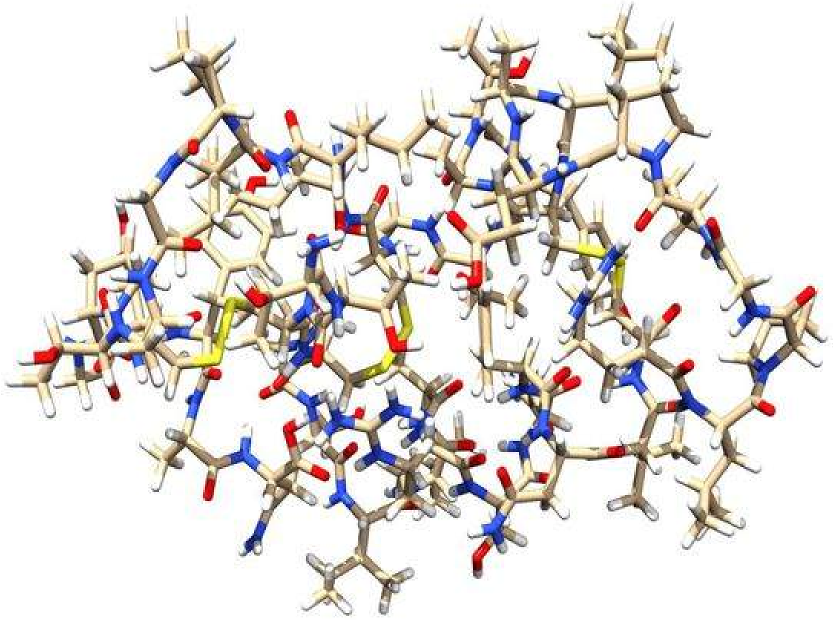
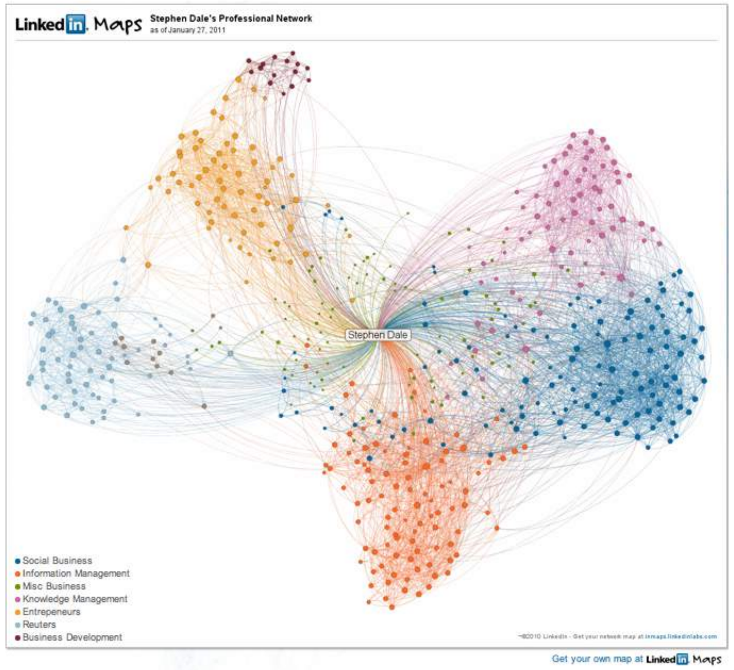
Geometry Understanding in Higher Dimensions
Collège de France, June 2017

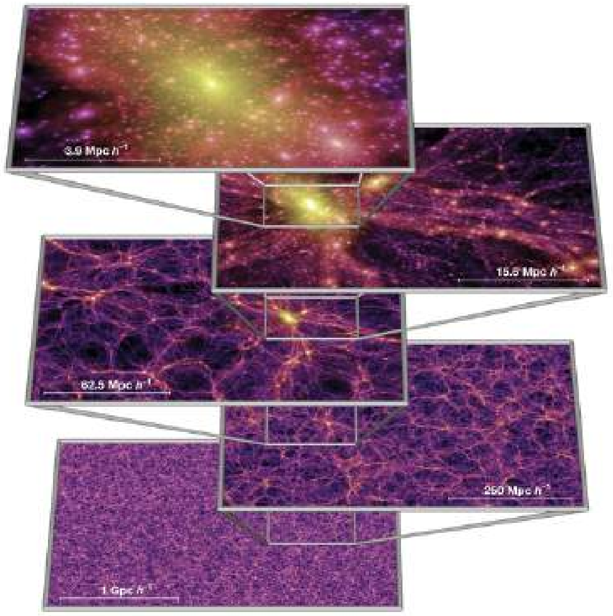
Topological Graphs for Data Analysis
Structure, Stability, and Statistics

Steve Oudot

Graph Structures in the Data Sciences

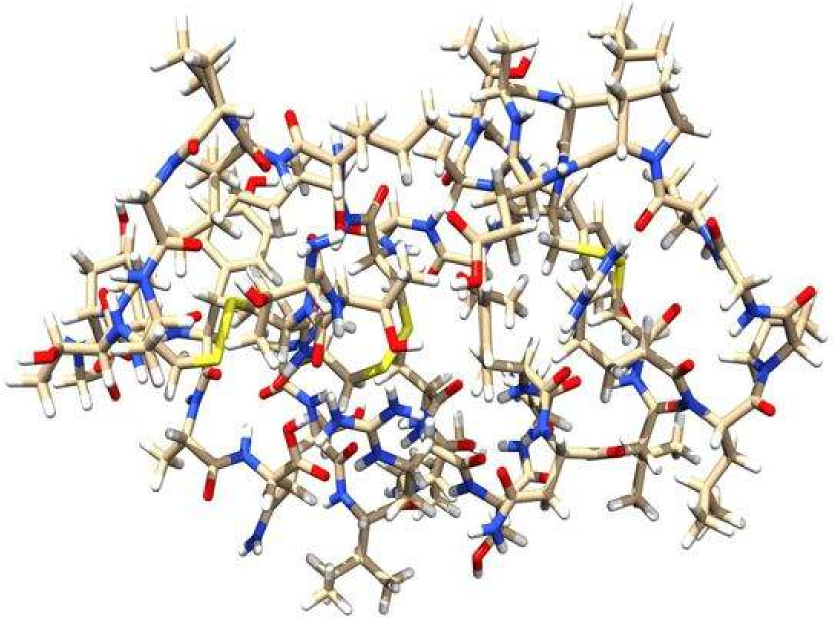


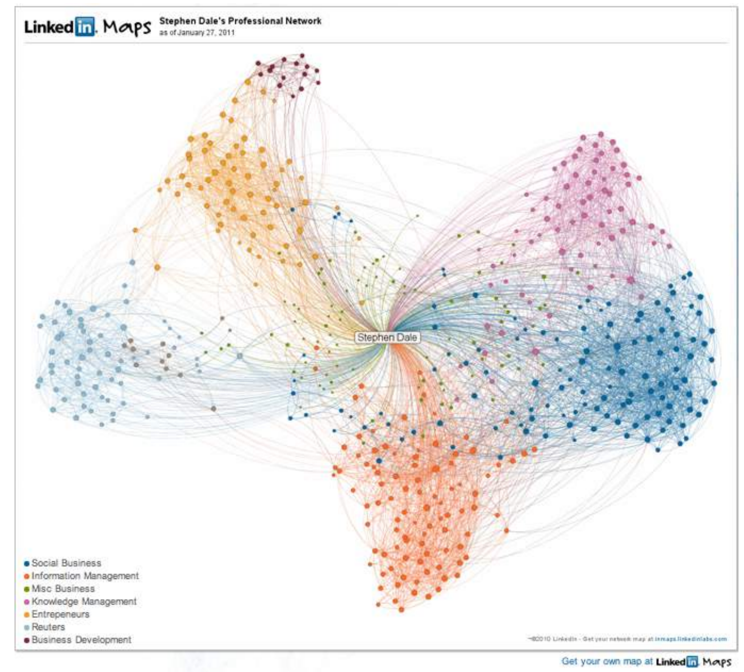


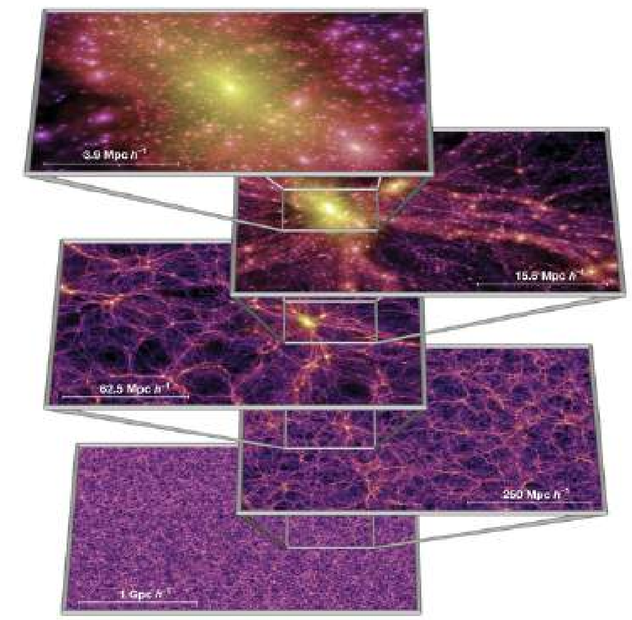


- Input: proteins, social networks, galaxies, etc.
- Statistical proxy: dendrograms, cluster trees, spanning trees, etc.
- Geometric proxy: neighborhood graphs, k-NN graphs, etc.
- Topological proxy: Reeb graphs, Joint Contour Nets, Mappers, etc.

Graph Structures in the Data Sciences

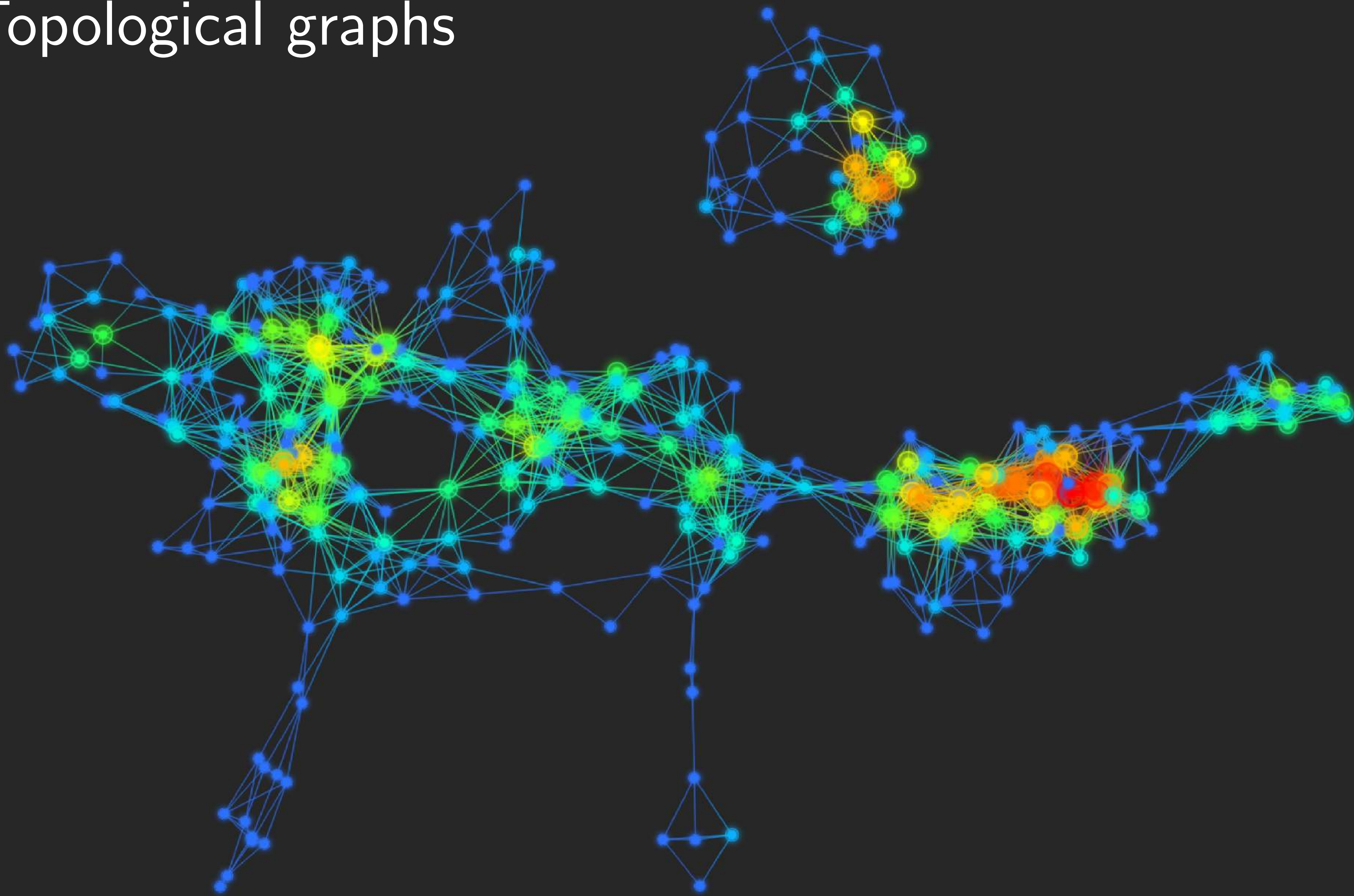






- Input: proteins, social networks, galaxies, etc.
- Statistical proxy: dendrograms, cluster trees, spanning trees, etc.
- Geometric proxy: neighborhood graphs, k-NN graphs, etc.
- Topological proxy: Reeb graphs, Joint Contour Nets, Mappers, etc.

Topological graphs

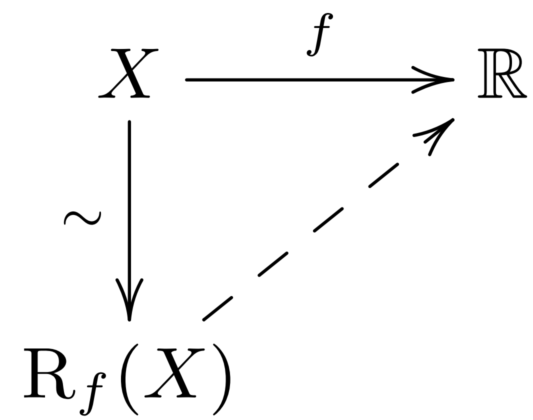


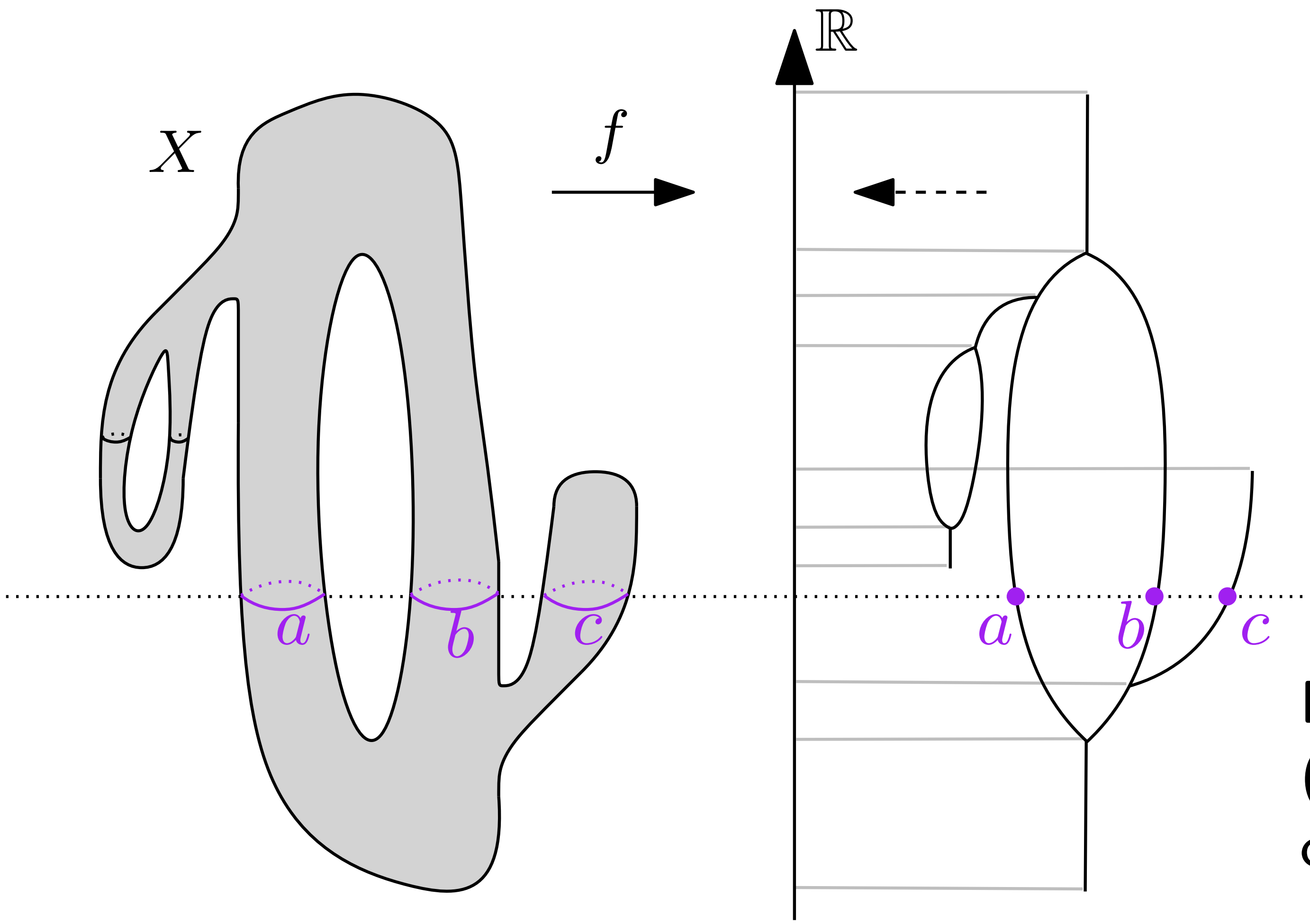
principle: summarize the topological structure of a map $f : X \to \mathbb{R}$ through a graph

Reeb Graph

$x \sim y \iff [\ f(x) = f(y)$ and x, y belong to same cc of $f^{-1}(\{f(x)\})\ ]$

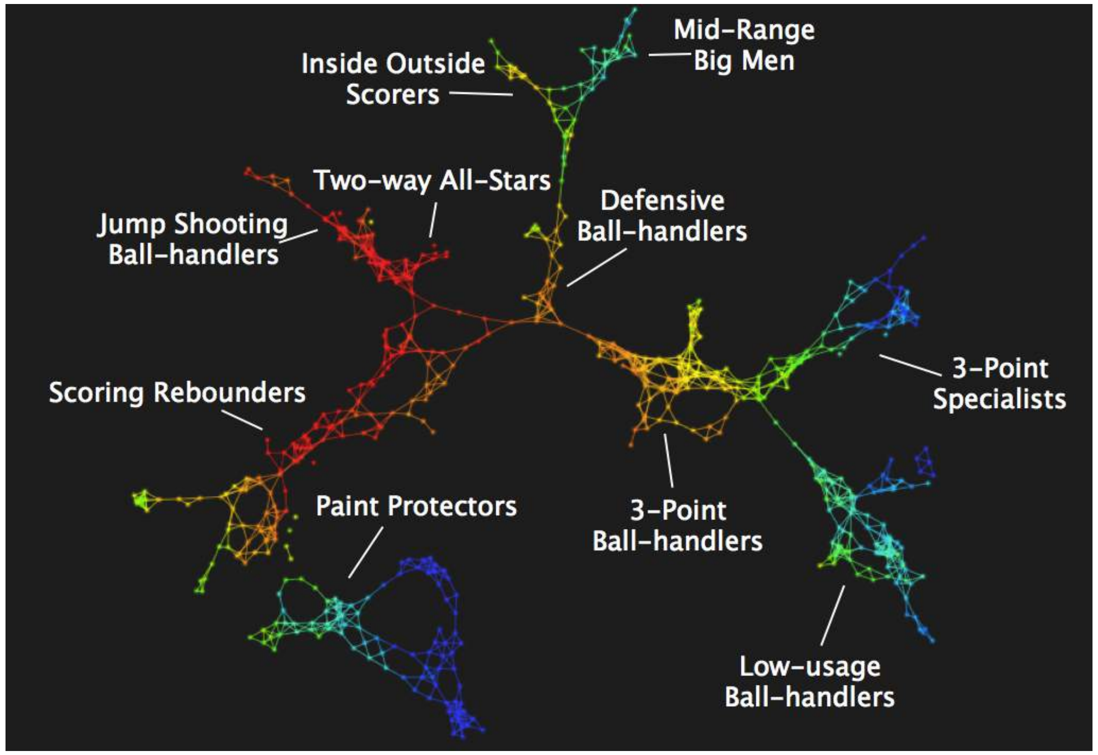
$\mathrm{R}_f(X) := X/\sim$



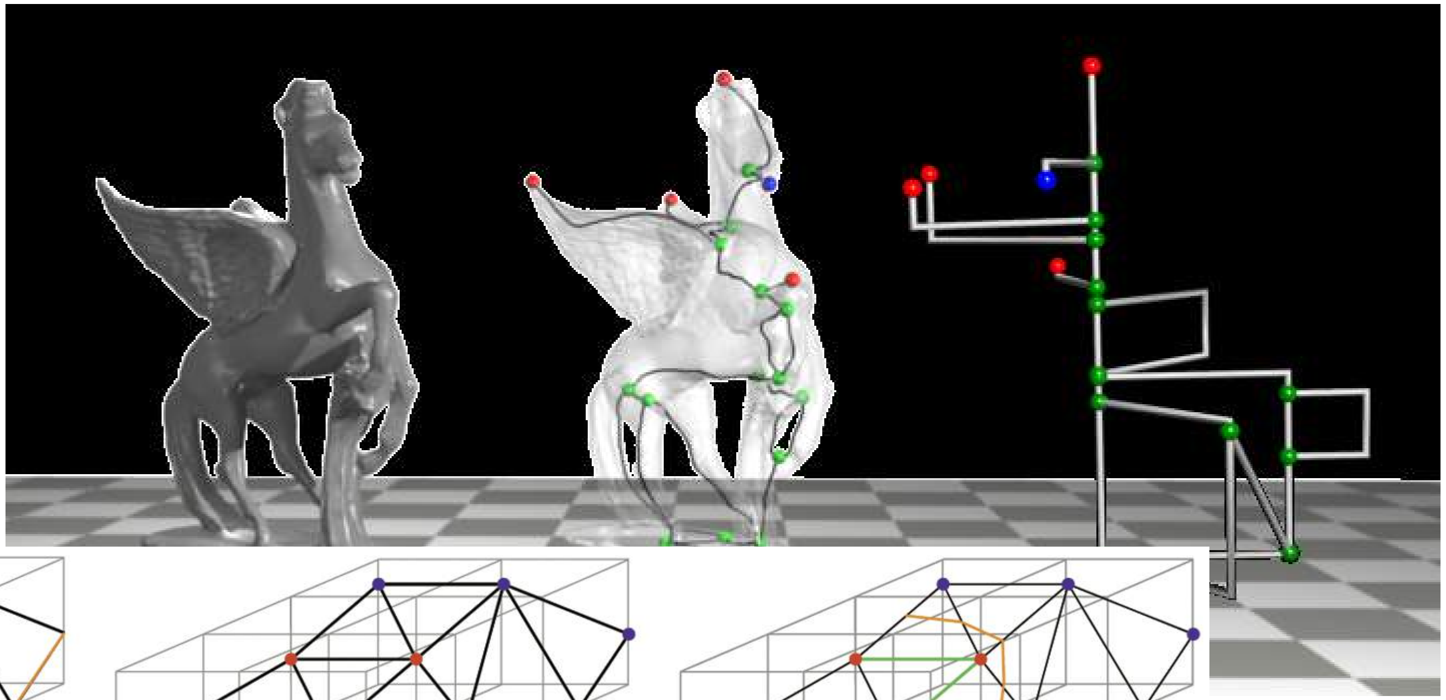


Prop: $\mathrm{R}_f(X)$ is a 1-d stratified space (*graph*) e.g. when (X, f) is Morse, or more generally of **Morse type**

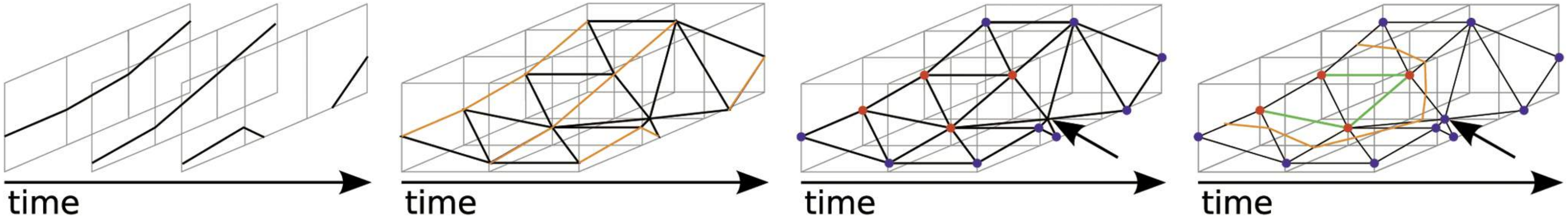
Applications of Reeb graphs



- Scientific visualization
- Skeletonization, parametrization
- Data comparison, segmentation, matching, property transfer, ...

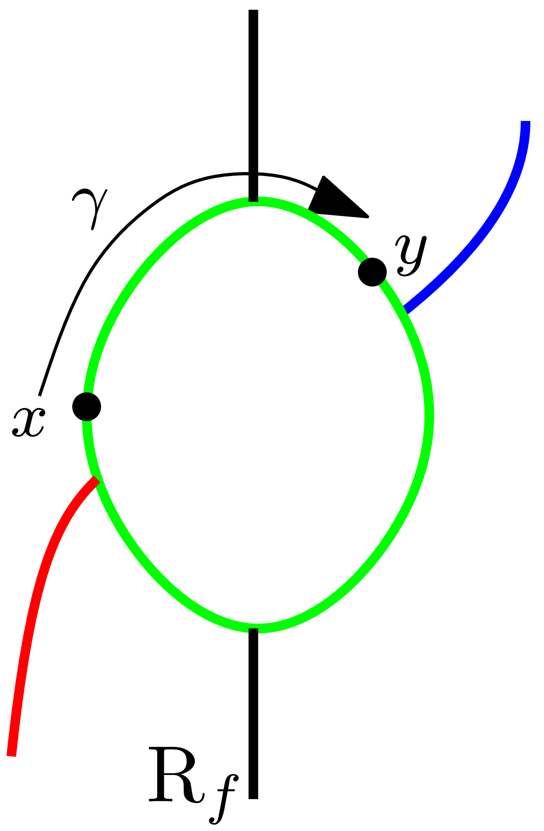


- Time-varying data
- ...



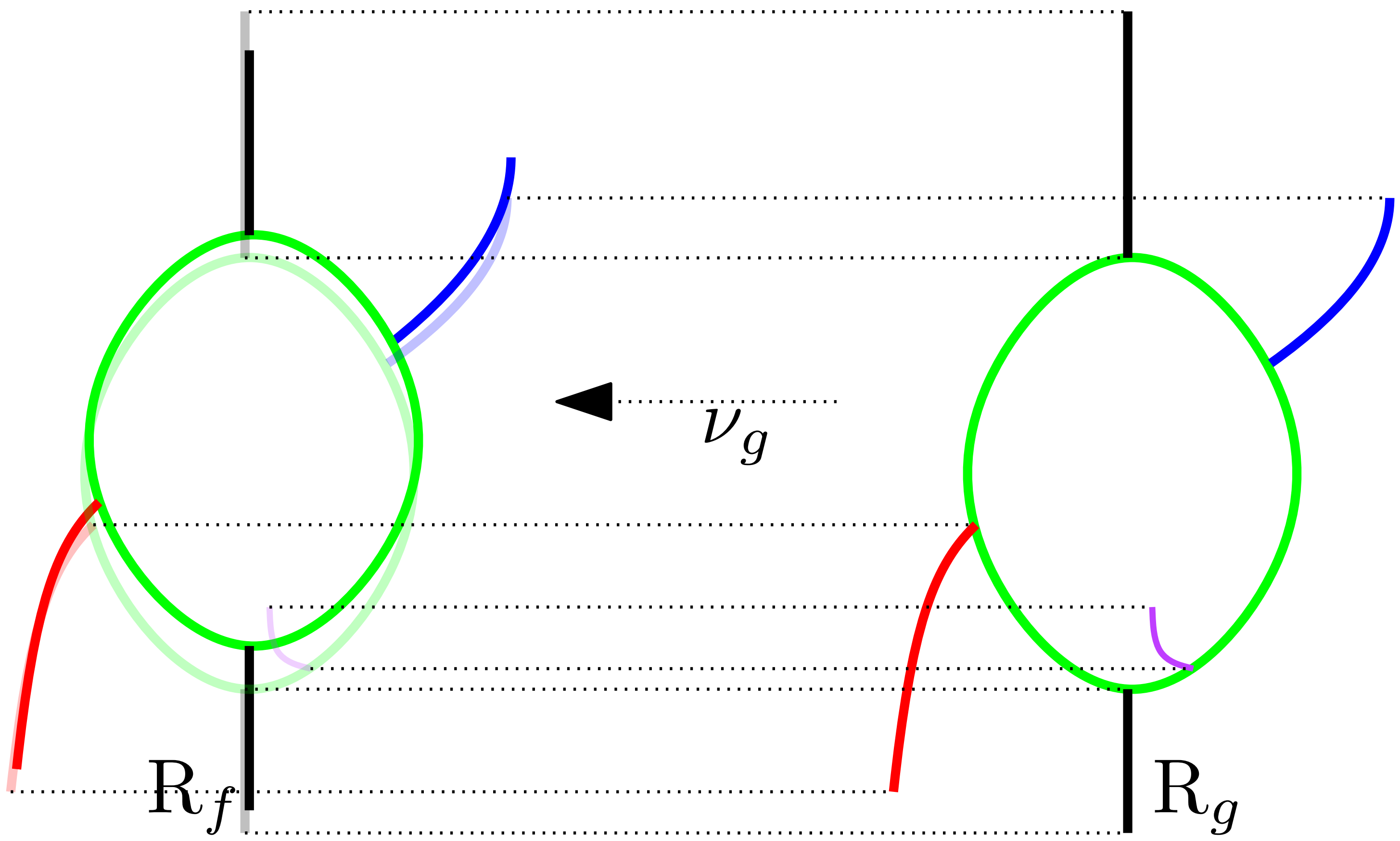
(a) (b) (c) (d)

Reeb graphs as metric spaces



Def: $\mathrm{d}_f(x,y) := \inf_{\gamma : x \rightsquigarrow y} \max f \circ \gamma - \min f \circ \gamma$

Reeb graphs as metric spaces

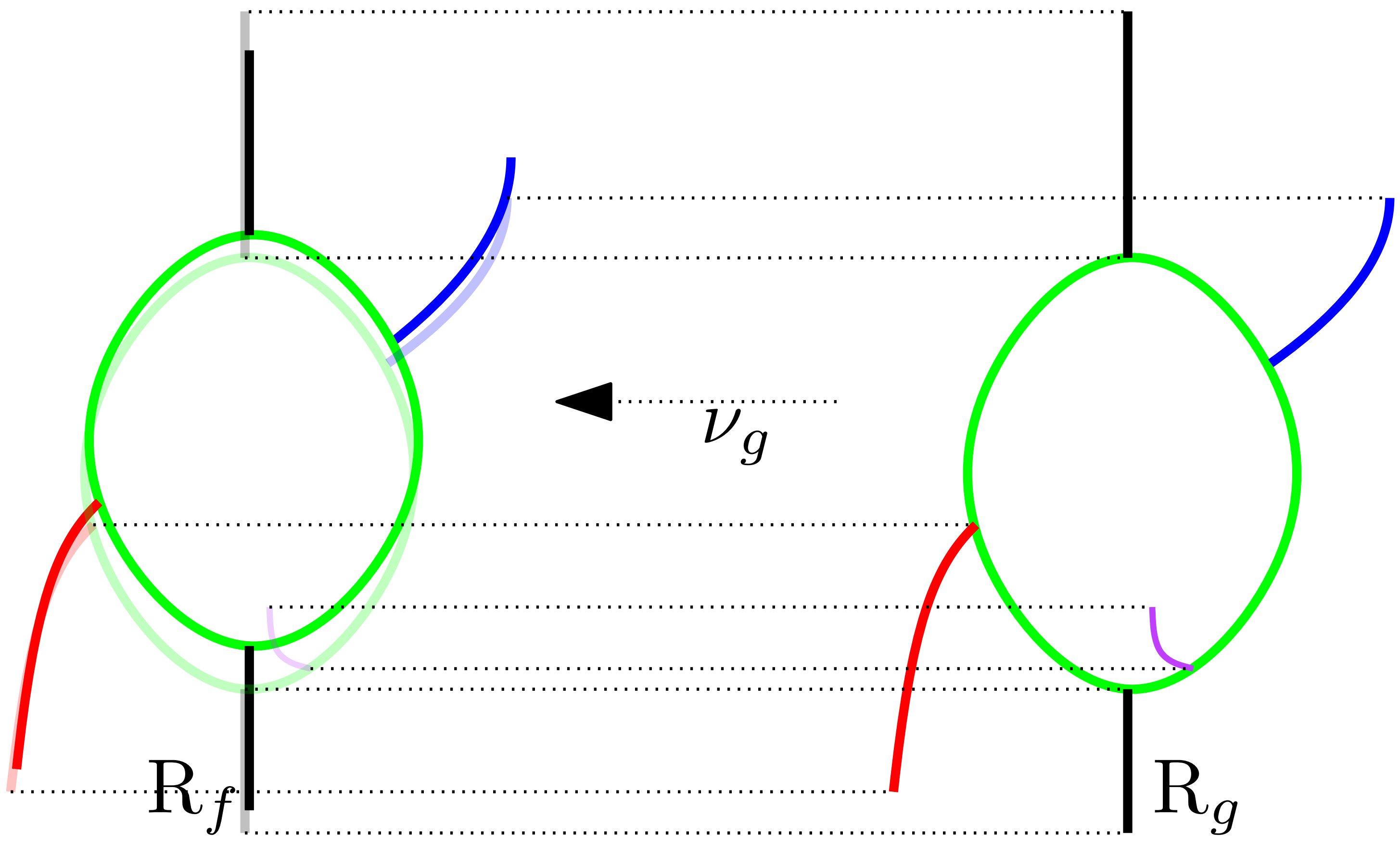


Gromov-Hausdorff distance:

Def: $\mathrm{d}_{\mathrm{GH}}(\mathrm{R}_f, \mathrm{R}_g) := \inf_{\nu_f, \nu_g} \mathrm{d}_{\mathrm{H}}(\nu_f(\mathrm{R}_f),\, \nu_g(\mathrm{R}_g))$

Note: $\mathrm{d}_{\mathrm{H}}(X, Y) = \inf\{\varepsilon \mid Y \subseteq \bigcup_{x \in X} B(x, \varepsilon)$ and $X \subseteq \bigcup_{y \in Y} B(y, \varepsilon)\}$

Reeb graphs as metric spaces

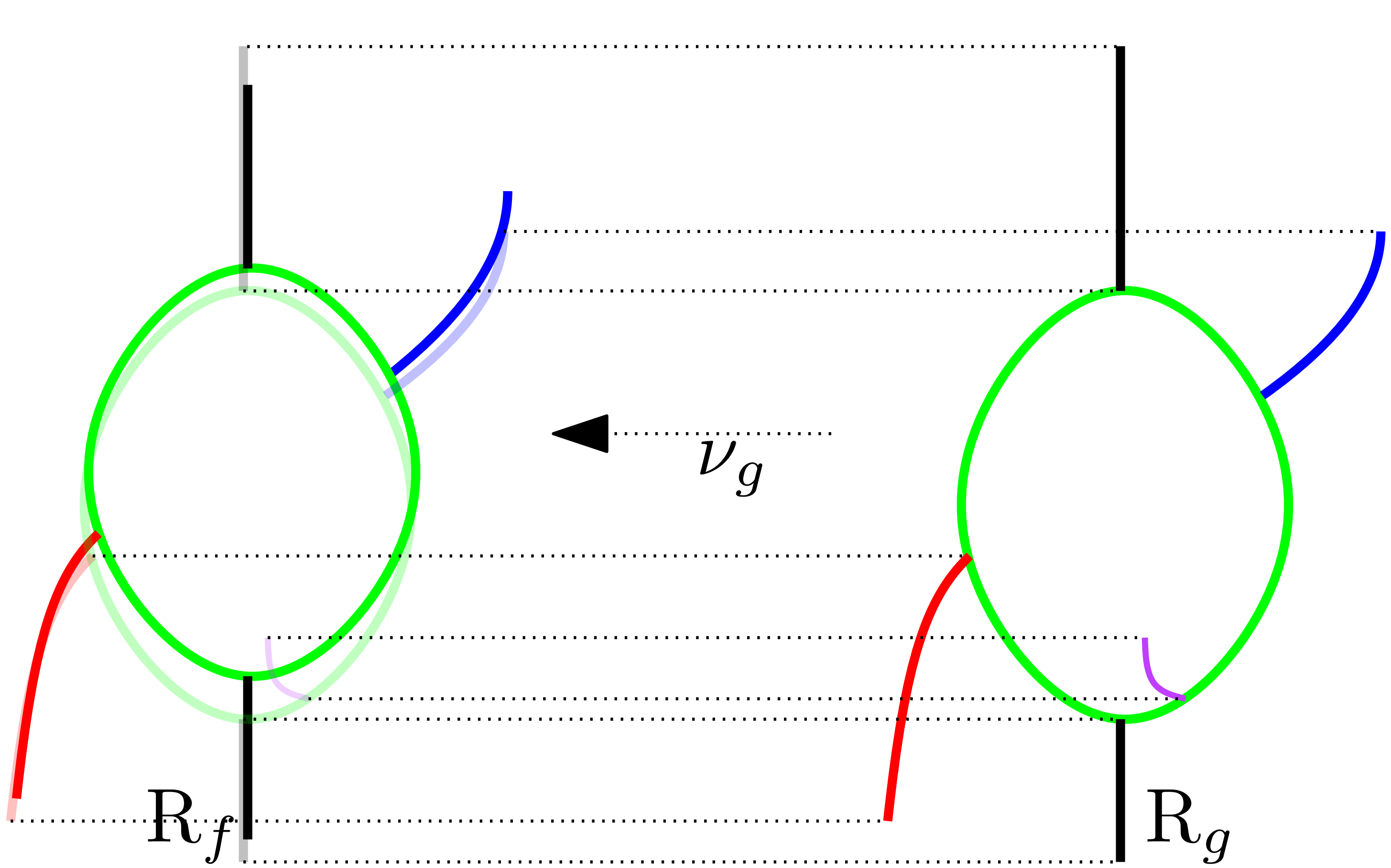


Gromov-Hausdorff distance:

