Metric and Topological Approaches to Network Analysis

Samir Chowdhury

The Ohio State University

Dissertation Defense

chowdhury.57@osu.edu

April 2, 2019

Papers compiled in this thesis

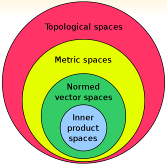
- "A functorial Dowker theorem and persistent homology of asymmetric networks", joint with F. Mémoli. Journal of Applied and Computational Topology, 2018.
- "Persistent Path Homology of Directed Networks", joint with F. Mémoli. SODA 2018.
- "Explicit geodesics in Gromov-Hausdorff space ", joint with F. Mémoli. Electronic Research Announcements of the AIMS, 2018.
- "Distances and Isomorphism between Networks and the Stability of Network Invariants", joint with F. Mémoli. arXiv preprint, 2018.
- "The Metric Space of Networks", joint with F. Mémoli. arXiv preprint, 2018.
- "Convergence of Hierarchical Clustering and Persistent Homology Methods on Directed Networks", joint with F. Mémoli. arXiv preprint, 2017.
- "The Gromov-Wasserstein distance between networks and stable network invariants", joint with F. Mémoli. arXiv preprint, 2017.

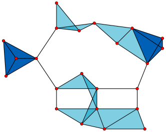
Preliminaries

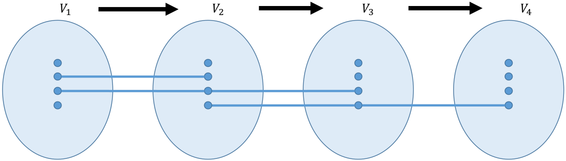
Metric space: a set of points X and a **distance function** $d_X : X \times X \to \mathbb{R}_+$.

Simplicial complex: a collection of points, line segments, triangles, tetrahedra, and higher-dimensional analogues that are "glued" together nicely. **Abstraction**: a collection of nonempty finite sets whose subsets are also in the collection.

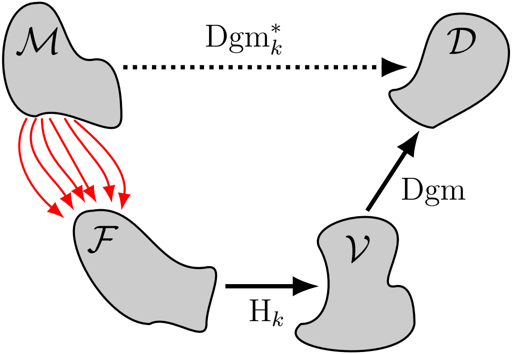
Persistent vector space: a sequence of vector spaces and liner maps.
Persistence diagram/barcode: a visual representation of how the basis vectors are mapped.







Quick intro to persistent homology



$\mathcal{M}$:= compact metric spaces, $\mathcal{F}$:= filtered simplicial complexes, $\mathcal{V}$:= persistent vector spaces, $\mathcal{D}$:= persistence diagrams

---

Figure from **"New families of simplicial filtration functors". S. Chowdhury, N. Clause, F. Mémoli, J.A. Sanchez, Z. Wellner.**

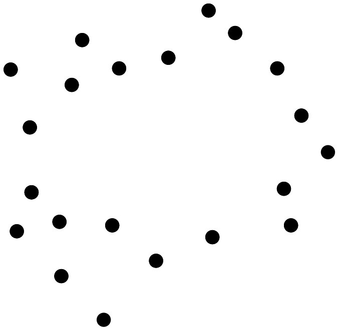
The Vietoris-Rips filtration

- (X, d_X) a metric space
- $\delta \in \mathbb{R}$

$$\mathrm{VR}_{\delta,X} := \{\sigma \text{ finite } \subseteq X : \max_{x,x' \in \sigma} d_X(x,x') \leq \delta\}.$$

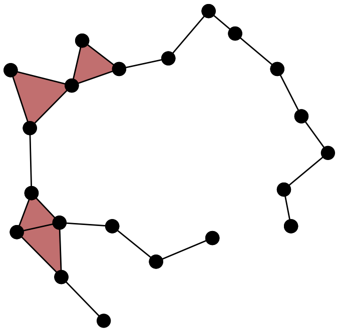
$$\text{VR filtration: } \{\mathrm{VR}_{\delta,X} \hookrightarrow \mathrm{VR}_{\delta',X}\}_{d \leq \delta'}.$$

The Vietoris-Rips filtration



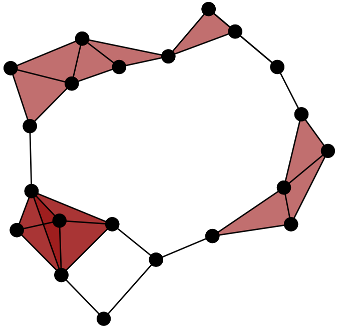
Figures courtesy of Henry Adams

The Vietoris-Rips filtration



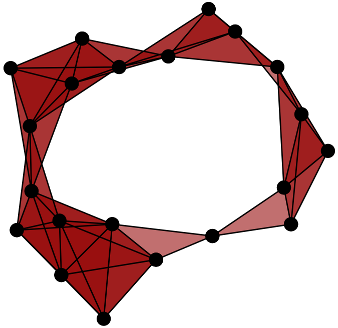
Figures courtesy of Henry Adams

The Vietoris-Rips filtration



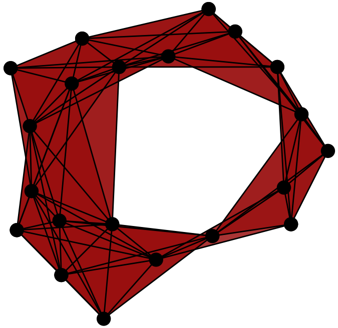
Figures courtesy of Henry Adams

The Vietoris-Rips filtration



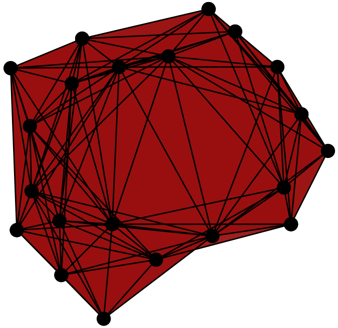
Figures courtesy of Henry Adams

The Vietoris-Rips filtration



Figures courtesy of Henry Adams

The Vietoris-Rips filtration



Figures courtesy of Henry Adams

The Vietoris-Rips filtration

- (X, d_X) a metric space
- $\delta \in \mathbb{R}$

$$\mathrm{VR}_{\delta,X} := \{\sigma \text{ finite } \subseteq X : \max_{x,x' \in \sigma} d_X(x,x') \leq \delta\}.$$

$$\text{VR filtration: } \{\mathrm{VR}_{\delta,X} \hookrightarrow \mathrm{VR}_{\delta',X}\}_{d \leq \delta'}.$$

- **Reconstructibility:** the metric space can be recovered from the filtration
- $d_X(x, x')$ is just the filtration value of $[x, x']$

Alternative filtrations

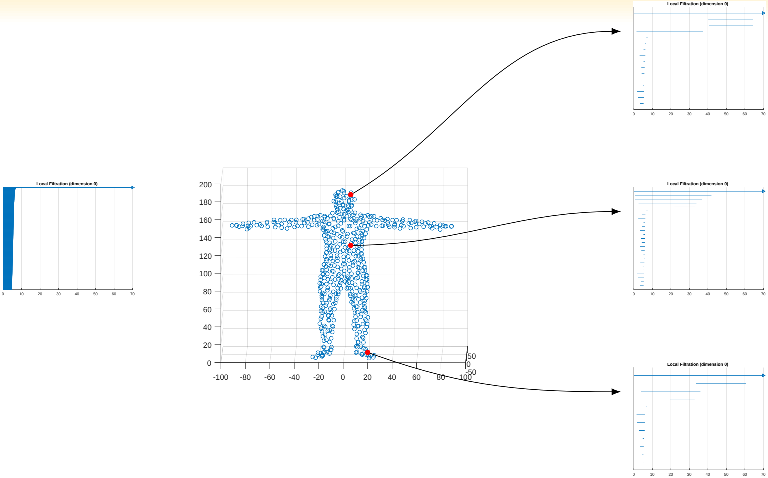


Figure from **"New families of simplicial filtration functors". S. Chowdhury, N. Clause, F. Mémoli, J.A. Sanchez, Z. Wellner.**

Alternative filtrations

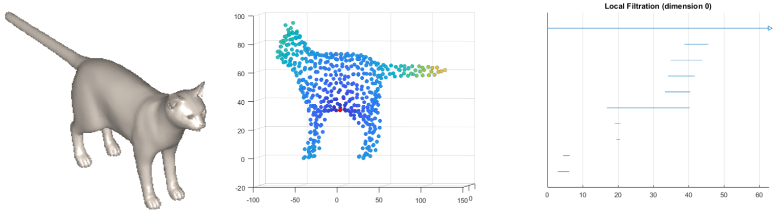
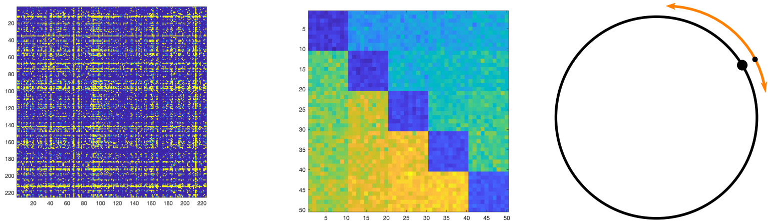


Figure from **"New families of simplicial filtration functors". S. Chowdhury, N. Clause, F. Mémoli, J.A. Sanchez, Z. Wellner.**

Non-metric datasets



Left: Global female bilateral migration in 2000, **Center:** a toy stochastic block model network, **Right:** a directed circle with finite reversibility.

Datasets in the wild are often non-metric!

Note that the migration dataset is a square, (possibly) asymmetric matrix, which in turn can be thought of as a weighted, directed graph. This justifies the **network** terminology that we will soon introduce.

