Using Topological Data Analysis for Insights into Biological and Artificial Neural Networks

Fabiana Ferracina

WSU Vancouver - Dept. of Math & Stats

December 10, 2021

Outline

About Me

1. B.S. in Mathematics from University of Rochester; M.S. in Mathematics from University of Washington.

About Me

1. B.S. in Mathematics from University of Rochester; M.S. in Mathematics from University of Washington.
2. WSU 2018-19: Mathematical modeling of forests, biogeochemistry, and distributed systems.

About Me

1. B.S. in Mathematics from University of Rochester; M.S. in Mathematics from University of Washington.
2. WSU 2018-19: Mathematical modeling of forests, biogeochemistry, and distributed systems.
3. WSU 2019-20: Transition to statistics, stochastic processes, state-space models.

About Me

1. B.S. in Mathematics from University of Rochester; M.S. in Mathematics from University of Washington.
2. WSU 2018-19: Mathematical modeling of forests, biogeochemistry, and distributed systems.
3. WSU 2019-20: Transition to statistics, stochastic processes, state-space models.
4. WSU 2020-21: Animal movement ecology, species distribution models, telemetry data.

About Me

1. B.S. in Mathematics from University of Rochester; M.S. in Mathematics from University of Washington.
2. WSU 2018-19: Mathematical modeling of forests, biogeochemistry, and distributed systems.
3. WSU 2019-20: Transition to statistics, stochastic processes, state-space models.
4. WSU 2020-21: Animal movement ecology, species distribution models, telemetry data.
5. WSU 2021-22: Topological data analysis, machine learning and neural networks.

About Me

1. B.S. in Mathematics from University of Rochester; M.S. in Mathematics from University of Washington.
2. WSU 2018-19: Mathematical modeling of forests, biogeochemistry, and distributed systems.
3. WSU 2019-20: Transition to statistics, stochastic processes, state-space models.
4. WSU 2020-21: Animal movement ecology, species distribution models, telemetry data.
5. WSU 2021-22: Topological data analysis, machine learning and neural networks.

Present: I have taken 90 credits of coursework, seminars, and independent studies...still missing 3, which I look forward to completing next semester.

About Me

1. B.S. in Mathematics from University of Rochester; M.S. in Mathematics from University of Washington.
2. WSU 2018-19: Mathematical modeling of forests, biogeochemistry, and distributed systems.
3. WSU 2019-20: Transition to statistics, stochastic processes, state-space models.
4. WSU 2020-21: Animal movement ecology, species distribution models, telemetry data.
5. WSU 2021-22: Topological data analysis, machine learning and neural networks.

Present: I have taken 90 credits of coursework, seminars, and independent studies...still missing 3, which I look forward to completing next semester.
Future: Dissertation, dissertation, dissertation!

What is Topological Data Analysis (TDA)?

Topological Data Analysis

What is Topological Data Analysis (TDA)?

- A subfield of algebraic topology that has found applications in many disciplines outside math.

Topological Data Analysis

What is Topological Data Analysis (TDA)?

- A subfield of algebraic topology that has found applications in many disciplines outside math.
- Based on the idea that data are sampled from an underlying manifold so we can study some algebraic invariants of that space.

Topological Data Analysis

What is Topological Data Analysis (TDA)?

- A subfield of algebraic topology that has found applications in many disciplines outside math.
- Based on the idea that data are sampled from an underlying manifold so we can study some algebraic invariants of that space.
- Important algebraic invariant class are homology groups, which arise from combinatorial representations of the manifold.

Simplicial Complexes

Definition. Let $\{a_0, \ldots, a_n\}$ be a geometrically independent set in $\mathbb{R}^N$. The *n-simplex* σ spanned by $a_0, \ldots, a_n$ is defined to be the set of all points x of $\mathbb{R}^N$ such that $x = \sum_{i=0}^{n} t_i a_i$, where $\sum_{i=0}^{n} t_i = 1$ and $t_i \geq 0$ for all i.

Simplicial Complexes

Definition. Let $\{a_0, \ldots, a_n\}$ be a geometrically independent set in $\mathbb{R}^N$. The *n-simplex* σ spanned by $a_0, \ldots, a_n$ is defined to be the set of all points x of $\mathbb{R}^N$ such that $x = \sum_{i=0}^{n} t_i a_i$, where $\sum_{i=0}^{n} t_i = 1$ and $t_i \geq 0$ for all i.