Def: $d_{GH}(R_f, R_g) := \inf_{\nu_f,\nu_g} d_H(\nu_f(R_f),\, \nu_g(R_g))$

d_{GH} is hard to compute, even for metric trees [Agarwal et al. 2015]

Reeb graphs as metric spaces



variants and simplifications:

- correspondences in product space [Gromov]
- correspondences from continuous maps [Bauer et al.]
- edit distances [di Fabio, Landi]
- interleaving distances [Morozov et al.] [de Silva et al.]
- descriptor distances

Gromov-Hausdorff distance:

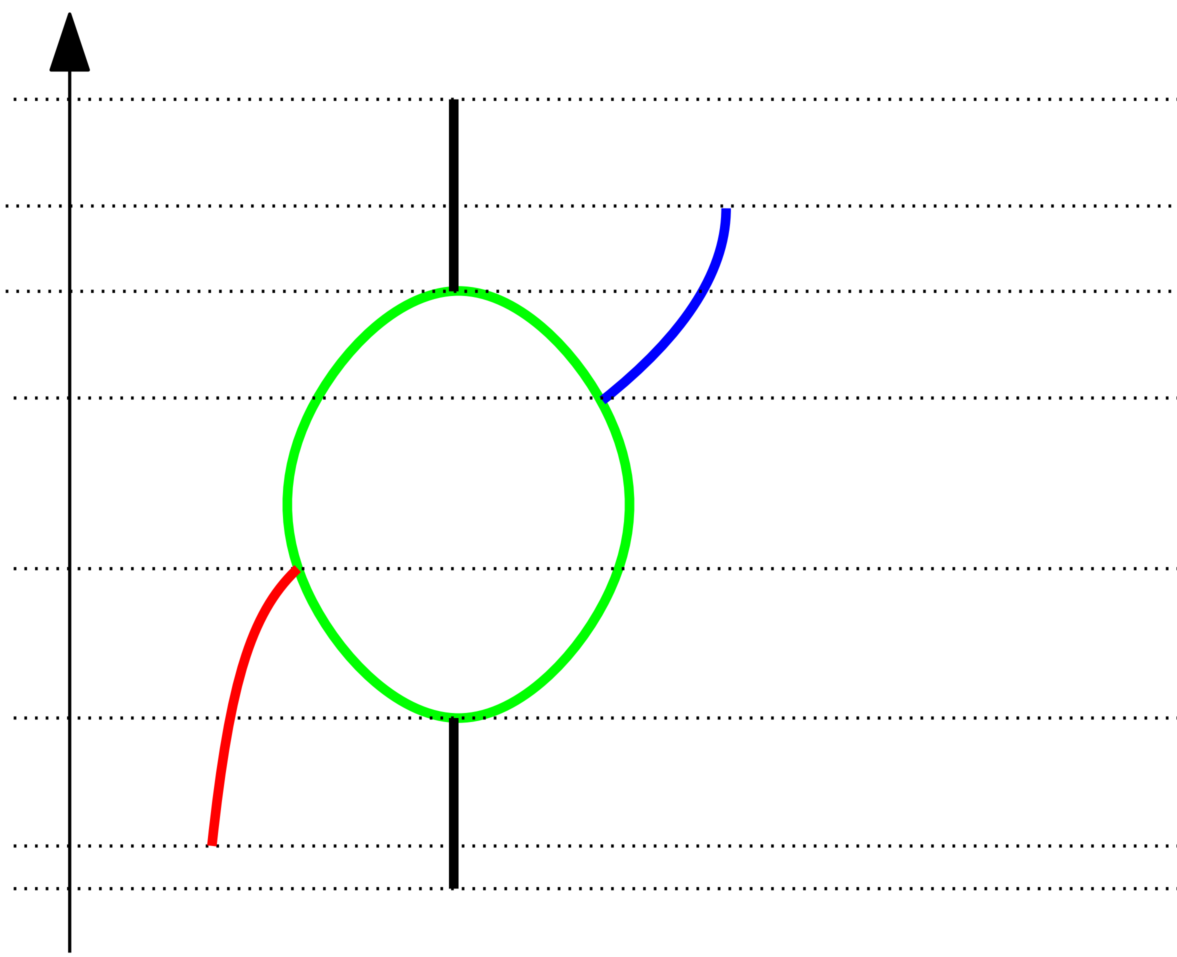
Def: $d_{GH}(R_f, R_g) := \inf_{\nu_f,\nu_g} d_H(\nu_f(R_f),\, \nu_g(R_g))$

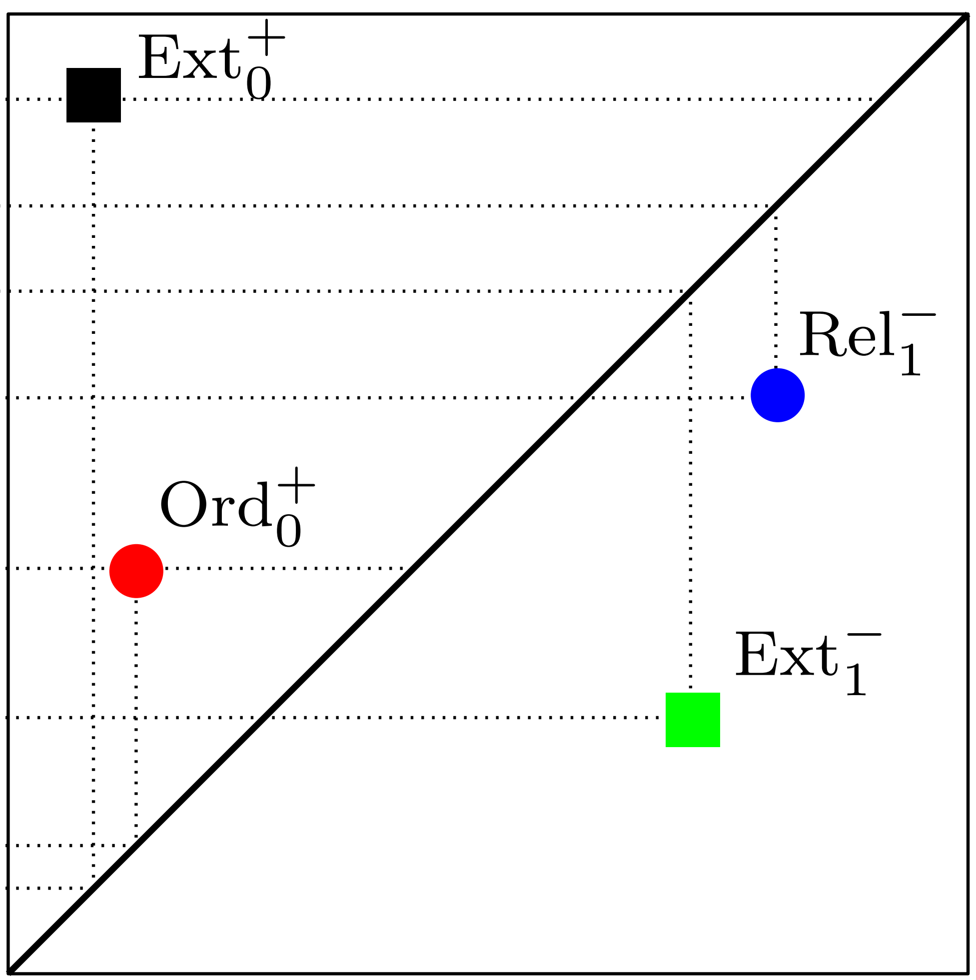
d_{GH} is hard to compute, even for metric trees [Agarwal et al. 2015]

Descriptors for Reeb graphs

$\mathrm{Dg}\,\mathrm{R}_f$: **bag-of-features** descriptor for $\mathrm{R}_f(X)$:

$\mathrm{Ord}_0\mathrm{R}_f \longleftrightarrow$ downward branches	$\mathrm{Ext}_0\mathrm{R}_f \longleftrightarrow$ trunks (cc)
$\mathrm{Rel}_1\mathrm{R}_f \longleftrightarrow$ upward branches	$\mathrm{Ext}_1\mathrm{R}_f \longleftrightarrow$ loops





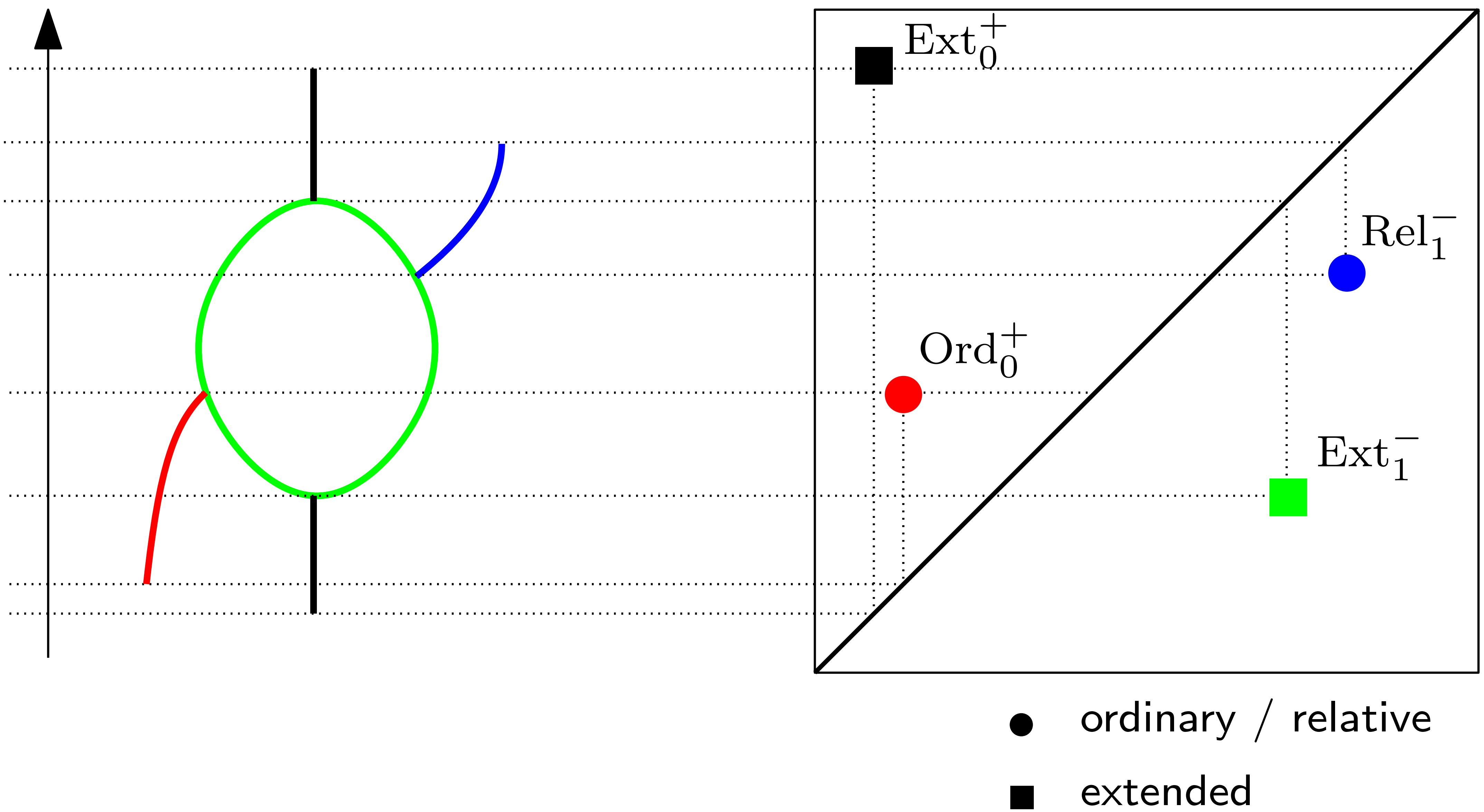
● ordinary / relative

■ extended

Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

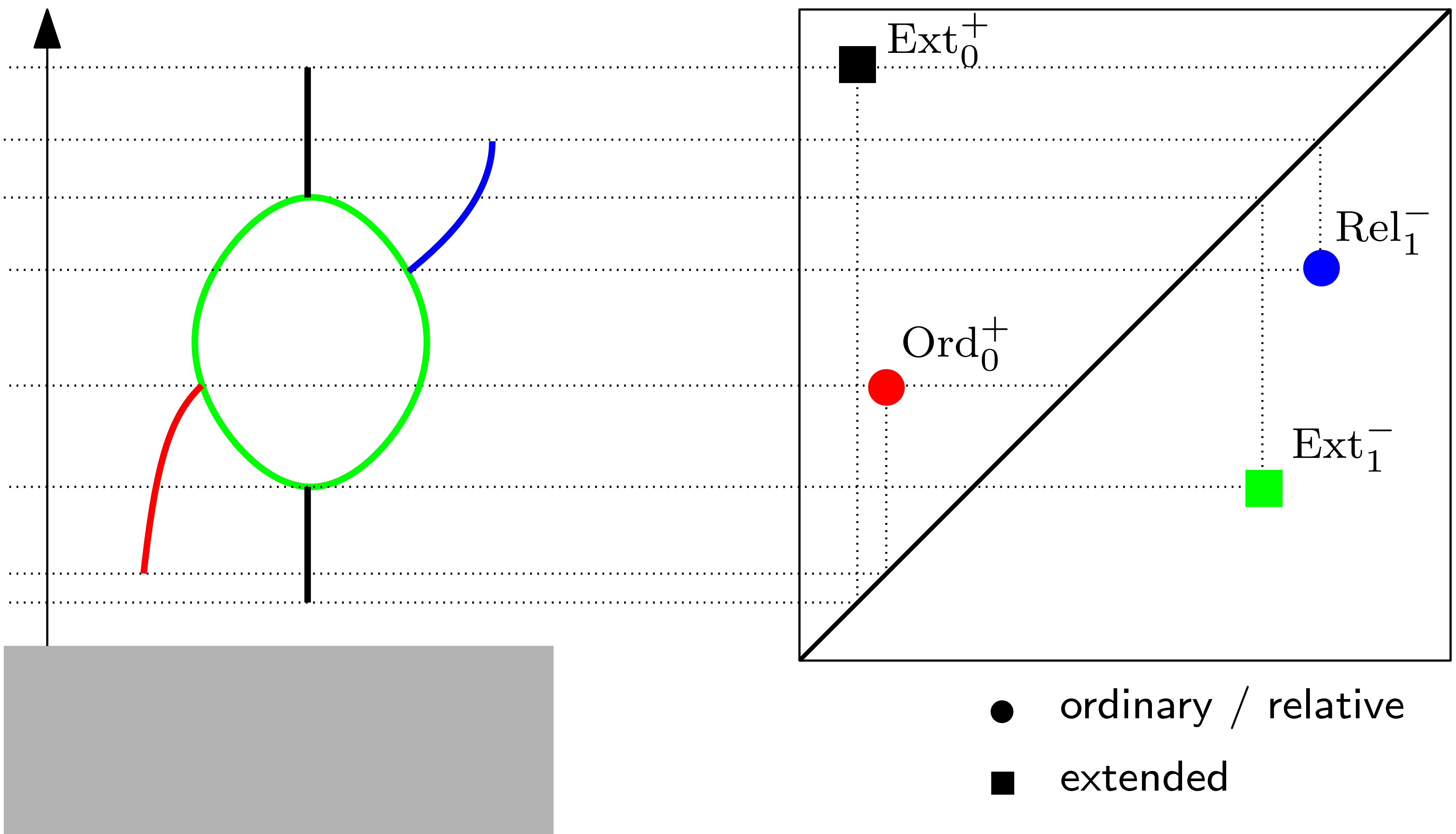
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

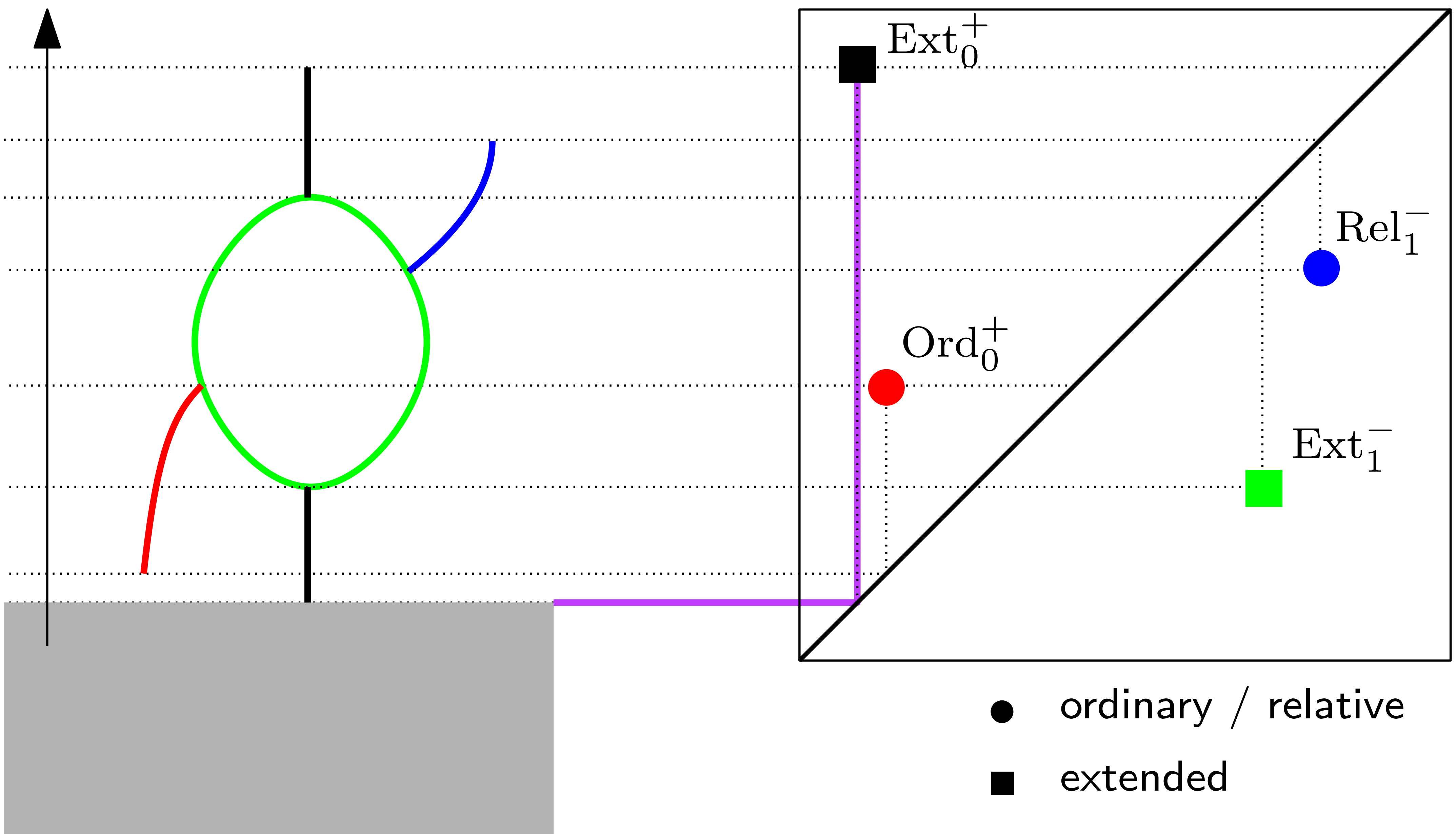
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

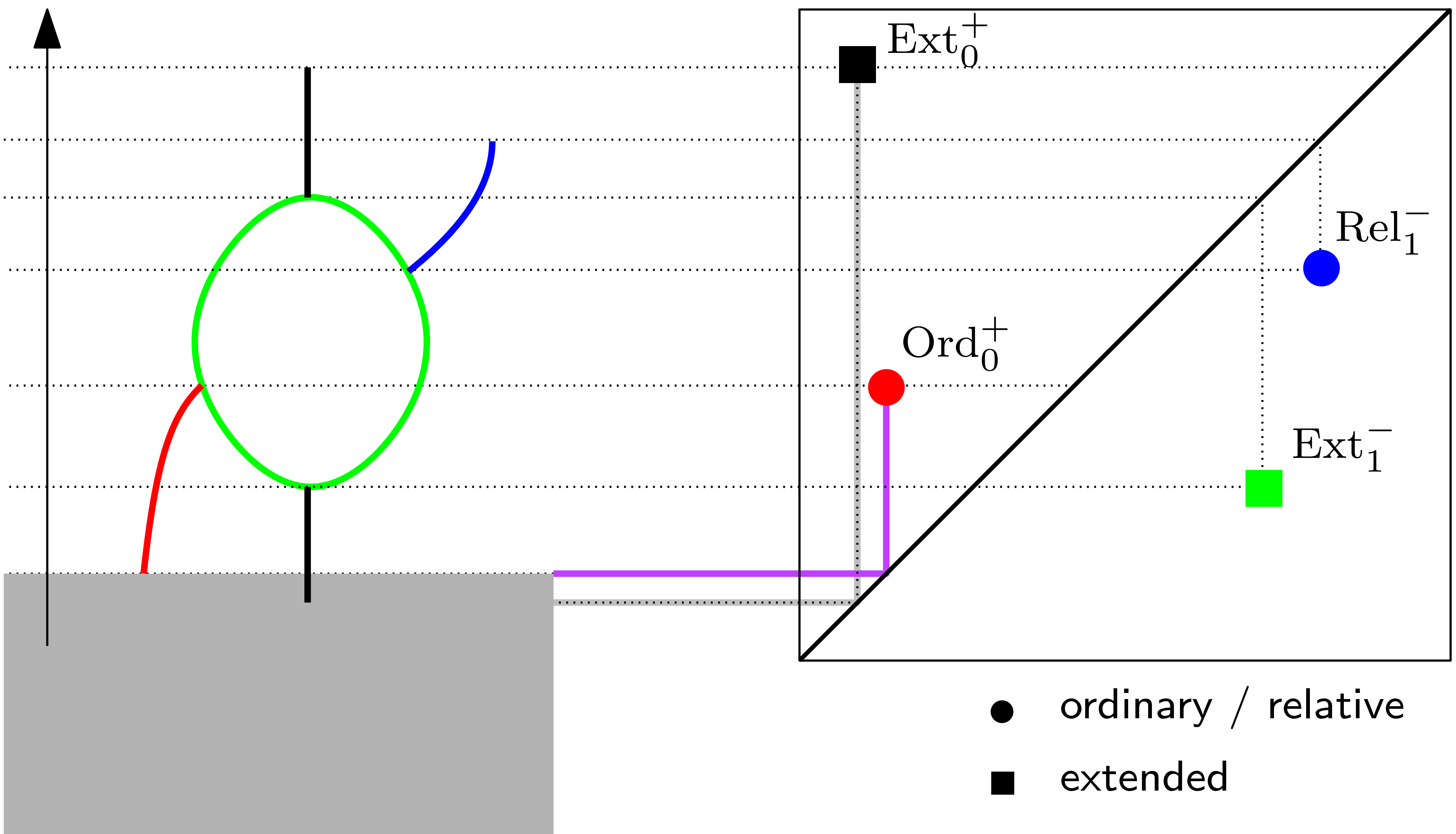
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

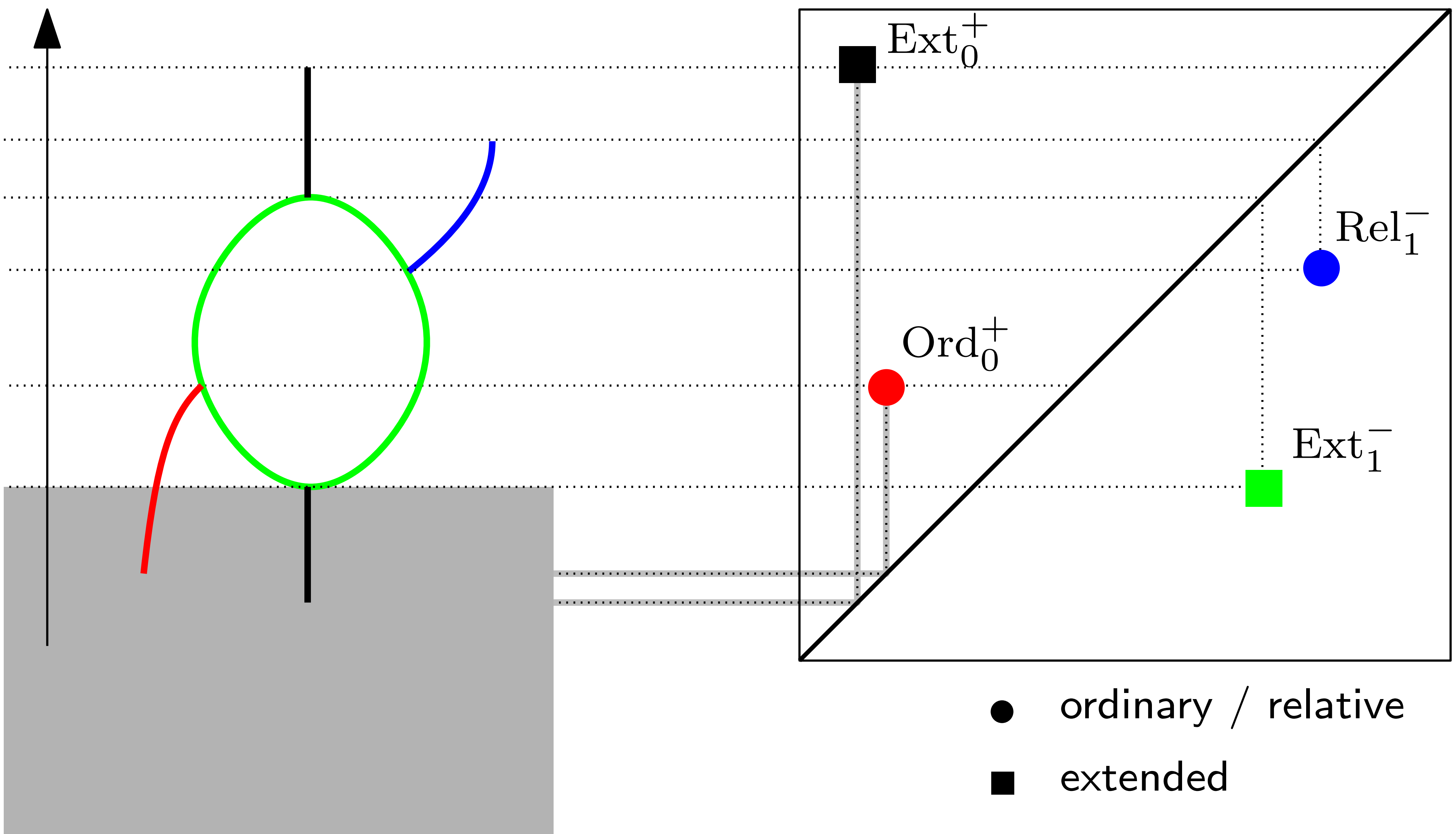
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

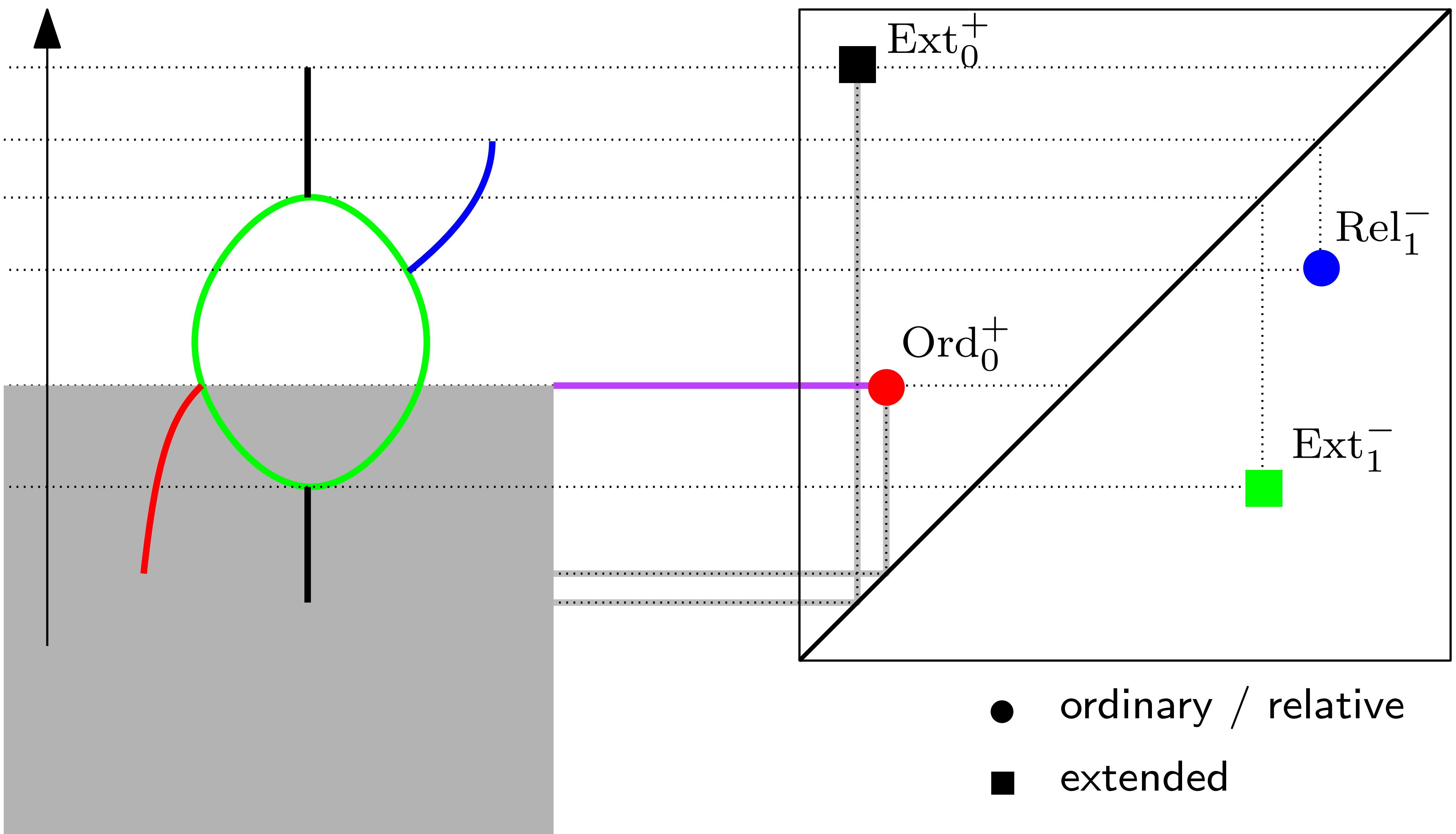
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

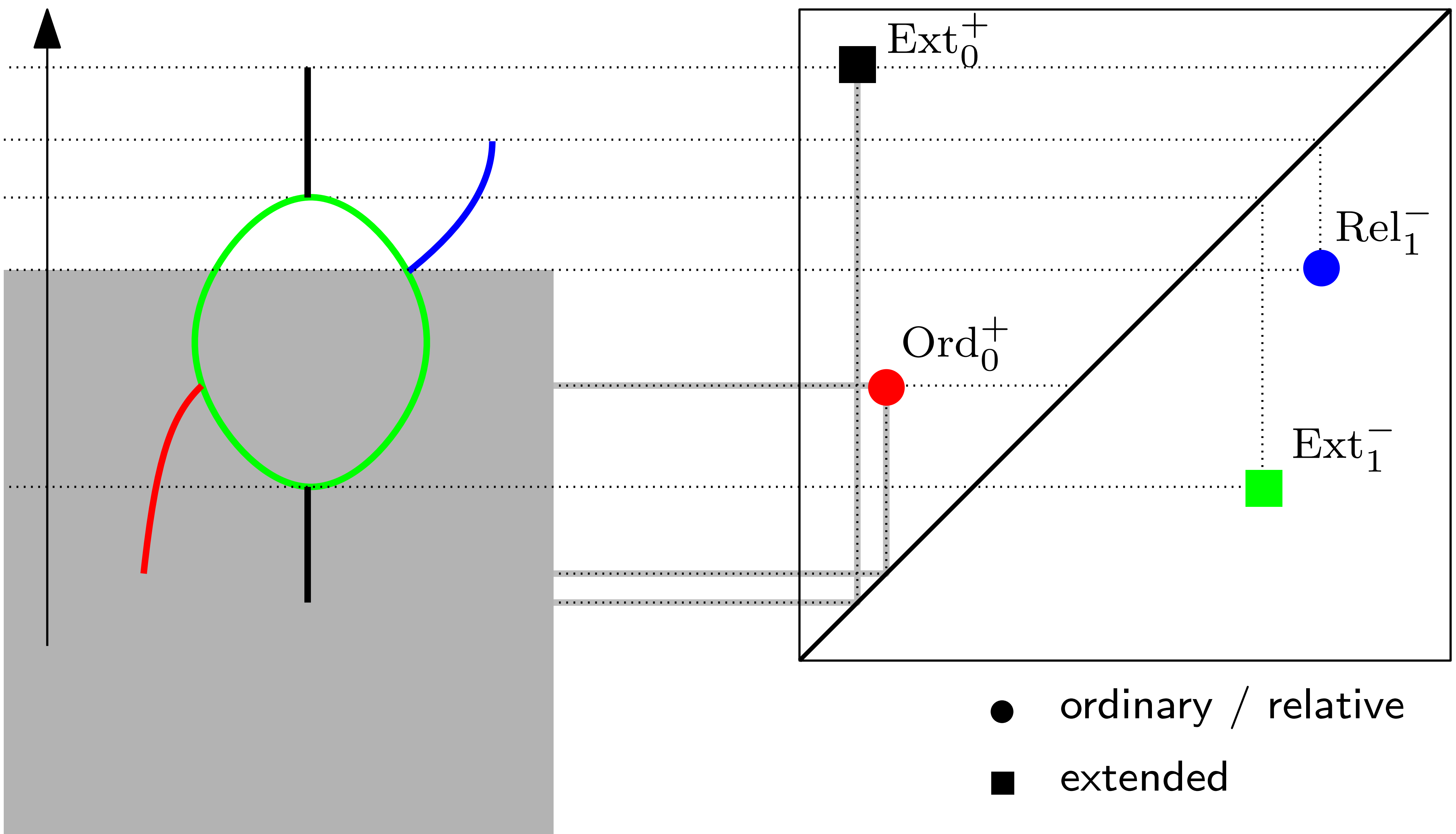
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