The Vietoris-Rips filtration on directed metric spaces

- (X, d_X) a **directed** metric space (symmetry relaxed)
- $\delta \in \mathbb{R}$

$$\mathrm{VR}_{\delta,X} := \{\sigma \text{ finite } \subseteq X : \max_{x,x' \in \sigma} d_X(x,x') \leq \delta\}.$$

$$\text{VR filtration: } \{\mathrm{VR}_{\delta,X} \hookrightarrow \mathrm{VR}_{\delta',X}\}_{d \leq \delta'}.$$

The Vietoris-Rips filtration on directed metric spaces

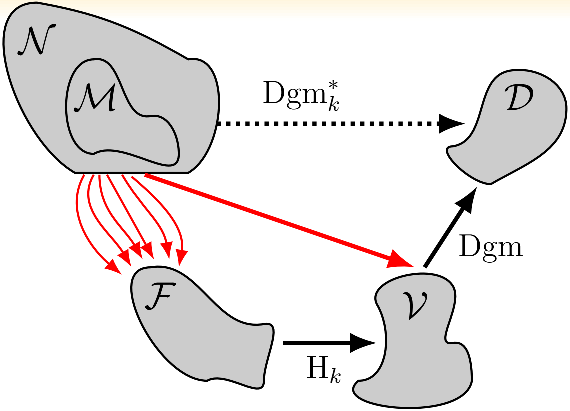
- (X, d_X) a **directed** metric space (symmetry relaxed)
- $\delta \in \mathbb{R}$

$$\mathrm{VR}_{\delta,X} := \{\sigma \text{ finite } \subseteq X : \max_{x,x' \in \sigma} d_X(x,x') \leq \delta\}.$$

$$\text{VR filtration: } \{\mathrm{VR}_{\delta,X} \hookrightarrow \mathrm{VR}_{\delta',X}\}_{d \leq \delta'}.$$

- **Reconstructibility?**
- At the (simplicial) chain complex level, have equality $[x, x'] = -[x', x]$
- One filtration value ($[x, x']$), two items to recover $(d_X(x,x'), d_X(x',x))$—**one** bit of information is **lost**

Our work



Topology: we studied methods of obtaining persistence diagrams from asymmetric data

Metric: we built a theoretically sound framework for "asymmetric data"

Our context

- (2012) Grigor'yan, Lin, Muranov, and Yau developed a **homology theory for digraphs**.
- (Summer 2014) We became interested in developing a "persistent homology theory for digraphs". Main obstructions: (1) computation, (2) stability. **Upside**: avoiding information loss!

Our context

- (2012) Grigor'yan, Lin, Muranov, and Yau developed a **homology theory for digraphs**.
- (Summer 2014) We became interested in developing a "persistent homology theory for digraphs". Main obstructions: (1) computation, (2) stability. **Upside**: avoiding information loss!
- (2013-2014) Carlsson, Mémoli, Ribeiro, Segarra studied hierarchical clustering on **directed networks** (X, A_X), where X is a finite set and $A_X : X \times X \to \mathbb{R}_+$ is a dissimilarity ($A_X(x, x') = 0$ iff $x = x'$). They also defined a **distance** $d_{\mathcal{N}}$, and proved stability w.r.t. $d_{\mathcal{N}}$.
- ($\sim$ December 2014) We restricted to the setting of **networks** $\mathcal{N}$, started developing "persistent homology of directed networks" using the **Dowker complex**, and simultaneously studied the structure of $\mathcal{N}$. **Upside/contribution**: theoretical setup + first practical implementation, software package.

More context

- Chazal, de Silva, Oudot (2014) observed that VR and Čech persistence diagrams could be defined for **infinite, totally bounded,** symmetric dissimilarity spaces.

More context

- Chazal, de Silva, Oudot (2014) observed that VR and Čech persistence diagrams could be defined for **infinite, totally bounded,** symmetric dissimilarity spaces.
- Turner (2016) considered several generalizations of the Vietoris-Rips complex to **finite** dissimilarity spaces where even symmetry is relaxed, proved that the persistence diagrams are stable. One generalization, the **directed flag complex**, is “morally non-simplicial”. Turner asked if this could be done in the infinite setting. (**Spoiler**: this is resolved by our study of the metric properties of $\mathcal{N}$.)

More context

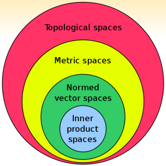
- Chazal, de Silva, Oudot (2014) observed that VR and Čech persistence diagrams could be defined for **infinite, totally bounded,** symmetric dissimilarity spaces.
- Turner (2016) considered several generalizations of the Vietoris-Rips complex to **finite** dissimilarity spaces where even symmetry is relaxed, proved that the persistence diagrams are stable. One generalization, the **directed flag complex**, is "morally non-simplicial". Turner asked if this could be done in the infinite setting. (**Spoiler**: this is resolved by our study of the metric properties of $\mathcal{N}$.)
- Lütgehetmann (2018) developed an efficient implementation of persistent homology using the directed flag complex. Dłotko, Hess, Levi, Markram, and others used the directed flag complex (w/ non-persistent homology) to study digital reconstructions of rat neuronal circuitry, and are currently pursuing this neuroscience application using the persistent counterpart.

Part I: Networks and $d_{\mathcal{N}}$

Question: what deserves to be called a network?

- Finite networks: X a finite set, $\omega_X : X \times X \to \mathbb{R}$ any function. We denote all finite networks by $\mathcal{FN}$.
- First guess at an infinite network: X any set, $\omega_X : X \times X \to \mathbb{R}$ any function. Too little structure!

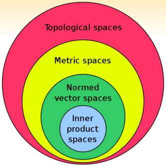
Question: what deserves to be called a network?



- Finite networks: X a finite set, $\omega_X : X \times X \to \mathbb{R}$ any function. We denote all finite networks by $\mathcal{FN}$.
- First guess at an infinite network: X any set, $\omega_X : X \times X \to \mathbb{R}$ any function. Too little structure!
- What works: X a **first countable topological space**, $\omega_X : X \times X \to \mathbb{R}$ **continuous** with respect to the product topology. We refer to the collection of all such networks as $\mathcal{N}$.
- An important subclass: (X, ω_X) as above, furthermore X is **compact**. We denote compact networks by $\mathcal{CN}$.

First countable + compact $\implies$ sequentially compact (sequences have convergent subsequences).
Continuity of ω_X ensures X has "enough" open sets; this imposes one-way control on topology.

Question: what deserves to be called a network?



- Finite networks: X a finite set, $\omega_X : X \times X \to \mathbb{R}$ any function. We denote all finite networks by $\mathcal{FN}$.
- First guess at an infinite network: X any set, $\omega_X : X \times X \to \mathbb{R}$ any function. Too little structure!
- What works: X a **first countable topological space**, $\omega_X : X \times X \to \mathbb{R}$ **continuous** with respect to the product topology. We refer to the collection of all such networks as $\mathcal{N}$.
- An important subclass: (X, ω_X) as above, furthermore X is **compact**. We denote compact networks by $\mathcal{CN}$.

First countable + compact $\implies$ sequentially compact (sequences have convergent subsequences).
Continuity of ω_X ensures X has "enough" open sets; this imposes one-way control on topology. **Possible problem**: can there be too many open sets?

Open sets, triangle inequality, and Axiom A2

Open sets are a topological measure of "nearness": two points x, x' are "nearby" if they are in many of the same open sets (we write $x' \to x$).

Open sets, triangle inequality, and Axiom A2

Open sets are a topological measure of "nearness": two points x, x' are "nearby" if they are in many of the same open sets (we write $x' \to x$).

Metric spaces:

- Metric controls topology; open sets are generated by open balls $B_\varepsilon(x) := \{x' \in X : d_X(x, x') < \varepsilon\}$.

Open sets, triangle inequality, and Axiom A2

Open sets are a topological measure of "nearness": two points x, x' are "nearby" if they are in many of the same open sets (we write $x' \to x$).

Metric spaces:

- Metric controls topology; open sets are generated by open balls $B_\varepsilon(x) := \{x' \in X : d_X(x, x') < \varepsilon\}$.

Networks:

- No triangle inequality/symmetry/non-degeneracy, so no open balls. A priori, ω_X places only a lower limit on the number of open sets.

Open sets, triangle inequality, and Axiom A2

Open sets are a topological measure of "nearness": two points x, x' are "nearby" if they are in many of the same open sets (we write $x' \to x$).

Metric spaces:

- Metric controls topology; open sets are generated by open balls $B_\varepsilon(x) := \{x' \in X : d_X(x, x') < \varepsilon\}$.

Networks:

- No triangle inequality/symmetry/non-degeneracy, so no open balls. A priori, ω_X places only a lower limit on the number of open sets.
- Could have x, x' topologically distinguished, even if $\omega_X(x, \cdot) = \omega_X(x', \cdot)$ and $\omega_X(\cdot, x) = \omega_X(\cdot, x')$.

Open sets, triangle inequality, and Axiom A2

Open sets are a topological measure of "nearness": two points x, x' are "nearby" if they are in many of the same open sets (we write $x' \to x$).

Metric spaces:

- Metric controls topology; open sets are generated by open balls $B_\varepsilon(x) := \{x' \in X : d_X(x, x') < \varepsilon\}$.

Networks:

- No triangle inequality/symmetry/non-degeneracy, so no open balls. A priori, ω_X places only a lower limit on the number of open sets.
- Could have x, x' topologically distinguished, even if $\omega_X(x, \cdot) = \omega_X(x', \cdot)$ and $\omega_X(\cdot, x) = \omega_X(\cdot, x')$.
- Two notions of "nearness": one coming from topology, and one coming from the weight function.

Open sets and Axiom A2

Two notions of "nearness": one coming from topology, and one coming from the weight function. How to reconcile these?

Axiom A1 is just first countability.

Open sets and Axiom A2

Two notions of "nearness": one coming from topology, and one coming from the weight function. How to reconcile these?

- One approach that doesn't work: pull back the topology on $\mathbb{R}$ to $X \times X$, project to X. (Can only work if pullback topology is very "regular")

Axiom A1 is just first countability.

Open sets and Axiom A2

Two notions of "nearness": one coming from topology, and one coming from the weight function. How to reconcile these?

- One approach that doesn't work: pull back the topology on $\mathbb{R}$ to $X \times X$, project to X. (Can only work if pullback topology is very "regular")

Idea

1. *Kuratowski embedding* $x \mapsto d_X(x, \cdot)$ *is an isometric embedding into* l^∞*; the collection* $\{d_X(x, \cdot) : x \in X\}$ *contains all the metric information.*

Axiom A1 is just first countability.

Open sets and Axiom A2

Two notions of "nearness": one coming from topology, and one coming from the weight function. How to reconcile these?

