...



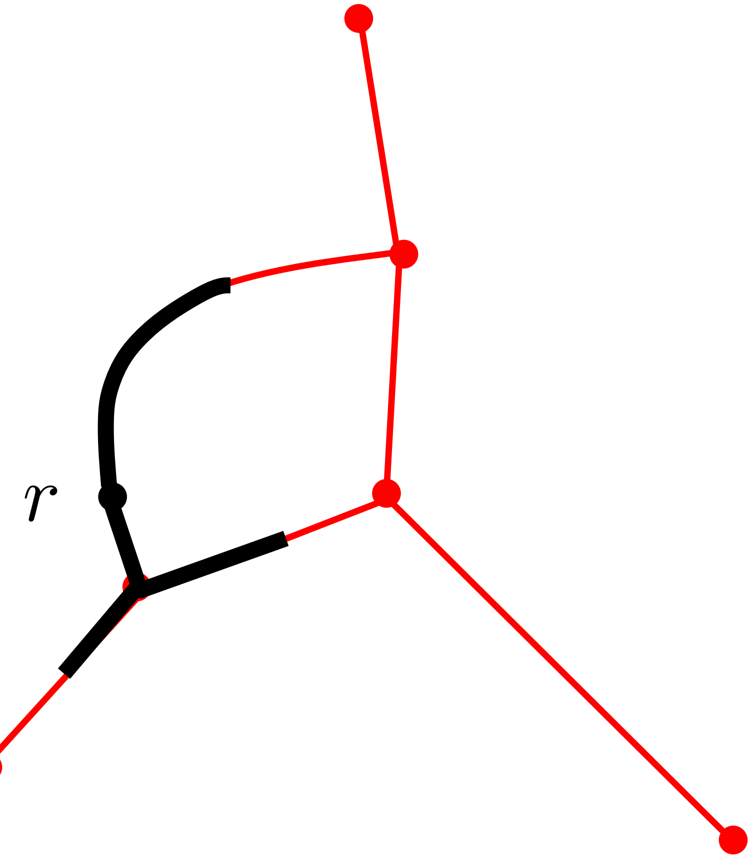
- $\mathbb{T}$, $\Pi : \mathbb{Y} \to \mathbb{T}$ and M can be computed efficiently using the 0-persistence algorithm in almost linear time (in the number of edges of $\mathbb{X}$)!
- δ can also be computed from the data but in $O(|Y|^4)$
  $\to$ can be improved to $O(|Y|^3)$; estimation from random samples of tetrahedra?
- Persistence can also be used to remove branches of the reconstructed tree.

$\Rightarrow$ An algorithm that always output a tree and an upperbound on the Gromov-Hausdorff distance between the reconstructed tree and the data (no parameter to choose except to build $\mathbb{X}$ from $\mathbb{Y}$).

Warning: even if $d_{GH}(\mathbb{X}, \mathbb{T}_0) << 1$, $\mathbb{T}$ is in general not homeomorphic to $\mathbb{T}_0$.