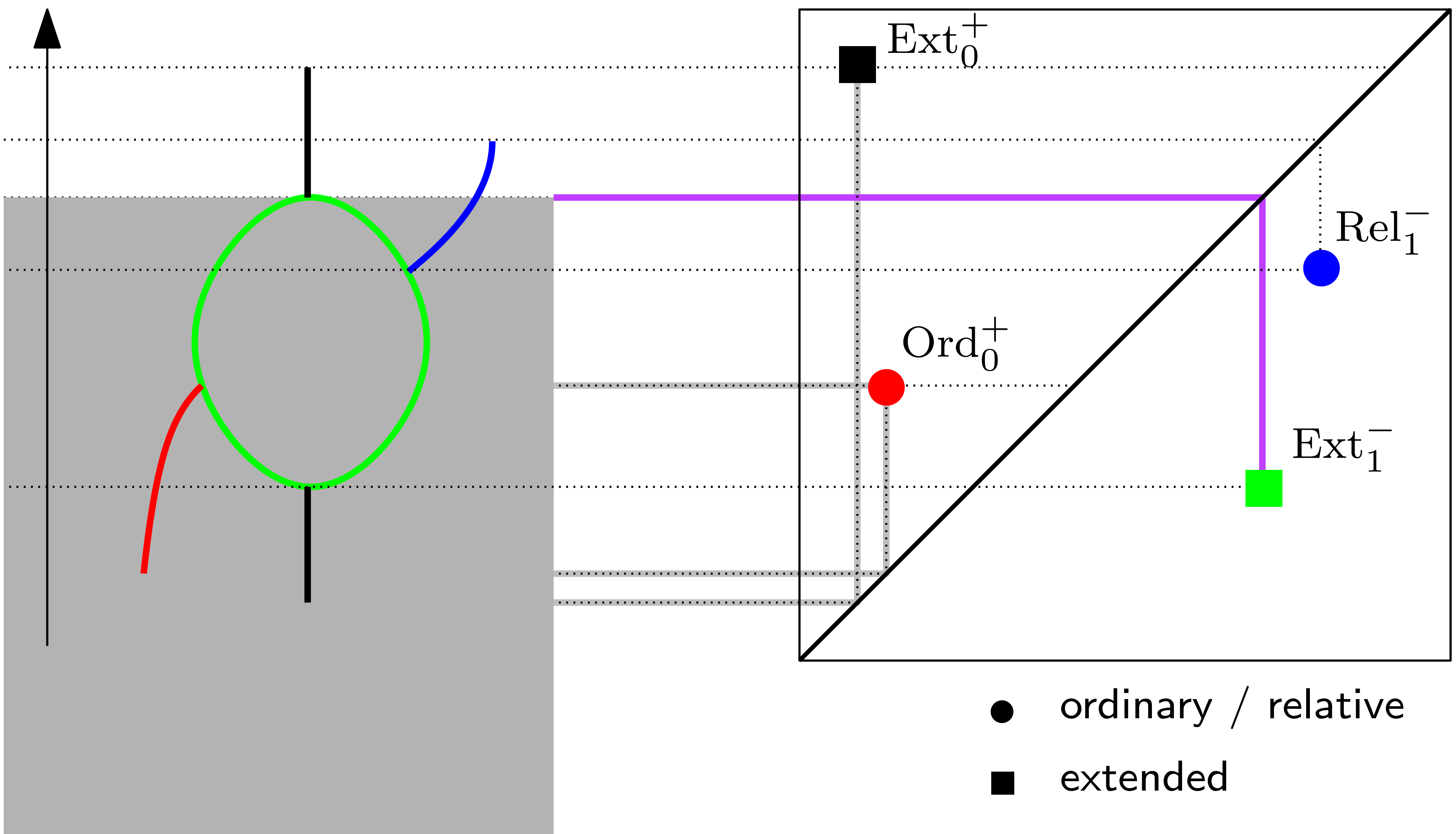
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family

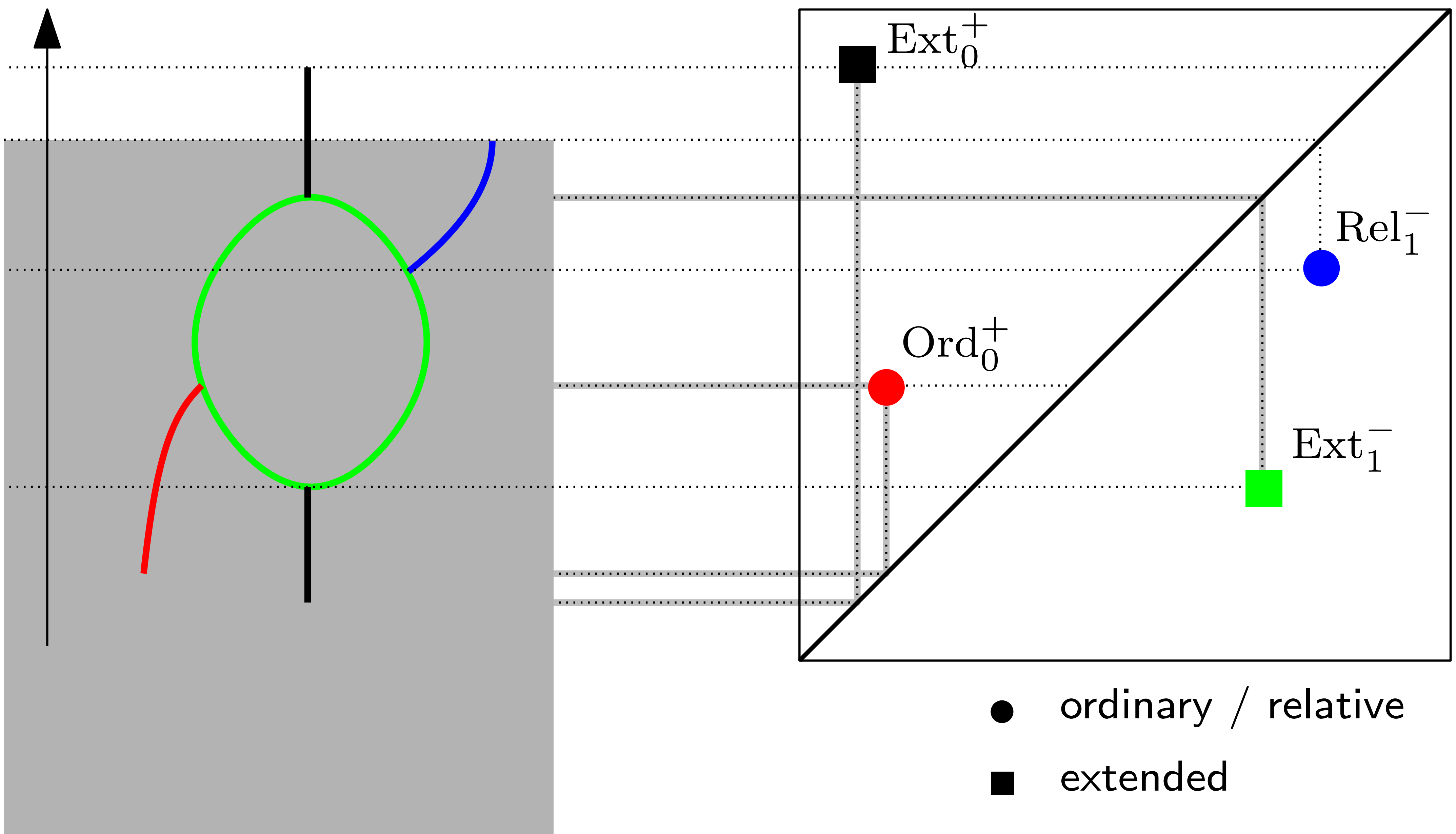


- ordinary / relative
- extended

Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

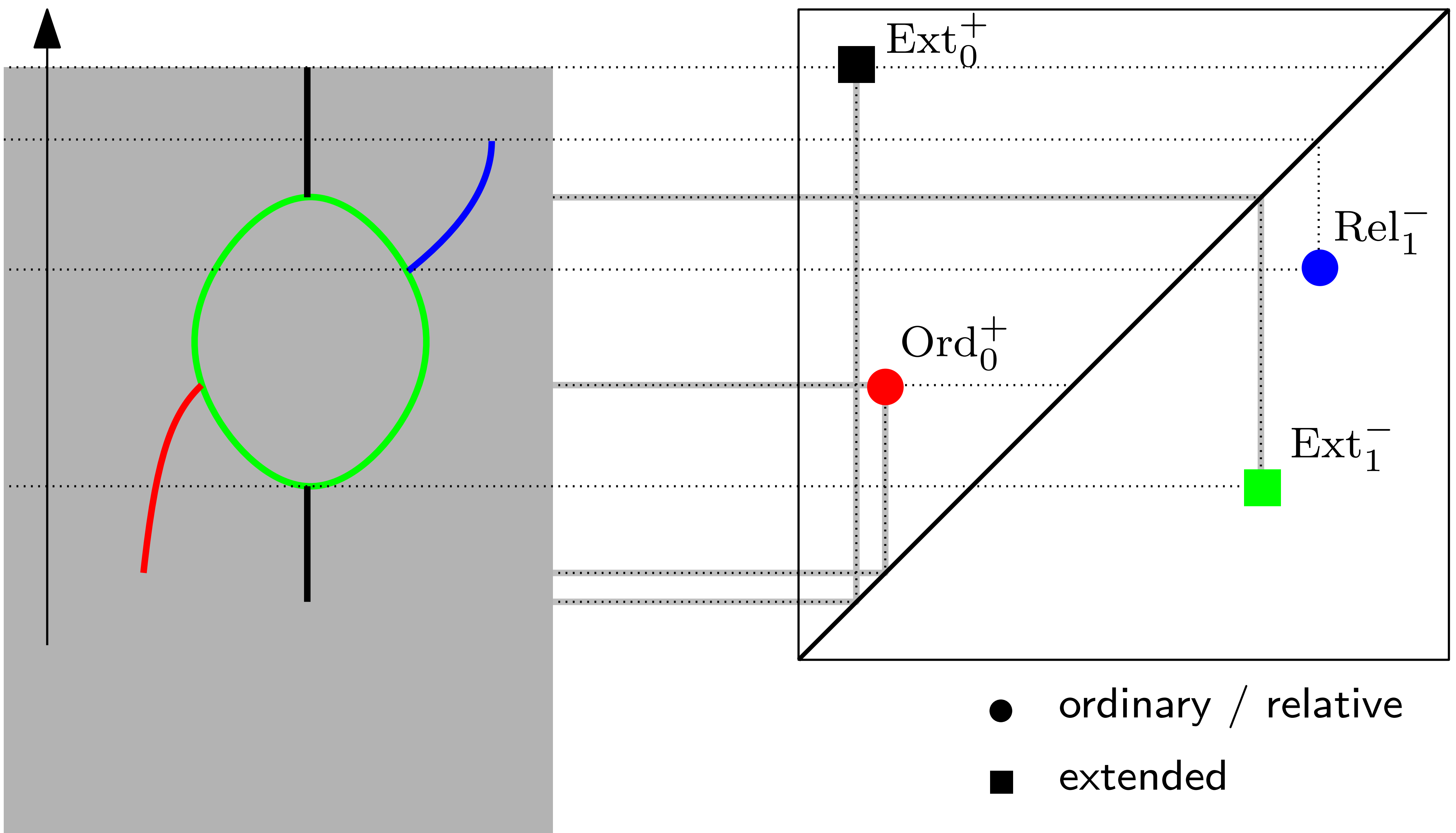
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

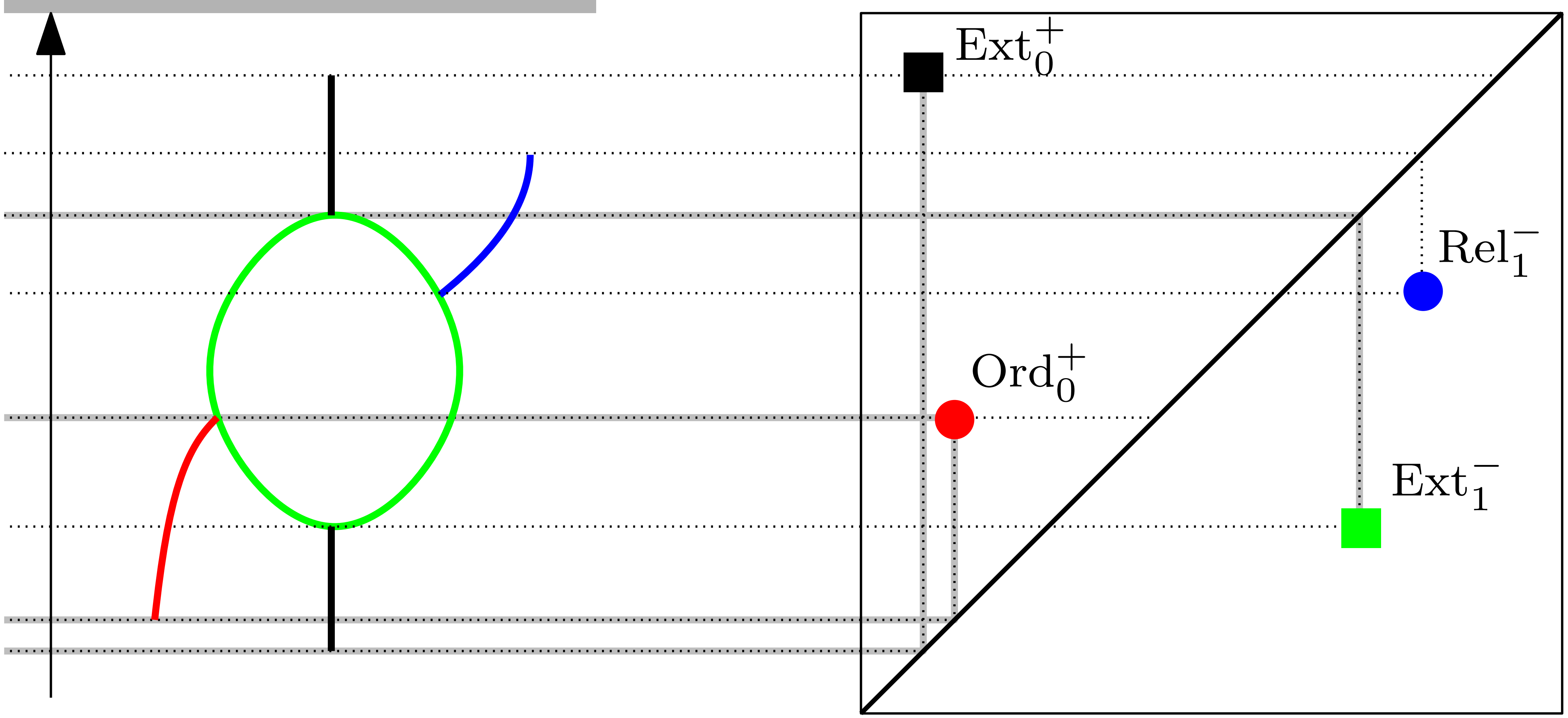
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



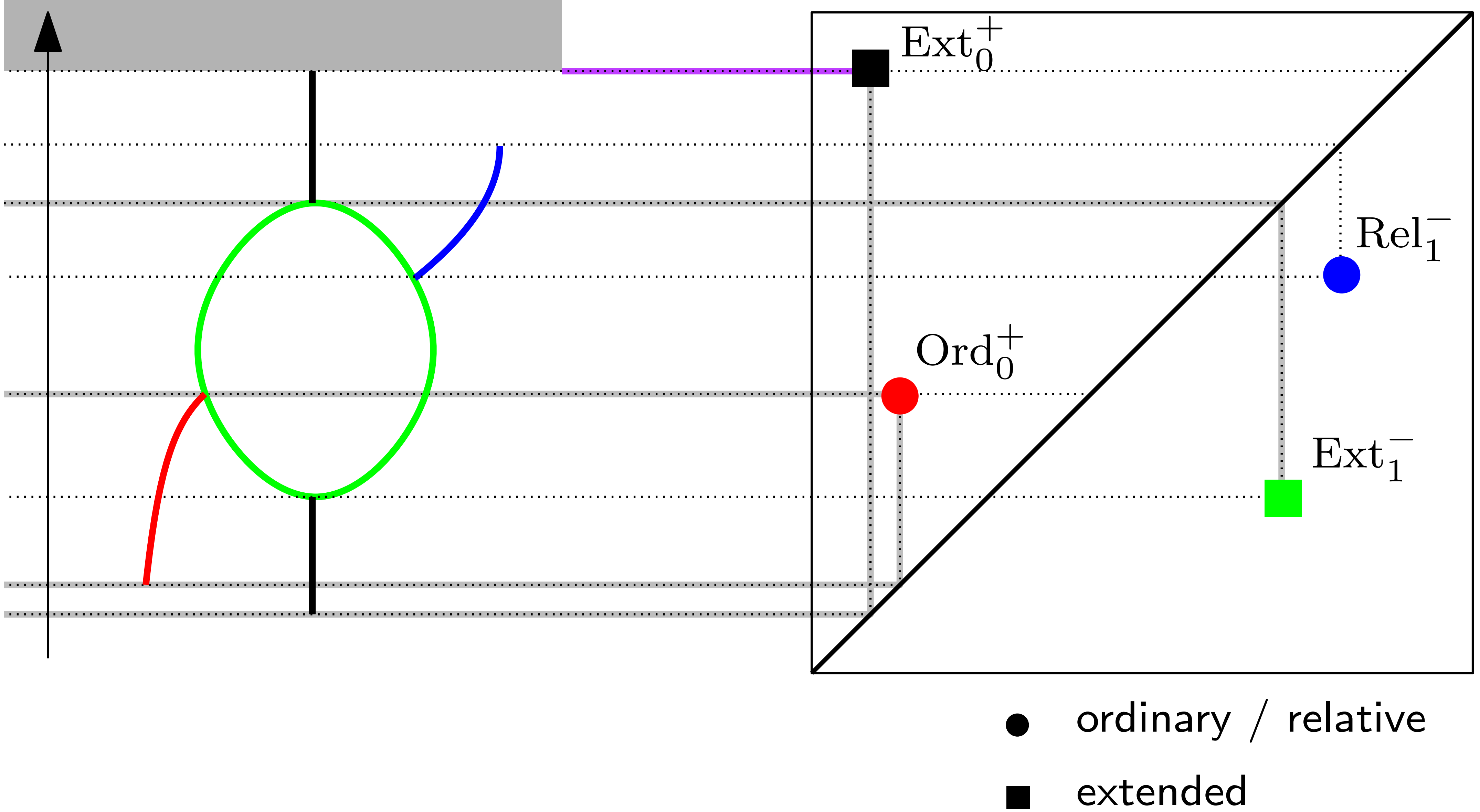
● ordinary / relative

■ extended

Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

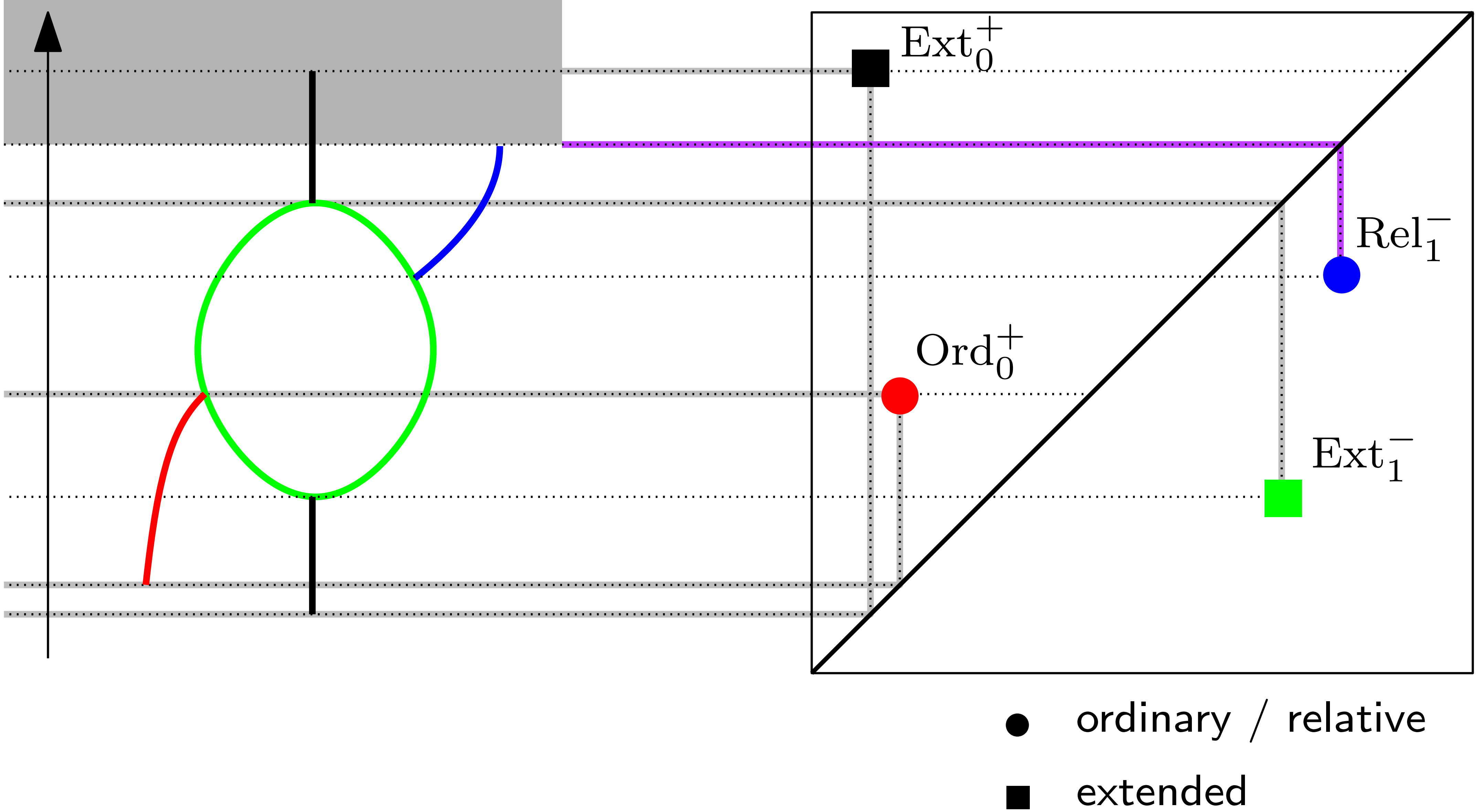
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

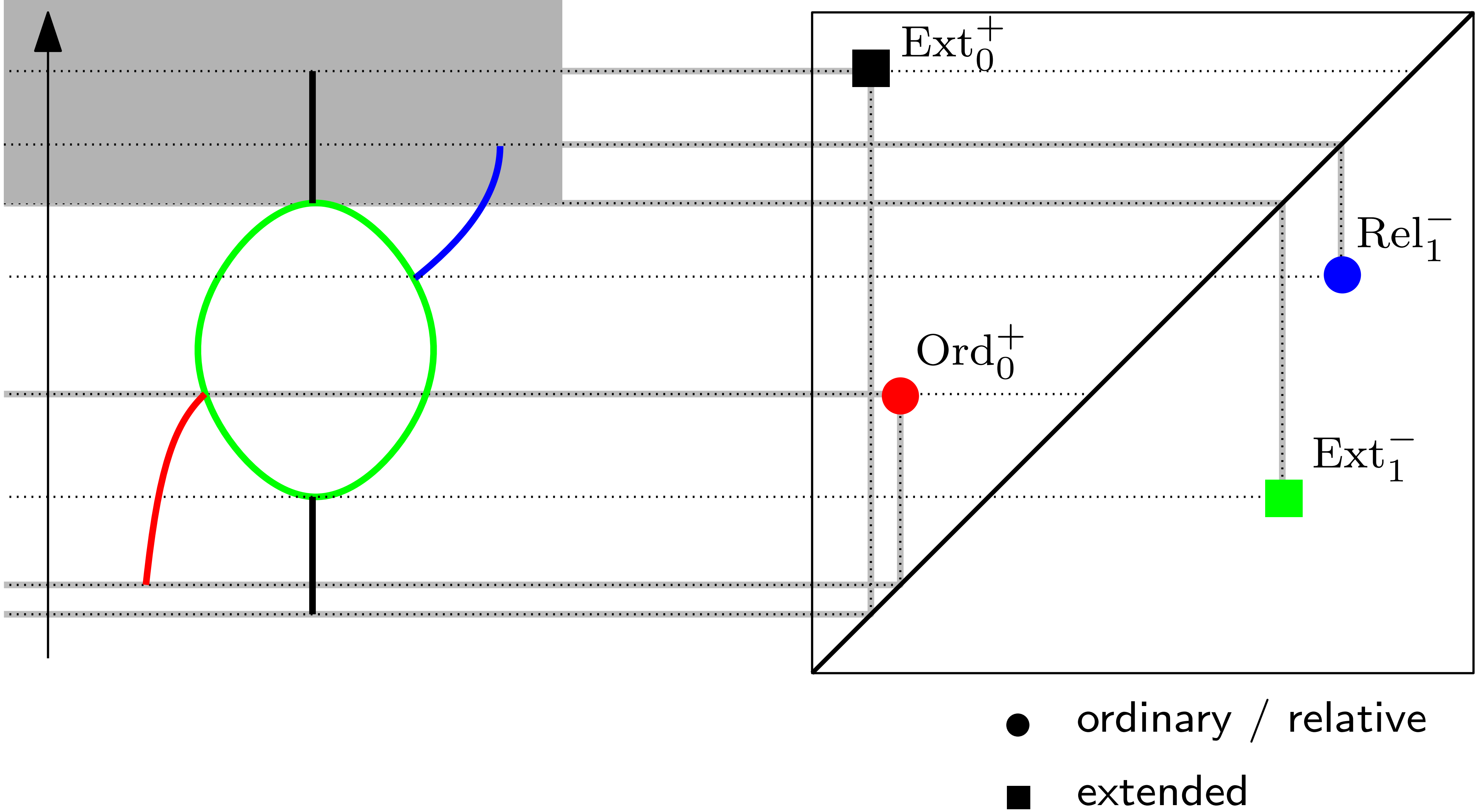
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

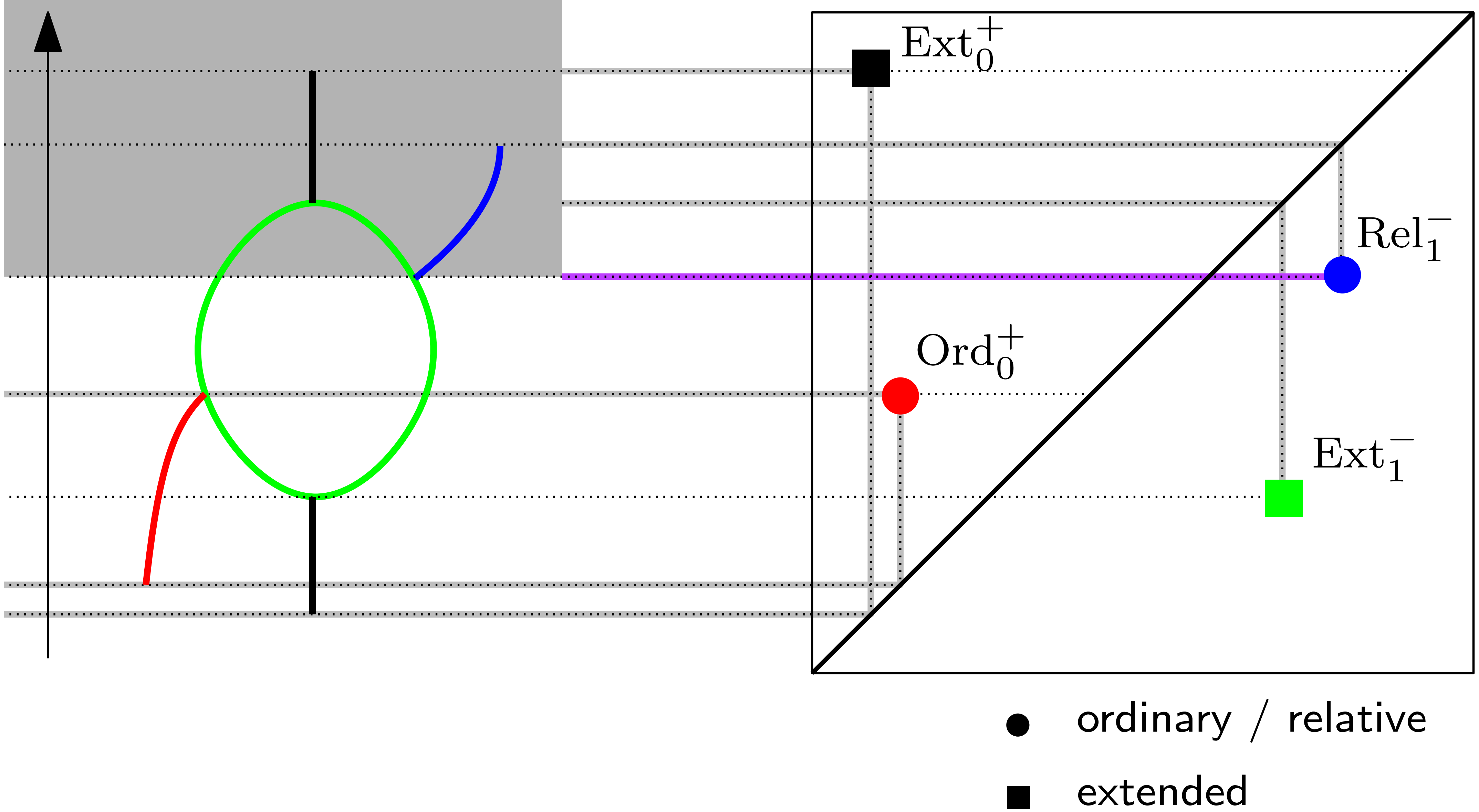
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

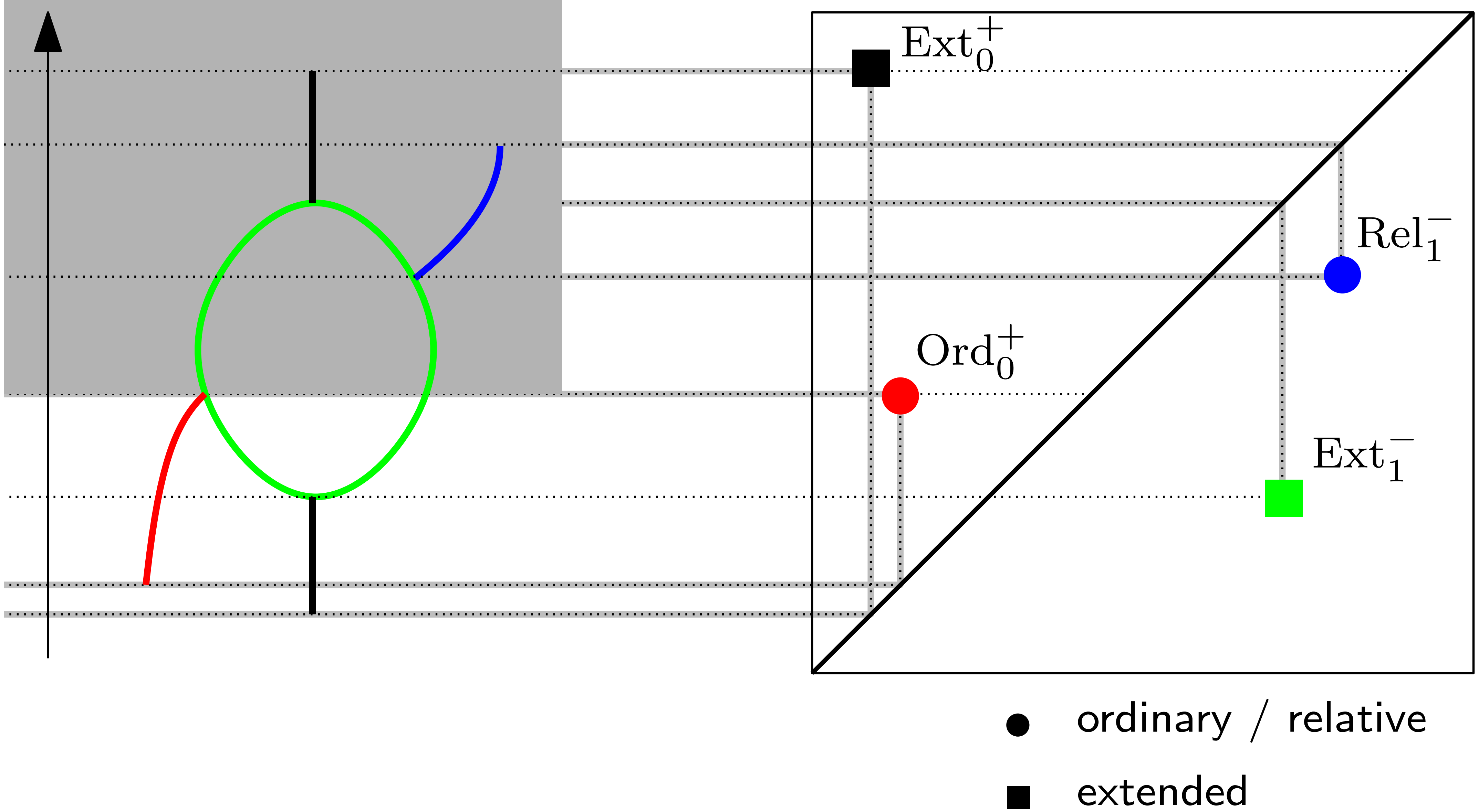
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

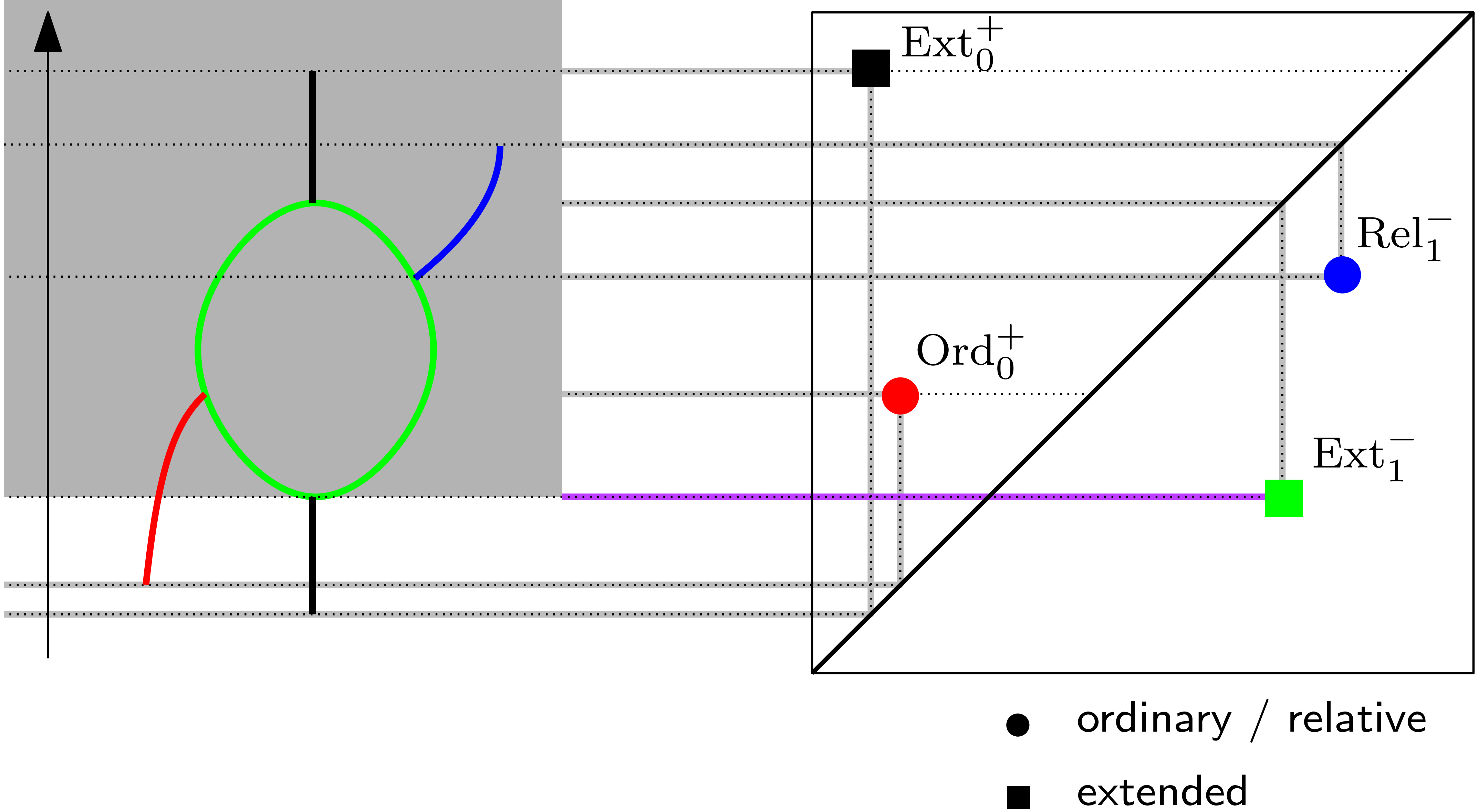
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