- One approach that doesn't work: pull back the topology on $\mathbb{R}$ to $X \times X$, project to X. (Can only work if pullback topology is very "regular")

Idea

1. *Kuratowski embedding* $x \mapsto d_X(x, \cdot)$ *is an isometric embedding into* l^∞*; the collection* $\{d_X(x, \cdot) : x \in X\}$ *contains all the metric information.*
2. *Axiom A2: say* $x' \to x$ *in a network if and only if*

$$\max(\|\omega_X(x, \cdot) - \omega_X(x', \cdot)\|_\infty, \|\omega_X(\cdot, x) - \omega_X(\cdot, x')\|_\infty) \to 0.$$

Axiom A2 guarantees that topological nearness agrees with ω_X-nearness.

Axiom A1 is just first countability.

Triangle inequality and Axiom A2

Large classes of networks satisfy A2:

- Finite networks (modulo a “contraction”)
- Metric spaces (by triangle inequality)
- Directed metric spaces with finite reversibility ($\sup_{x \neq x'} \frac{d_X(x,x')}{d_X(x',x)} < \infty$)

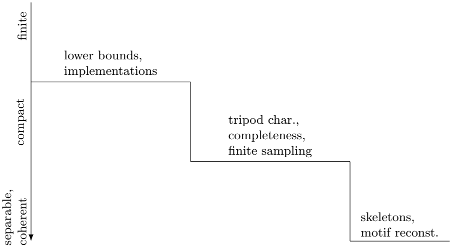
Intuitively, A2 says that $x' \to x$ if they have the same “perception” (of other nodes and by other nodes).

When a network satisfies A2, we say that it has a **coherent** topology.

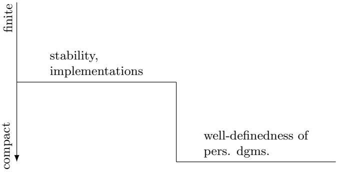
Summary: A network is a first countable topological space X with a continuous weight function $\omega_X : X \times X \to \mathbb{R}$. Useful additional assumptions include compactness and coherence.

Dependence on assumptions

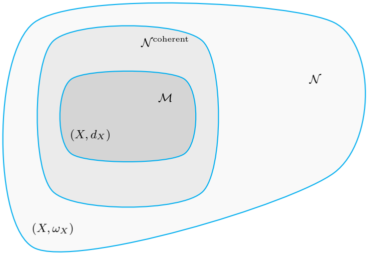
Metric structures on networks

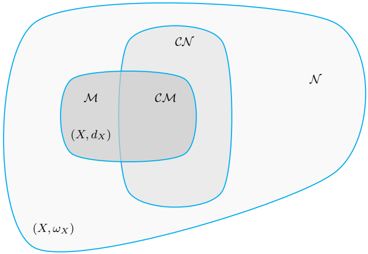


Persistent homology on networks

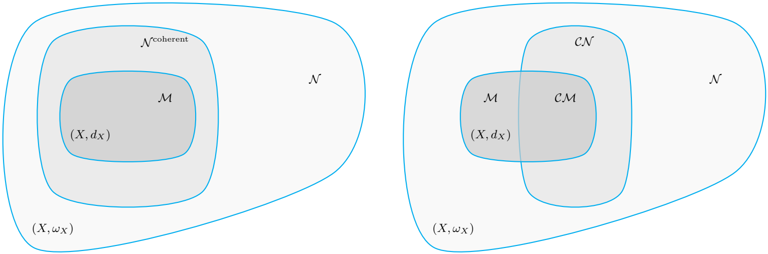


$\mathcal{M}$ vs $\mathcal{N}$



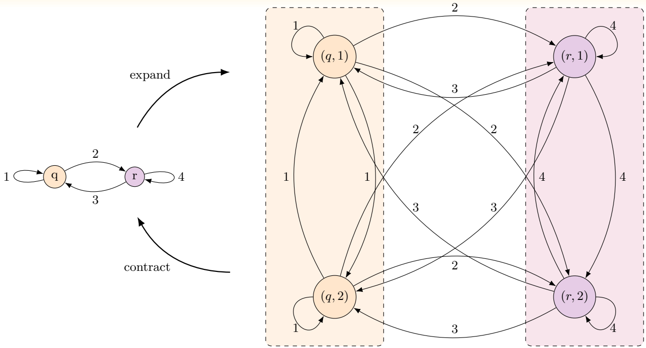


$\mathcal{M}$ vs $\mathcal{N}$



We now know about the objects of $\mathcal{N}$. What about the relationships between the objects? Which networks are "equal"?

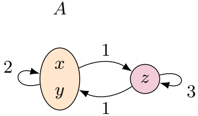
Networks and isomorphism



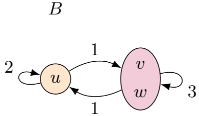
$\omega = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ $\qquad$ $\omega = \begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \\ 3 & 3 & 4 & 4 \\ 3 & 3 & 4 & 4 \end{pmatrix}$

These networks should be the same; they contain the same information.

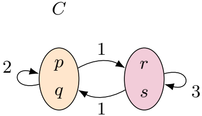
Networks and isomorphism: Tripod characterization



$$\omega_A(x,y,z) = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$



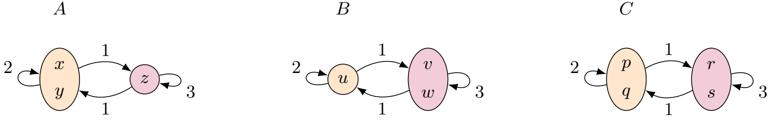
$$\omega_B(u,v,w) = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 3 \end{pmatrix}$$



$$\omega_C(p,q,r,s) = \begin{pmatrix} 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 3 & 3 \\ 1 & 1 & 3 & 3 \end{pmatrix}$$

A and B are not immediately related by a surjection, but they are both related to C through a surjection.

Networks and isomorphism: Tripod characterization



$\omega_A(x,y,z) = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ $\omega_B(u,v,w) = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 3 & 3 \\ 1 & 3 & 3 \end{pmatrix}$ $\omega_C(p,q,r,s) = \begin{pmatrix} 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 3 & 3 \\ 1 & 1 & 3 & 3 \end{pmatrix}$

A and B are not immediately related by a surjection, but they are both related to C through a surjection. This leads to the following:

Strong isomorphism: ϕ_X, ϕ_Y injective and surjective

$$\begin{tikzcd} & Z \arrow[dl, "\phi_X"'] \arrow[dr, "\phi_Y"] & \\ X \arrow[rr, phantom, "\cong^s"] & & Y \end{tikzcd} \qquad \begin{tikzcd} & Z \arrow[dl, "\phi_X"'] \arrow[dr, "\phi_Y"] & \\ X \arrow[rr, phantom, "\cong^w"] & & Y \end{tikzcd}$$

Weak isomorphism: ϕ_X, ϕ_Y only surjective

Poset of weak isomorphism and skeletons

For any compact network (X, ω_X), define its **poset of weak isomorphism** as

$$\mathfrak{p}(X) := \{(Y, \omega_Y) \in \mathcal{CN} : \exists \text{ surjective, weight preserving map } \varphi : X \to Y\}.$$

Poset of weak isomorphism and skeletons

For any compact network (X, ω_X), define its **poset of weak isomorphism** as

$$\mathfrak{p}(X) := \{(Y, \omega_Y) \in \mathcal{CN} : \exists \text{ surjective, weight preserving map } \varphi : X \to Y\}.$$

$$\begin{tikzcd} X \arrow[d, bend right, "f"'] \arrow[d, bend left, "g"] \\ Z \arrow[loop below, dashed, "\varphi"] \end{tikzcd} \qquad \begin{tikzcd} X \arrow[r, two heads] \arrow[drrr] & V \arrow[r, two heads] \arrow[drr] & Y \arrow[r, two heads] \arrow[dr] & \cdots \arrow[d] \\ & & & Z \end{tikzcd}$$

Definition

A compact network Z is **terminal** in $\mathfrak{p}(X)$ if any element of $\mathfrak{p}(X)$ surjects onto Z, and any two surjections are related by a permutation.

Constructing a terminal network

Consider the "contraction" relation on a compact network:

$$x \sim x' \text{ iff } \omega_X(x,z) = \omega_X(x',z) \text{ and } \omega_X(z,x) = \omega_X(z,x') \text{ for all } z \in X.$$

- A priori, $X/\sim$ with the quotient topology is compact, but not necessarily first countable
- If X has a coherent topology, then $X/\sim$ is first countable.
- **(Skeleton construction)** $\mathrm{sk}(X) := X/\sim$ equipped with $(\omega_X)_\#$.
- $\mathrm{sk}(X)$ is **terminal**.
- **(Hausdorffification)** $\mathrm{sk}(X)$ is coherent, and moreover it is Hausdorff!

Moral: Every compact network lives on a fiber above its skeleton; we can "expand" or "contract" parts of a network, and $\cong^w$ realizes that these operations do not change the information in the network.

Question: how do we extract information? Or pieces thereof?

Motif extraction and reconstruction

A common technique is to extract "circuits" or "motifs": these are the $3, 4, \ldots, n$-node subnetworks:

$$\mathrm{M}_n(X, \omega_X) := \left\{ \left(\left(\omega_X(x_i, x_j) \right) \right)_{i,j=1}^n : x_1, \ldots, x_n \in X \right\}.$$

- Repeats are allowed.
- Therefore, if $X \cong^w X'$, we have $\mathrm{M}_n(X) = \mathrm{M}_n(X')$ for all $n \in \mathbb{N}$.
- In the metric space setting, motif sets admit the following "reconstruction theorem":
- **(Gromov)** Two compact metric spaces X, Y are isometric if and only if $\mathrm{M}_n(X) = \mathrm{M}_n(Y)$ for all $n \in \mathbb{N}$.
    - Proof: take countable dense $S_X \subseteq X$. $K_n(X) = K_n(Y)$ means every n-point subset embeds isometrically; extend to an embedding $S_X \hookrightarrow Y$ and then back $\hookrightarrow X$. By completeness, extend to isometric endomorphism $f : X \to X$.
    - NTS f surjective. (**Idea: f preserves volumes.**) Suppose $f(X) \neq X$, let $x \in X \setminus f(X)$ and $\varepsilon = d_X(x, f(X))$. Let S be a maximally ε-separated in X. But then $f(S) \cup \{x\}$ is ε-separated, contradiction.