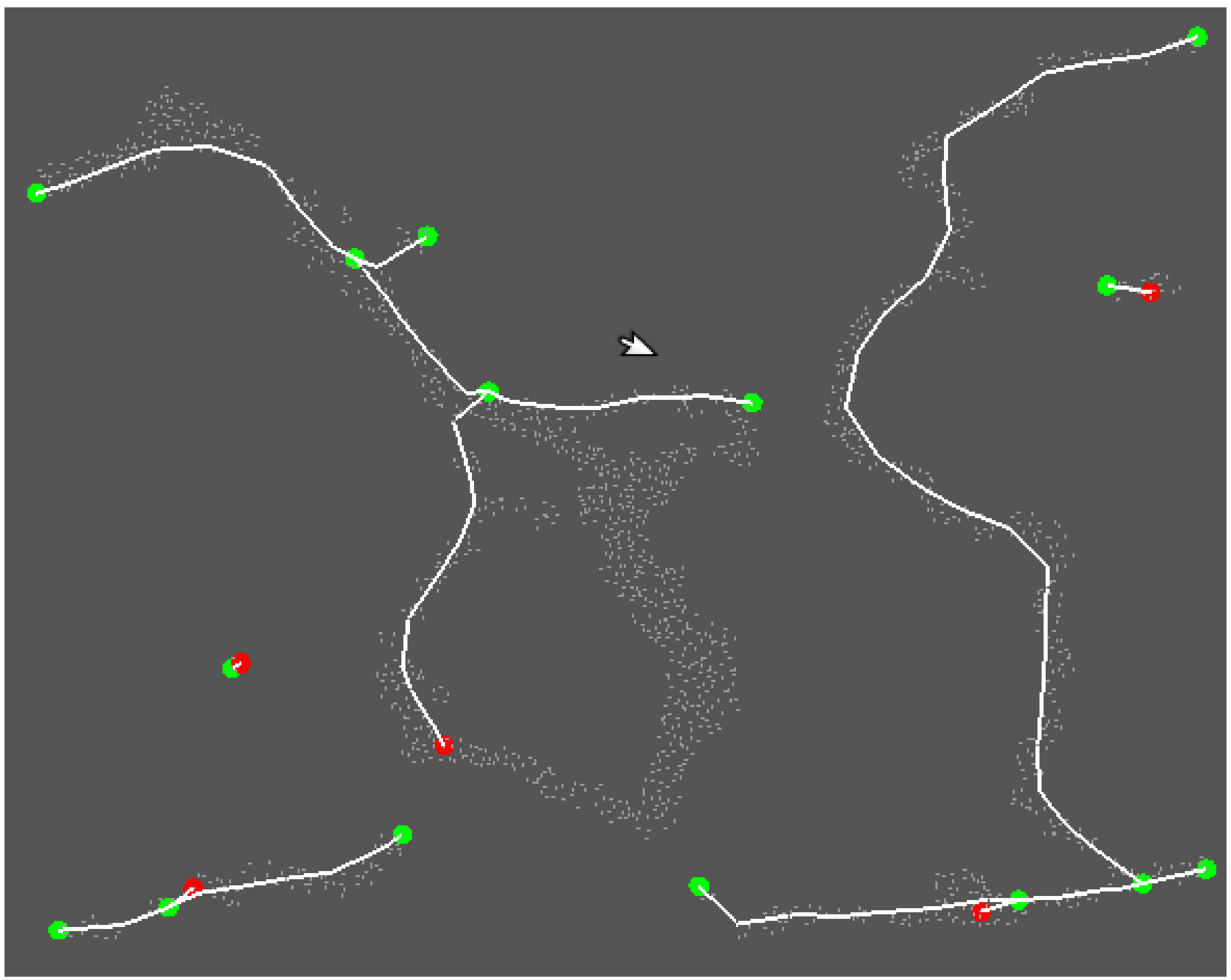
From trees to graphs

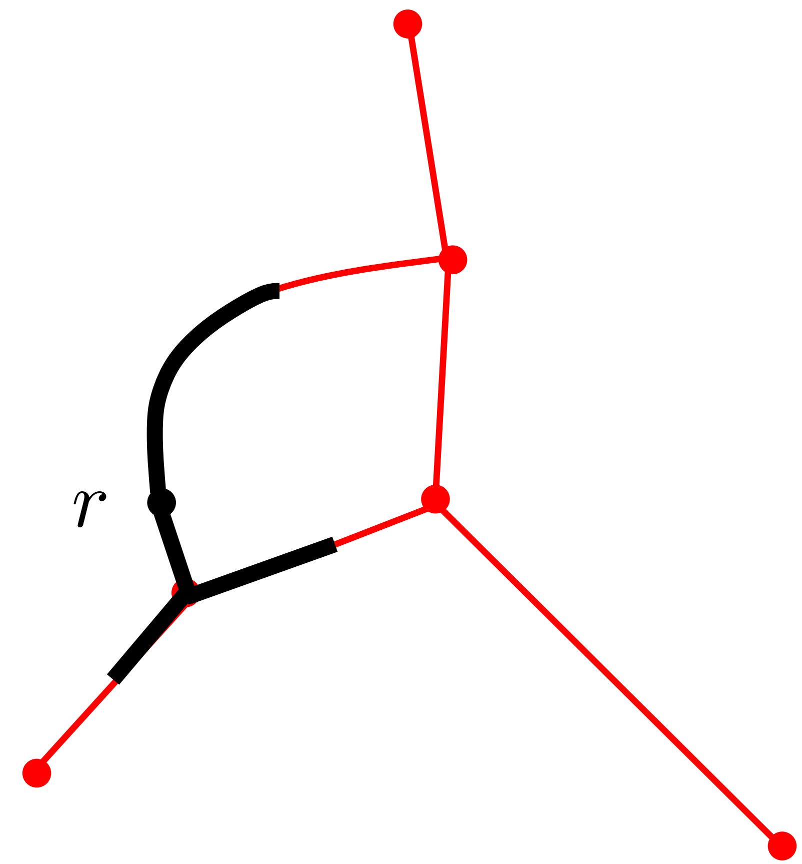
- Locally, metric graphs are trees: if $\mathbb{G}$ is a metric graph and $l(\mathbb{G})$ is the length of the shortest non-null homotopic simple path then for any $r \in \mathbb{G}$, $(B(r, l(\mathbb{G})/4), d_{\mathbb{G}})$ is a metric tree.