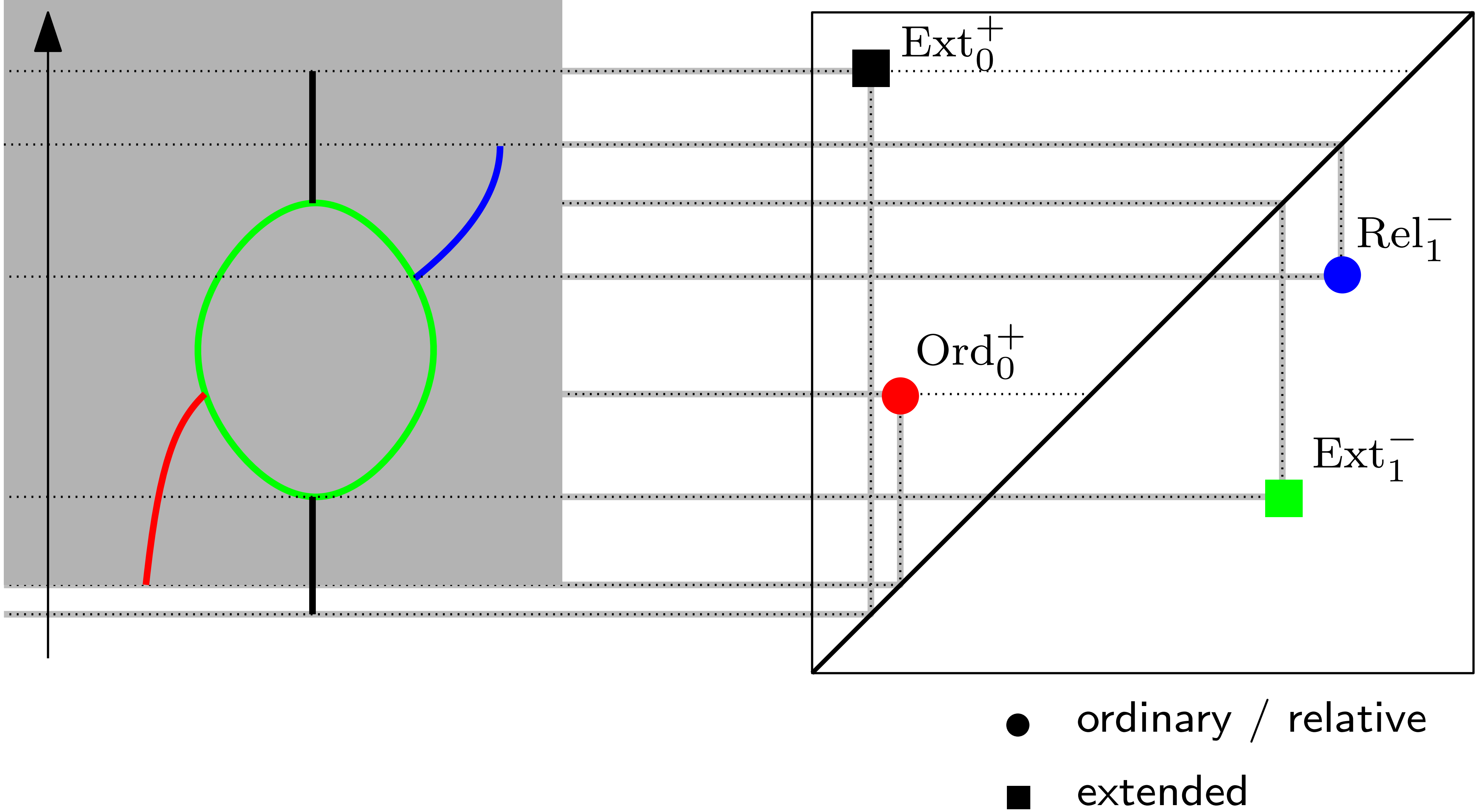
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

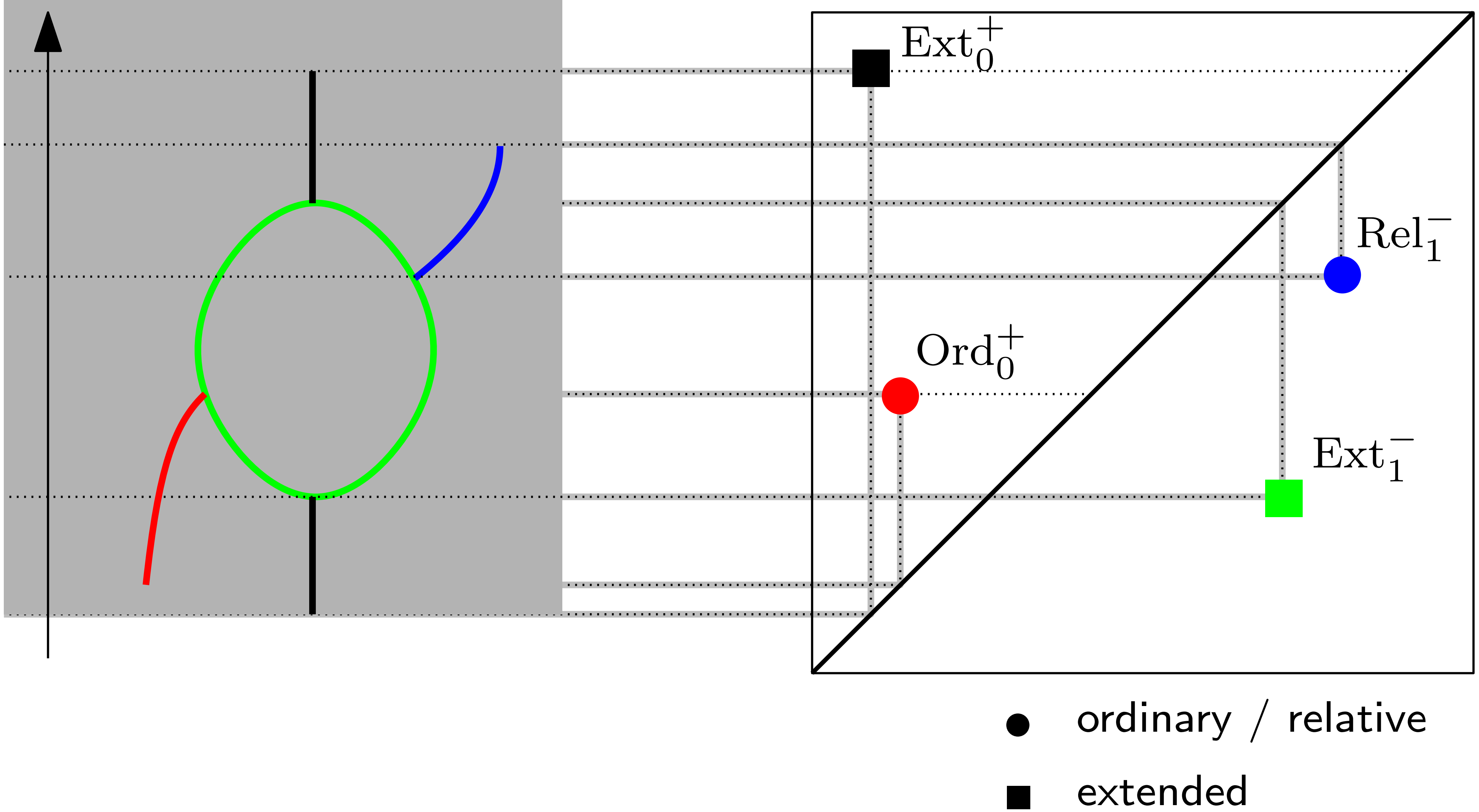
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

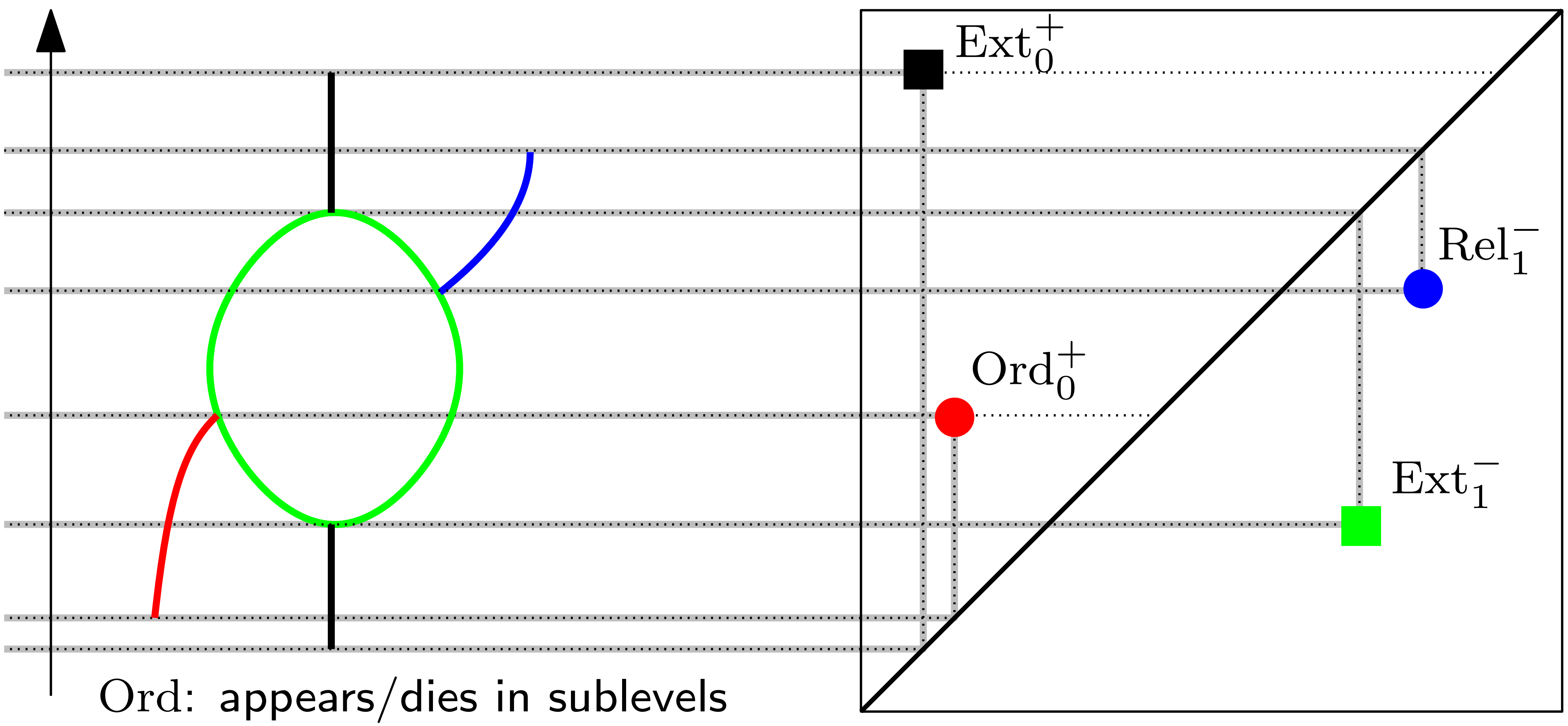
- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Descriptors for Reeb graphs

Construction uses **extended persistence**: [Cohen-Steiner, Edelsbrunner, Harer 2009]

- family of *excursion sets* (sublevel then superlevel sets) of Reeb graph
- use *homological algebra* to encode the evolution of the topology of the family



Ord: appears/dies in sublevels

Rel: appears/dies in superlevels

Ext: appears in sublevels, dies in superlevels

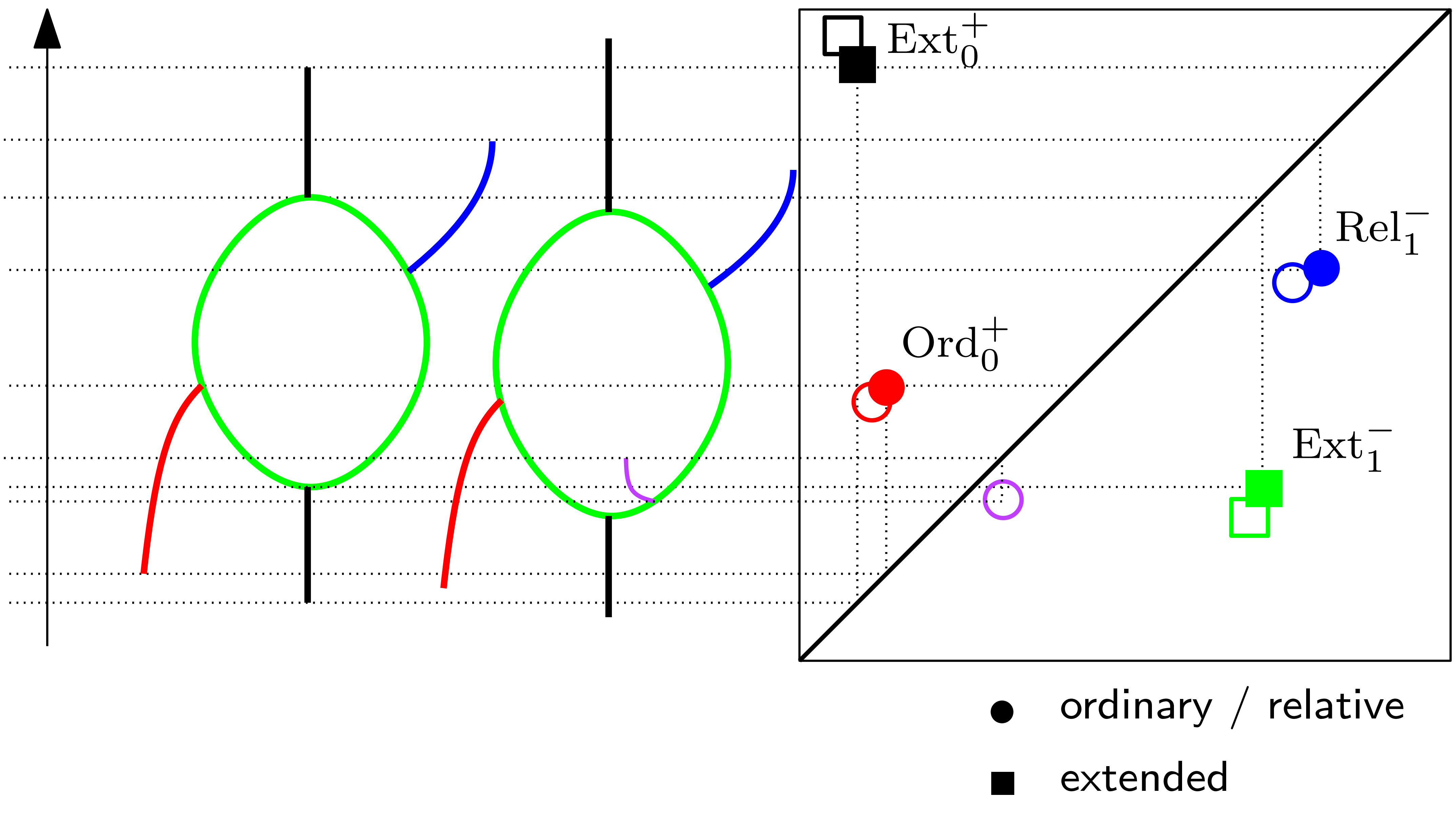
● ordinary / relative

■ extended

Descriptors for Reeb graphs

Theorem (stability): [Bauer, Ge, Wang 2014]

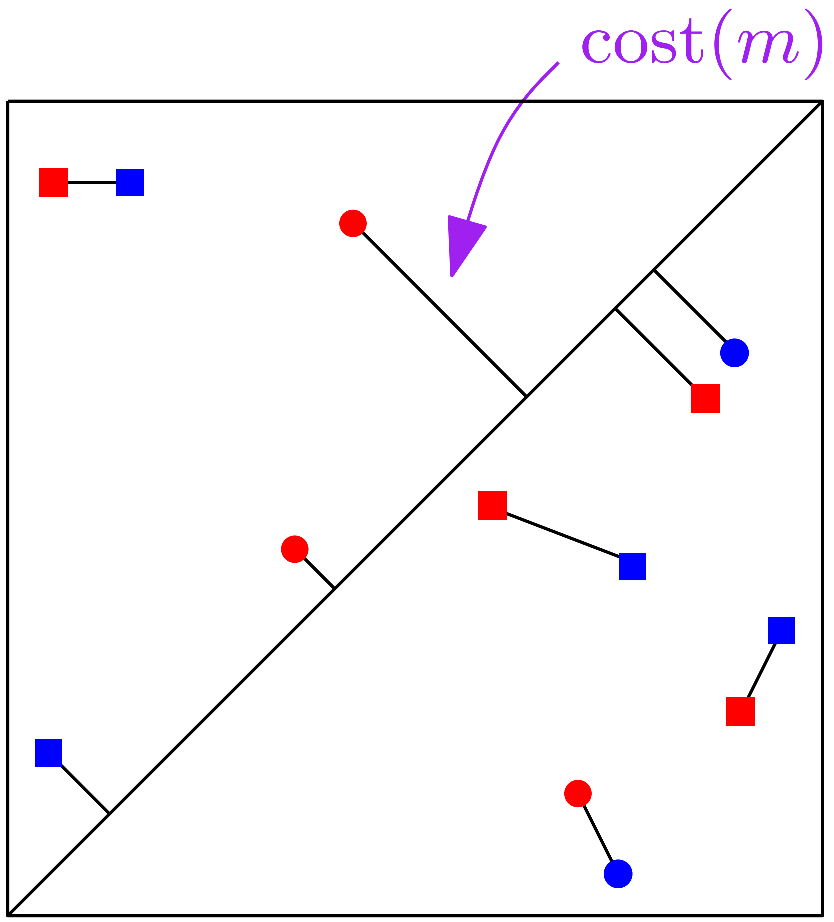
$$\mathrm{d}_B(\mathrm{Dg}\,\mathrm{R}_f, \mathrm{Dg}\,\mathrm{R}_g) \leq 6\,\mathrm{d}_{\mathrm{GH}}(\mathrm{R}_f, \mathrm{R}_g)$$



Descriptors for Reeb graphs

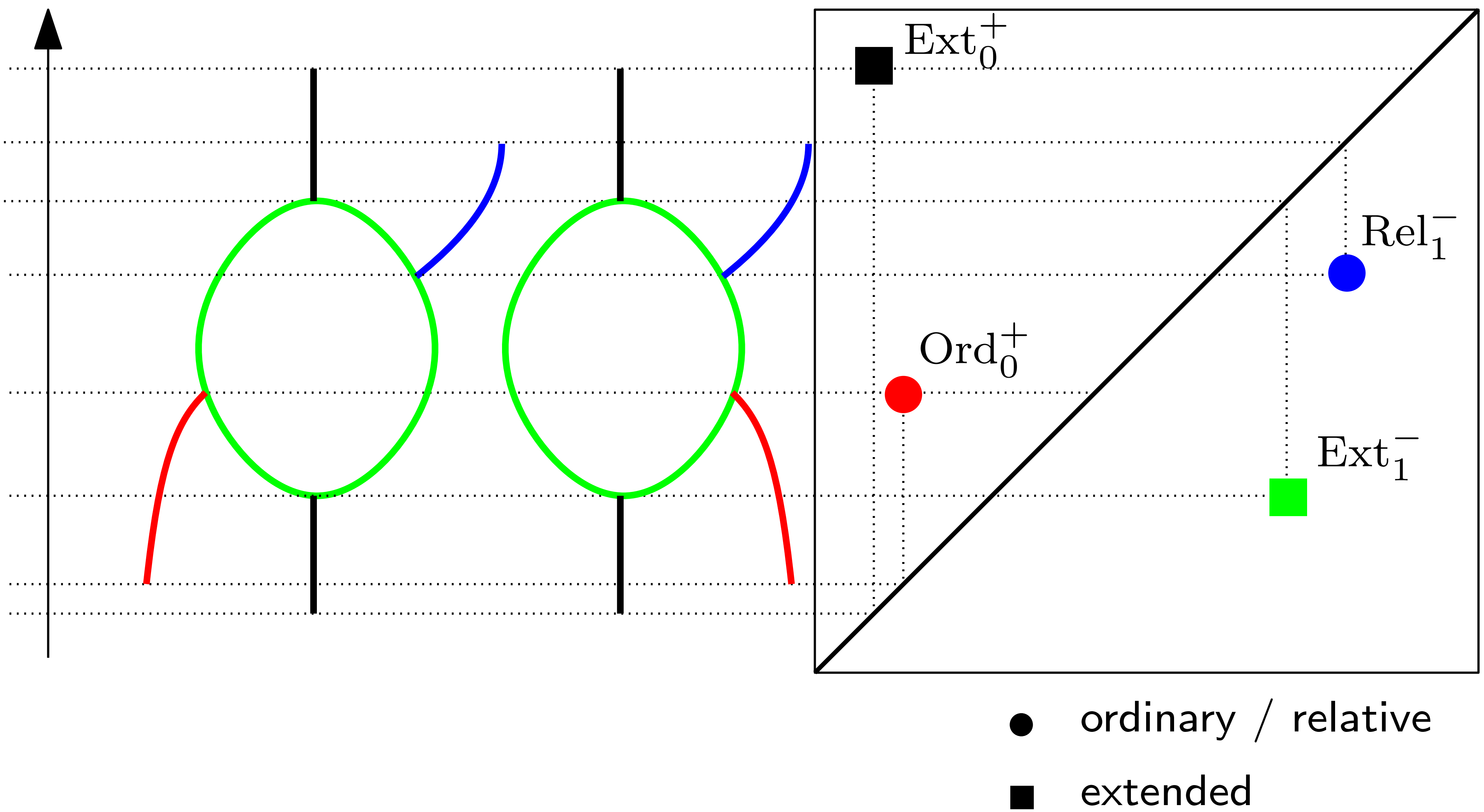
Theorem (stability): [Bauer, Ge, Wang 2014]

$$\mathrm{d}_B(\mathrm{Dg}\,\mathrm{R}_f, \mathrm{Dg}\,\mathrm{R}_g) \leq 6\,\mathrm{d}_{\mathrm{GH}}(\mathrm{R}_f, \mathrm{R}_g)$$



Descriptors for Reeb graphs

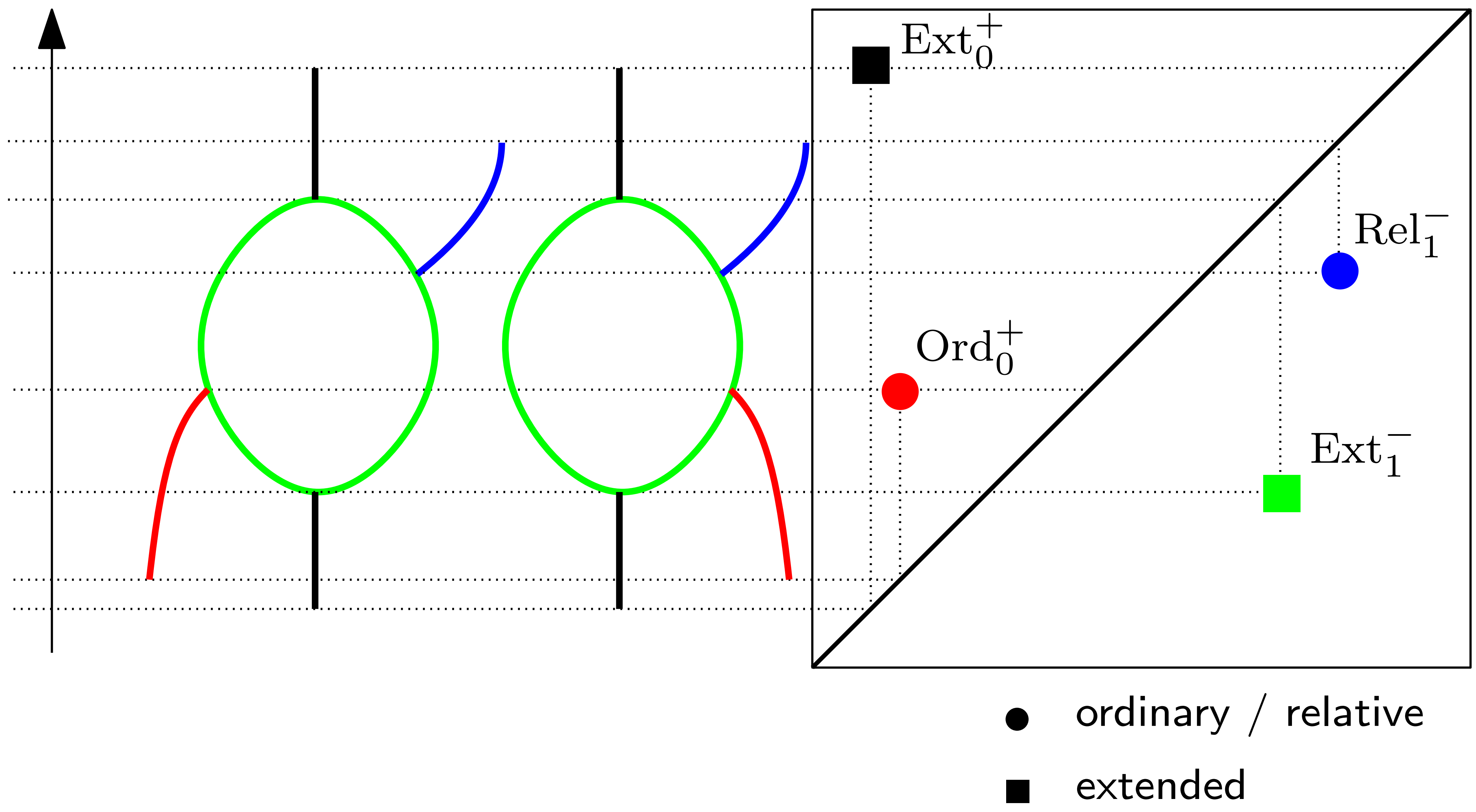
Note: $\mathrm{d}_B(\mathrm{Dg}\cdot, \mathrm{Dg}\cdot)$ is only a pseudometric on Reeb graphs



Descriptors for Reeb graphs

Note: $\mathrm{d}_B(\mathrm{Dg}\cdot, \mathrm{Dg}\cdot)$ is only a pseudometric on Reeb graphs

Thm: [Carrière, O. 2017]
$\mathrm{d}_B(\mathrm{Dg}\cdot, \mathrm{Dg}\cdot)$ is *locally* a metric equivalent to d_{GH}



Computing Reeb graphs

Procedure given a point cloud P and a filter $f : P \to \mathbb{R}$:

1. build a (possibly non-manifold) 2-d simplicial complex X on top of P
2. compute the Reeb graph of $(X, \bar{f})$, where $\bar{f}$ is the PL interpolation of f

Computing Reeb graphs

Procedure given a point cloud P and a filter $f : P \to \mathbb{R}$:

1. build a (possibly non-manifold) 2-d simplicial complex X on top of P

2. compute the Reeb graph of $(X, \bar{f})$, where $\bar{f}$ is the PL interpolation of f

Complexity: $O(nm)$ for n points and m simplices

Q: convergence?

Thm: [Dey, Wang 2013]
For P an ε-*sample* of M a *sufficiently regular manifold*, and for X a Rips complex on P of *appropriate parameter* r, $\mathrm{d}_B(\mathrm{Dg}\,\mathrm{R}_{\bar{f}}(X), \mathrm{Dg}\,\mathrm{R}_f(M)) \leq cr$.

Approximations to Reeb graphs

- α-Reeb graphs [Chazal, Huang, Sun 2015]

- Joint Contour Nets [Carr, Duke 2014]

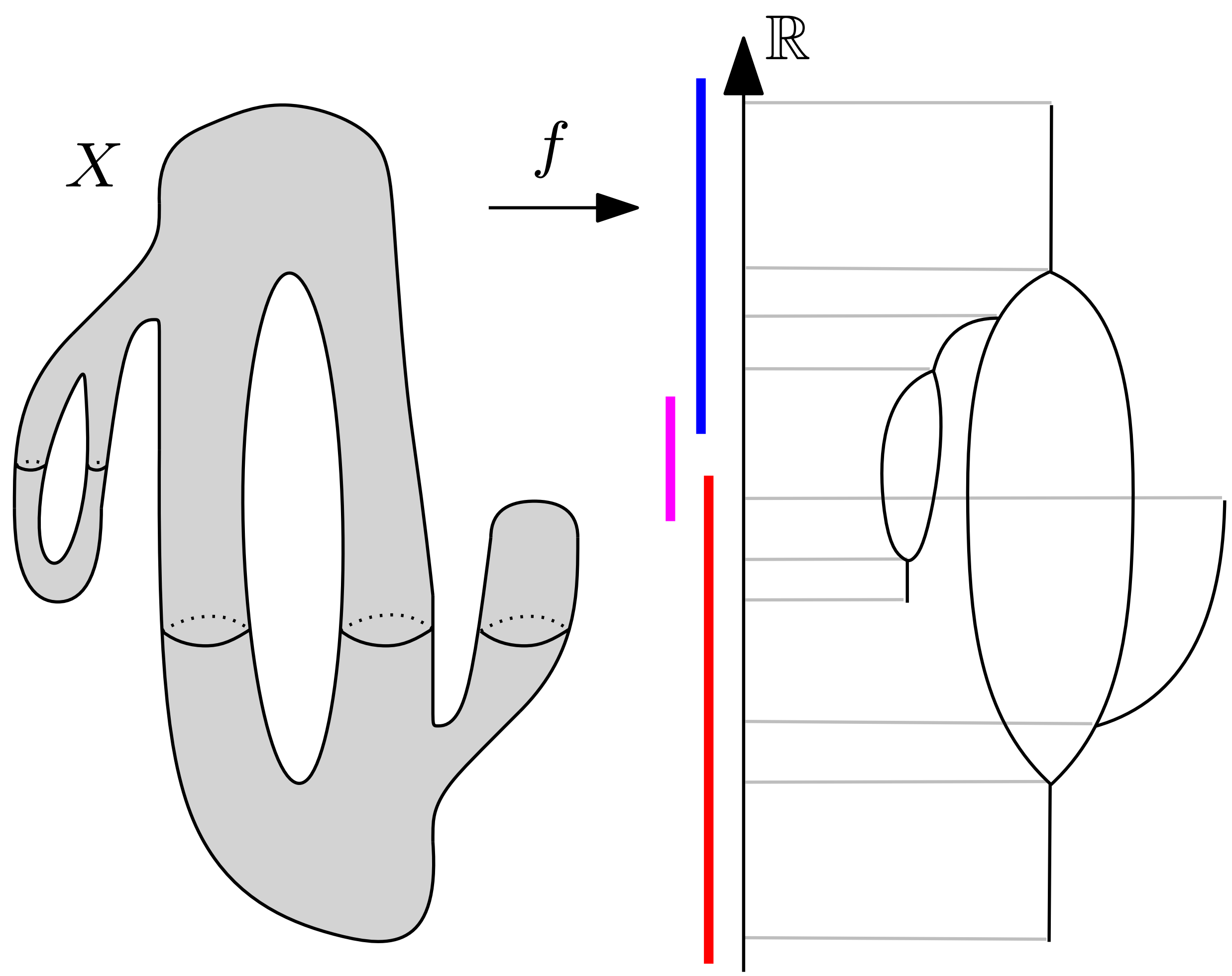
- Mappers [Singh, Mémoli, Carlsson 2007]

- etc.

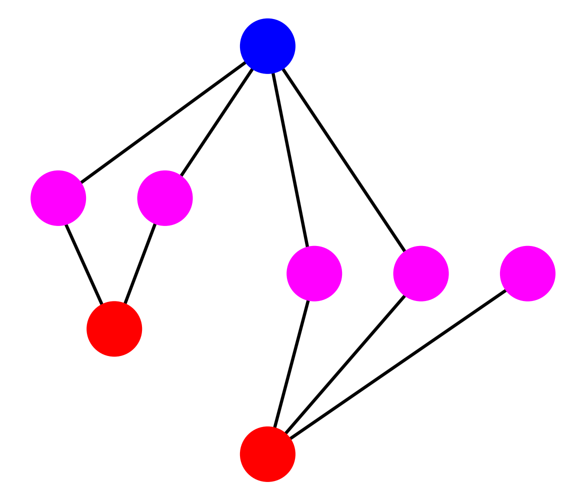
Reeb Graph vs. Mapper

$x \sim y \iff [\ f(x) = f(y)$ and x, y belong to same cc of $f^{-1}(\{f(x)\})\ ]$