In other words, it is the convex hull of $n+1$ affinely independent points $a_0, \ldots, a_n$ in an Euclidean space. From simplices we can build simplicial complexes:

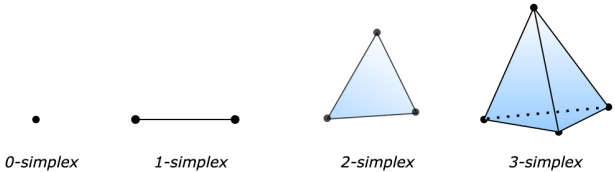
Simplicial Complexes

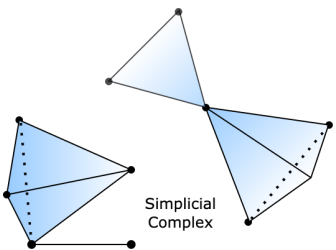
Definition. Let $\{a_0, \ldots, a_n\}$ be a geometrically independent set in $\mathbb{R}^N$. The *n-simplex* σ spanned by $a_0, \ldots, a_n$ is defined to be the set of all points x of $\mathbb{R}^N$ such that $x = \sum_{i=0}^{n} t_i a_i$, where $\sum_{i=0}^{n} t_i = 1$ and $t_i \geq 0$ for all i.

In other words, it is the convex hull of $n+1$ affinely independent points $a_0, \ldots, a_n$ in an Euclidean space. From simplices we can build simplicial complexes:

Definition. A *simplicial complex* K in $\mathbb{R}^N$ is a collection of simplices in $\mathbb{R}^N$ such that (1) every face of a simplex of K is in K, and (2) the intersection of any two simplices of K is a face of each of them.

Examples





Given any topological manifold $\mathbb{X}$, we can model it using simplices via homology groups.

Homology and Betti-Numbers

- The quotient group $H_p(K) := Z_p(K)/B_p(K)$ is the p-**th homology group of** K.

Homology and Betti-Numbers

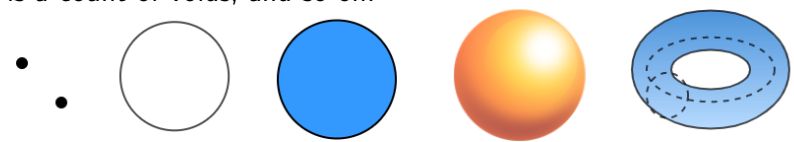
- The quotient group $H_p(K) := Z_p(K)/B_p(K)$ is the p-**th homology group of** K.
- Elements of $H_p(K)$ are p-cycles which do not bound any thing in the middle, and thus describe disconnections, holes, voids and higher-dimensional "absence of stuff".

Homology and Betti-Numbers

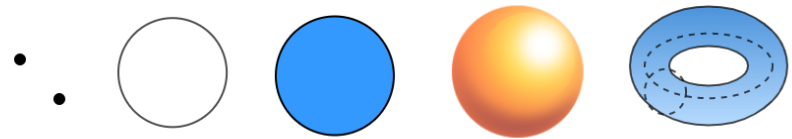
- The quotient group $H_p(K) := Z_p(K)/B_p(K)$ is the p-**th homology group of** K.
- Elements of $H_p(K)$ are p-cycles which do not bound any thing in the middle, and thus describe disconnections, holes, voids and higher-dimensional "absence of stuff".
- Homology groups remain invariant under homeomorphisms, i.e. continuous deformations, and thus they encode intrinsic information on the topology of the given manifold $\mathbb{X}$.

Homology and Betti-Numbers

Betti-number: the dimension of its corresponding homology group, that is $\beta_p := \dim H_p(K)$. For instance β_0 is a count of connected components, β_1 is a count of holes, β_2 is a count of voids, and so on.



Homology and Betti-Numbers



Betti-number	(a)	(b)	(c)	(d)	(e)
β_0	2	1	1	1	1
β_1	0	1	0	0	2
β_2	0	0	0	1	1

Filtrations and Persistent Homology

- When we look at a particular dataset, it is not immediately clear at what scale interesting topological features occur.

Filtrations and Persistent Homology

- When we look at a particular dataset, it is not immediately clear at what scale interesting topological features occur.
- Different topological features occur at different scales, and not clear at what scale these features are born and die.