From trees to graphs

- Locally, metric graphs are trees: if $\mathbb{G}$ is a metric graph and $l(\mathbb{G})$ is the length of the shortest non-null homotopic simple path then for any $r \in \mathbb{G}$, $(B(r, l(\mathbb{G})/4), d_{\mathbb{G}})$ is a metric tree.

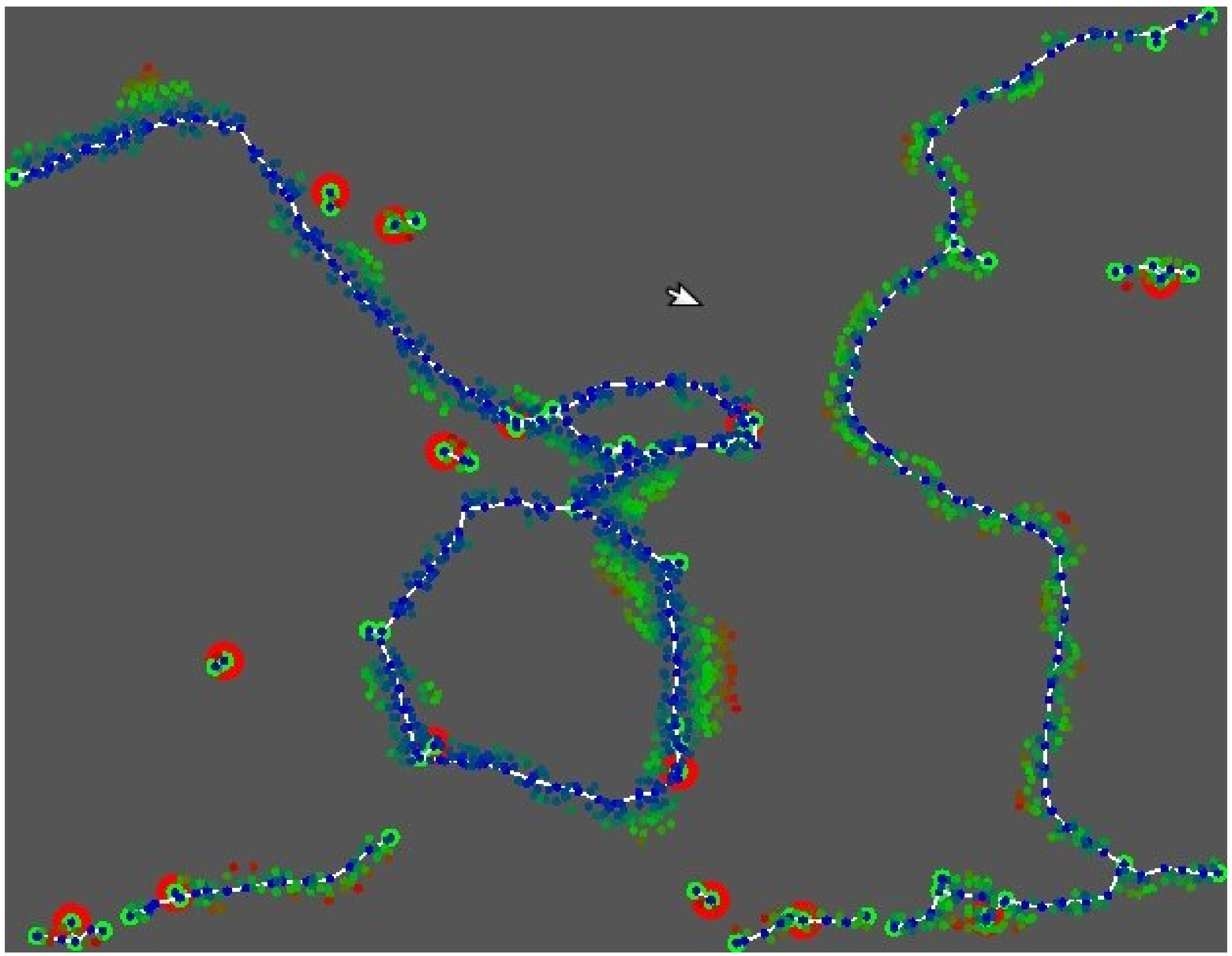


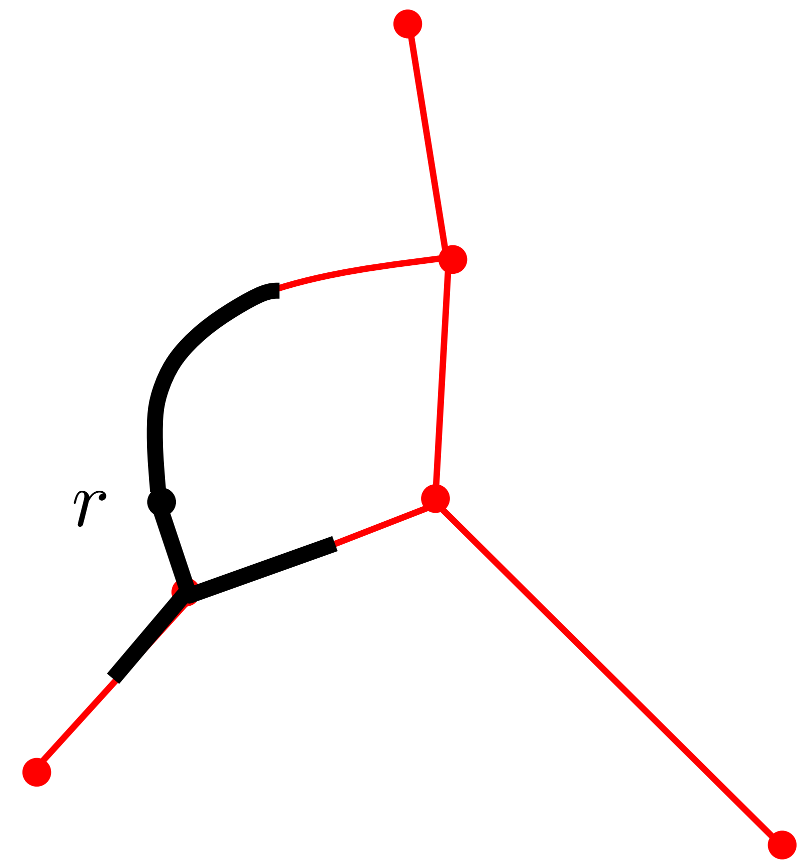


$\rightarrow$ Data can be "semi-locally" approximated by trees.

From trees to graphs

- Locally, metric graphs are trees: if $\mathbb{G}$ is a metric graph and $l(\mathbb{G})$ is the length of the shortest non-null homotopic simple path then for any $r \in \mathbb{G}$, $(B(r, l(\mathbb{G})/4), d_{\mathbb{G}})$ is a metric tree.





$\rightarrow$ Data can be "semi-locally" approximated by trees.