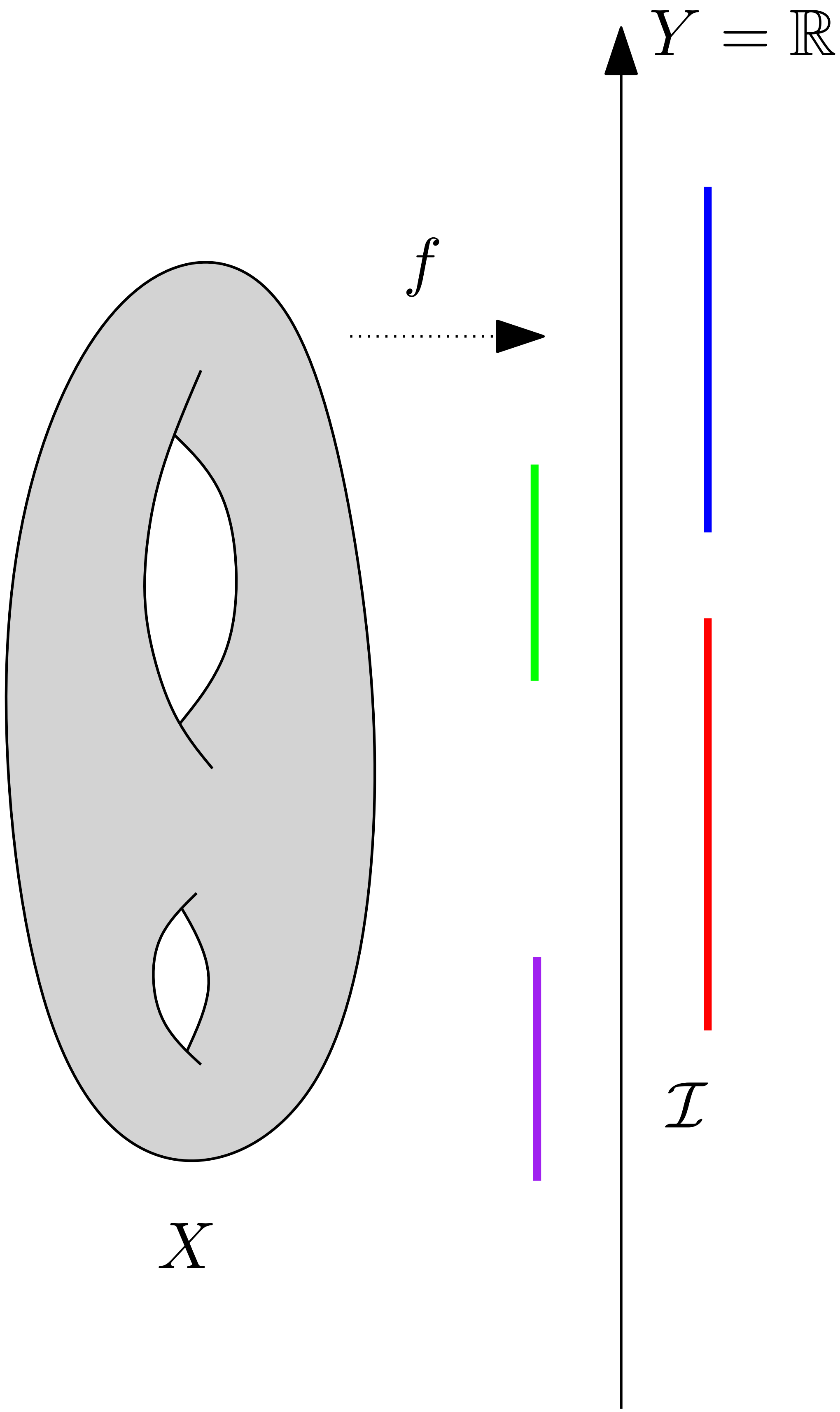
$\mathrm{R}_f(X) := X/\sim$



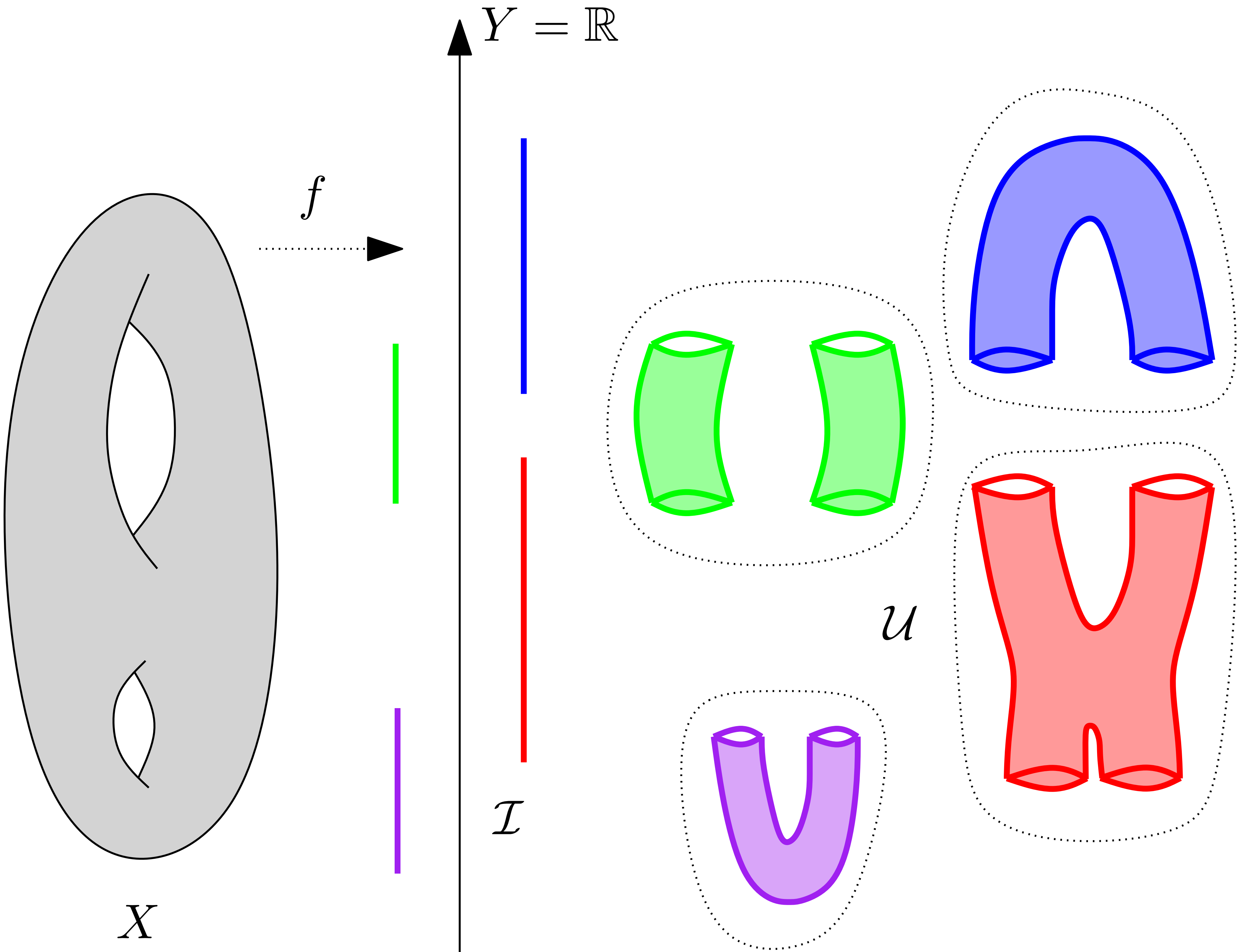
mapper $\equiv$ *pixelized* Reeb graph



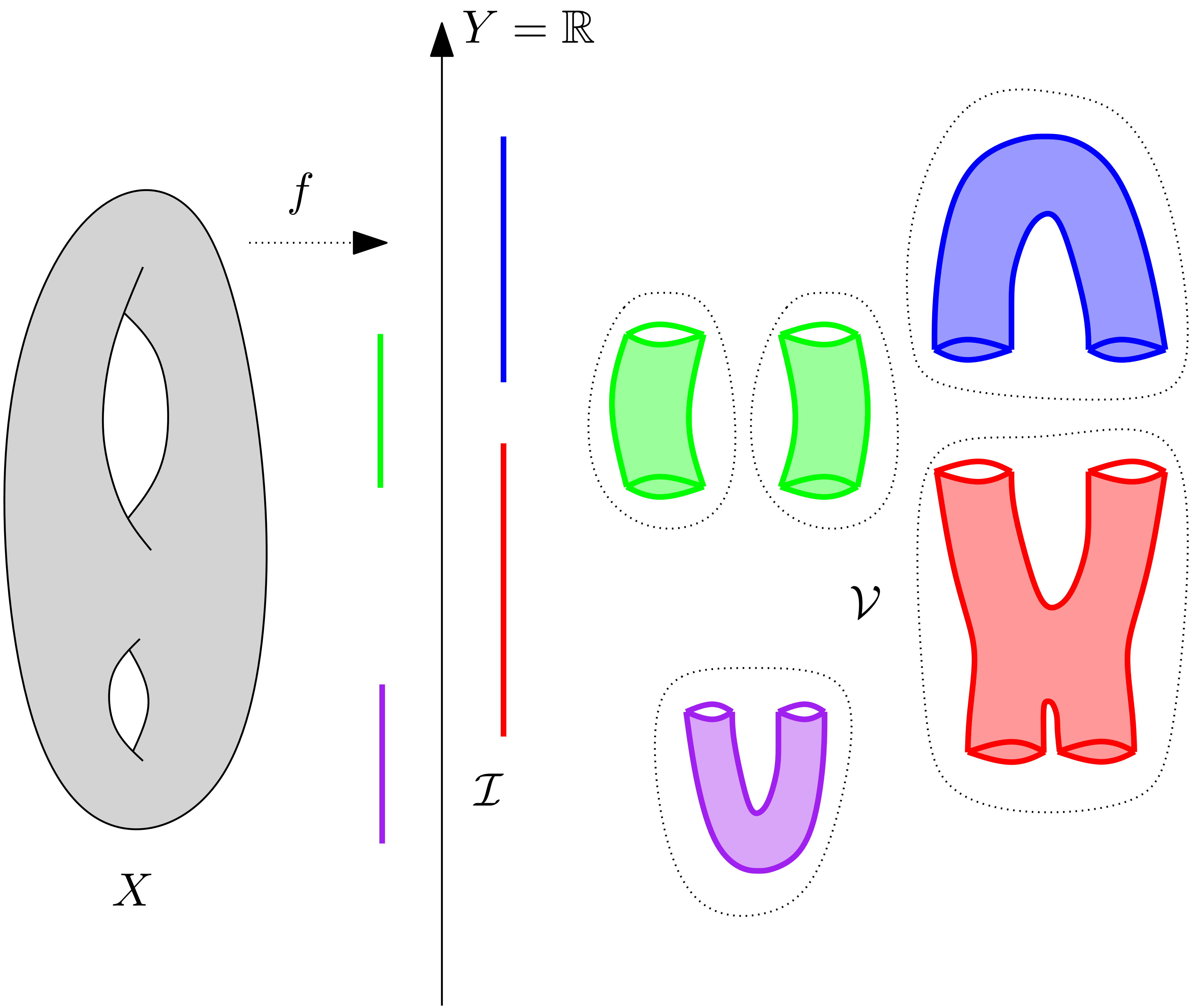
Mapper in the continuous setting



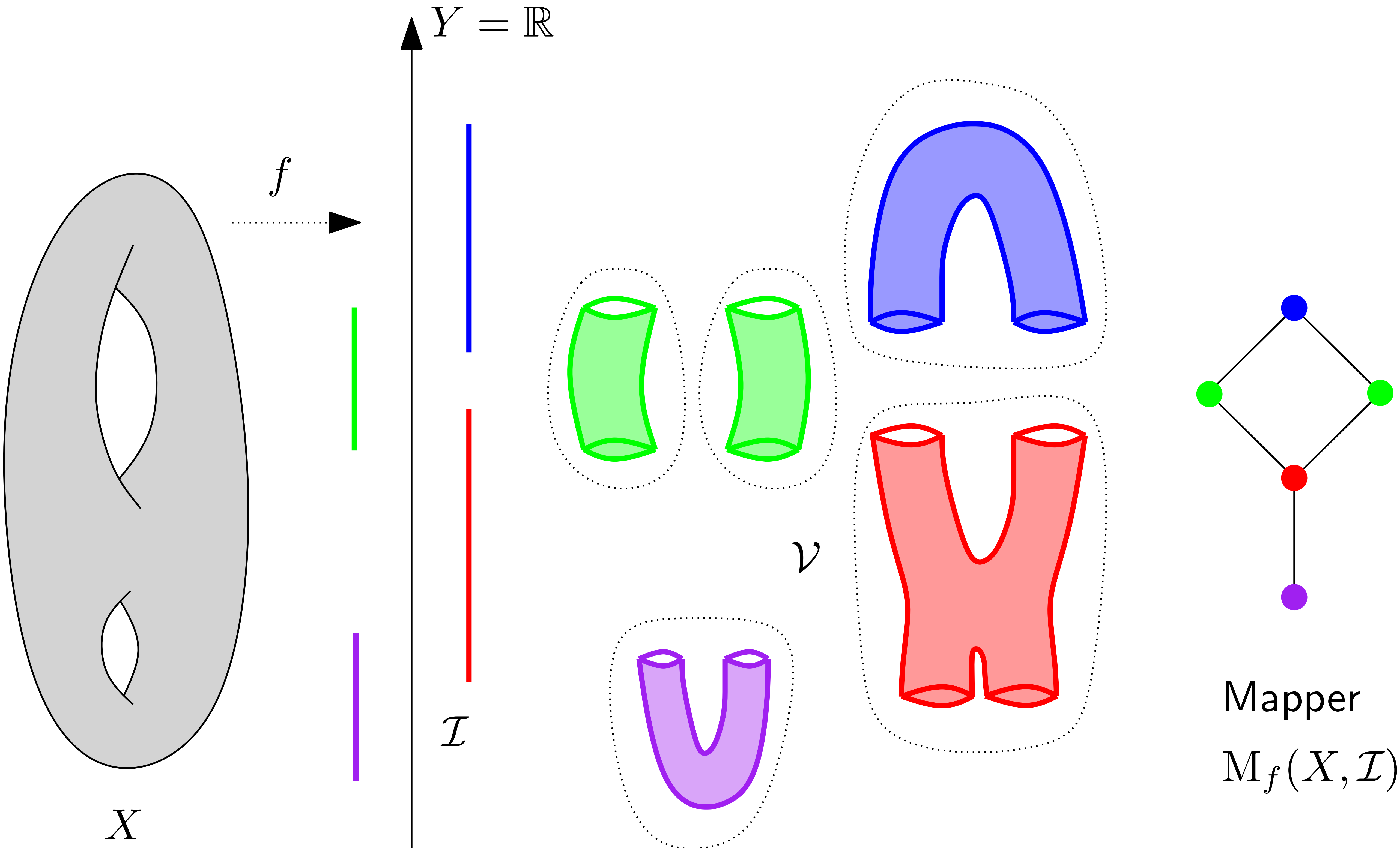
Mapper in the continuous setting



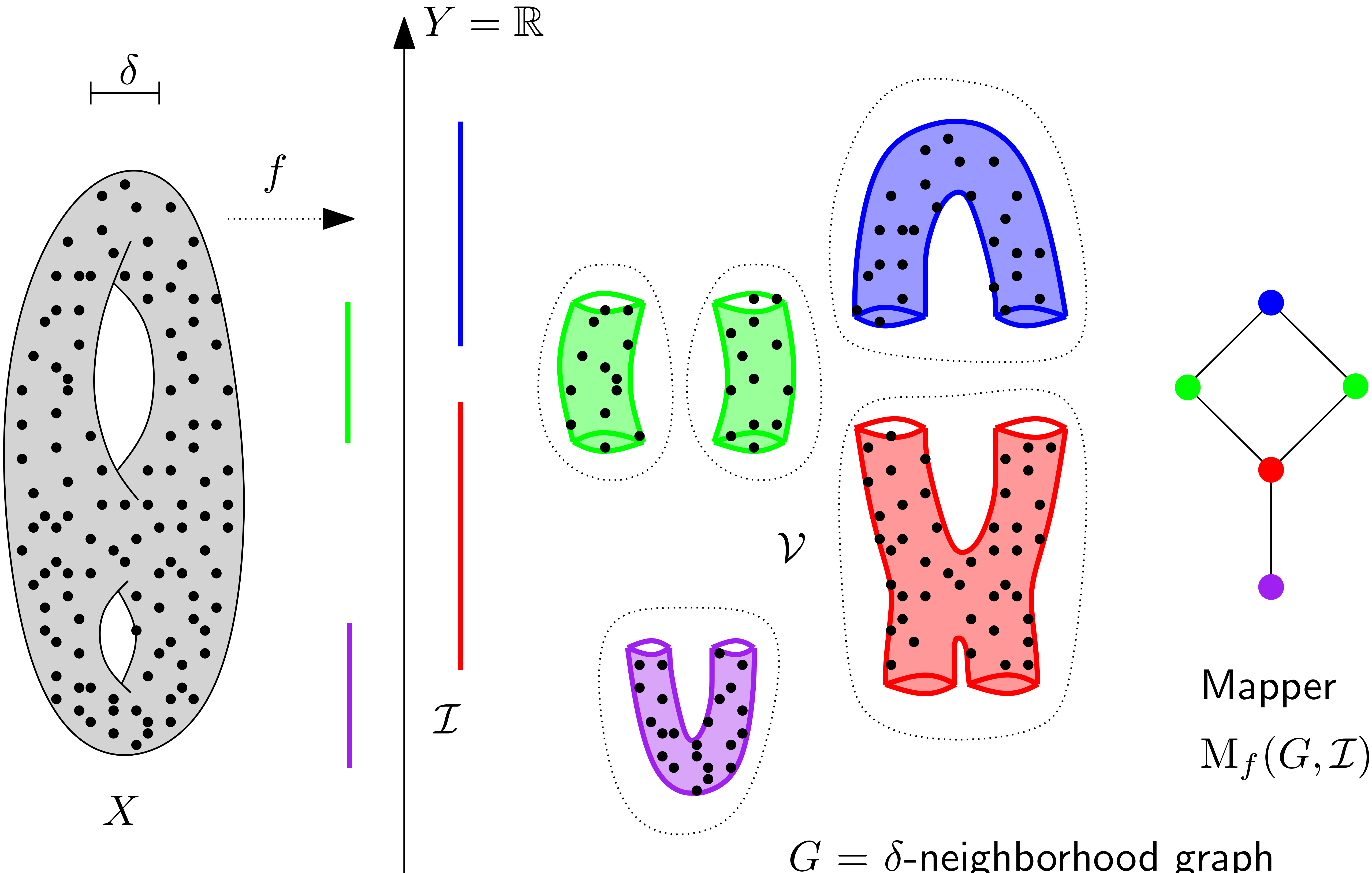
Mapper in the continuous setting



Mapper in the continuous setting



Mapper in practice

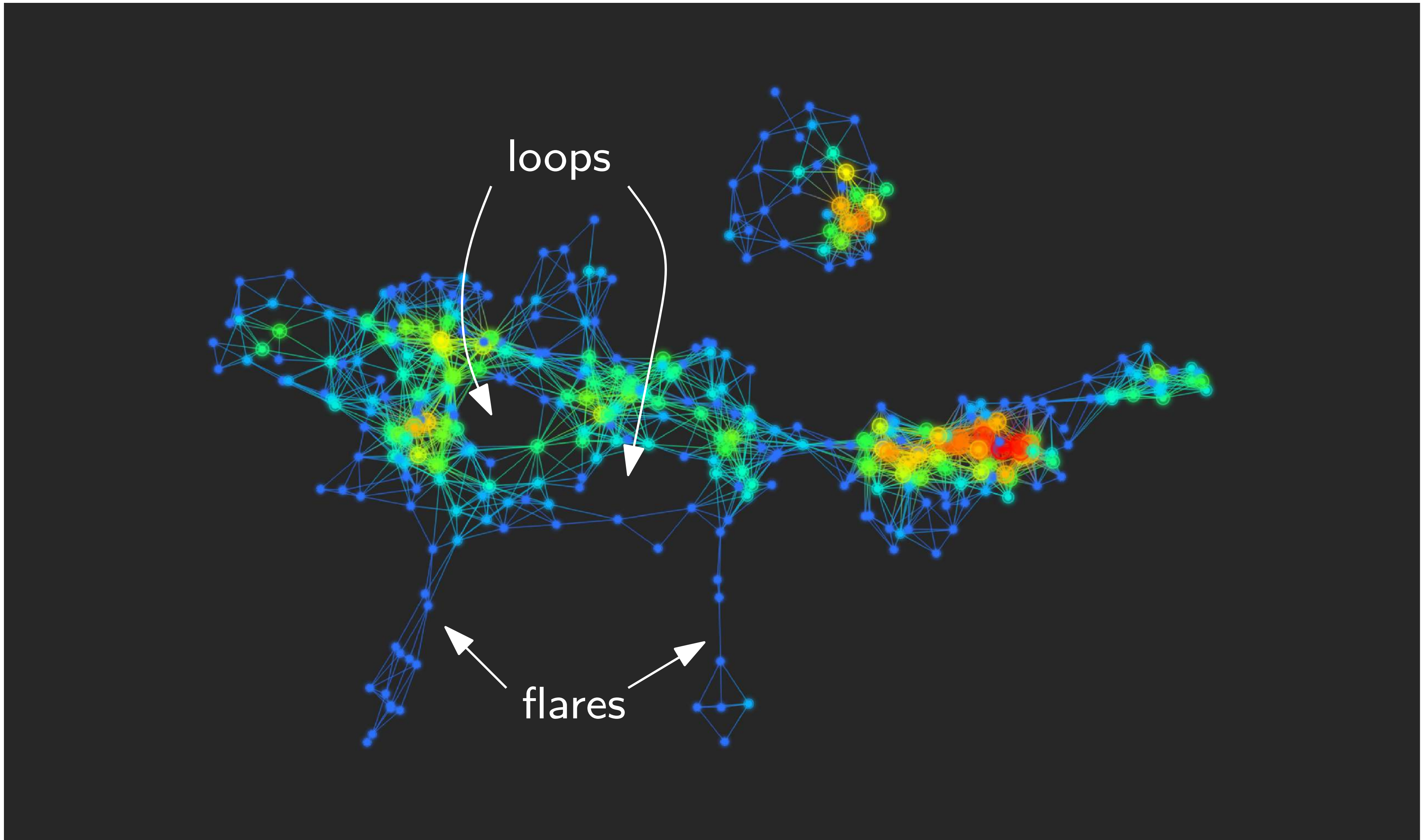


Mapper in applications

Two types of applications:

- clustering
- feature selection

principle: identify statistically relevant sub-populations through patterns (flares, loops)



Choice of parameters

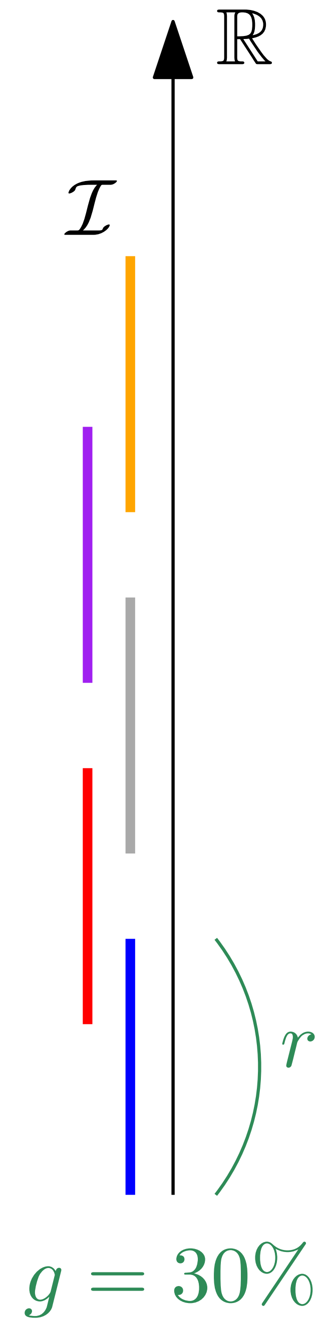
Parameters:

- function $f : P \to \mathbb{R}$ ← lens | filter

- cover $\mathcal{I}$ of $\mathrm{im}(f)$ by open intervals ← range scale

- neighborhood size δ ← geometric scale

Choice of parameters



Parameters:

- function $f : P \to \mathbb{R}$ ← lens | filter

- cover $\mathcal{I}$ of $\mathrm{im}(f)$ by open intervals ← range scale

- neighborhood size δ ← geometric scale

$\to$ uniform cover $\mathcal{I}$: ← range scale

- resolution / granularity: r (diameter of intervals)
- gain: g (percentage of overlap)

Choice of parameters

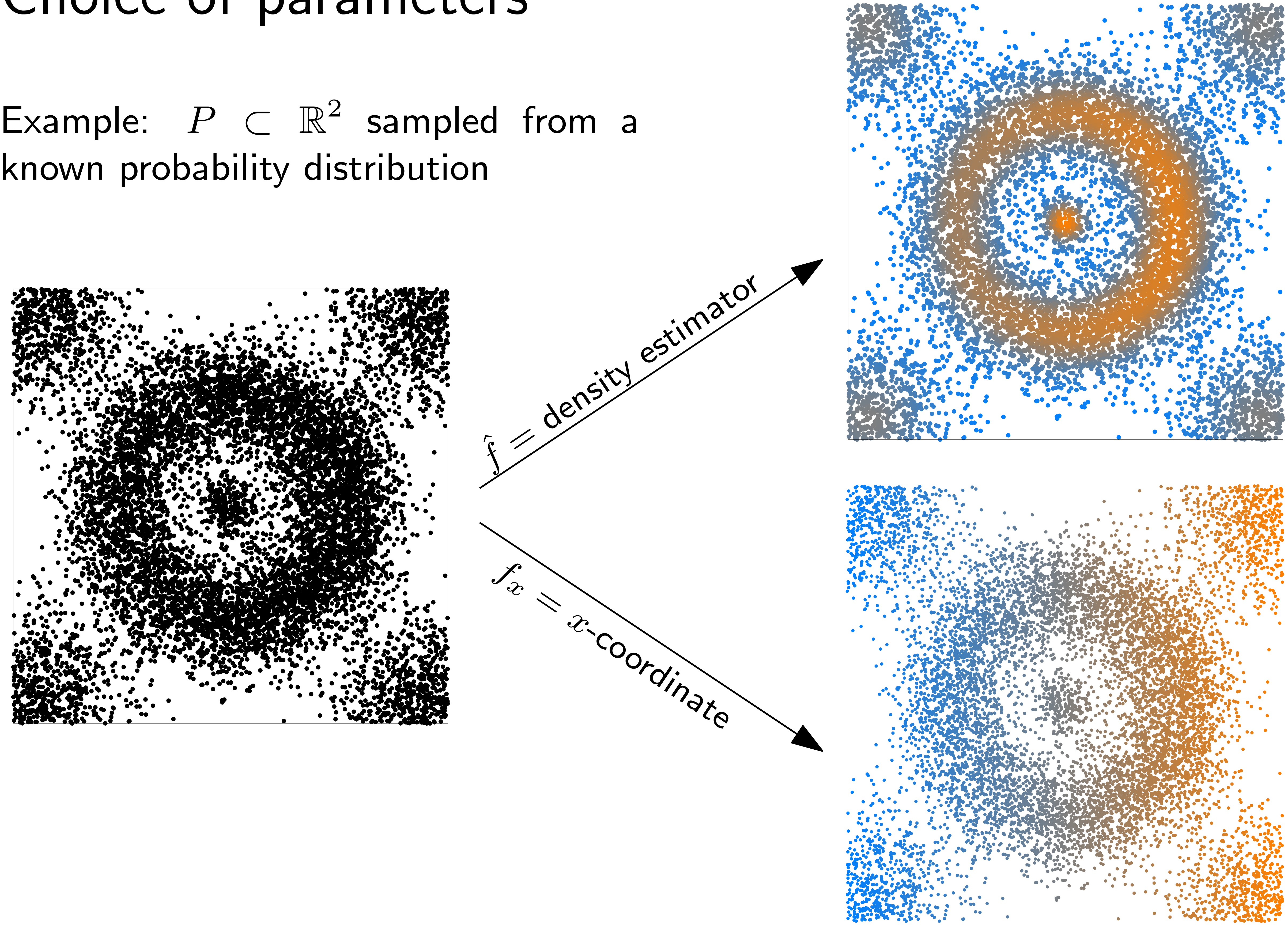
→ in practice: trial-and-error

> high-dimensional data sets[40,48]. This is performed automatically within the software, by deploying an ensemble machine learning algorithm that iterates through overlapping subject bins of different sizes that resample the metric space (with replacement), thereby using a combination of the metric location and similarity of subjects in the network topology. After performing millions of iterations, the algorithm returns the most stable, consensus vote for the resulting 'golden network' (Reeb graph), representing the multidimensional data shape[12,40].

Nielson et al.: *Topological Data Analysis for Discovery in Preclinical Spinal Cord Injury and Traumatic Brain Injury*, Nature, 2015

Choice of parameters

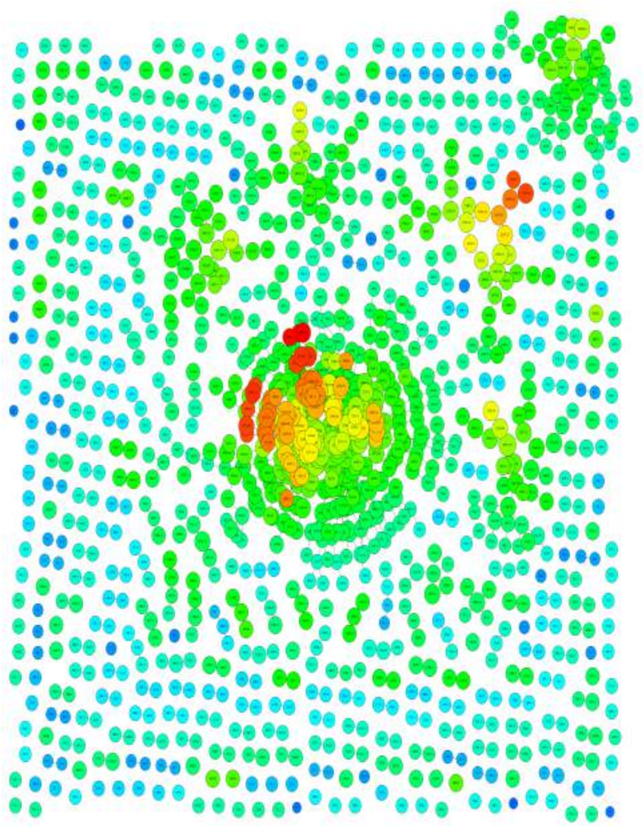
Example: $P \subset \mathbb{R}^2$ sampled from a known probability distribution

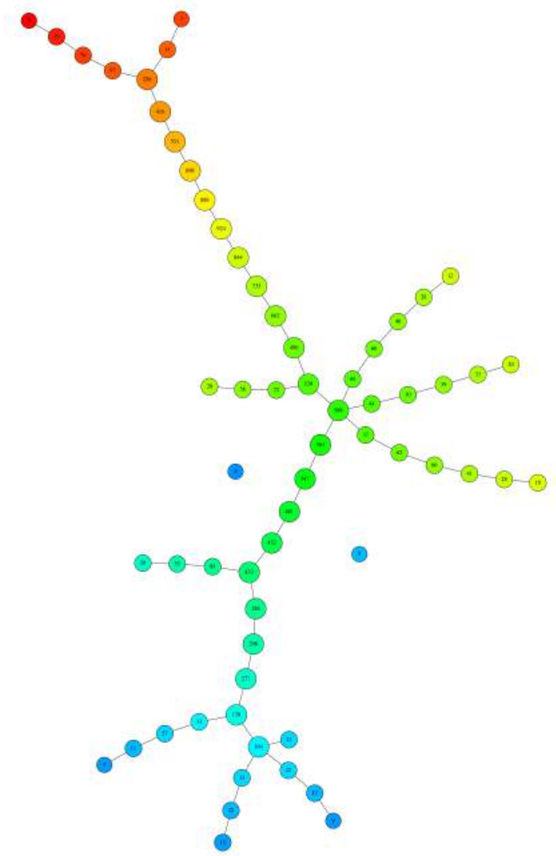


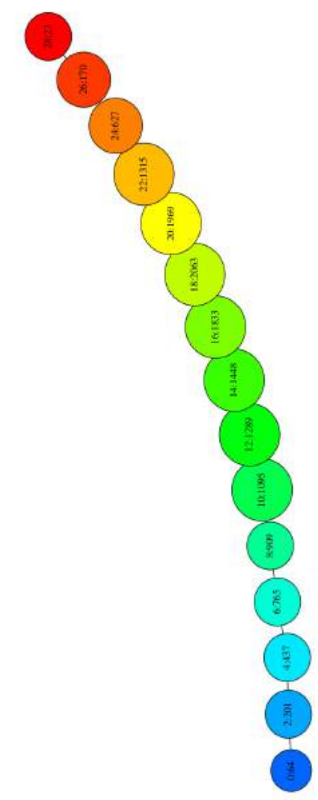
Choice of parameters

$r = 0.3,\ g = 20\%$

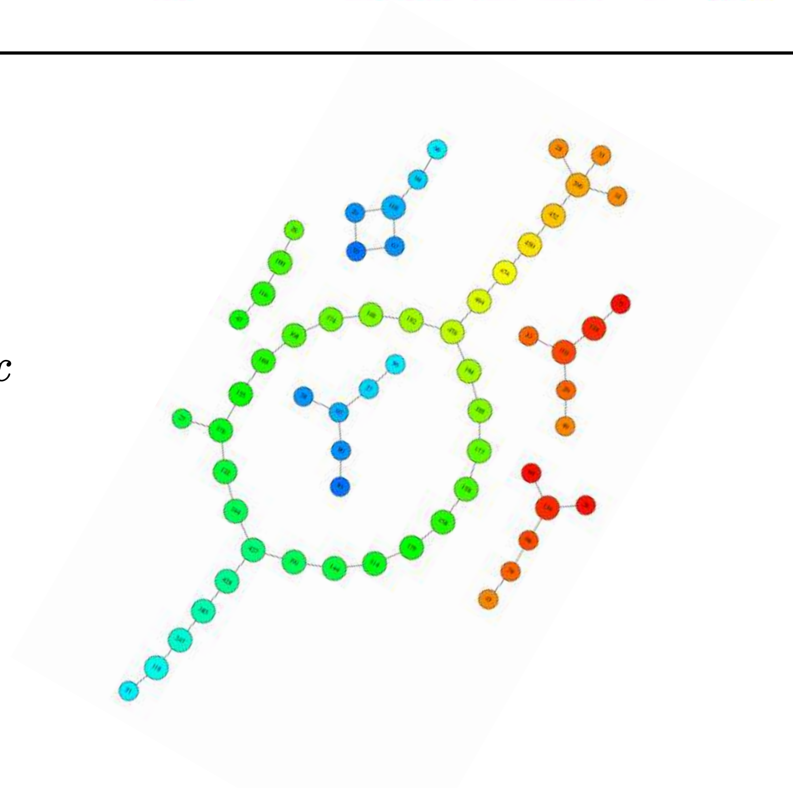
$f = \hat{f}$

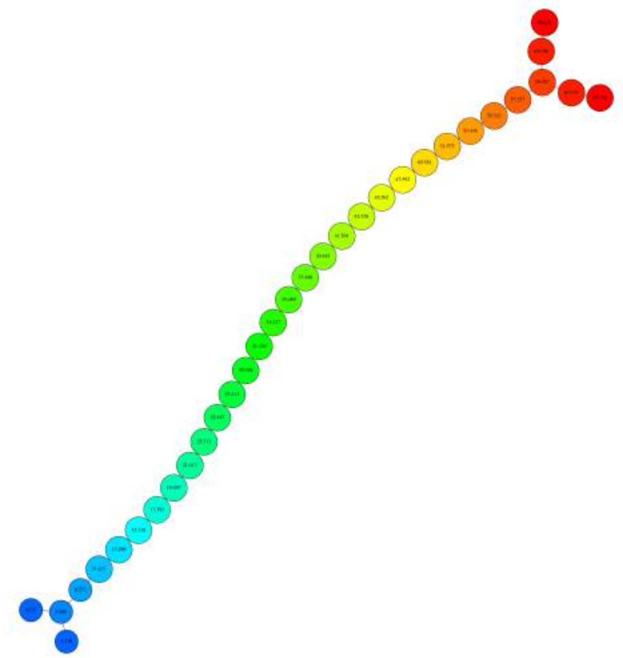


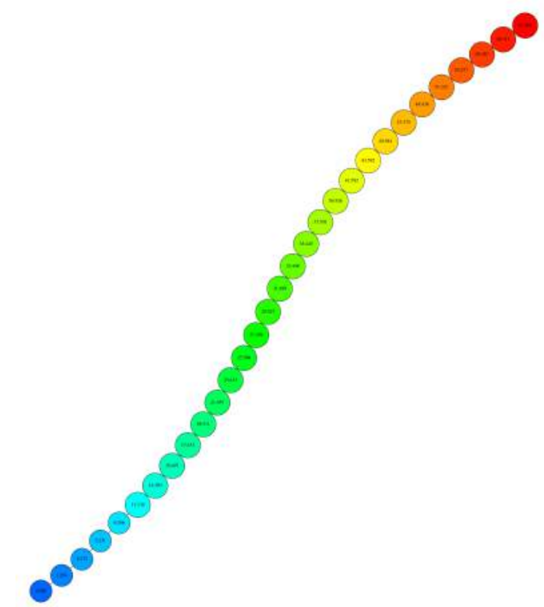




$f = f_x$



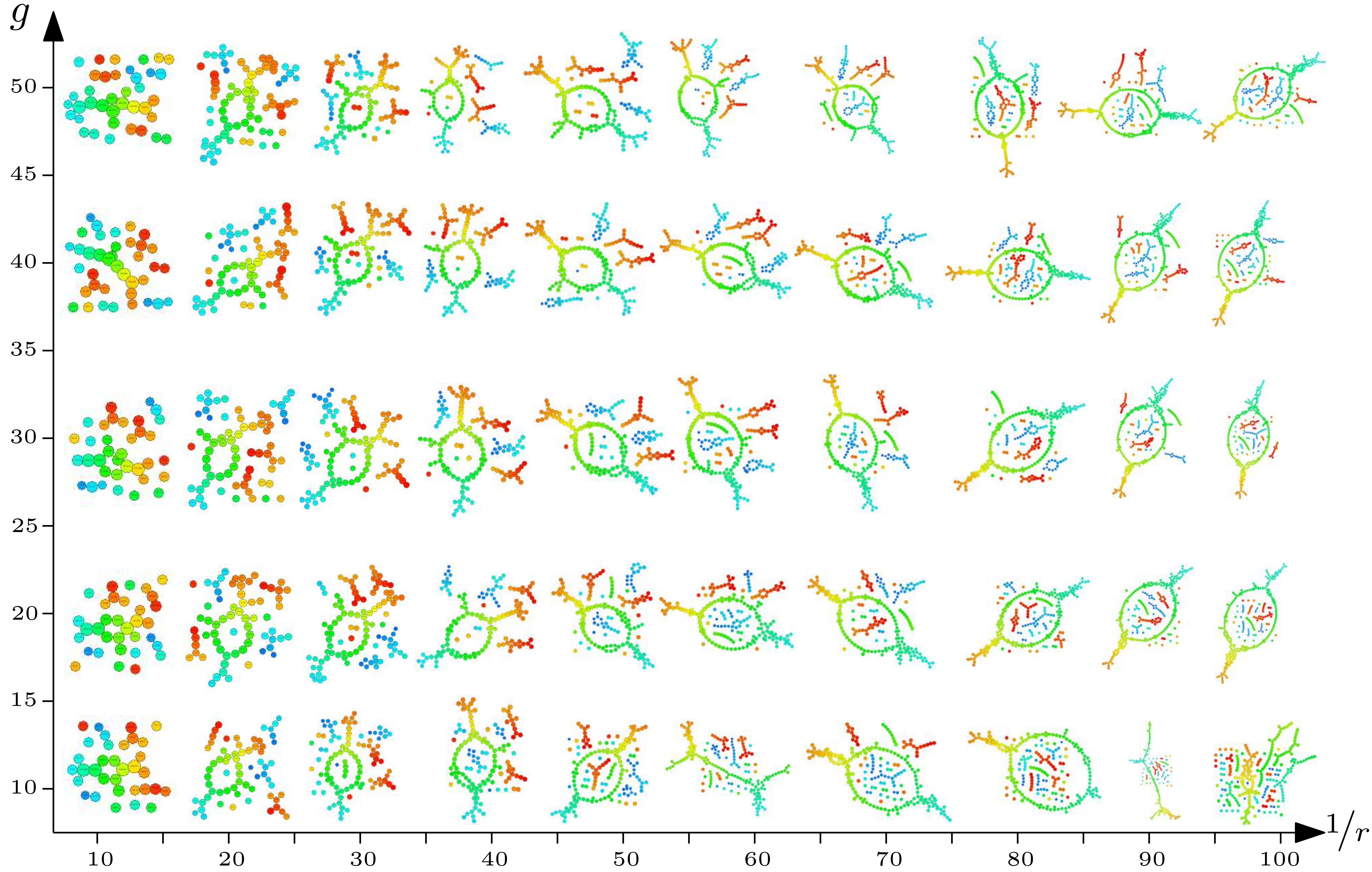




$\delta = 1\%$ $\delta = 10\%$ $\delta = 25\%$

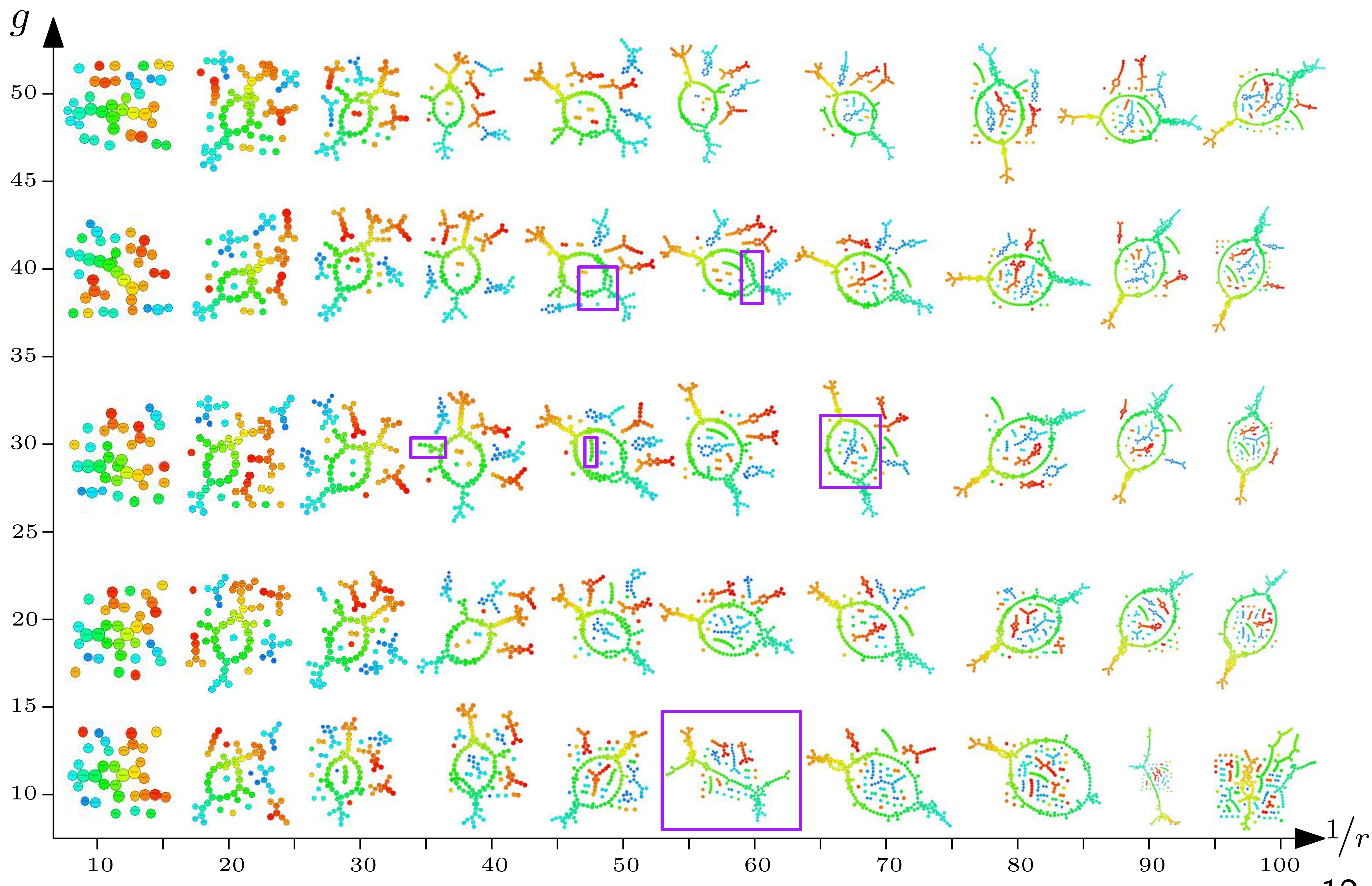
Choice of parameters

$f = f_x$, $\delta = 1\%$



Choice of parameters

$f = f_x$, $\delta = 1\%$



Choice of parameters

Recent contributions:

$\rightarrow$ clarify the roles of r and g in the continuous setting

$\rightarrow$ introduce metrics between mappers

$\rightarrow$ establish stability and convergence results for Mappers

$\rightarrow$ relate discrete and continuous Mappers under conditions on δ

---

2 approaches:

• connection to topological persistence and representation theory
[Carrière, O. 2016] $<$ [Bauer, Ge, Wang 2013] [Cohen-Steiner, Edelsbrunner, Harer 2009]

• connection to constructible cosheaves in Sets and stratification theory
[Munch, Wang 2016] $<$ [de Silva, Munch, Patel 2016]

Choice of parameters

Recent contributions:

$\rightarrow$ clarify the roles of r and g in the continuous setting

$\rightarrow$ introduce metrics between mappers

$\rightarrow$ establish stability and convergence results for Mappers

$\rightarrow$ relate discrete and continuous Mappers under conditions on δ

---

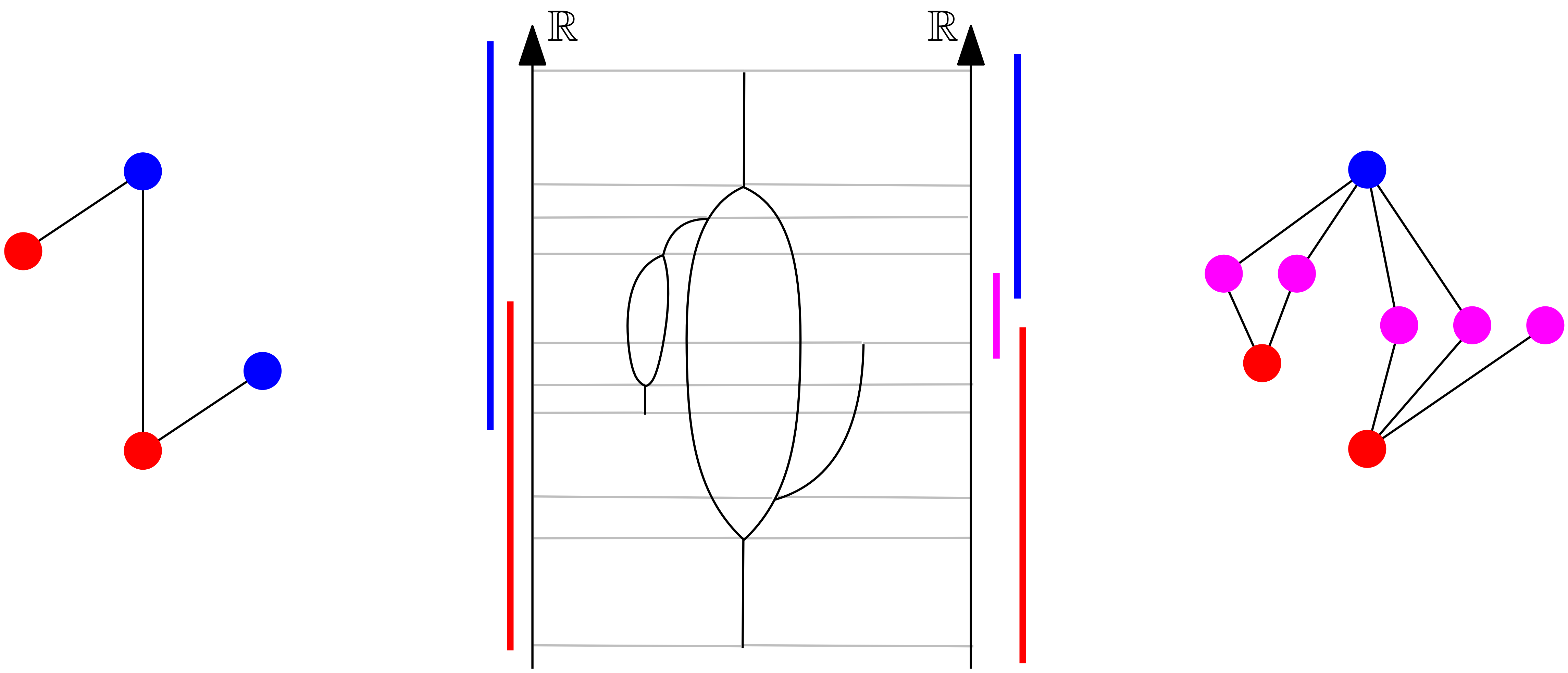
2 approaches:

• connection to topological persistence and representation theory
[Carrière, O. 2016] $<$ [Bauer, Ge, Wang 2013] [Cohen-Steiner, Edelsbrunner, Harer 2008]

• connection to constructible cosheaves in Sets and stratification theory
[Munch, Wang 2016] $<$ [de Silva, Munch, Patel 2016]

Descriptor for Mapper

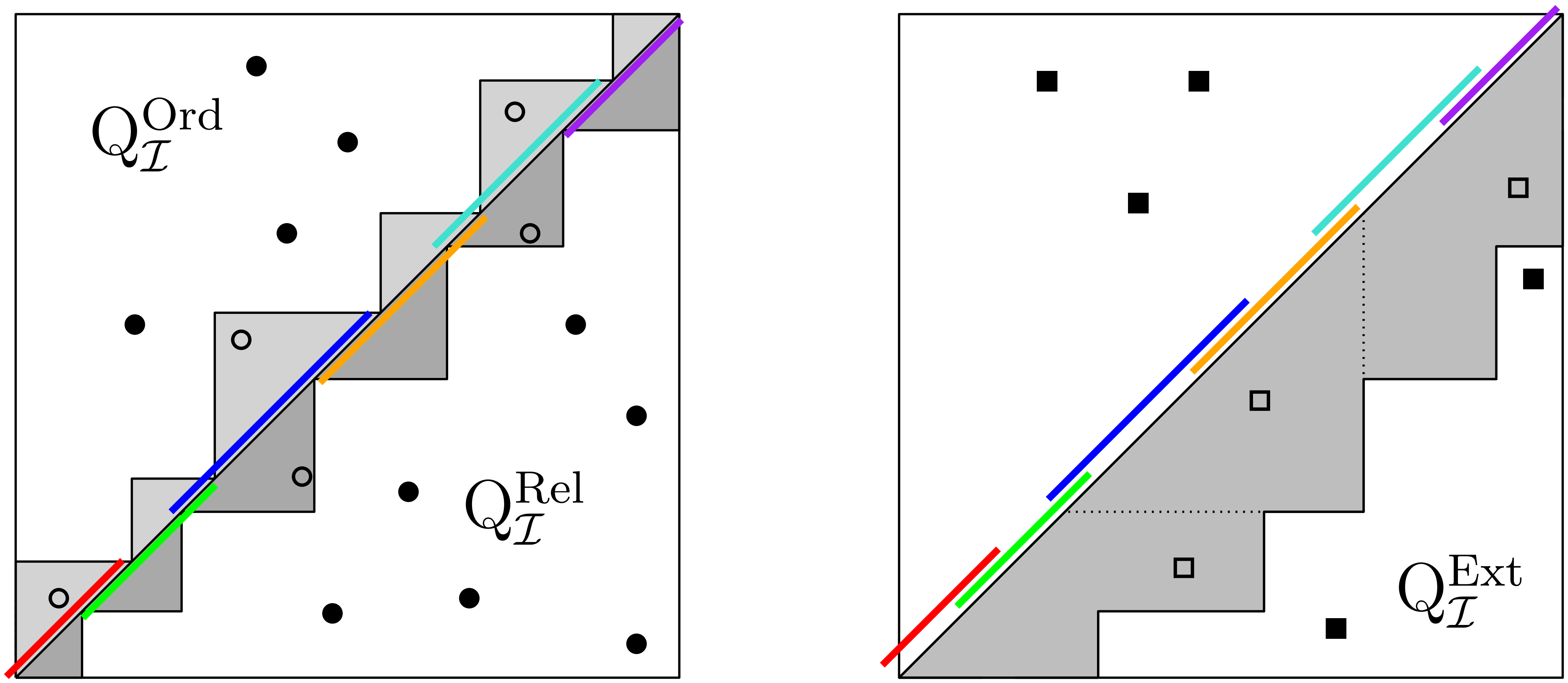
Reminder: mapper $\equiv$ *pixelized* Reeb graph



Descriptor for Mapper

Def: Given $X, f, \mathcal{I}$:

$$\mathrm{Dg}\,\mathrm{M}_f := \left(\mathrm{Ord}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ord}}\right) \cup \left(\mathrm{Rel}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Rel}}\right) \cup \left(\mathrm{Ext}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ext}}\right)$$

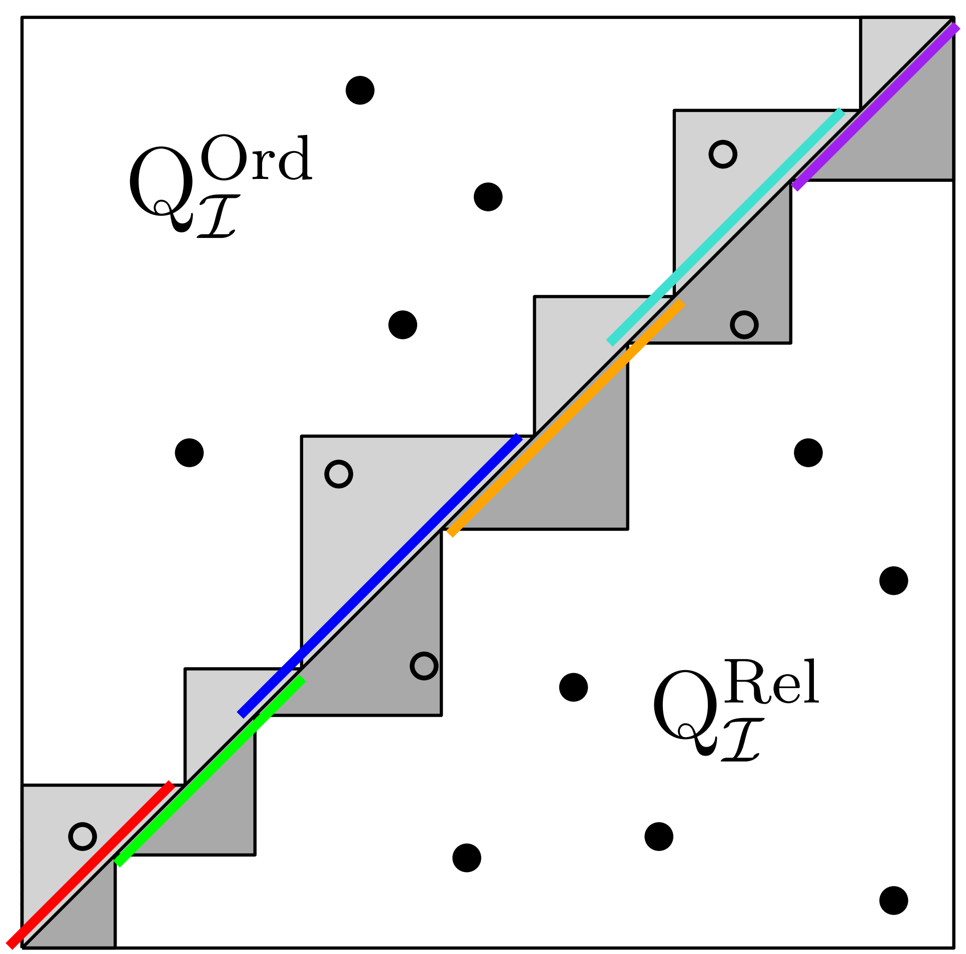


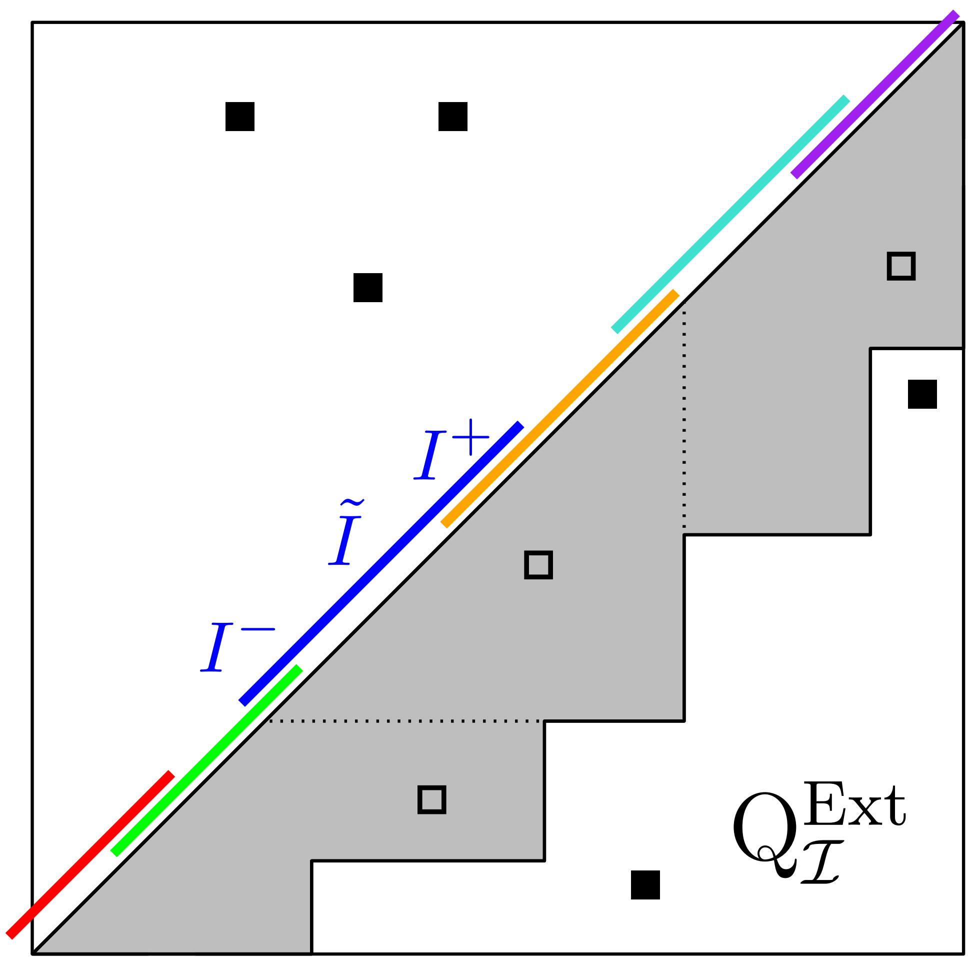
Descriptor for Mapper

Def: Given $X, f, \mathcal{I}$:

$$\mathrm{Dg}\,\mathrm{M}_f := \left(\mathrm{Ord}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ord}}\right) \cup \left(\mathrm{Rel}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Rel}}\right) \cup \left(\mathrm{Ext}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ext}}\right)$$

$$Q_{\mathcal{I}}^{\mathrm{Ord}} = \bigcup_{I \in \mathcal{I}} Q^{+}_{\tilde{I} \cup I^{+}} \qquad Q_{\mathcal{I}}^{\mathrm{Rel}} = \bigcup_{I \in \mathcal{I}} Q^{-}_{I^{-} \cup \tilde{I}} \qquad Q_{\mathcal{I}}^{\mathrm{Ext}} = \bigcup_{\substack{I, J \in \mathcal{I} \\ I \cap J \neq \emptyset}} Q^{-}_{I \cup J}$$

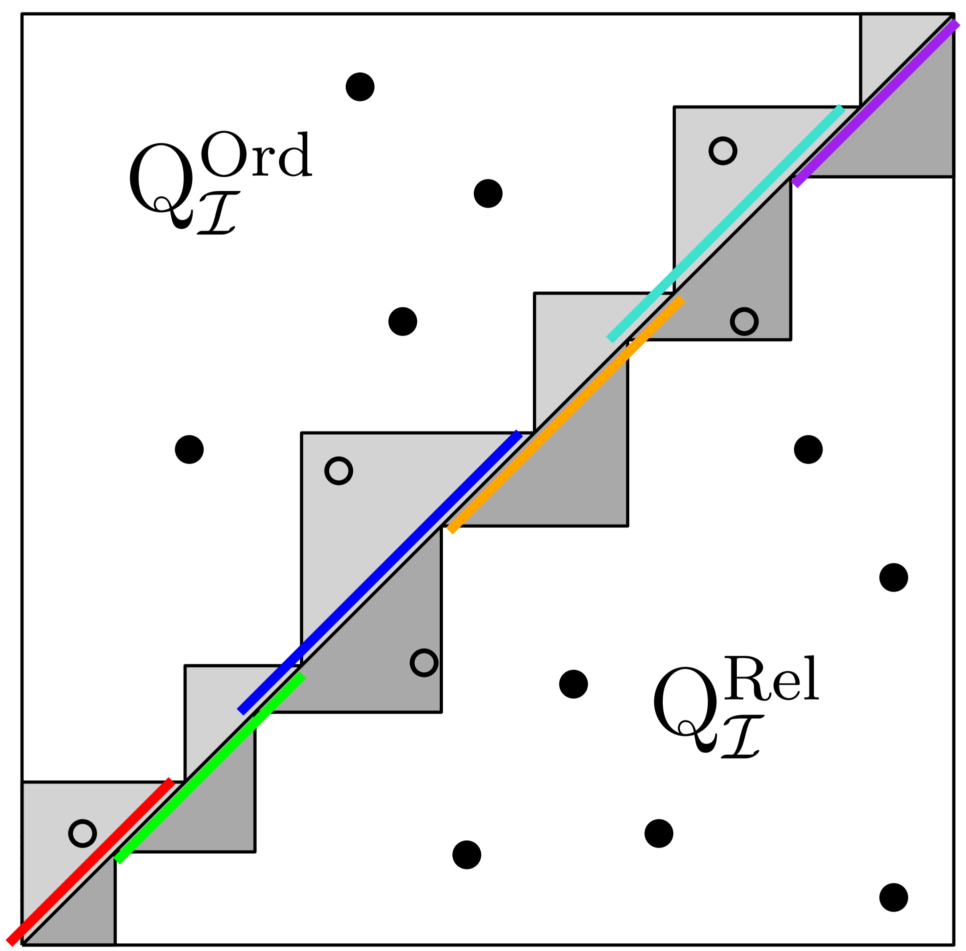


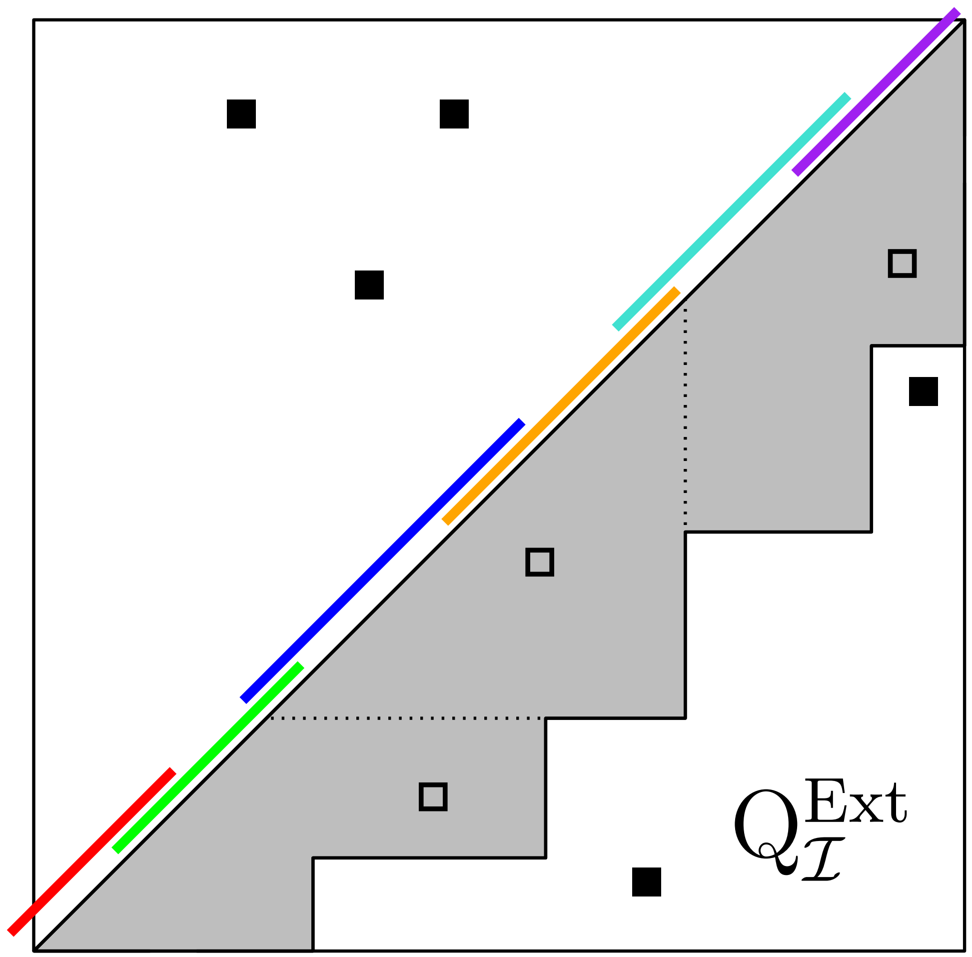


Descriptor for Mapper

Thm: [Carrière, O. 2016]
$\mathrm{Dg}\ \mathrm{M}_f$ provides a **bag-of-features** descriptor for $\mathrm{M}_f(X, \mathcal{I})$:

$\mathrm{Ord}_0 \longleftrightarrow$ downward branches	$\mathrm{Ext}_0 \longleftrightarrow$ trunks (cc)
$\mathrm{Rel}_1 \longleftrightarrow$ upward branches	$\mathrm{Ext}_1 \longleftrightarrow$ loops

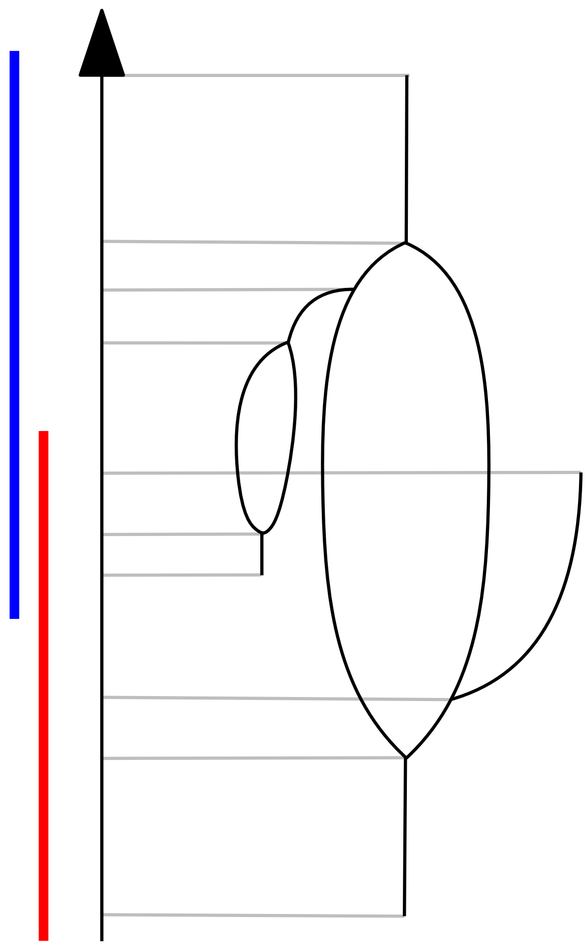


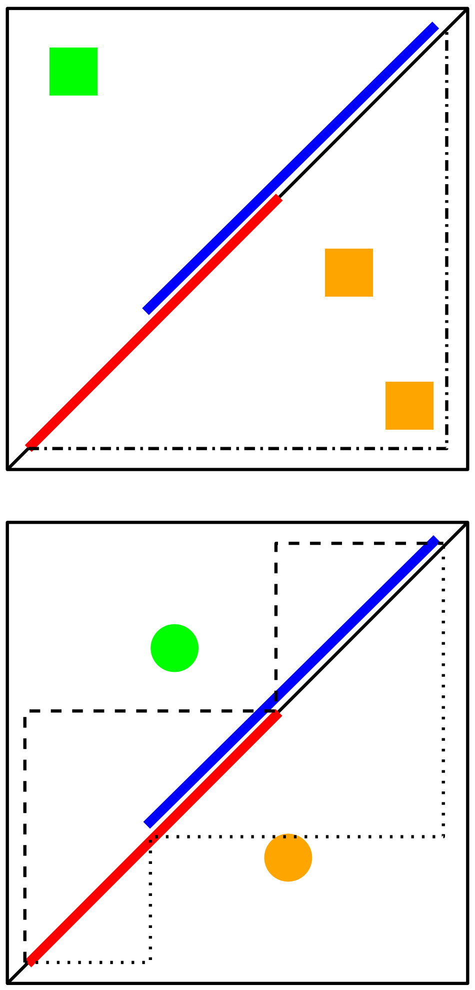


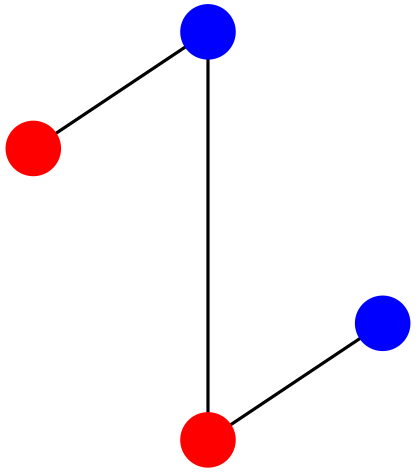
Descriptor for Mapper

Thm: [Carrière, O. 2016]
$\mathrm{Dg}\ \mathrm{M}_f$ provides a **bag-of-features** descriptor for $\mathrm{M}_f(X, \mathcal{I})$:

$\mathrm{Ord}_0 \longleftrightarrow$ downward branches	$\mathrm{Ext}_0 \longleftrightarrow$ trunks (cc)
$\mathrm{Rel}_1 \longleftrightarrow$ upward branches	$\mathrm{Ext}_1 \longleftrightarrow$ loops





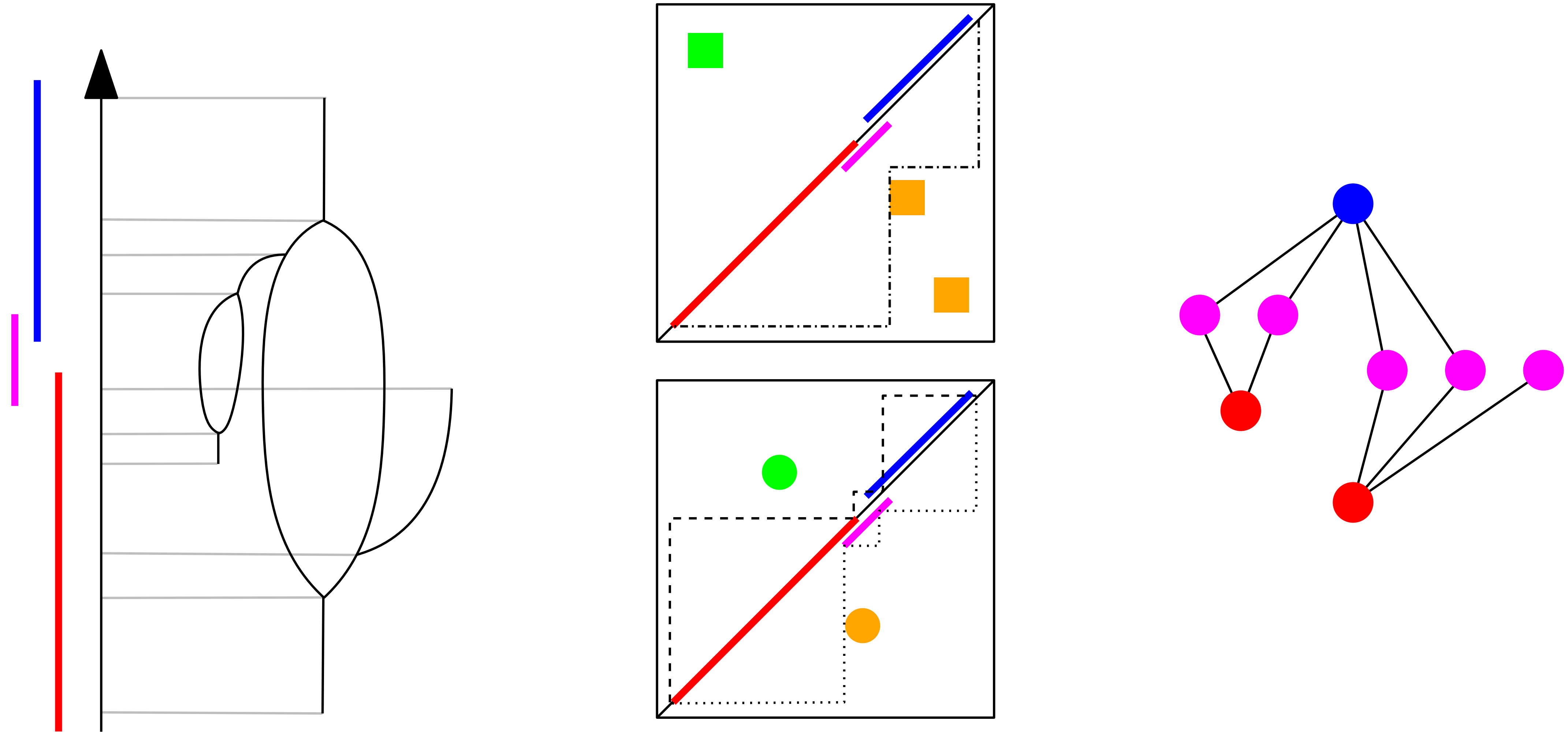


Descriptor for Mapper

Thm: [Carrière, O. 2016]

$\mathrm{Dg}\ \mathrm{M}_f$ provides a **bag-of-features** descriptor for $\mathrm{M}_f(X, \mathcal{I})$:

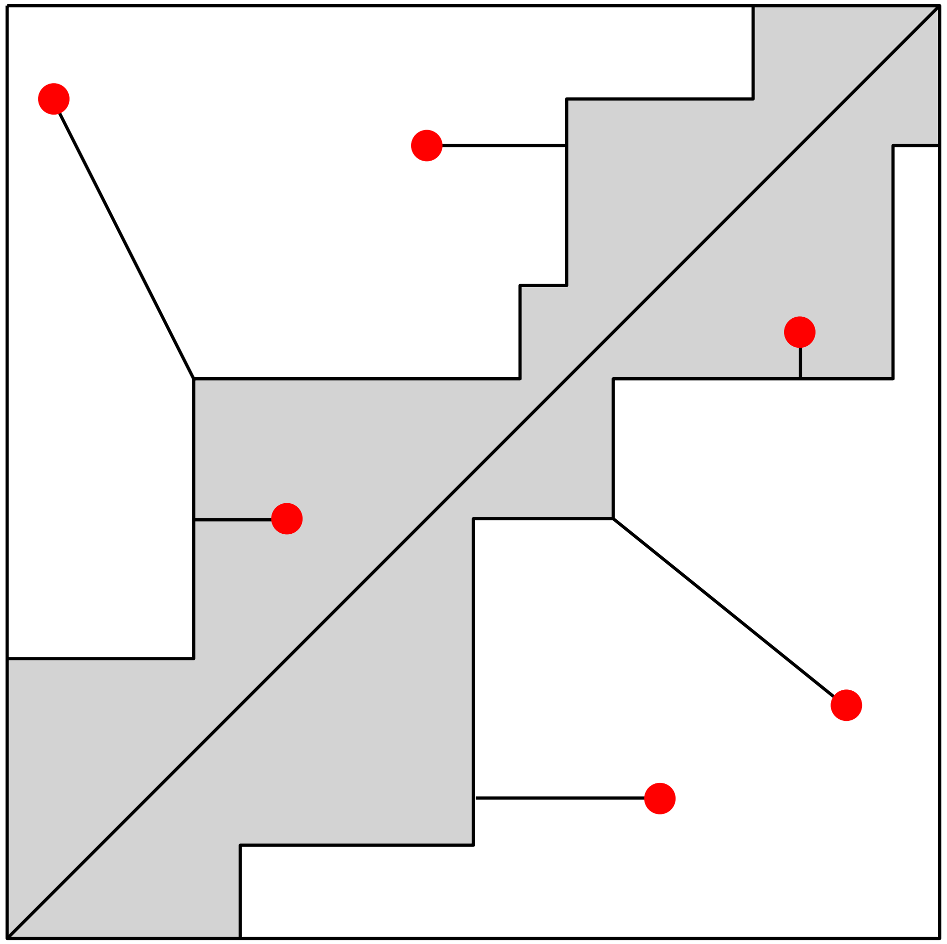
$\mathrm{Ord}_0 \longleftrightarrow$ downward branches	$\mathrm{Ext}_0 \longleftrightarrow$ trunks (cc)
$\mathrm{Rel}_1 \longleftrightarrow$ upward branches	$\mathrm{Ext}_1 \longleftrightarrow$ loops



Stability of Mapper

Definition: $\mathrm{Dg}\ \mathrm{M}_f := \left(\mathrm{Ord}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ord}}\right) \cup \left(\mathrm{Rel}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Rel}}\right) \cup \left(\mathrm{Ext}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ext}}\right)$

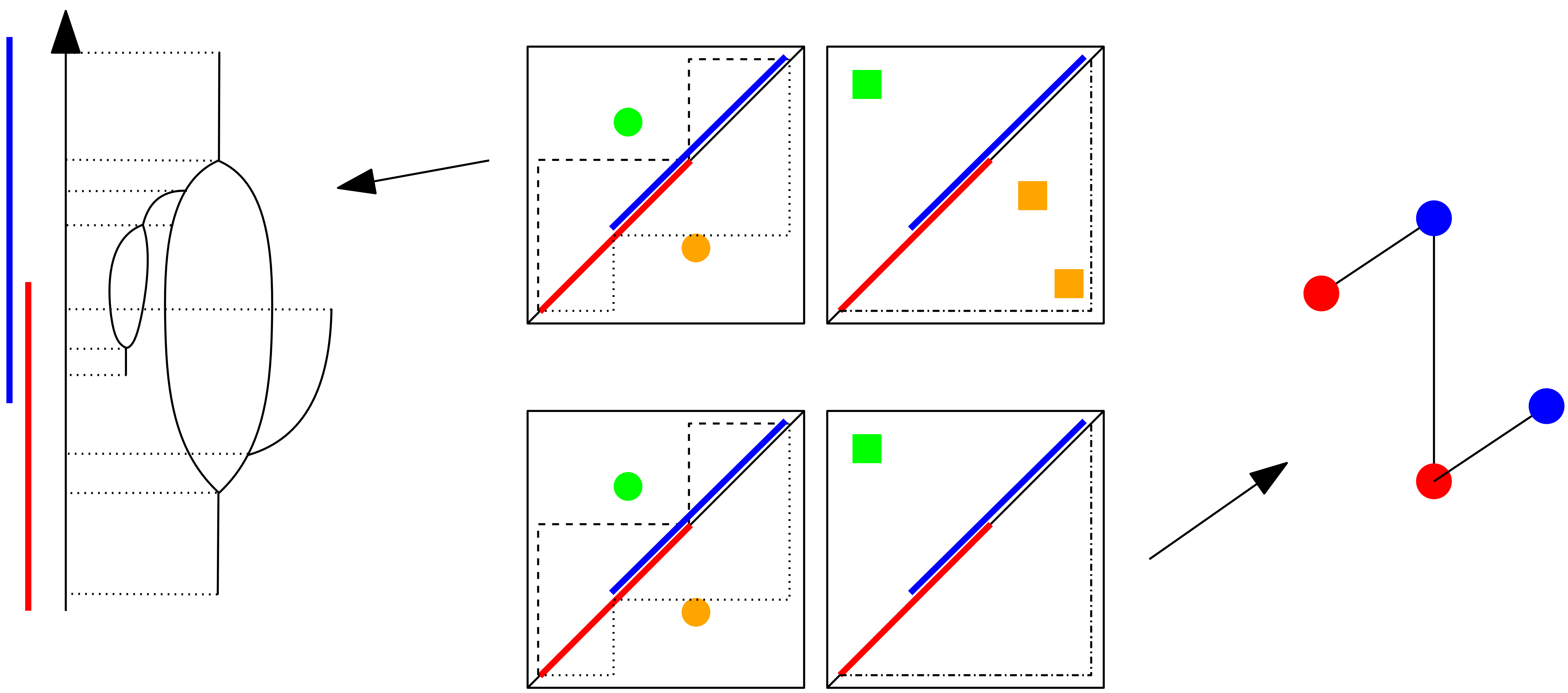
Observation: distance to staircase boundary measures (in-)stability of each feature of $\mathrm{M}_f(X, \mathcal{I})$ w.r.t. perturbations of $(X, f, \mathcal{I})$