Motif extraction and reconstruction

Theorem (C., Mémoli '17)

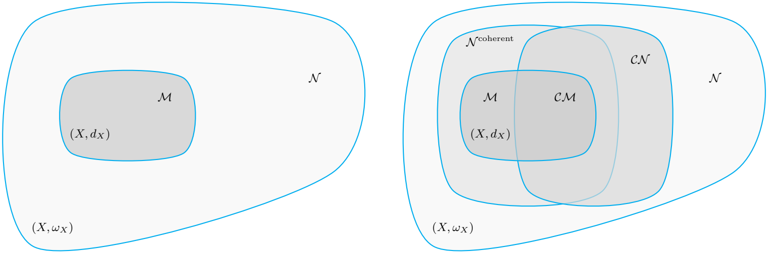
Let $X, Y \in \mathcal{CN}$ be separable and coherent. Then,
$\mathrm{M}_n(X) = \mathrm{M}_n(Y) \forall n \iff \mathrm{sk}(X) \cong^s \mathrm{sk}(Y)$.

Application: Reconstruction holds for compact **directed** metric spaces with finite reversibility.

Proof sketch.

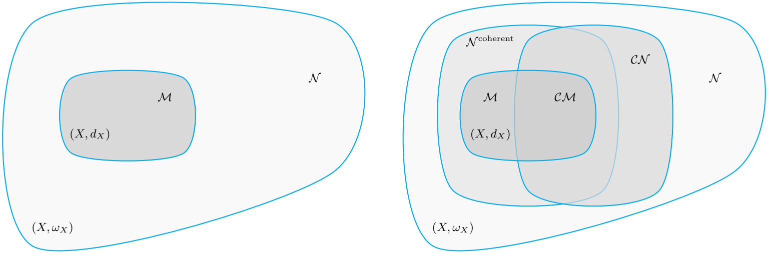
- Pass to skeletons to get Hausdorff spaces; coherence, compactness, first countability preserved.
- Take countable dense $S_X \hookrightarrow X$; compact+first countable $\implies f : X \to X$ weight-preserving.
- Hausdorff $\implies f(X)$ closed. Suppose $f(X) \neq X$, let $x \in X \setminus f(X)$.
- $X \supset f(X) \supset f^2(X) \supset \ldots$, $Z := \cap_n f^n(X)$. Look at $(f^n(x))_n$; sequential compactness $\implies$ there is a convergent subsequence tending to some $z \in Z$.
- Pass to Kuratowski embedding; have $\|\omega_X(Z, \cdot) - \omega_X(x, \cdot)\|_\infty = 0$. Get sequence $(z_n)_n \to x$. Z closed, so $x \in Z$. Contradiction. $\square$

$\mathcal{M}$ vs $\mathcal{N}$



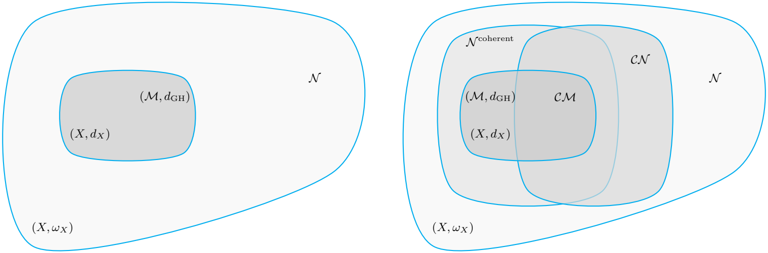
We now have assertions for the intrinsic structure of networks.

$\mathcal{M}$ vs $\mathcal{N}$



We now have assertions for the intrinsic structure of networks.
Question: Can we get **global** structural results?

$\mathcal{M}$ vs $\mathcal{N}$



We now have assertions for the intrinsic structure of networks.
Question: Can we get **global** structural results?

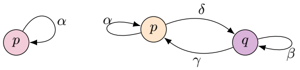
Construction of $d_{\mathcal{N}}$ I

Idea

Carlsson, Mémoli, Ribeiro, Segarra (2014) observed that the Gromov-Hausdorff distance between metric spaces has an equivalent reformulation that can be applied to (finite) networks; they named this network distance $d_{\mathcal{N}}$.

"Hierarchical Quasi-Clustering Methods for Asymmetric Networks", ICML 2014

Intuitive description of $d_{\mathcal{N}}$: find the optimal matching (not necessarily bijective) between nodes that minimizes the maximal amount by which edges are "crushed".



(X, ω_X) (Y, ω_Y)

Construction of $d_{\mathcal{N}}$ II

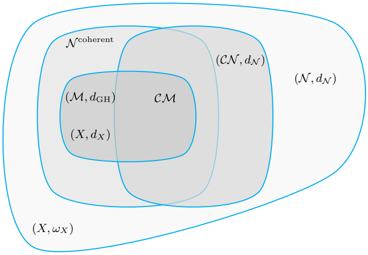
- The (not necessarily bijective) matchings are obtained via **correspondences**: sets $R \subseteq X \times Y$ such that $\pi_X(R) = X$ and $\pi_Y(R) = Y$.
- Distortion of correspondence:
  $\mathrm{dis}(R) = \sup\{|\omega_X(x, x') - \omega_Y(y, y')| : (x, y), (x', y') \in R\}$.
- $d_{\mathcal{N}}(X, Y) := \frac{1}{2} \inf\{\mathrm{dis}(R) : R$ **a correspondence**$\}$**.**
- Already in the finite setting, $d_{\mathcal{N}}$ can be used to produce "stable persistent homology methods."
- We further studied $d_{\mathcal{N}}$ in the setting of infinite networks.

$\mathcal{CN}$ and $d_{\mathcal{N}}$

Theorem (C., Mémoli 2017)

*$(\mathcal{CN}, d_{\mathcal{N}})$ is a **complete** pseudometric space; its zero sets are precisely those given by $\cong^w$. Moreover, it is precisely the metric completion of $\mathcal{FN}$.*

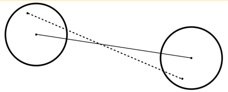
First countability is critically used in relating $\cong^w$ to $d_{\mathcal{N}}$: we use arguments involving sequential compactness.



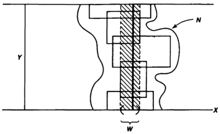
Among many other uses, $d_{\mathcal{N}}$ suggests an approach for formulating a "finite approximation to a network".

Finite approximation

A natural way to approximate a compact metric space X is to choose an ε-net S. Then by triangle inequality, d_X can be viewed as $d_X|_{S\times S}$ with some "blur".



Correspondence/"Voronoi" viewpoint: Partition X into ε-balls around S. Distance between blocks $\approx$ distance between corresponding points in S. $\implies d_{\mathrm{GH}}(X,S) < \varepsilon$.



Emulating this for **compact networks**, we do the following:

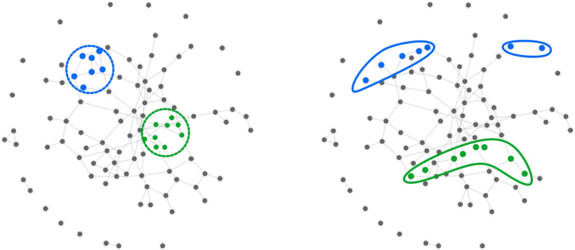
- For a compact network (X,ω_X), take a cover of $\mathbb{R}$ by $\varepsilon/4$-balls, pullback via ω_X to get an open cover $\mathcal{U}$ of $X \times X$.
- Use compactness + multiple applications of the Tube Lemma to get a cover $\{G_1, G_2, \ldots, G_d\}$ such that each $G_i \times G_j$ belongs to an element of $\mathcal{U}$. So there exists $r_{ij} \in \mathbb{R}$ s.t. $\omega_X(G_i, G_j) \subseteq B_{\varepsilon/4}(r_{ij})$. (**This idea used later for SBM-networks**)
- Pick a representative s_i from each G_i, call the collection S.
- Using correspondences, we have $d_{\mathcal{N}}(S,X) < \varepsilon$.

We summarize this as:

Finite approximation

Theorem (C., Mémoli '17)

Any compact network can be ε-approximated in the $d_{\mathcal{N}}$-sense by a finite network.

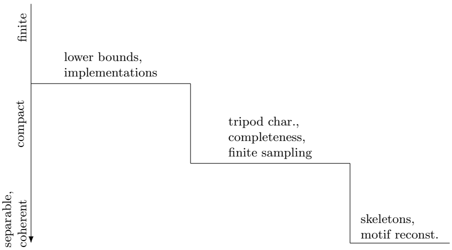


Applications: probabilistic convergence results, stochastic block model networks, well-definedness of persistence diagrams in the compact setting.

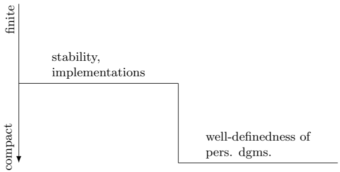
"Lose the quantitative guarantees, keep the qualitative."

Dependence on assumptions

Metric structures on networks



Persistent homology on networks



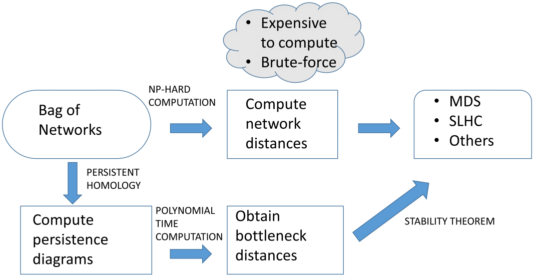
Part II: Persistent homology on (directed) networks

Motivation

- d_{GH} (and therefore $d_{\mathcal{N}}$) is NP-hard to compute (Schmiedl 2015), so need alternative means of pairwise network comparison.
- At the individual level, useful to have different lenses through which to view data.

Motivation

- d_{GH} (and therefore $d_{\mathcal{N}}$) is NP-hard to compute (Schmiedl 2015), so need alternative means of pairwise network comparison.
- At the individual level, useful to have different lenses through which to view data.



Constructing simplicial complexes on a network

Let (X, ω_X) be a network.
We already saw the Vietoris-Rips filtration:

$$\mathrm{VR}_{\delta,X} := \{\sigma \text{ finite } \subseteq X : \max_{x,x' \in \sigma} d_X(x,x') \leq \delta\}; \qquad \{\mathrm{VR}_{\delta,X} \hookrightarrow \mathrm{VR}_{\delta',X}\}_{d \leq \delta'}.$$

- Once a simplicial complex is constructed, existing PH implementations can take over.
- It is possible (and desirable) to construct many different simplicial filtrations on a dataset. Each filtration gives a different "projection" of the data.
- VR relies on checking a two-point condition; it is "easy" to generate filtrations relying on n-point conditions.
- On the "$n \to \infty$" end, there is a construction called the **Dowker complex**. This appears to be superior to the Vietoris-Rips complex in several aspects.