From trees to graphs

- Locally, metric graphs are trees: if $\mathbb{G}$ is a metric graph and $l(\mathbb{G})$ is the length of the shortest non-null homotopic simple path then for any $r \in \mathbb{G}$, $(B(r, l(\mathbb{G})/4), d_{\mathbb{G}})$ is a metric tree.

$l(\mathbb{G})$ can be infered from the data:

Proposition: For any metric space $\mathbb{Y}$ such that $d_{GH}(\mathbb{G}, \mathbb{Y}) < \frac{1}{16} l(\mathbb{G})$ and any $d_{GH}(\mathbb{G}, \mathbb{Y}) < \alpha < \frac{3}{16} l(\mathbb{G})$, the first Betti number of $\mathbb{G}$ is given by

$$b_1(\mathbb{G}) = \text{rank}\left(H_1(Rips(\mathbb{Y}, \alpha)) \to H_1(Rips(\mathbb{Y}, 3\alpha)\right)$$

where the homomorphism between the homology groups is the one induced by the inclusion maps between the Rips complexes.

"Proof": [C., de Silva, Oudot 2012] + [Haussmann 95]

Then a shortest persistent homology basis containing the shortest loop can be computed from [T. Dey, J. Sun, Y. Wang 2010]

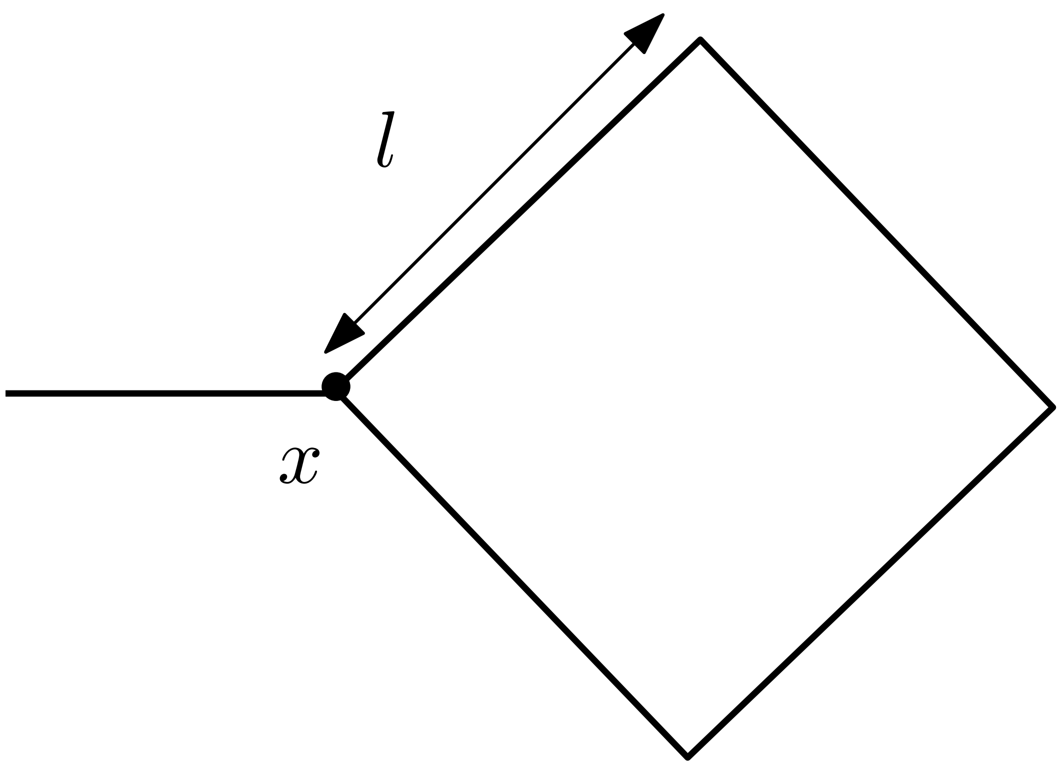
From trees to graphs

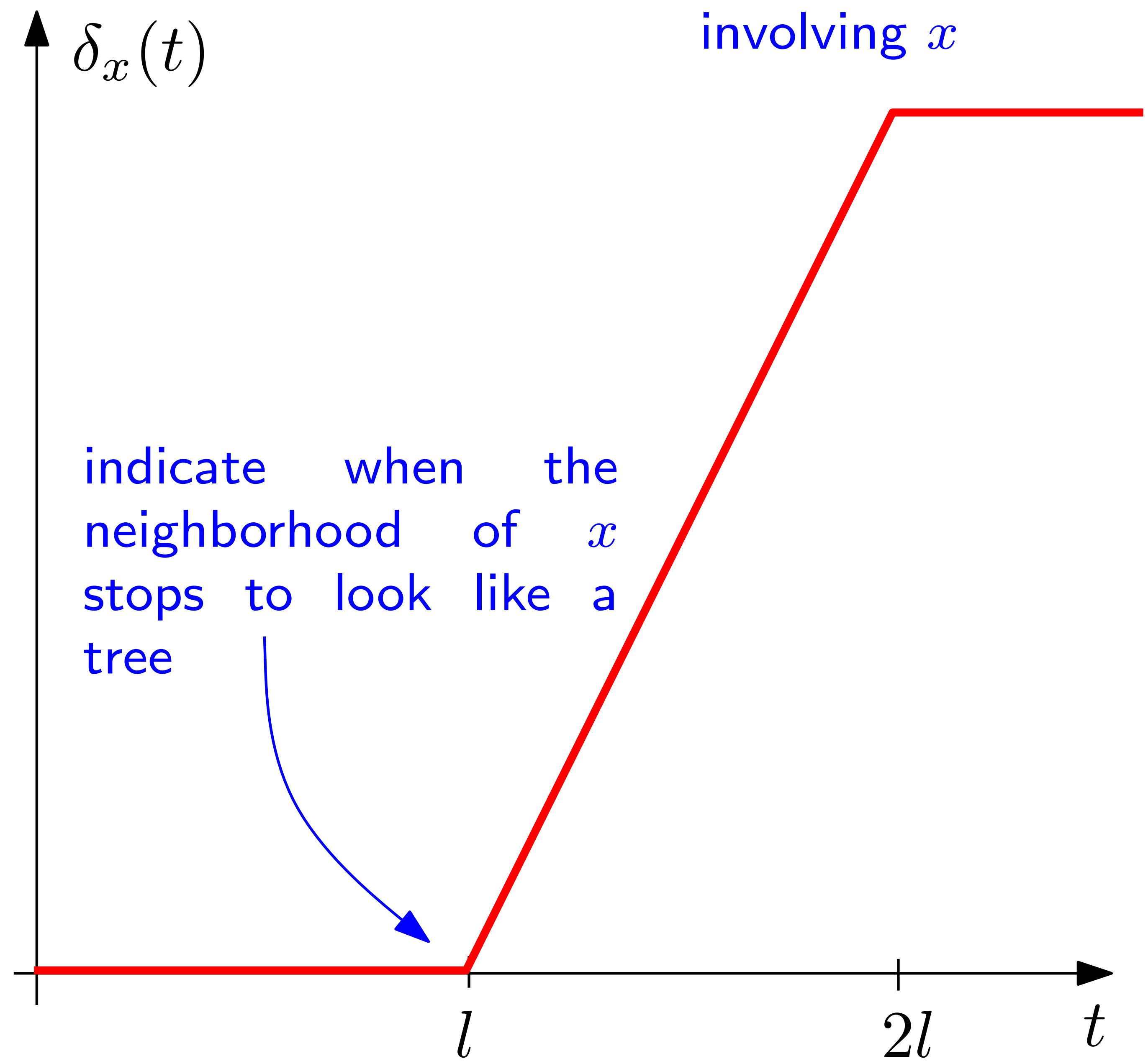
"Local" hyperbolicity can also be used to explore the "linearity" of the data.

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a metric space. For any $x \in \mathbb{X}$ define local hyperbolicity function at x, $\delta_x : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$\delta_x(t) = \inf\{\delta \geq 0 : (B(x,t), d_{\mathbb{X}}) \text{ is } \delta\text{-hyperbolic at } x\}$$

test only tetrahedra involving x





- $\delta_x(.)$ is stable w.r.t. d_{GH}.
- Can be made more robust to "noise" by replacing $\min$ by something like median,...