Stability of Mapper

Definition: $\mathrm{Dg}\ \mathrm{M}_f := \left(\mathrm{Ord}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ord}}\right) \cup \left(\mathrm{Rel}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Rel}}\right) \cup \left(\mathrm{Ext}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ext}}\right)$

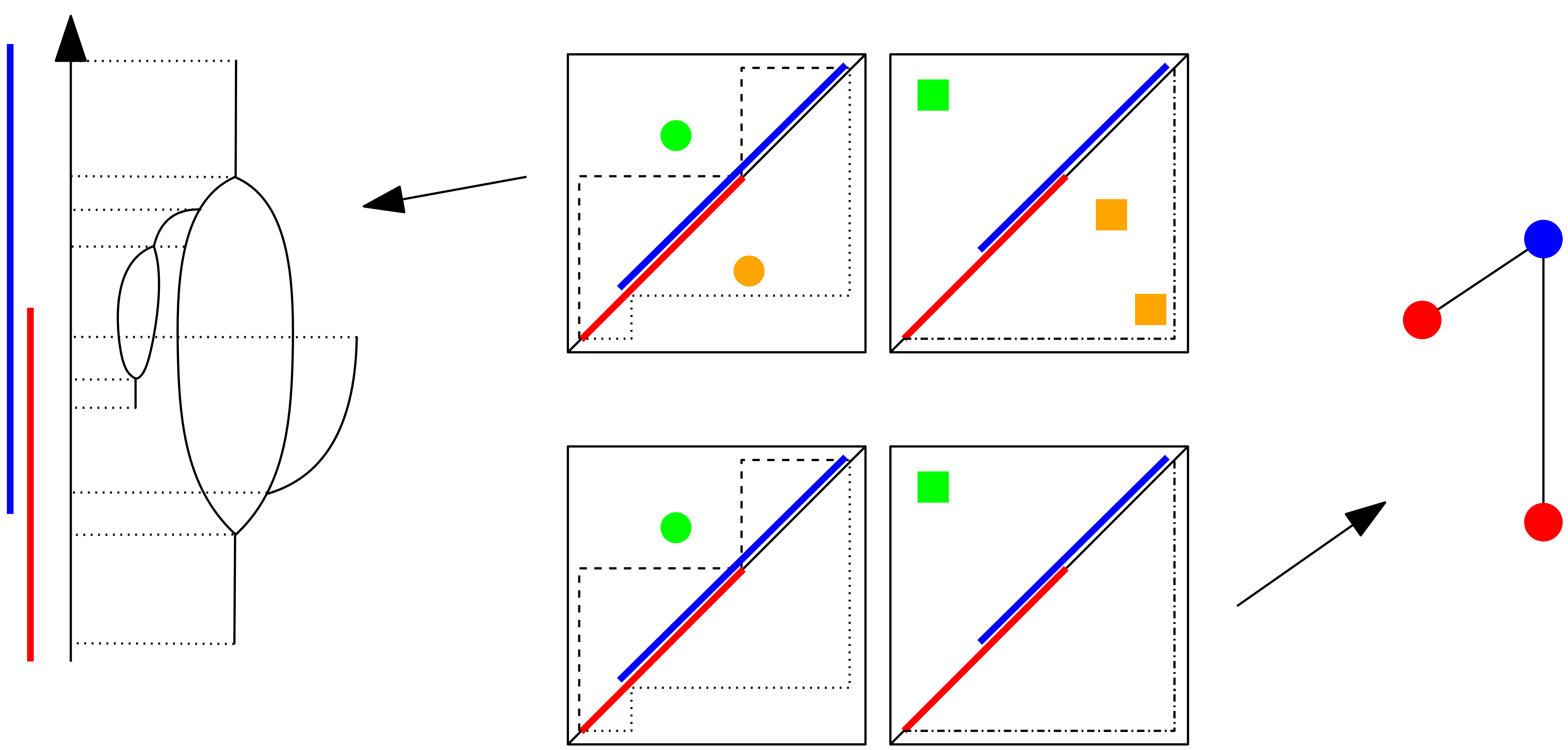
Observation: distance to staircase boundary measures (in-)stability of each feature of $\mathrm{M}_f(X, \mathcal{I})$ w.r.t. perturbations of $(X, f, \mathcal{I})$



Stability of Mapper

Definition: $\mathrm{Dg}\ \mathrm{M}_f := \left(\mathrm{Ord}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ord}}\right) \cup \left(\mathrm{Rel}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Rel}}\right) \cup \left(\mathrm{Ext}\,\mathrm{R}_f \setminus Q_{\mathcal{I}}^{\mathrm{Ext}}\right)$

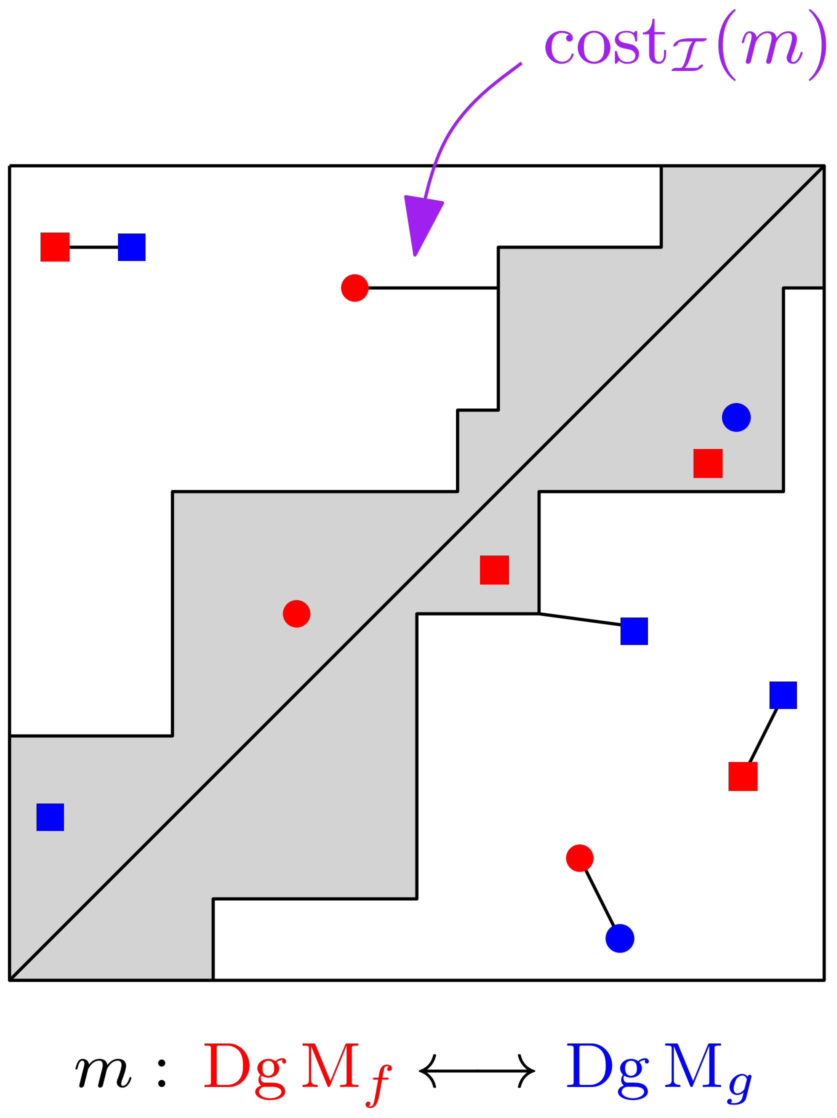
Observation: distance to staircase boundary measures (in-)stability of each feature of $\mathrm{M}_f(X, \mathcal{I})$ w.r.t. perturbations of $(X, f, \mathcal{I})$



Stability of Mapper

Definition: Given $X, \mathcal{I}$:

$$\mathrm{d}_{\mathcal{I}}(\mathrm{Dg}\,\mathrm{M}_f,\ \mathrm{Dg}\,\mathrm{M}_g) := \inf_m \mathrm{cost}_{\mathcal{I}}(\mathrm{m})$$



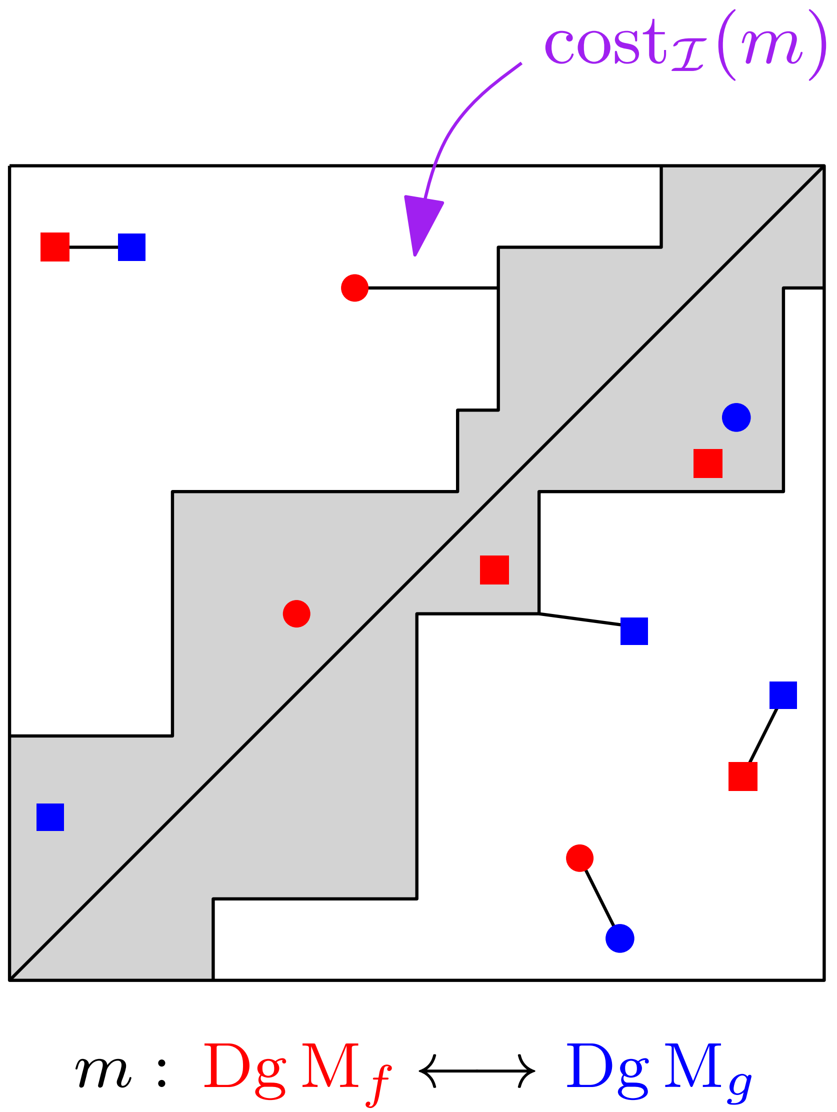
Stability of Mapper

Definition: Given $X, \mathcal{I}$:

$$\mathrm{d}_{\mathcal{I}}(\mathrm{Dg\,M}_f,\ \mathrm{Dg\,M}_g) := \inf_m \mathrm{cost}_{\mathcal{I}}(\mathrm{m})$$

Thm: [Carrière, O. 2016]
For any Morse-type functions $f, g : X \to \mathbb{R}$:

$$\mathrm{d}_{\mathcal{I}}(\mathrm{Dg\,M}_f(X, \mathcal{I}),\ \mathrm{Dg\,M}_g(X, \mathcal{I})) \leq \|f - g\|_\infty$$



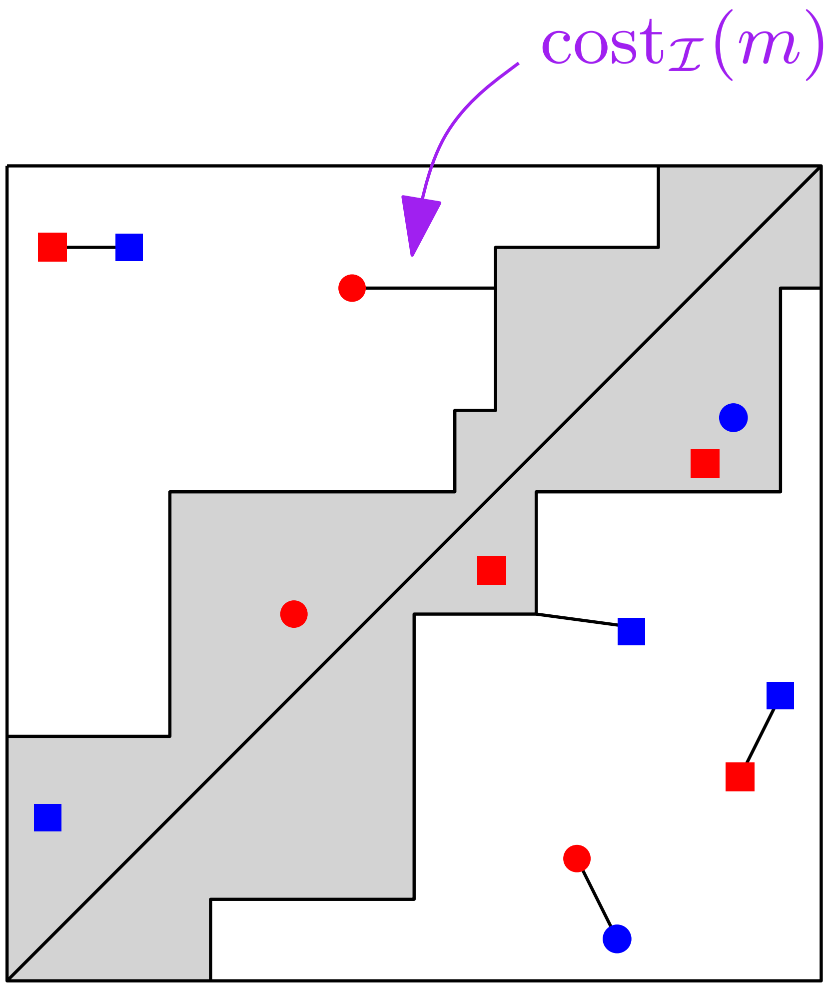
Stability of Mapper

Definition: Given $X, \mathcal{I}$:

$$\mathrm{d}_{\mathcal{I}}(\mathrm{Dg}\,\mathrm{M}_f,\ \mathrm{Dg}\,\mathrm{M}_g) := \inf_m \mathrm{cost}_{\mathcal{I}}(\mathrm{m})$$

Thm: [Carrière, O. 2016]
For any Morse-type functions $f, g : X \to \mathbb{R}$:

$$\mathrm{d}_{\mathcal{I}}(\mathrm{Dg}\,\mathrm{M}_f(X,\mathcal{I}),\ \mathrm{Dg}\,\mathrm{M}_g(X,\mathcal{I})) \leq \|f - g\|_\infty$$

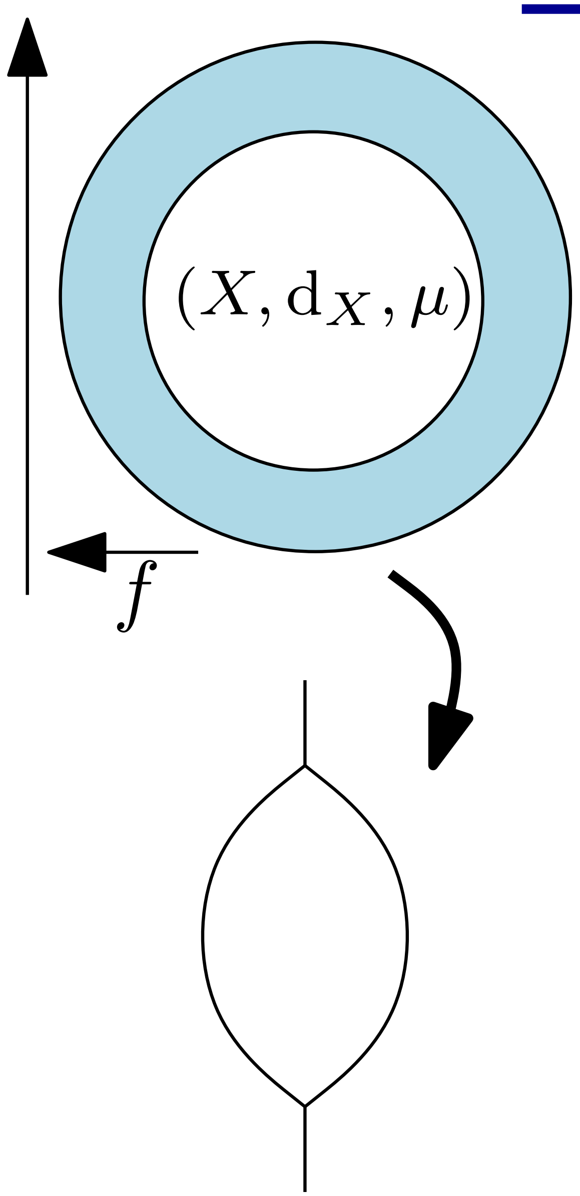


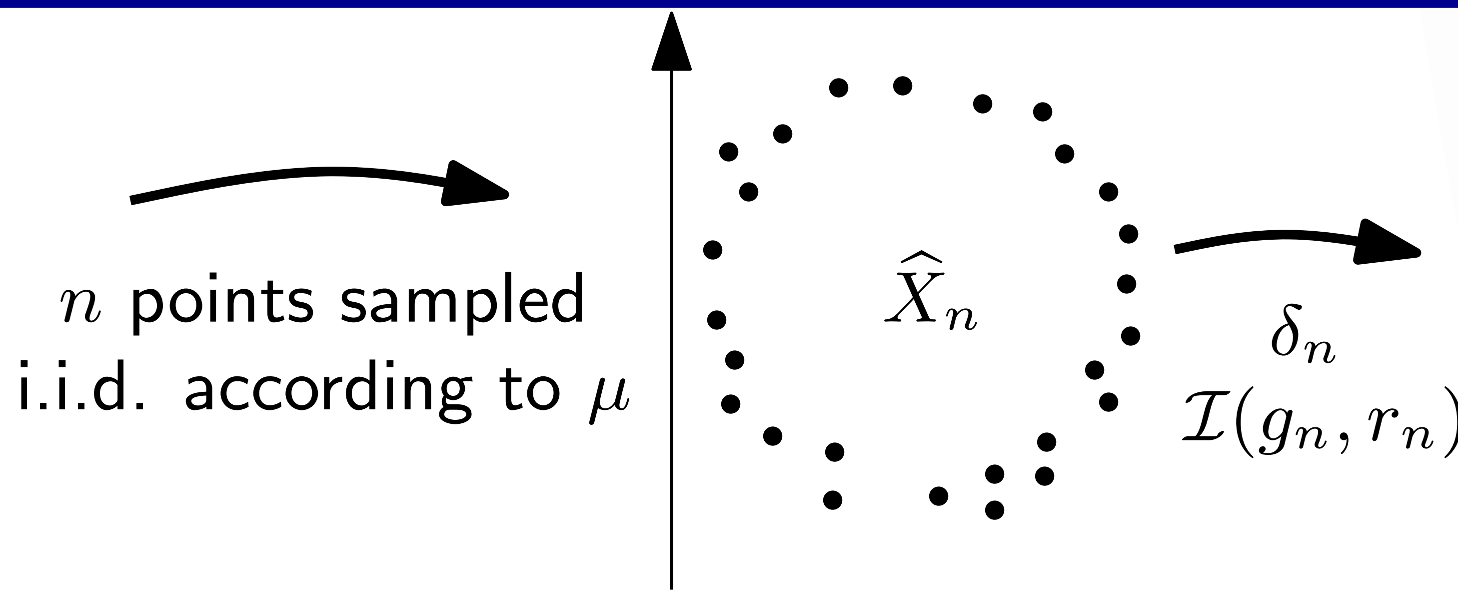
$m : \mathrm{Dg}\,\mathrm{M}_f \longleftrightarrow \mathrm{Dg}\,\mathrm{M}_g$

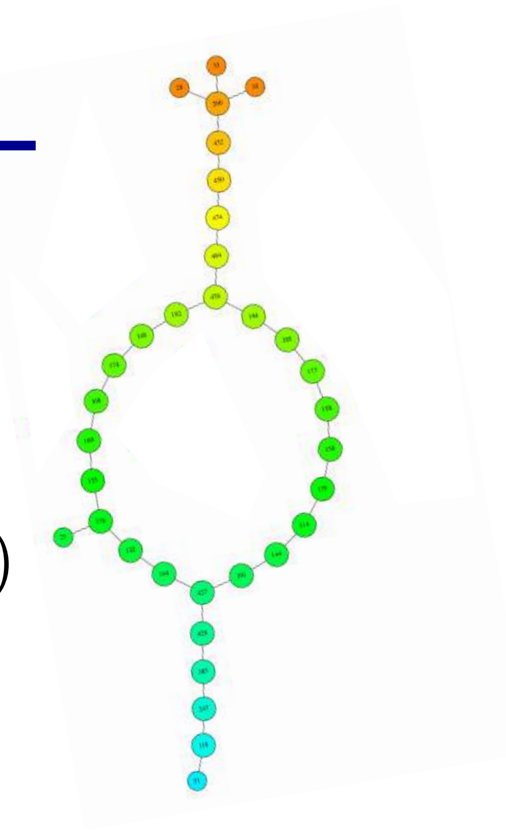
Extensions to:

- perturbations of X
- perturbations of $\mathcal{I}$

Convergence of Mapper





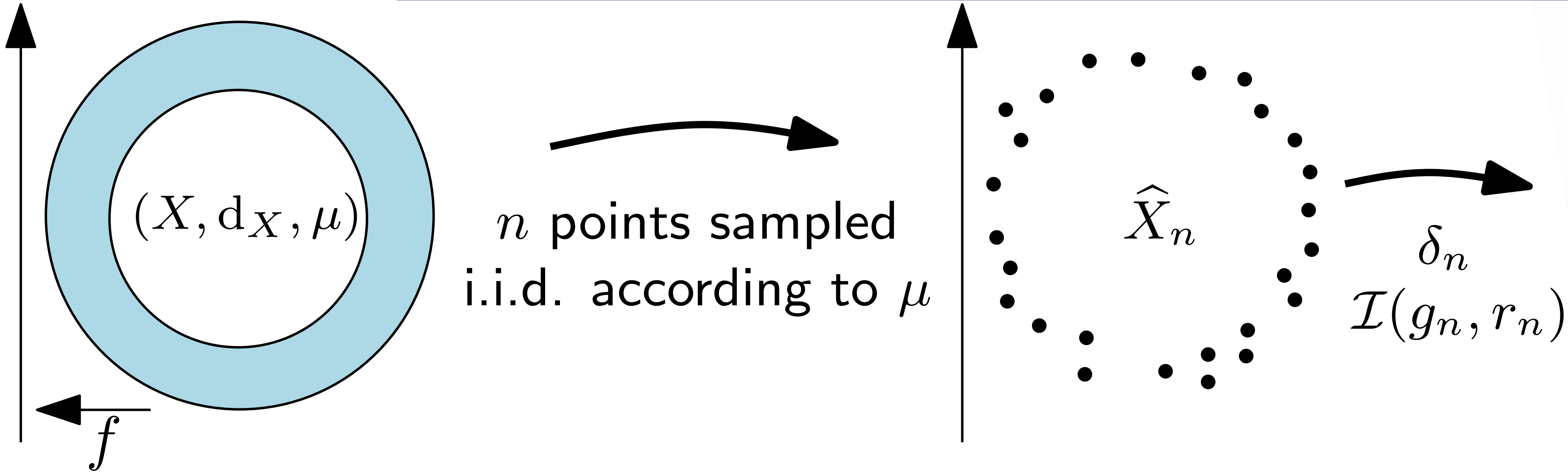


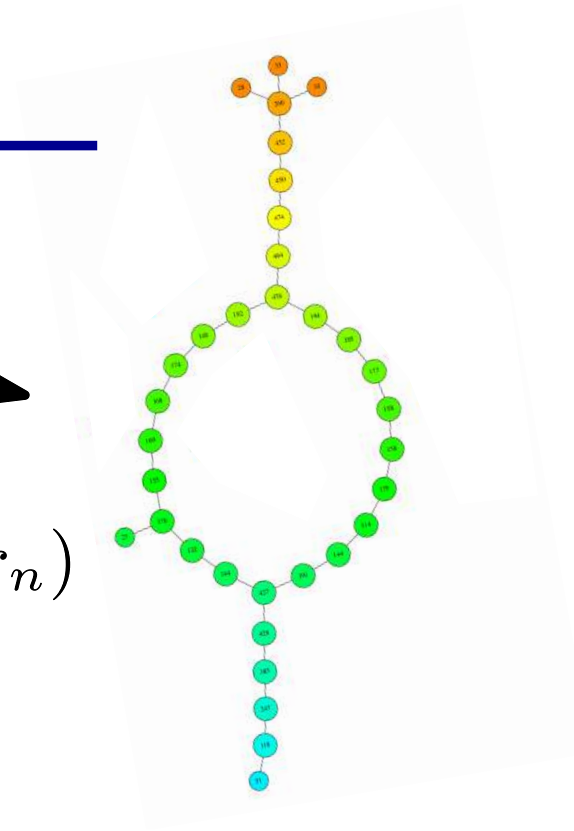
Proposition [Carrière, Michel, O. 2017]:
Both maps are measurable.

Questions:

- Statistical properties of the estimator $\mathrm{M}_f(\widehat{X}_n, \delta_n, \mathcal{I}(g_n, r_n))$?
- Convergence to the ground truth $\mathrm{R}_f(X)$ in d_B? Deviation bounds?

Convergence of Mapper





$$V_n(\delta_n) := \max\{f(X_i) - f(X_j) \mid \|X_i - X_j\| \le \delta_n\}$$

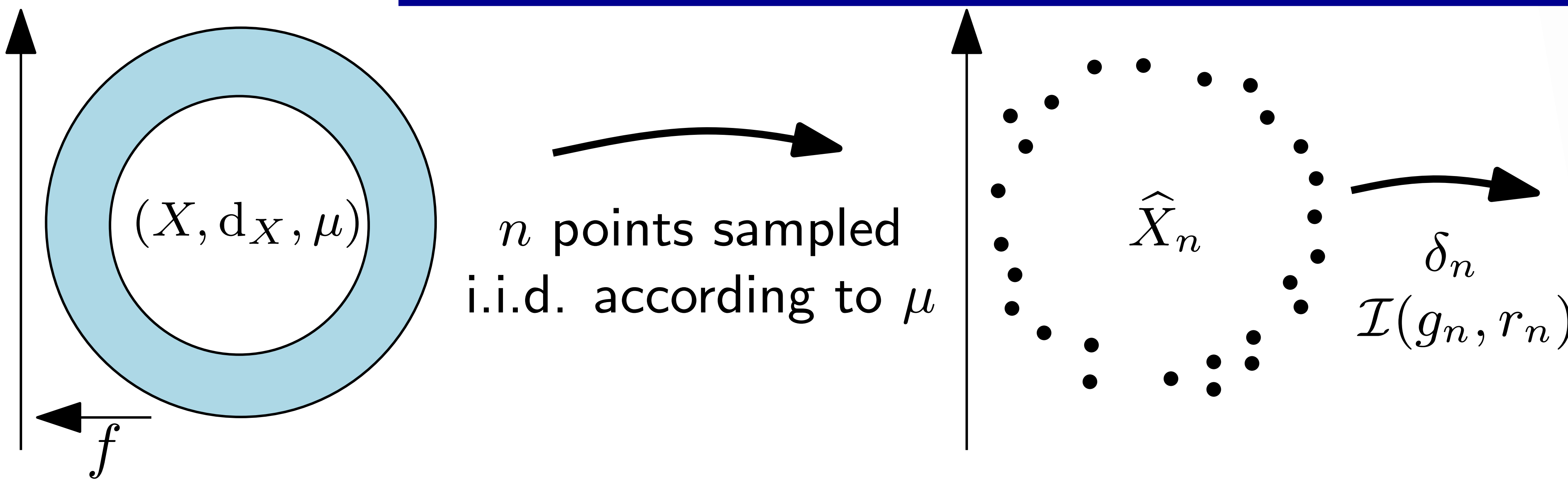
Theorem [Carrière, Michel, O. 2017]:

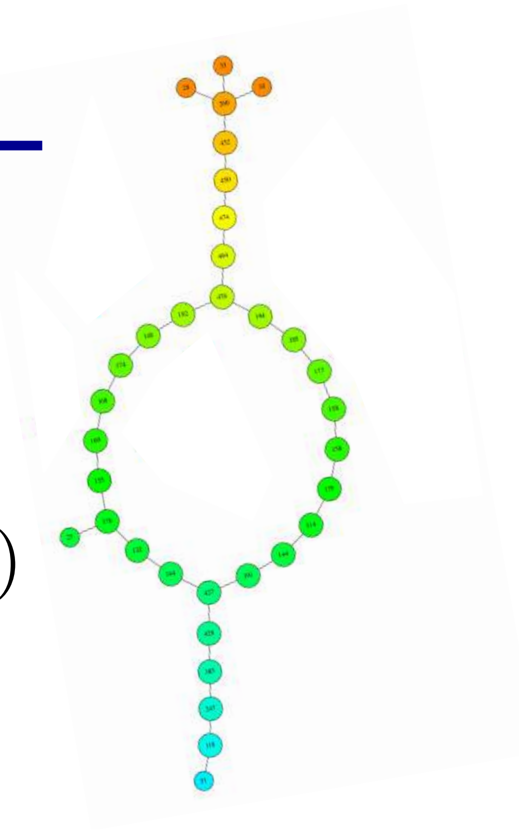
If μ is (a, b)-standard, then for $g \in \left(0, \frac{1}{2}\right)$, $\delta_n = 8\left(\frac{2\log n}{an}\right)^{1/b}$, $r_n = \frac{V_n(\delta_n)}{g}$:

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}\left[\mathrm{d}_B\left(\mathrm{Dg}\,\mathrm{R}_f(X)\right),\ \mathrm{Dg}\,\mathrm{M}_f(\widehat{X}_n, \delta_n, \mathcal{I}(g, r_n))\right] \le C\,\omega(\delta_n),$$

where ω is the *modulus of continuity* of f and C depends only on a, b. Moreover, the estimator $\mathrm{Dg}\,\mathrm{M}_f(\widehat{X}_n, \delta_n, \mathcal{I}(g, r_n))$ is **minimax optimal** (up to $\log n$ factors).

Convergence of Mapper





$$V_n(\delta_n) := \max\{f(X_i) - f(X_j) \mid \|X_i - X_j\| \le \delta_n\}$$

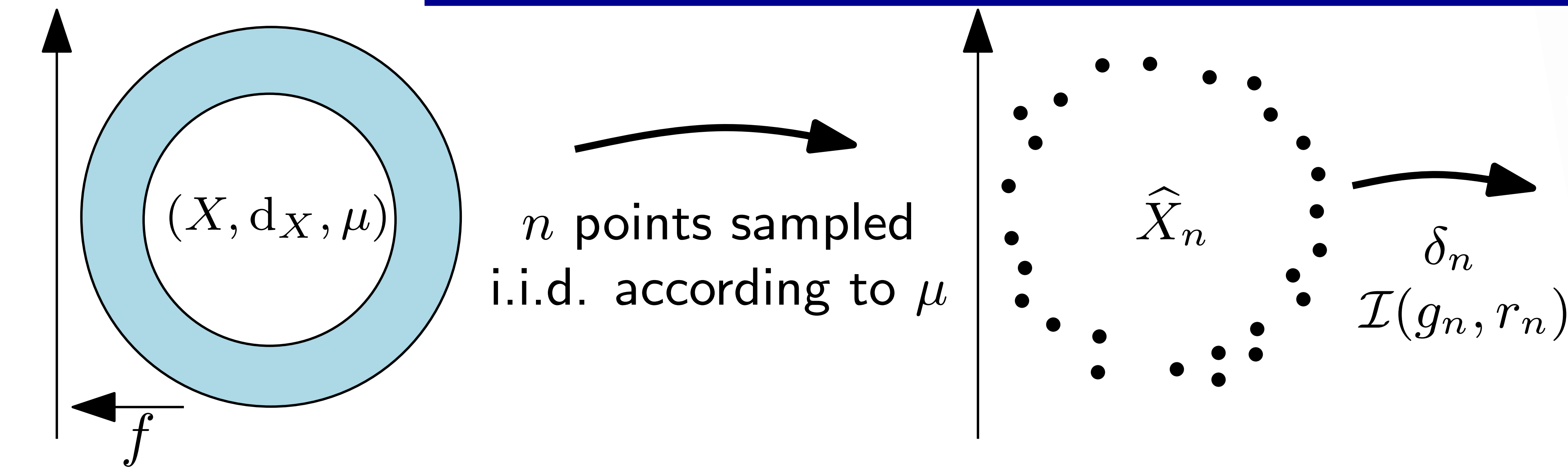
Theorem [Carrière, Michel, O. 2017]: known generative model

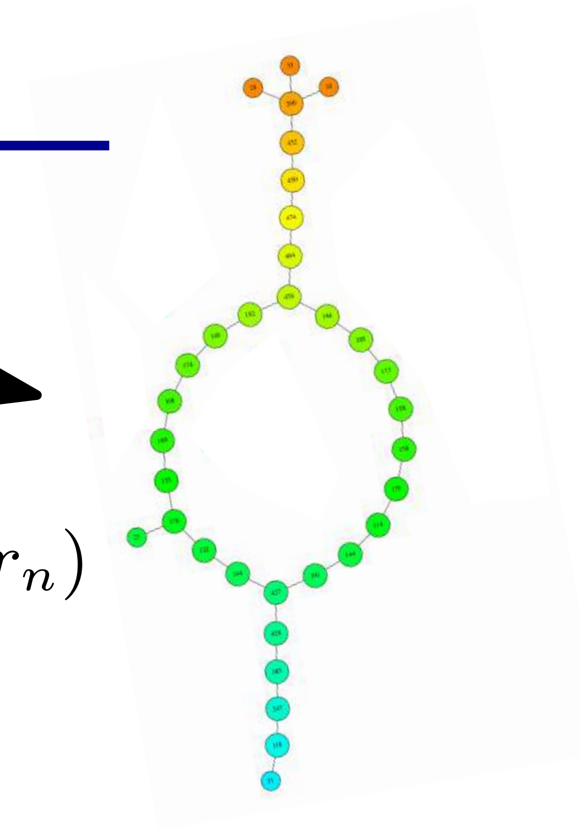
If μ is (a, b)-standard, then for $g \in \left(0, \frac{1}{2}\right)$, $\delta_n = 8\left(\frac{2\log n}{an}\right)^{1/b}$, $r_n = \frac{V_n(\delta_n)}{g}$:

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}\left[\mathrm{d}_B\left(\mathrm{Dg}\,\mathrm{R}_f(X)\right),\ \mathrm{Dg}\,\mathrm{M}_f(\widehat{X}_n, \delta_n, \mathcal{I}(g, r_n))\right] \le C\,\omega(\delta_n),$$

where ω is the *modulus of continuity* of f and C depends only on a, b. Moreover, the estimator $\mathrm{Dg}\,\mathrm{M}_f(\widehat{X}_n, \delta_n, \mathcal{I}(g, r_n))$ is **minimax optimal** (up to $\log n$ factors).