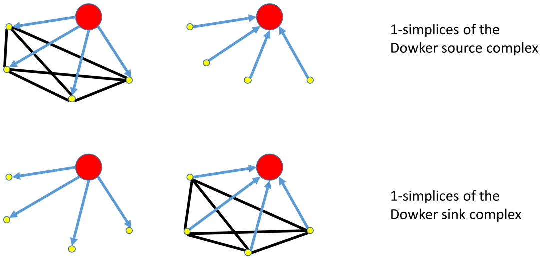
"Towards an observer-oriented theory of shape comparison", P. Frosini, and **"New families of simplicial filtration functors". S. Chowdhury, N. Clause, F. Mémoli, J.A. Sanchez, Z. Wellner.**

The Dowker complex of a network

We define the **Dowker δ-sink complex** associated to X as follows:

$$\mathfrak{D}^{\mathrm{si}}_{\delta,X} := \{\sigma \subseteq X : \text{ there exists } x' \in X \text{ such that } \max_{x\in\sigma} \omega_X(x,x') \leq \delta\}.$$

The dual construction is called the **source complex**.



The (special) Functorial Dowker Theorem I

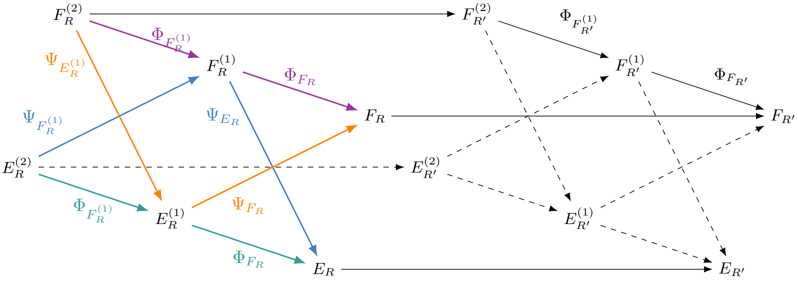
Theorem (C., Mémoli '18)

There exist homotopy equivalences $\Gamma_\delta : |\mathfrak{D}^{\mathrm{si}}_{\delta,X}| \to |\mathfrak{D}^{\mathrm{so}}_{\delta,X}|$ *and* $\Gamma_{\delta'} : |\mathfrak{D}^{\mathrm{si}}_{\delta',X}| \to |\mathfrak{D}^{\mathrm{so}}_{\delta',X}|$ *such that the following diagram commutes up to homotopy:*

$$\begin{tikzcd}
|\mathfrak{D}^{\mathrm{si}}_{\delta,X}| \arrow[r, "|\iota_{\mathfrak{D}^{\mathrm{si}}}|"] \arrow[d, "\Gamma_\delta"', "\simeq"] & |\mathfrak{D}^{\mathrm{si}}_{\delta',X}| \arrow[d, "\simeq"', "\Gamma_{\delta'}"] \\
|\mathfrak{D}^{\mathrm{so}}_{\delta,X}| \arrow[r, "|\iota_{\mathfrak{D}^{\mathrm{so}}}|"] & |\mathfrak{D}^{\mathrm{so}}_{\delta',X}|
\end{tikzcd}$$

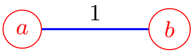
Application: Source and sink Dowker complexes give the same persistence diagram! (When the diagram is defined...)

The (special) Functorial Dowker Theorem II

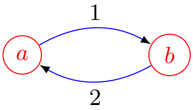


A limitation/blessing of simplicial methods

Given this two-point metric space, the standard linearization process used in PH maps $\{a, b\}$ to the vector $[a, b] = -[b, a]$ "at time $t = 1$".



However, in the asymmetric network setting, this (simplicial) linearization may be unsatisfactory. The equality $[a, b] = -[b, a]$ at the vector space level ignores the unequal weights on the arrows $a \to b$ and $b \to a$.



Path homology solves this problem by assigning $[a, b]$ and $[b, a]$ to **linearly independent** components at the vector space level.

Construction of path homology

Let $\mathbb{K}$ be a field.

Given a set X and any integer $p \in \mathbb{Z}_+$, an **elementary p-path over X** is a sequence $[x_0, \ldots, x_p]$ of $p+1$ elements of X.

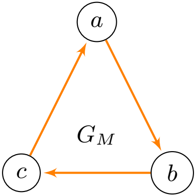
A boundary map is then defined in the standard way:

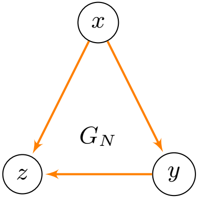
$$\partial_p([x_0, \ldots, x_p]) := \sum_{i=0}^{p} (-1)^i [x_0, \ldots, \widehat{x_i}, \ldots, x_p].$$

$$\text{E.g. } \partial_2(abc) = [bc] - [ac] + [ab].$$
$$\partial_1(ab) = [b] - [a].$$

Let $G = (X, E)$ be a digraph. For each $p \in \mathbb{Z}_+$, an elementary p-path $[x_0, \ldots, x_p]$ on X is **allowed** if $(x_i, x_{i+1}) \in E$ for each $0 \leq i \leq p - 1$.

For each $p \in \mathbb{Z}_+$, the free vector space on the collection of allowed p-paths on (X, E) is denoted $\mathcal{A}_p = \mathcal{A}_p(G) = \mathcal{A}_p(X, E, \mathbb{K})$, and is called the **space of allowed p-paths**.





$$\begin{aligned}
\mathcal{A}_0(G_M) &= \mathbb{K}[\{a, b, c\}] \\
\mathcal{A}_1(G_M) &= \mathbb{K}[\{ab, bc, ca\}] \\
\mathcal{A}_2(G_M) &= \mathbb{K}[\{abc, bca, cab\}] \\
\mathcal{A}_3(G_M) &= \mathbb{K}[\{abca, bcab, cabc\}]
\end{aligned}$$

$$\begin{aligned}
\mathcal{A}_0(G_N) &= \mathbb{K}[\{x, y, z\}] \\
\mathcal{A}_1(G_N) &= \mathbb{K}[\{xy, yz, xz\}] \\
\mathcal{A}_2(G_N) &= \mathbb{K}[\{xyz\}] \\
\mathcal{A}_3(G_N) &= \{0\}
\end{aligned}$$

The **space of ∂-invariant p-paths on G** is the following subspace of $\mathcal{A}_p(G)$:

$$\Omega_p = \Omega_p(G) = \Omega_p(X, E, \mathbb{K}) := \{c \in \mathcal{A}_p : \partial_p(c) \in \mathcal{A}_{p-1}\}.$$

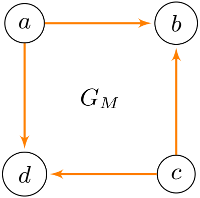
Thus we have a chain complex:

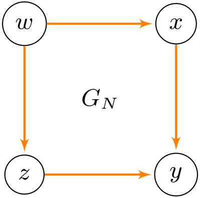
$$\ldots \xrightarrow{\partial_3} \Omega_2 \xrightarrow{\partial_2} \Omega_1 \xrightarrow{\partial_1} \Omega_0 \xrightarrow{\partial_0} \mathbb{K} \xrightarrow{\partial_{-1}} 0$$

For each $p \in \mathbb{Z}_+$, the **p-dimensional path homology groups of $G = (X, E)$** are defined as:

$$H_p^{\Xi}(G) = H_p^{\Xi}(X, E, \mathbb{K}) := \ker(\partial_p)/\operatorname{im}(\partial_{p+1}).$$

An example (with biological motivations) I





$\Omega_0(G_M) = \mathbb{K}[\{a, b, c, d\}]$
$\Omega_1(G_M) = \mathbb{K}[\{ab, cb, cd, ad\}]$
$\Omega_2(G_M) = \{0\}$

$\Omega_0(G_N) = \mathbb{K}[\{w, x, y, z\}]$
$\Omega_1(G_N) = \mathbb{K}[\{wx, xy, zy, wz\}]$
$\Omega_2(G_N) = \mathbb{K}[\{wxy - wzy\}]$

Note $\partial_2(wxy) = xy - wy + wx$,
$\partial_2(wzy) = zy - wy + wz$

An example (with biological motivations) II

The crux of the $\Omega_\bullet$ construction lies in understanding $\Omega_2(G_N)$. Note that even though $\partial_2^{G_N}(wxy), \partial_2^{G_N}(wzy) \notin \mathcal{A}_2(G_N)$ (because $wy \notin \mathcal{A}_1(G_N)$), we still have:

$$\partial_2^{G_N}(wxy - wzy) = xy - wy + wx - zy + wy - wz \in \mathcal{A}_1(G_N).$$

One can then verify that

$$\ker(\partial_1^{G_M}) = \mathbb{K}[\{ab - cb + cd - ad\}] \neq \{0\} = \mathrm{im}(\partial_2^{G_M}),$$

$$\ker(\partial_1^{G_N}) = \mathbb{K}[\{wx + xy - zy - wz\}] = \mathrm{im}(\partial_2^{G_N}).$$

Thus $\dim(H_1^{\Xi}(G_M)) = \mathbf{1}$, and $\dim(H_1^{\Xi}(G_N)) = \mathbf{0}$.

- In systems biology, G_M, G_N are referred to as the **bi-fan** and **bi-parallel motifs**, respectively. Distinguishing between these two motifs is an important task in that domain.
- A construction closely related to path homology—ordered homology on a directed flag complex—does not distinguish between these two motifs.

An example (with biological motivations) III

- **The challenge of finding a natural basis for $\Omega_\bullet$.** G_N is a minimal example showing that it is nontrivial to compute bases for the vector spaces $\Omega_\bullet$. Specifically, while it is trivial to read off bases for the allowed paths $\mathcal{A}_\bullet$ from a digraph, one needs to consider **linear combinations** of allowed paths in a systematic manner to obtain bases for the ∂-invariant paths. **This raises a red flag for computations!**

Persistent path homology

Now suppose we have a **digraph filtration** $\{G_X^{\delta} \hookrightarrow G_X^{\delta'}\}_{\delta \leq \delta' \in \mathbb{R}_+}$.