Advantages and drawbacks of the previous approach

Good points:

- A simple and efficient algorithm for tree recontruction/approximation.
- No parameter (except for the construction of the neighboring graph) and upper bounds!
- "Semi-local" exploration of the structure of metric data.

Advantages and drawbacks of the previous approach

Good points:

- A simple and efficient algorithm for tree recontruction/approximation.
- No parameter (except for the construction of the neighboring graph) and upper bounds!
- "Semi-local" exploration of the structure of metric data.

Bad points:

- No guarantees on the topology (homeomorphism).
- No global graph reconstruction:
  - covering the data by trees and gluying them together might lead to choice of parameters issues and pretty bad bounds on the metric approximation.

Advantages and drawbacks of the previous approach

Good points:

- A simple and efficient algorithm for tree recontruction/approximation.
- No parameter (except for the construction of the neighboring graph) and upper bounds!
- "Semi-local" exploration of the structure of metric data.

Bad points:

- No guarantees on the topology (homeomorphism).
- No global graph reconstruction:
  - covering the data by trees and gluying them together might lead to choice of parameters issues and pretty bad bounds on the metric approximation.

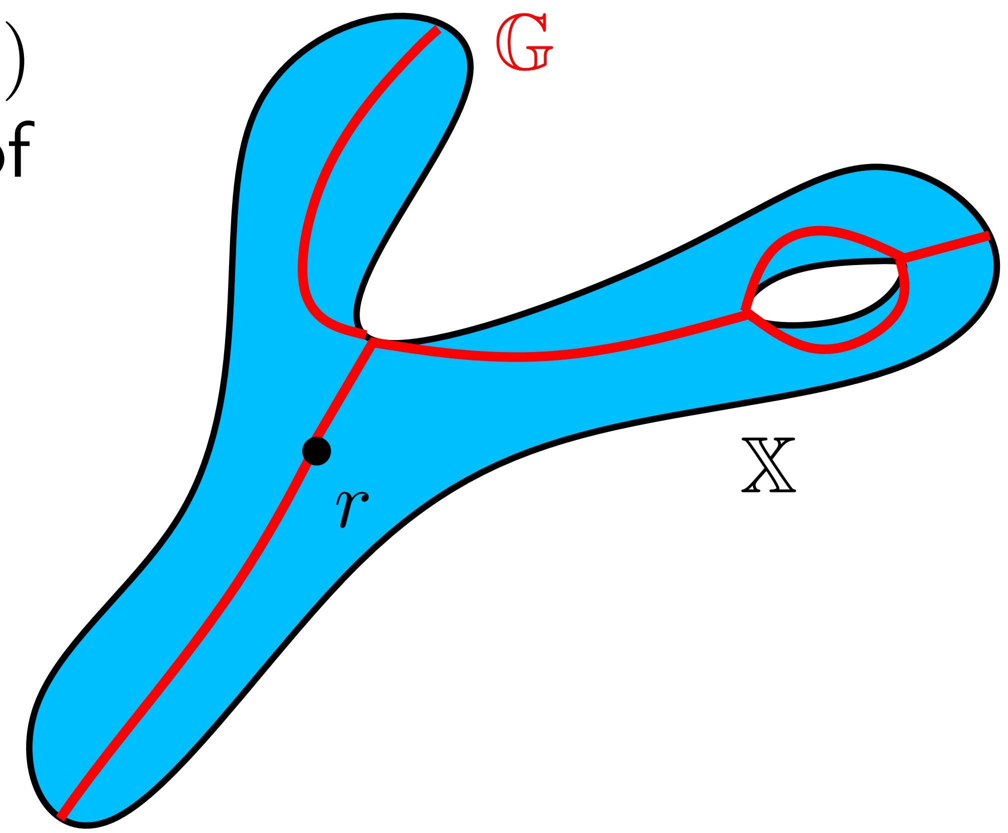
However...

Perspectives

The previous ideas and results can be used to prove some approximation results for the Reeb graph of the distance to a point in a path metric space.