Convergence of Mapper

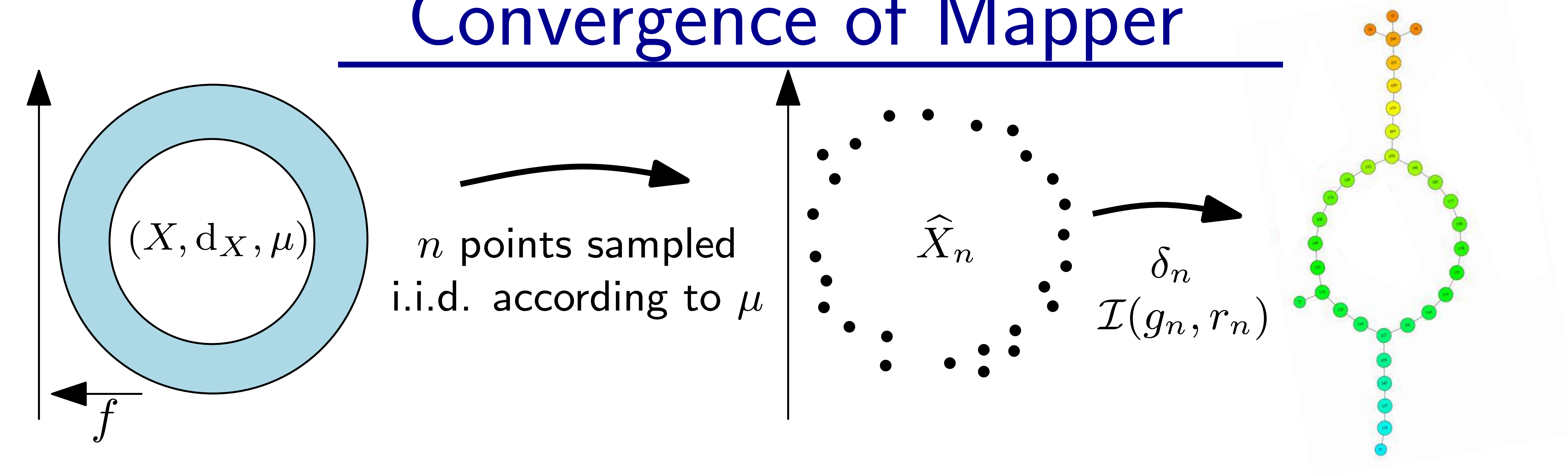




→ subsampling to tune δ_n: let $\beta > 0$ and take $s(n) = \frac{n}{\log(n)^{1+\beta}}$

$\delta_n := \mathrm{d}_{\mathrm{H}}(\hat{X}_n^{s(n)}, \hat{X}_n)$ where $\hat{X}_n^{s(n)}$ is a subset of $\hat{X}_n$ of size $s(n)$

Convergence of Mapper



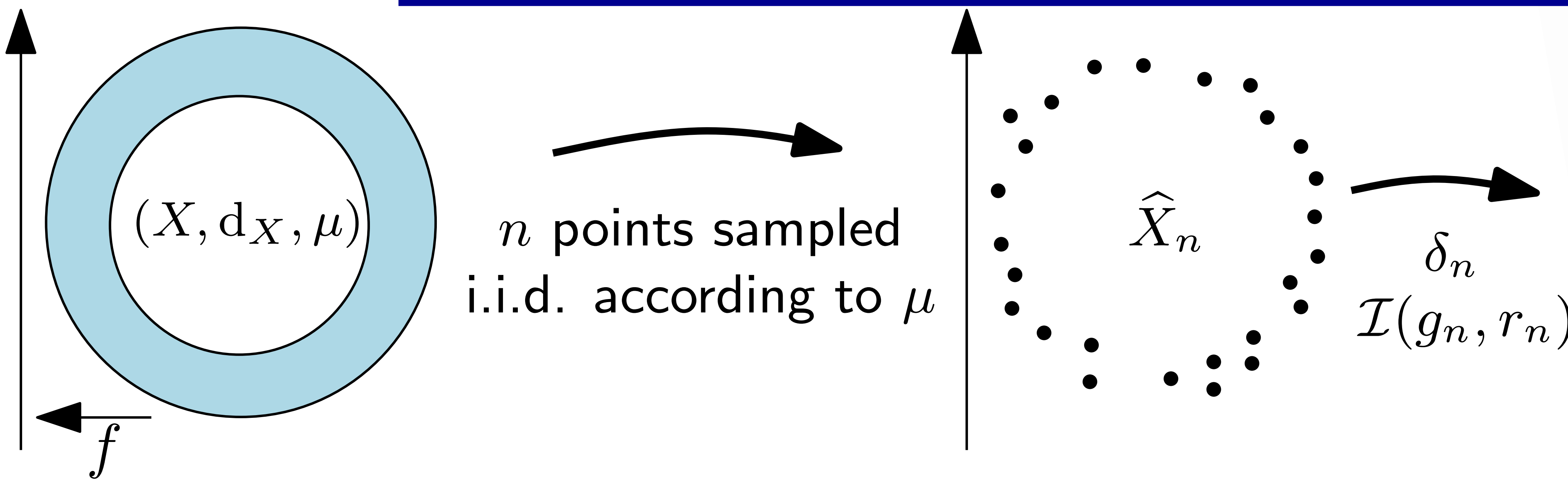
$\rightarrow$ subsampling to tune δ_n: let $\beta > 0$ and take $s(n) = \frac{n}{\log(n)^{1+\beta}}$

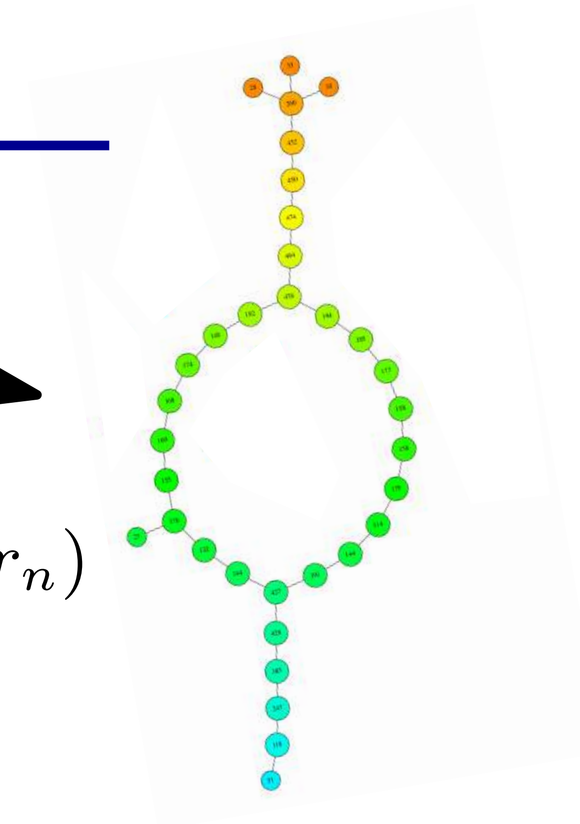
$\delta_n := \mathrm{d}_{\mathrm{H}}(\hat{X}_n^{s(n)}, \hat{X}_n)$ where $\hat{X}_n^{s(n)}$ is a subset of $\hat{X}_n$ of size $s(n)$

Theorem [Carrière, Michel, O. 2017]:
If μ is (a,b)-standard, then for $g \in \left(0, \frac{1}{2}\right)$, δ_n a above, $r_n = \frac{V_n(\delta_n)}{g}$:

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}\left[\mathrm{d}_B\left(\mathrm{Dg}\,\mathrm{R}_f(X)\right),\ \mathrm{Dg}\,\mathrm{M}_f(\widehat{X}_n, \delta_n, \mathcal{I}(g, r_n))\right] \leq C\,\omega\left(\left(\frac{\log(n)^{2+\beta}}{n}\right)^{1/b}\right)$$

Convergence of Mapper





$\rightarrow$ subsampling to tune δ_n: let $\beta > 0$ and take $s(n) = \frac{n}{\log(n)^{1+\beta}}$

$\delta_n := \mathrm{d_H}(\hat{X}_n^{s(n)}, \hat{X}_n)$ where $\hat{X}_n^{s(n)}$ is a subset of $\hat{X}_n$ of size $s(n)$

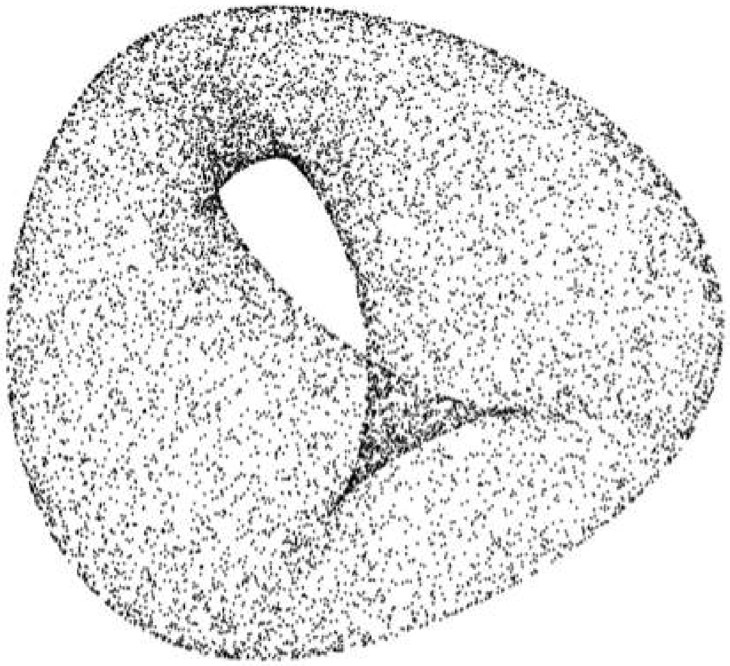
Theorem [Carrière, Michel, O. 2017]:
If μ is (a, b)-standard, then for $g \in \left(0, \frac{1}{2}\right)$, δ_n a above, $r_n = \frac{V_n(\delta_n)}{g}$:

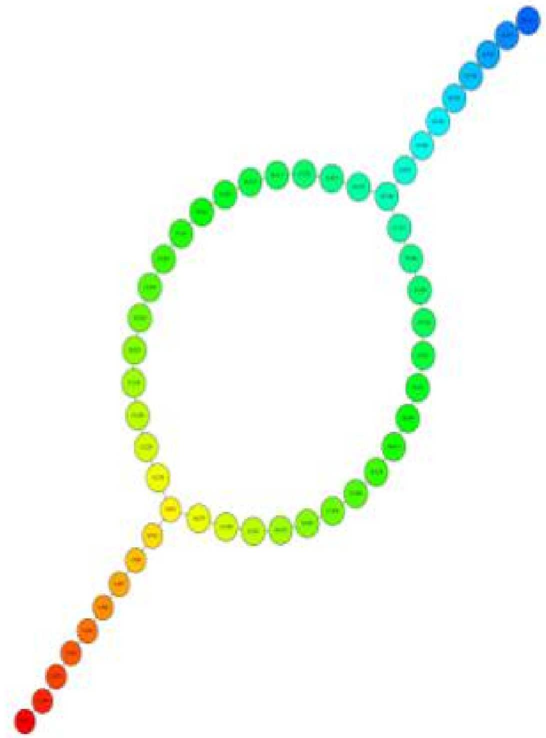
$$\sup_{\mu \in \mathcal{P}} \mathbb{E}\left[\mathrm{d}_B\left(\mathrm{Dg}\,\mathrm{R}_f(X)\right),\ \mathrm{Dg}\,\mathrm{M}_f(\widehat{X}_n, \delta_n, \mathcal{I}(g, r_n))\right] \leq C\,\omega\left(\left(\frac{\log(n)^{2+\beta}}{n}\right)^{1/b}\right)$$

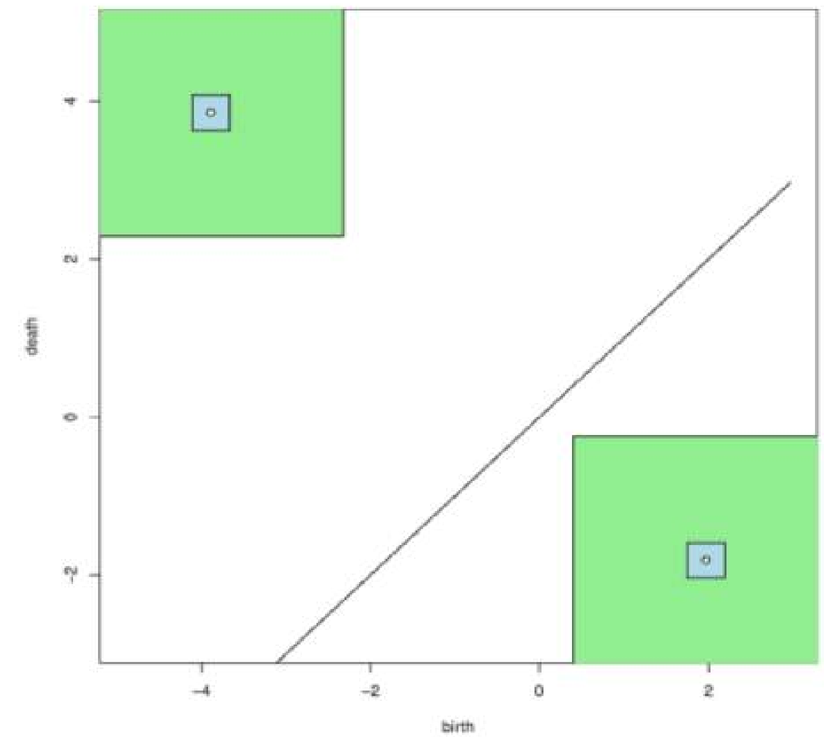
$\rightarrow$ iterate subsampling to get confidence regions

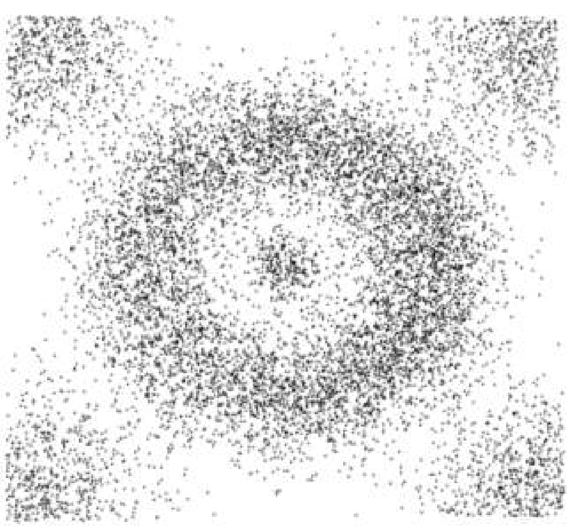
Experiments

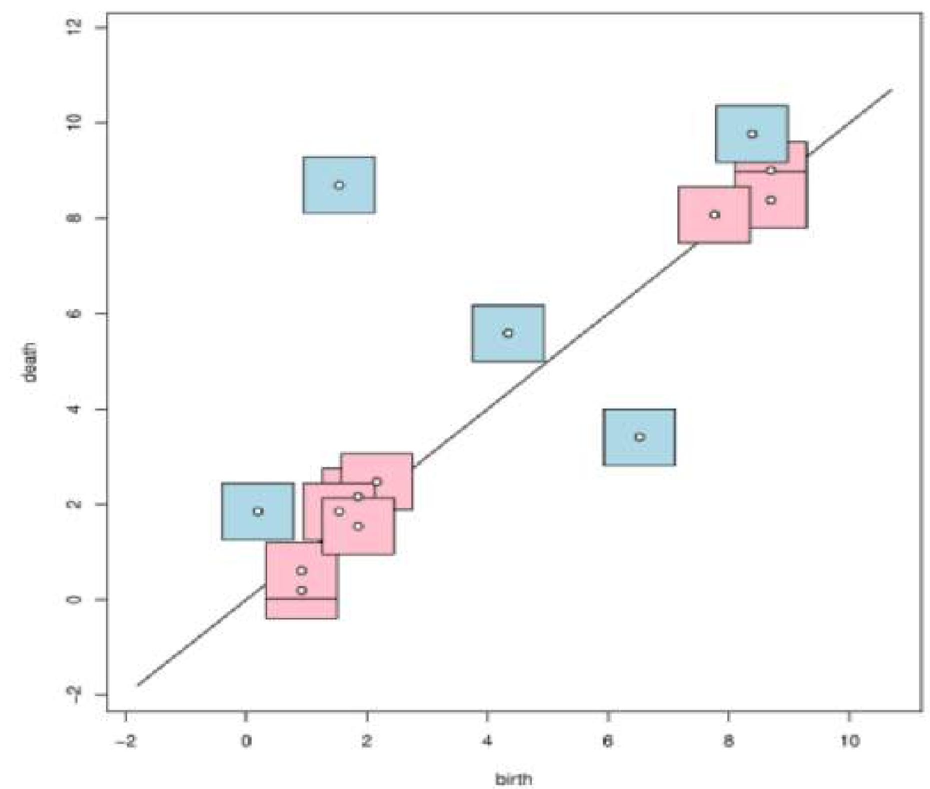
confidence level: 85%





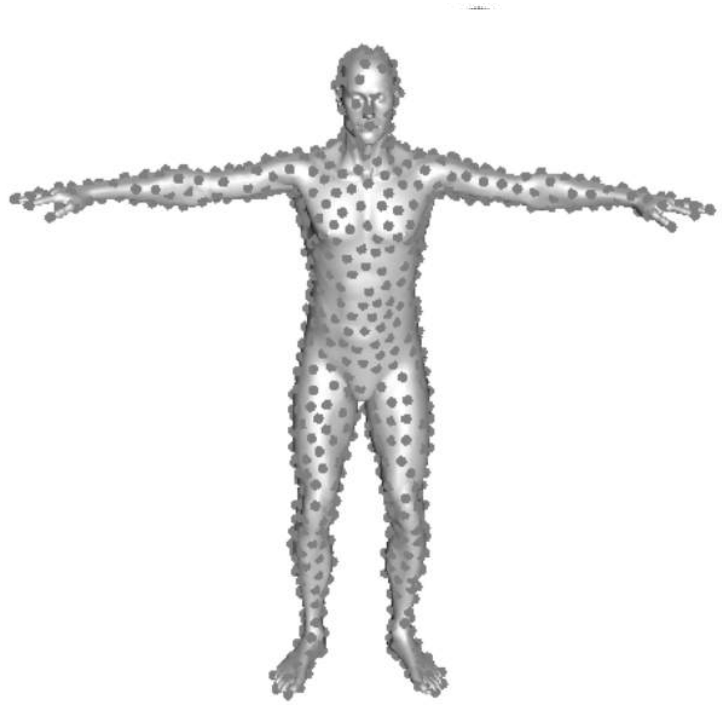


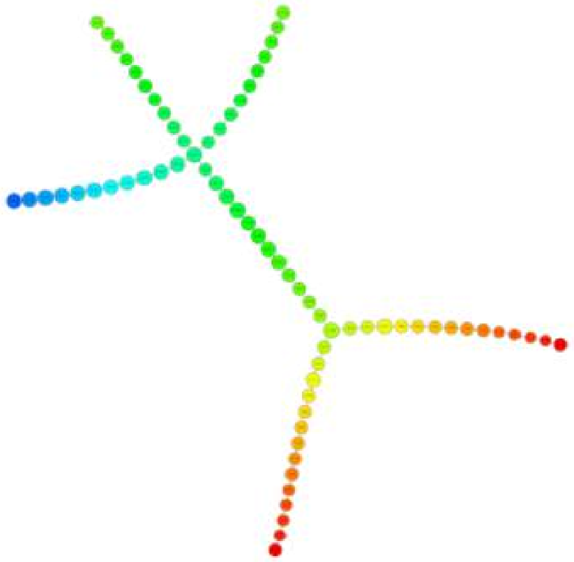


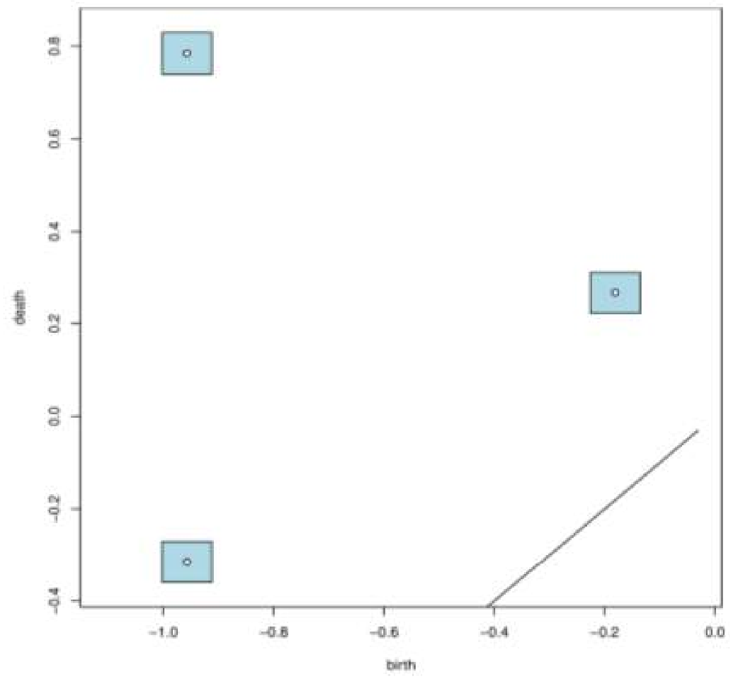


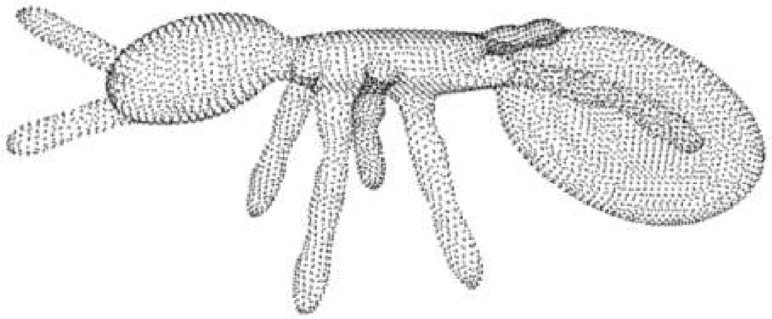
Experiments

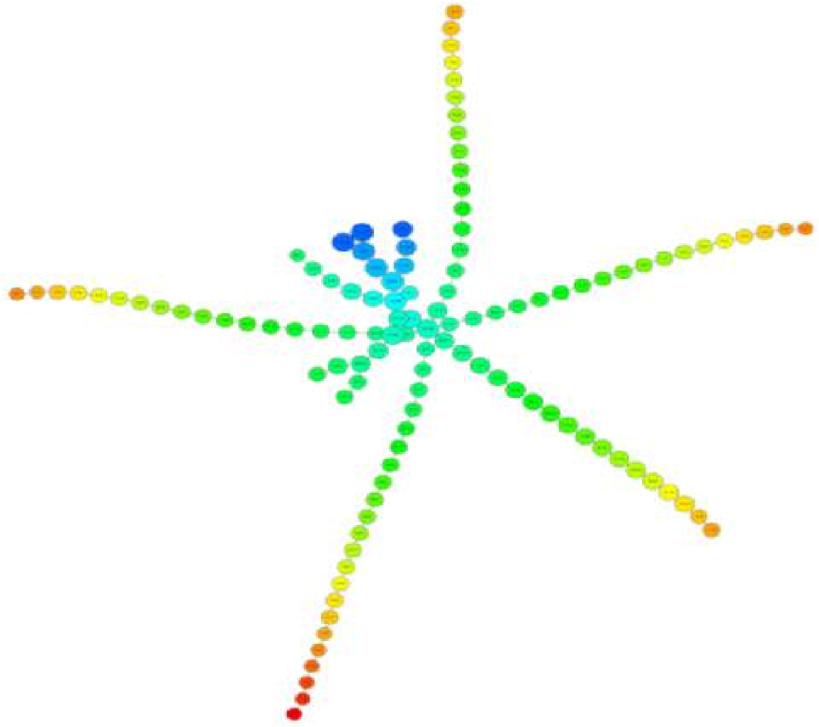
confidence level: 85%

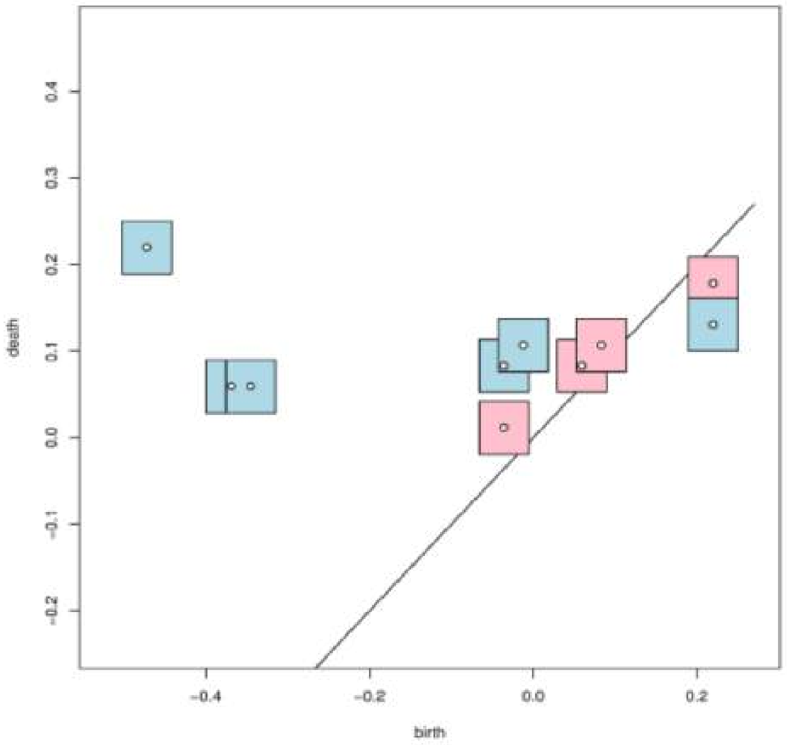






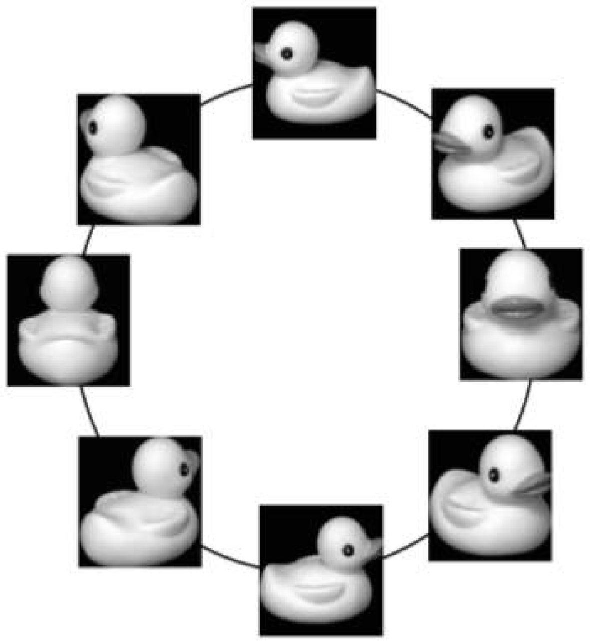


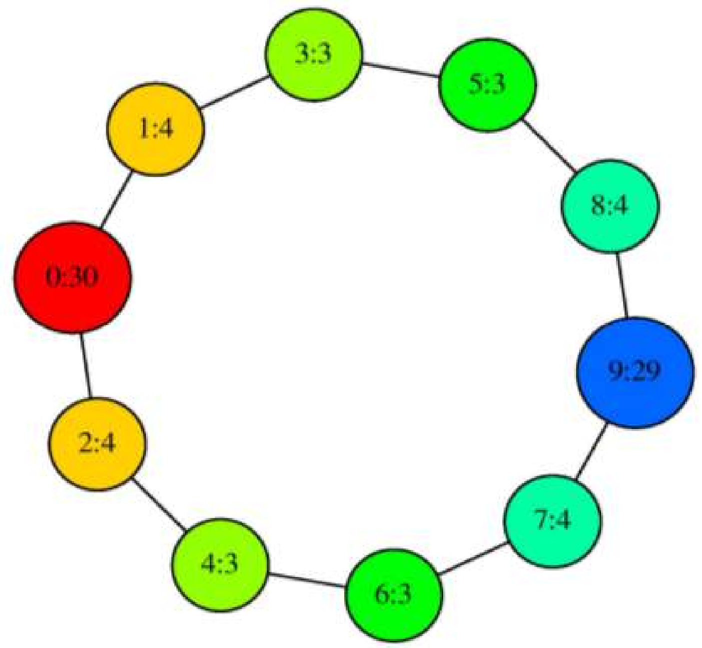


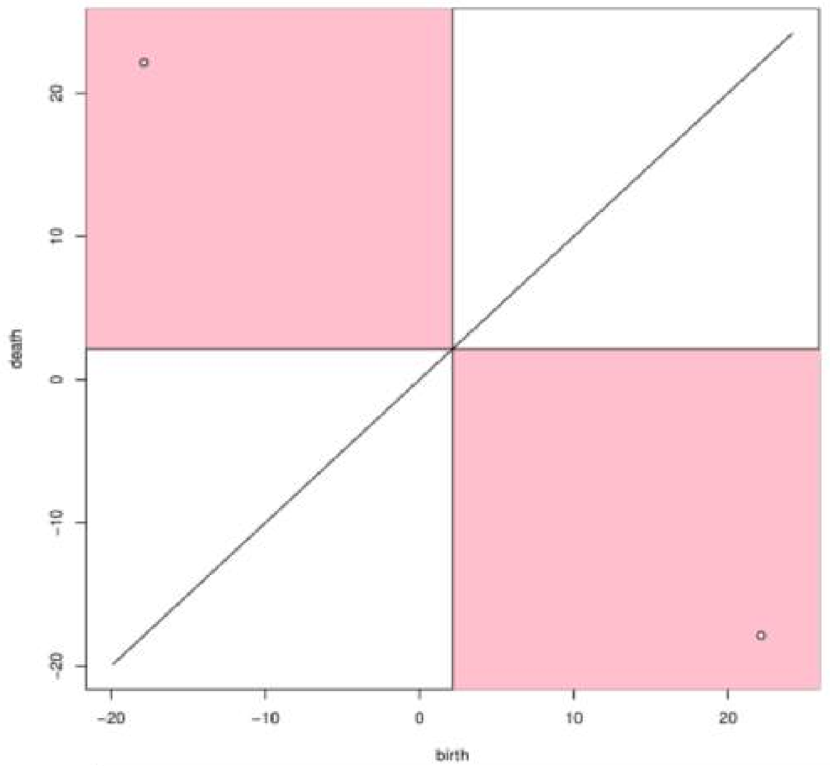


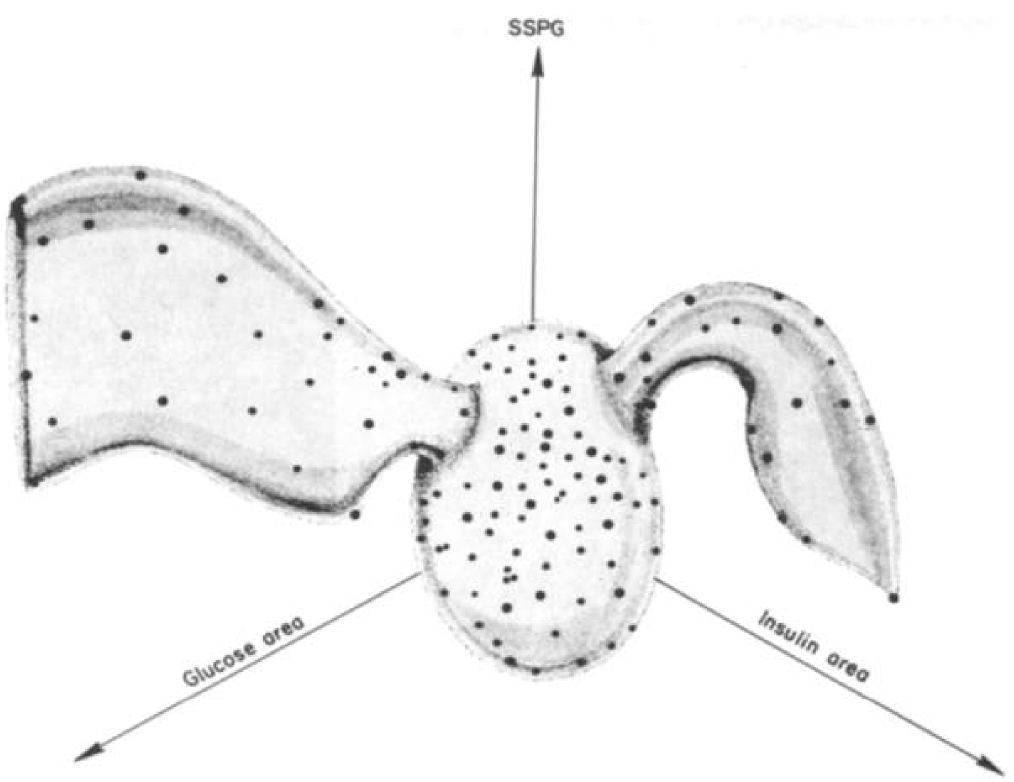
Experiments

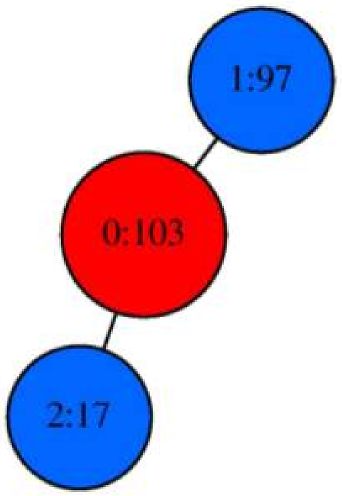
confidence level: 85%

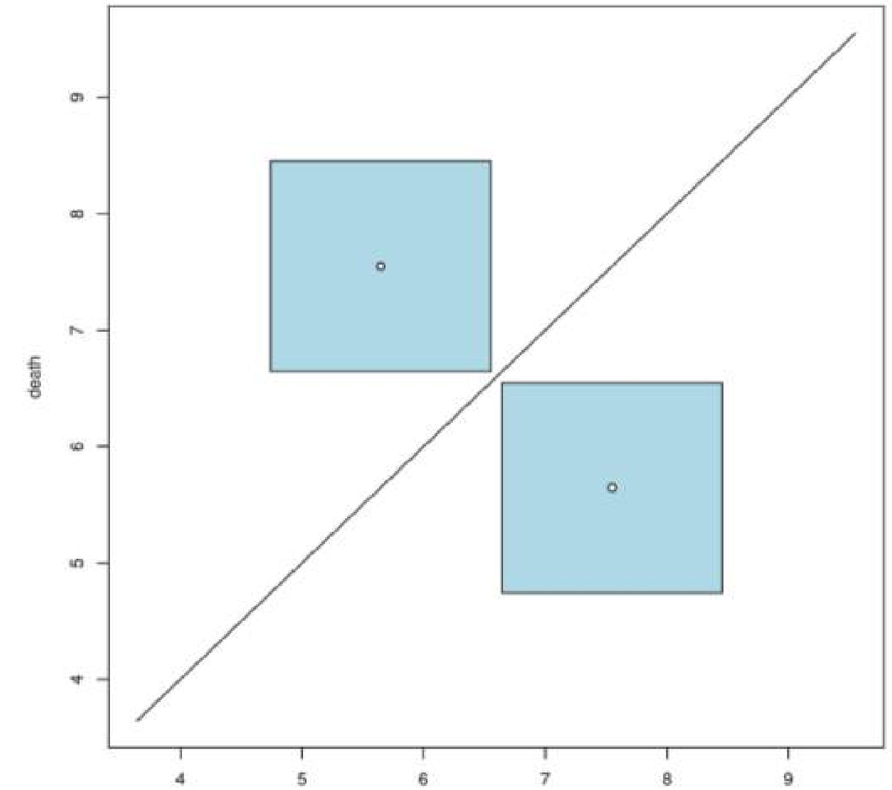












References

Metrics on Reeb graphs:

- M. Carrière, S. O. (2017): Proc. Sympos. Comput. Geom.
- de Silva, Munch, Patel (2016): Discrete Comput. Geometry, 55(4):854–906
- di Fabio, Landi (2016): Discrete Comput. Geometry, 55(2):423–461
- F. Chazal, R. Huang, J. Sun (2015): Discrete Comput. Geometry, 53(3):621–649
- U. Bauer, E. Munch, Y. Wang (2015): Proc. Sympos. Comput. Geom.
- U. Bauer, X. Ge, Y. Wang (2014): Proc. Sympos. Comput. Geom.
- Morozov, Weber (2014): Proc. TopoinVis

Mapper: structure, stability, statistics:

- M. Carrière, B. Michel, S. O. (2017): arXiv 1511.05823 [math.AT]
- M. Carrière, S. O. (2016): Proc. Sympos. Comput. Geom.
- E. Munch, B. Wang (2016): Proc. Sympos. Comput. Geom.
- G.Singh, F. Mémoli, G. Carlsson (2007): Proc. Sympos. Point-Based Graphics

Miscellaneous:

- Agarwal, Fox, Nath, Sidiropoulos, Wang (2015): Proc. Sympos. Algorithms Comput.
- Carr, Duke (2014): IEEE Trans Vis Comput Graph.
- Cohen-Steiner, Edelsbrunner, Harer (2009): J. Found. Comput. Mathematics
- M. Gromov (1981):*Structures métriques pour les variétés riemanniennes*. CEDIC