—perhaps we are adding edges as time progresses, or as we raise a scale parameter, or by some other rule

Here X is the set of nodes/vertices over which we add edges.

Applying path homology then yields a **persistent vector space**

$$\mathbf{PVec}_p^{\Xi}(X, \omega_X) := \{H_p^{\Xi}(G_X^{\delta}) \xrightarrow{(\iota_{\delta,\delta'})\#} H_p^{\Xi}(G_X^{\delta'})\}_{\delta \leq \delta' \in \mathbb{R}_+}.$$

We denote the associated **persistence diagram** by $\mathrm{Dgm}_p^{\Xi}(X)$.

(At least when X is finite, the diagram is well-defined by "standard results" in the literature)

Network PPH: threshold the weights of (X, ω_X) to get a digraph filtration, apply the above.

Two challenges

- Computation? Harder than in the simplicial case:
  - n^p vs $\binom{n}{p}$ to compute $(p-1)$-dimensional diagram
  - Additional linear algebra overhead to obtain basis Ω from data $\mathcal{A}$

Two challenges

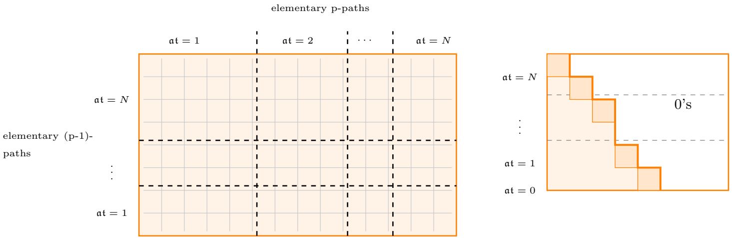
- Computation? Harder than in the simplicial case:
  - n^p vs $\binom{n}{p}$ to compute $(p-1)$-dimensional diagram
  - Additional linear algebra overhead to obtain basis Ω from data $\mathcal{A}$
  - C., Mémoli (2018)—showed that the above step can be incorporated into a standard persistent homology algorithm
  - Implementations (Matlab, Python, C++) on `https://github.com/samirchowdhury`

Two challenges

- Computation? Harder than in the simplicial case:
  - n^p vs $\binom{n}{p}$ to compute $(p-1)$-dimensional diagram
  - Additional linear algebra overhead to obtain basis Ω from data $\mathcal{A}$
  - C., Mémoli (2018)—showed that the above step can be incorporated into a standard persistent homology algorithm
  - Implementations (Matlab, Python, C++) on `https://github.com/samirchowdhury`
- Stability? If two digraph filtrations are similar (in what metric?), their persistence diagrams should be similar (in **bottleneck distance**)...
  - Needed to wait until a homotopy theory of digraphs was released by the original path homology authors.

Theorem (C., Mémoli '18)

Left-to-right column reduction is sufficient to obtain compatible bases for $\Omega_\bullet$. This is precisely the operation used in the classical PH algorithm, so the two steps (obtaining bases and computing PH) can be combined into one.



The running time for the algorithm is the same as that of Gaussian elimination, i.e. is cubic in the number of paths.

Well-definedness and stability

Thus far we have a variety of simplicial constructions on networks, as well as a construction coming from path homology.

Question: are persistence diagrams on **infinite** networks well-defined? (yes for finite networks, a priori no for infinite networks)

Well-definedness and stability

Thus far we have a variety of simplicial constructions on networks, as well as a construction coming from path homology.

Question: are persistence diagrams on **infinite** networks well-defined? (yes for finite networks, a priori no for infinite networks)

Theorem (C., Mémoli 2017-2018)

Any of the Vietoris-Rips, Dowker, and PPH persistence diagrams are well-defined on compact networks. Moreover, for any $(X, \omega_X), (Y, \omega_Y) \in \mathcal{CN}$ *and any* $p \in \mathbb{Z}_+$,

$$d_{\mathrm{B}}(\mathrm{Dgm}_p^{\bullet}(X), \mathrm{Dgm}_p^{\bullet}(Y)) \leq 2 d_{\mathcal{N}}(X, Y),$$

where $\bullet$ *stands for any of the preceding persistent homology methods.*

Actually, our result shows that **any** stable (up to interleaving) persistent homology method on networks will have a well-defined persistence diagram in the compact case.

Well-definedness

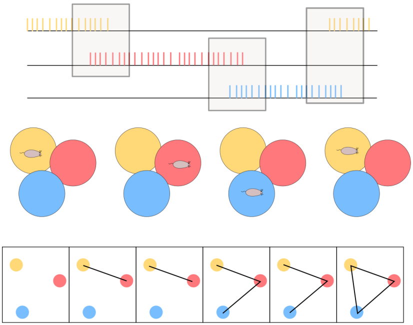
Proof.

Let (X, ω_X) be a compact network.

- Obtain the persistent vector space $\mathbf{PVec}_p^\bullet(X) = \{V^\delta \xrightarrow{\nu_{\delta,\delta'}} V^{\delta'}\}_{\delta \leq \delta'}$.
- For the persistence diagram to be well-defined, **need** $\nu_{\delta,\delta'}$ to have finite rank.
- Let $\varepsilon = (\delta' - \delta)/2$. By the **finite sampling** theorem, take a finite subnetwork X' such that $d_{\mathcal{N}}(X, X') < \varepsilon/2$.
- Write $\mathbf{PVec}_p^\bullet(X') = \{U^\delta \xrightarrow{\mu_{\delta,\delta'}} U^{\delta'}\}_{\delta \leq \delta'}$; each vector space is finite dimensional.
- By "interleaving stability" (this is proved separately for each of the methods we consider), The map $V^\delta \to V^{\delta'}$ factorizes as $V^\delta \to U^{\delta+\varepsilon} \to V^{\delta+2\varepsilon} = V^{\delta'}$. Thus $\nu_{\delta,\delta'}$ has finite rank! $\square$

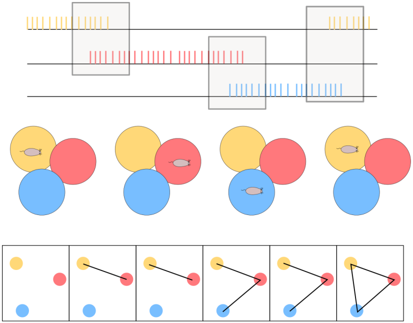
Part III: An application to hippocampal map formation

Place cells, spike trains, and simplicial complex formation



“Cell groups reveal structure of stimulus space”, Curto, Itskov and “A topological paradigm for hippocampal spatial map formation using persistent homology”, Dabaghian, Mémoli, Frank, Carlsson

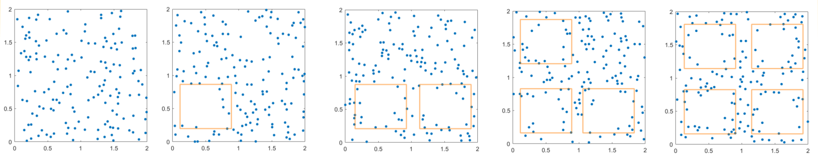
Place cells, spike trains, and simplicial complex formation



This is a symmetric "fire together, wire together" model—can it help to incorporate asymmetry?

"Cell groups reveal structure of stimulus space", Curto, Itskov and **"A topological paradigm for hippocampal spatial map formation using persistent homology", Dabaghian, Mémoli, Frank, Carlsson**

Experiment setup



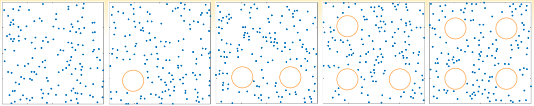
- 5 arenas with 0-4 obstacles, 6 random walk trajectories for each arena.
- Uniform random choice of 150 place field centers (with rejection sampling).
- Firing pattern for each place cell sampled from an inhomogeneous Poisson process; more firing when trajectory is close to place field.
- 20 spike trains for each trajectory; total of 120 spike trains generated for each arena (also have experiment with 600 spike trains for each arena)
- Network built as:

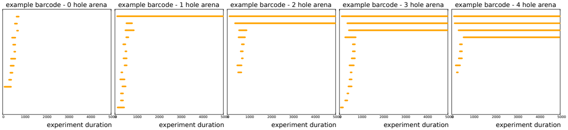
$$\omega_X(x_i, x_j) = 1 - \frac{N_{i,j}(5)}{\sum_{i=1}^{n_k} N_{i,j}(5)},$$

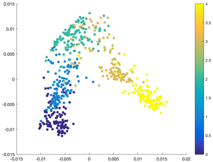
where $N_{i,j}(5) \approx$ number of times cell j spiked shortly after cell i spiked.

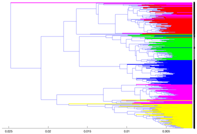
"Topology of stimulus space via directed network persistent homology", Chowdhury, Dai, Mémoli, Cosyne 2017

Results



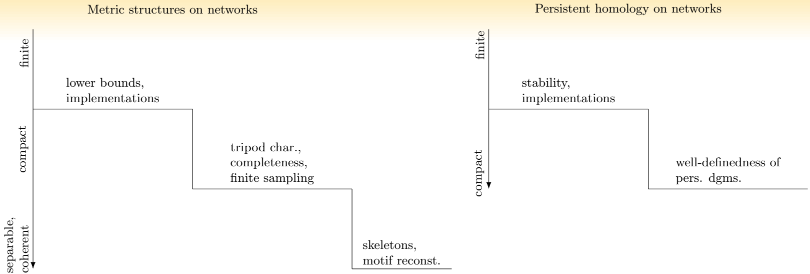


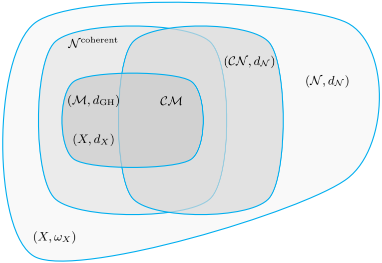




On the same dataset, 1-nn error rate improved from 0.55 (symmetric) to 0.19 (asymmetric).