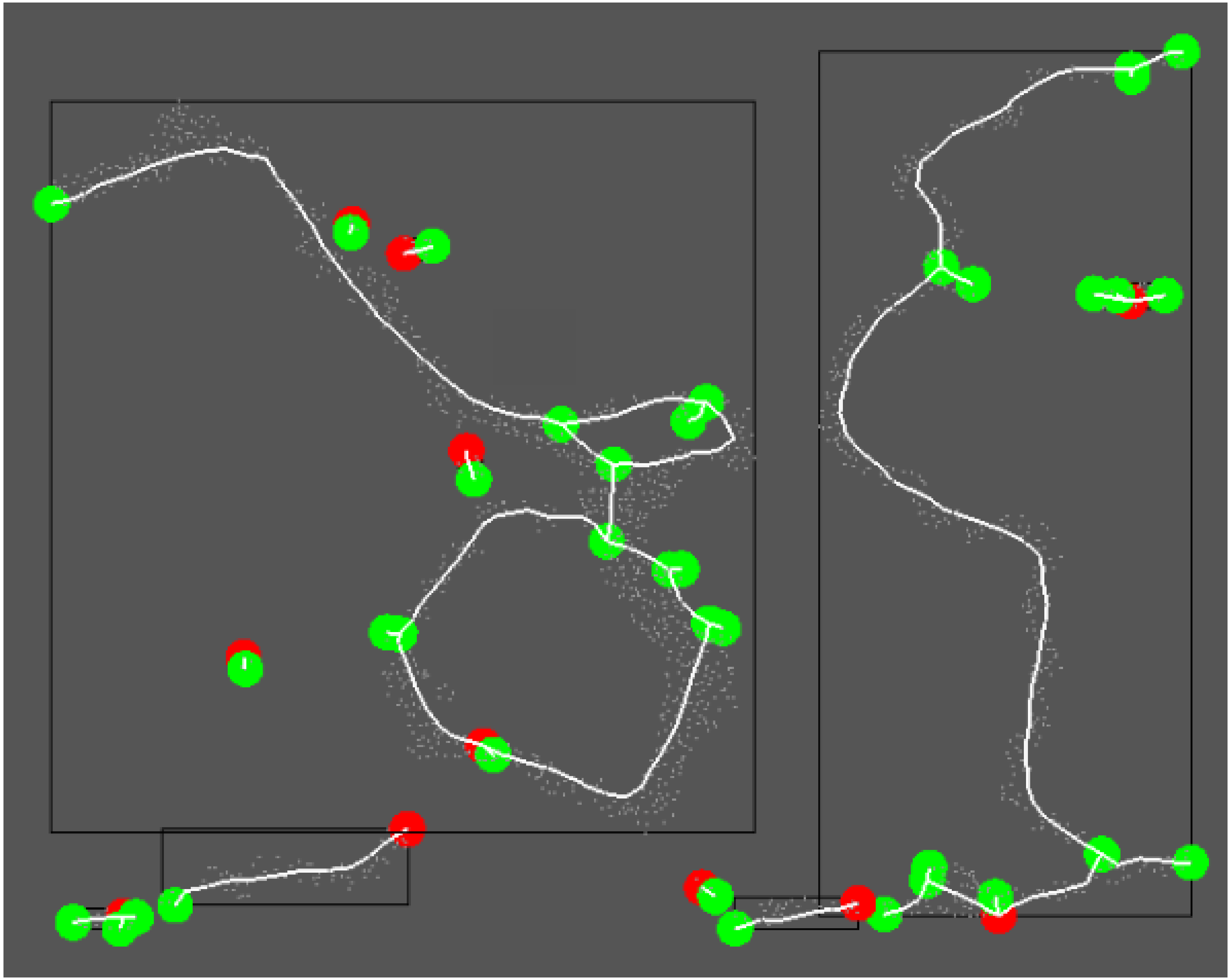
Equivalence relation: $x \sim y$ iff $d(x) = d(y)$ and x and y are contained in the same c.c. of $d^{-1}(d(x))$.

Reeb graph: $\mathbb{G} = \mathbb{X}/\sim$



Perspectives

The previous ideas and results can be used to prove some approximation results for the Reeb graph of the distance to a point in a path metric space.



Thank you for your attention!