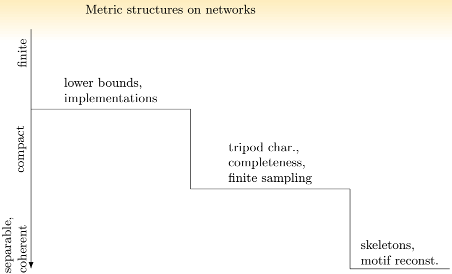
Summary

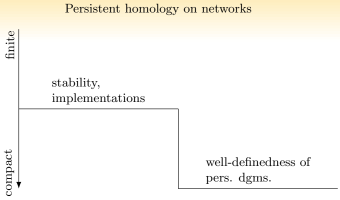


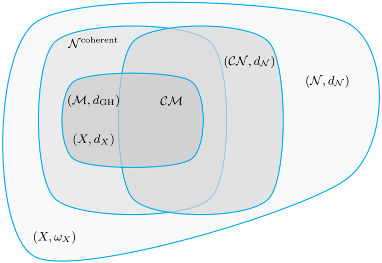


- The setting of networks is highly relevant to data analysis.
- Work in this thesis shows that this setting has sufficiently rich mathematical structure to allow a wide range of useful result.

Summary







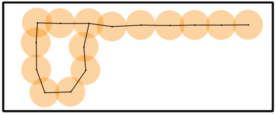
- The setting of networks is highly relevant to data analysis.
- Work in this thesis shows that this setting has sufficiently rich mathematical structure to allow a wide range of useful result.

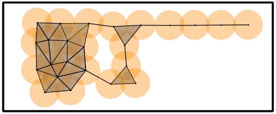
Thank you!

Extensions

- Further applications (related to neuroscience)
- Geodesics in $d_{\mathcal{N}}$: construction, exotic examples
- Measure networks and the network Gromov-Wasserstein distance
- Functorial Dowker Theorem: proof structure and relations to simplicial Functorial Nerve Theorems
- Probabilistic convergence of PH methods

Limiting memory improves generalization

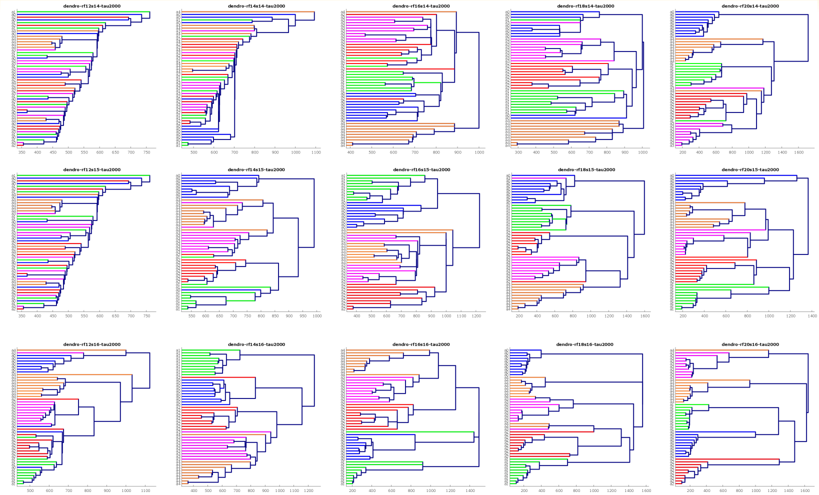




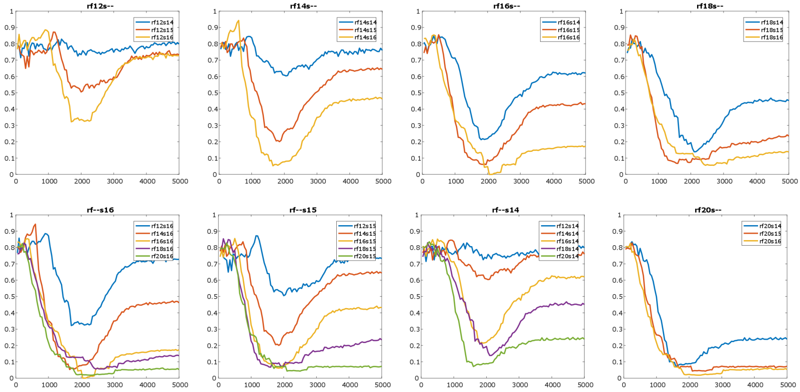
Topological learning and forgetting:

- Initialize the abstract simplicial complex with one 0-simplex for each place cell
- When a pair or a triple of place cells cofires, add the corresponding 1 or 2-simplex to the abstract simplicial complex
- Remove an edge or a face if the corresponding place cells have not cofired in the past T steps
- T is our "memory parameter"

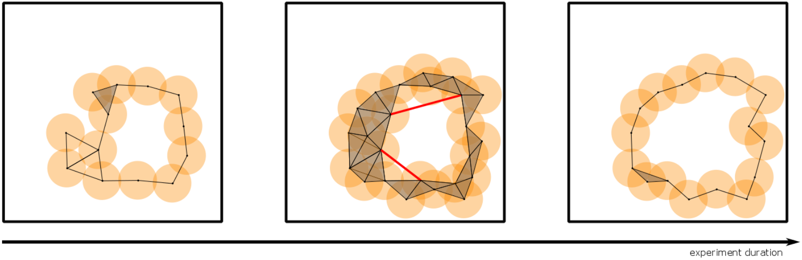
Effect of parameters



1-nearest neighbor error rate vs T



An explanation for the error rate dip



Geodesics in $\mathcal{CN}$ I

Theorem (C., Mémoli '18)

The complete metric space $(\mathcal{CN}, d_{\mathcal{N}})$ is geodesic.

Proof.

Let $X, Y \in \mathcal{CN}$.

- Using finite sampling, choose X_n, Y_n finite networks converging to X, Y.
- Get geodesic between each X_n, Y_n, let R_n be an optimal correspondence with the "midpoint weight function".
- Show that the collection $\{R_n\}$ is precompact; it limits then to a true midpoint of X and Y.
- Fact: metric spaces with midpoints are geodesic. $\square$

- The above proof follows the ideas used by Ivanov, Nikolaeva, and Tuzhilin (2017) to prove that the space of compact metric spaces equipped with d_{GH} is geodesic.

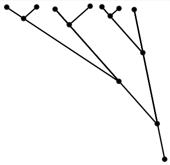
Geodesics in $\mathcal{CN}$ II

- We had concurrently been working on the same result, and provided a proof showing that optimal correspondences **always** exist between compact metric spaces, and hence a geodesic can be constructed explicitly.

- In all these existence constructions, geodesics are given by a convex combination form. Vladimir Zolotov pointed out the existence of branching geodesics in $\mathcal{FM}$ to us. Following this, we became interested in finding exotic geodesics in $\mathcal{FM}$ (and hence $\mathcal{CM}$ and $\mathcal{CN}$).

Theorem (C., Mémoli '18)

$\mathcal{FM}$ contains infinitely many branching and non-convex-combination geodesics between the 1-point space and the n-point space, for any $n \geq 2$.



"The Gromov-Hausdorff metric on the space of compact metric spaces is strictly intrinsic". A. Ivanov, N. Nikolaeva, and A. Tuzhilin.

Measure networks and the Gromov-Wasserstein distance I

- One of the limitations of $d_{\mathcal{N}}$ is that it is very sensitive to outliers: even a single bad sample can throw it off.
- To account for this, we consider measure networks: triples (X, ω_X, μ_X), where X is a Polish space, ω_X is a measurable, integrable function, and μ_X is a Borel probability measure.
- Sturm has shown that the symmetric measure networks (gauged spaces in his language) enjoy very rich geometric structure, including directional derivatives and exponential maps.
- We are interested in these results, and also in using these ideas for computation.
- Specifically, we are interested in approximate computation of the network Gromov-Wasserstein distance, as well as its lower bounds.

The Network Gromov-Wasserstein distance

$(X, \omega_X, \mu_X), (Y, \omega_Y, \mu_Y)$ two measure networks, $p \in [1, \infty)$.

- Consider **measure couplings** $\mu \in \mathrm{Prob}(X \times Y)$ with marginals μ_X and μ_Y.
-
$$\mathrm{dis}_p(\mu) := \left(\int_{X \times Y} \int_{X \times Y} |\omega_X(x, x') - \omega_Y(y, y')|^p \, d\mu(x, y) \, d\mu(x', y') \right)^{1/p}.$$
- Infimize the distortion over all couplings:
$$d_{\mathcal{N},p}(X, Y) := \frac{1}{2} \inf_{\mu \in \Pi(\mu_X, \mu_Y)} \mathrm{dis}_p(\mu)$$
- For some of our computational purposes, we study an **entropically regularized** form:
$$d^{\varepsilon}_{\mathcal{N},p}(X, Y) := \frac{1}{2} \inf_{\mu \in \Pi(\mu_X, \mu_Y)} \mathrm{dis}_p(\mu) + \varepsilon H(\mu),$$
where H is the entropy function. Corresponds to "smoother" optimal couplings.
- In the discrete case, $H(\mu) = -\sum_{ij} \mu_{ij} \log(\mu_{ij})$.

Functorial Dowker Theorem and simplicial Functorial Nerve Theorems I

Let X, Y be two totally ordered sets, and let $R \subseteq X \times Y$ be a nonempty relation.

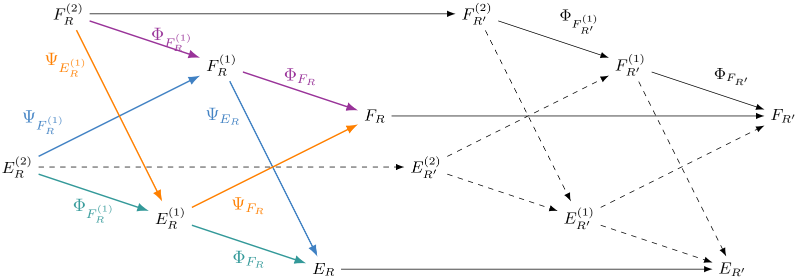
- Dowker (1952) defined two simplicial complexes E_R and F_R as follows.
- A finite subset $\sigma \subseteq X$ belongs to E_R whenever there exists $y \in Y$ such that $(x, y) \in R$ for each $x \in \sigma$.
- Similarly a finite subset $\tau \subseteq Y$ belongs to F_R whenever there exists $x \in X$ such that $(x, y) \in R$ for each $y \in \tau$

Barycentric subdivision: For any simplicial complex Σ, one constructs

$$\Sigma^{(1)} := \{[\sigma_1, \sigma_2, \ldots, \sigma_p] : \sigma_1 \subseteq \sigma_2 \subseteq \ldots \subseteq \sigma_p, \text{ each } \sigma_i \in \Sigma\}.$$

Vertices of $\Sigma^{(1)}$ are the simplices of Σ, and the simplices of $\Sigma^{(1)}$ are nested sequences of simplices of Σ.

Functorial Dowker Theorem and simplicial Functorial Nerve Theorems II



$\Phi_{E_R} : E_R^{(1)} \to E_R$ defined as follows:

- First define Φ_{E_R} on vertices of $E_R^{(1)}$ by $\Phi_{E_R}(\sigma) = s_\sigma$, where s_σ is the least vertex of σ with respect to the total order.
- Next, for any simplex $[\sigma_1, \ldots, \sigma_k]$ of $E_R^{(1)}$, where $\sigma_1 \subseteq \ldots \subseteq \sigma_k$, we have $\Phi_{E_R}(\sigma_i) = s_{\sigma_i} \in \sigma_k$ for all $1 \leq i \leq k$.

Functorial Dowker Theorem and simplicial Functorial Nerve Theorems III

- Thus $[\Phi_{E_R}(\sigma_1), \ldots, \Phi_{E_R}(\sigma_k)] = [s_{\sigma_1}, s_{\sigma_2}, \ldots, s_{\sigma_k}]$ is a face of σ_k, hence a simplex of Σ.
- This defines Φ_{E_R} as a simplicial map $E_R^{(1)} \to E_R$.
- Also define $\Psi_{F_R} : E_R^{(1)} \to F_R$ as follows. Given a vertex $\sigma = [x_0, \ldots, x_k] \in E_R^{(1)}$, $\Psi_{F_R}(\sigma) = y_\sigma$, for some $y_\sigma \in Y$ such that $(x_i, y_\sigma) \in R$ for each i.
- Ψ_{F_R} is also simplicial.

Functorial Dowker Theorem and simplicial Functorial Nerve Theorems IV

Theorem (The Functorial Dowker Theorem (FDT))

Let X, Y be two totally ordered sets, let $R \subseteq R' \subseteq X \times Y$ be two nonempty relations, and let $E_R, F_R, E_{R'}, F_{R'}$ be their associated simplicial complexes. Then there exist homotopy equivalences $\Gamma_{|E_R|} : |F_R| \to |E_R|$ and $\Gamma_{|E_{R'}|} : |F_{R'}| \to |E_{R'}|$ such that the following diagram commutes up to homotopy:

$$\begin{tikzcd}
|F_R| \arrow[r, "|\iota_E|"] \arrow[d, "\Gamma_{|E_R|}"', "\simeq"] & |F_{R'}| \arrow[d, "\simeq"', "\Gamma_{|E_{R'}|}"] \\
|E_R| \arrow[r, "|\iota_F|"] & |E_{R'}|
\end{tikzcd}$$

Functorial Dowker Theorem and simplicial Functorial Nerve Theorems V

Definition

Let $\mathcal{A} = \{A_i\}_{i \in I}$ be a family of nonempty sets indexed by I. The **nerve** of $\mathcal{A}$ is the simplicial complex
$\mathcal{N}(\mathcal{A}) := \{\sigma \in \mathrm{pow}(I) : \sigma \text{ is finite, nonempty, and } \cap_{i \in \sigma} A_i \neq \varnothing\}$.

Definition (Covers of simplices and subcomplexes)

Let Σ be a simplicial complex. Then a collection of subcomplexes $\mathcal{A}_\Sigma = \{\Sigma_i\}_{i \in I}$ is said to be a **cover of subcomplexes** for Σ if $\Sigma = \cup_{i \in I} \Sigma_i$. Furthermore, $\mathcal{A}_\Sigma$ is said to be a **cover of simplices** if each $\Sigma_i \in \mathcal{A}_\Sigma$ has the property that $\Sigma_i = \mathrm{pow}(V(\Sigma_i))$. In this case, each Σ_i has precisely one top-dimensional simplex, consisting of the vertex set $V(\Sigma_i)$.

We present two **simplicial** formulations of the Functorial Nerve Theorem that turn out to be equivalent; the statements differ in that one is about covers of simplices and the other is about covers of subcomplexes.

Functorial Dowker Theorem and simplicial Functorial Nerve Theorems VI

Theorem (Functorial Nerve I)

Let $\Sigma \subseteq \Sigma'$ be two simplicial complexes, and let $\mathcal{A}_\Sigma = \{\Sigma_i\}_{i \in I}$, $\mathcal{A}_{\Sigma'} = \{\Sigma'_i\}_{i \in I'}$ be finite covers of simplices for Σ and Σ' such that $I \subseteq I'$ and $\Sigma_i \subseteq \Sigma'_i$ for each $i \in I$. In particular, $\mathrm{card}(I') < \infty$. Suppose that for each finite subset $\sigma \subseteq I'$, the intersection $\cap_{i \in \sigma} \Sigma'_i$ is either empty or contractible (and likewise for $\cap_{i \in \sigma} \Sigma_i$). Then $|\Sigma| \simeq |\mathcal{N}(\mathcal{A}_\Sigma)|$ and $|\Sigma'| \simeq |\mathcal{N}(\mathcal{A}_{\Sigma'})|$, via maps that commute up to homotopy with the canonical inclusions.

Functorial Dowker Theorem and simplicial Functorial Nerve Theorems VII

Theorem (Functorial Nerve II)

Let $\Sigma \subseteq \Sigma'$ be two simplicial complexes, and let $\mathcal{A}_\Sigma = \{\Sigma_i\}_{i \in I}$, $\mathcal{A}_{\Sigma'} = \{\Sigma'_i\}_{i \in I'}$ be finite covers of subcomplexes for Σ and Σ' such that $I \subseteq I'$ and $\Sigma_i \subseteq \Sigma'_i$ for each $i \in I$. In particular, $\mathrm{card}(I') < \infty$. Suppose that for each finite subset $\sigma \subseteq I'$, the intersection $\cap_{i \in \sigma} \Sigma'_i$ is either empty or contractible (and likewise for $\cap_{i \in \sigma} \Sigma_i$). Then $|\Sigma| \simeq |\mathcal{N}(\mathcal{A}_\Sigma)|$ and $|\Sigma'| \simeq |\mathcal{N}(\mathcal{A}_{\Sigma'})|$, via maps that commute up to homotopy with the canonical inclusions.

Functorial Dowker Theorem and simplicial Functorial Nerve Theorems VIII

Theorem (C., Mémoli '18)

The finite FDT, the FNT I, and the FNT II are all equivalent. Moreover, all of these results are implied by the FDT, as below:

$$\begin{tikzcd} & & \text{FNT I} \arrow[dr, Rightarrow] & \\ \text{FDT} \arrow[r, Rightarrow] & \text{Finite FDT} \arrow[ur, Rightarrow] & & \text{FNT II} \arrow[ll, Rightarrow] \end{tikzcd}$$

Probabilistic convergence I

Definition

Let (X, ω_X) be a network with a Borel probability measure μ_X. Let $\mathcal{U}$ be any ε-system on X. We define the **minimal mass function** $\mathfrak{m}(\mathcal{U}) := \min\{\mu_X(U) : U \in \mathcal{U},\ \mu_X(U) > 0\}$. Note that $\mathfrak{m}$ returns the minimal non-zero mass of an element in $\mathcal{U}$.
Next let $\varepsilon > 0$. Define a function $\mathfrak{M}_\varepsilon : \mathcal{CN} \to (0, 1]$ as follows:

$$\mathfrak{M}_\varepsilon(X) := \sup\{\mathfrak{m}(\mathcal{U}) : \mathcal{U} \text{ an } \varepsilon\text{-system on } X\}.$$

Probabilistic convergence II

Theorem (Probabilistic network approximation, C., Mémoli '17)

Let (X, ω_X) *be a network equipped with a Borel probability measure* μ_X*. For each* $i \in \mathbb{N}$*, let* $x_i : \Omega \to X$ *be an independent random variable defined on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with distribution* μ_X*. For each* $n \in \mathbb{N}$*, let* $\mathbb{X}_n = \{x_1, x_2, \ldots, x_n\}$*. Let* $\varepsilon > 0$*. Then we have:*

$$\mathbb{P}\big(\{\omega \in \Omega : d_{\mathcal{N}}(\mathrm{supp}(\mu_X), \mathbb{X}_n(\omega)) \geq \varepsilon\} \big) \leq \frac{\big(1 - \mathfrak{M}_{\varepsilon/2}(\mathrm{supp}(\mu_X))\big)^n}{\mathfrak{M}_{\varepsilon/2}(\mathrm{supp}(\mu_X))},$$

where $\mathbb{X}_n(\omega)$ *is the subnetwork induced by* $\{x_1(\omega), \ldots, x_n(\omega)\}$*. In particular, the subnetwork* $\mathbb{X}_n$ *converges almost surely to* X *in the* $d_{\mathcal{N}}$*-sense.*

Probabilistic convergence III

Theorem (C., Mémoli '17)

Let (X, ω_X) *be a network equipped with a Borel probability measure* μ_X*. For each* $i \in \mathbb{N}$*, let* $x_i : \Omega \to X$ *be an independent random variable defined on some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *with distribution* μ_X*. For each* $n \in \mathbb{N}$*, let* $\mathbb{X}_n = \{x_1, x_2, \ldots, x_n\}$*. Let* $\varepsilon > 0$*. Then we have:*

$$\mathbb{P}\big(\{\omega \in \Omega : d_{\mathrm{B}}(\mathrm{Dgm}^{\bullet}(\mathrm{supp}(\mu_X)), \mathrm{Dgm}^{\bullet}(\mathbb{X}_n(\omega))) \geq \varepsilon\} \big) \leq \frac{\big(1 - \mathfrak{M}_{\varepsilon/4}(\mathrm{supp}(\mu_X))\big)^n}{\mathfrak{M}_{\varepsilon/4}(\mathrm{supp}(\mu_X))},$$

where $\mathbb{X}_n(\omega)$ *is the subnetwork induced by* $\{x_1(\omega), \ldots, x_n(\omega)\}$ *and* $\mathrm{Dgm}^{\bullet}$ *is either of* $\mathrm{Dgm}^{\mathfrak{R}}$ *and* $\mathrm{Dgm}^{\mathfrak{D}}$*. In particular, either of the Rips and Dowker persistent vector spaces of the subnetwork* $\mathbb{X}_n$ *converges almost surely to that of* $\mathrm{supp}(\mu_X)$ *in bottleneck distance.*