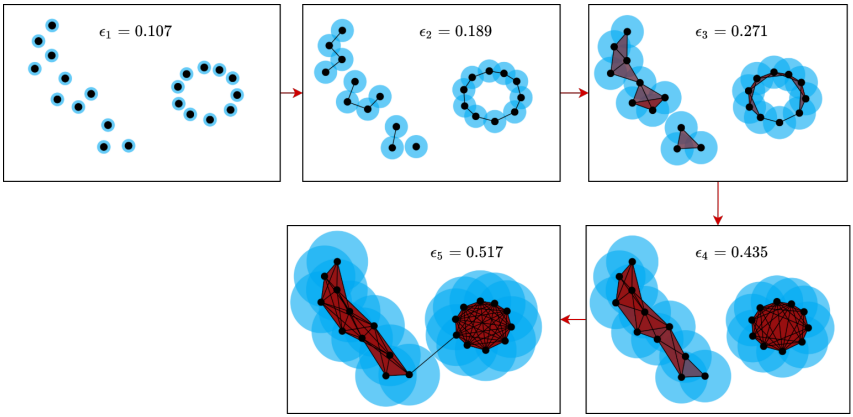
Filtrations and Persistent Homology

- When we look at a particular dataset, it is not immediately clear at what scale interesting topological features occur.
- Different topological features occur at different scales, and not clear at what scale these features are born and die.
- **Persistent Homology** is a tool of TDA used to measure the scale of a topological feature via **filtrations**, sequences of chain complexes along with inclusion maps.

Filtrations and Persistent Homology

- When we look at a particular dataset, it is not immediately clear at what scale interesting topological features occur.
- Different topological features occur at different scales, and not clear at what scale these features are born and die.
- **Persistent Homology** is a tool of TDA used to measure the scale of a topological feature via **filtrations**, sequences of chain complexes along with inclusion maps.
- We use a *persistence diagram* to store information of the topological features in each filtration.

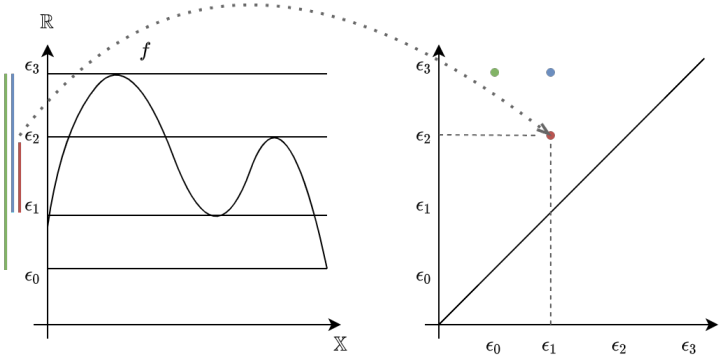
Data and Homology



source: Henry Adams.

Sublevel Sets

$\mathbb{X}$ a connected topological space; $f : \mathbb{X} \to \mathbb{R}$ a continuous function. The *sublevel sets* $\mathbb{X}^\epsilon := f^{-1}(\mathbb{R}_{<\epsilon})$ form a family of nested subspaces for parameters ϵ_i, that is $\mathbb{X}^{\epsilon_0} \subset \mathbb{X}^{\epsilon_1} \subset \mathbb{X}^{\epsilon_2} \subset \dots$. Let $\epsilon_i < \epsilon_j$ we define $\beta(\epsilon_i, \epsilon_j)$ to be the number of components in $\mathbb{X}^{\epsilon_j}$ that have a non-empty intersection with $\mathbb{X}^{\epsilon_i}$.



Persistence Diagrams

The persistence diagram encodes the (birth, death) of persistent homology features of a particular rank.

Persistence Diagrams

The persistence diagram encodes the (birth, death) of persistent homology features of a particular rank.

The space of persistence diagrams can be endowed with a metric induced by Wasserstein distances. We define the distance between two persistence diagrams X and Y as

$$W_q(X, Y) := \left[\inf_{\eta: X \to Y} \sum_{x \in X} \|x - \eta(x)\|_\infty^q \right]^{1/q}.$$

Persistence Diagrams

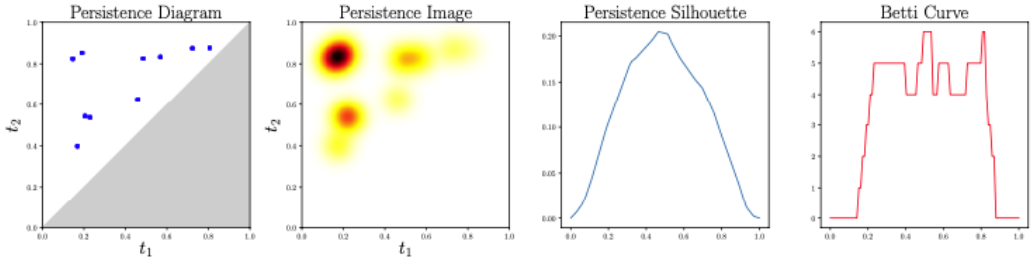
Besides comparing persistence diagrams Wasserstein distances, we can also compute their summary statistics, such as total persistence and the p-norm.

Persistence Diagrams

Besides comparing persistence diagrams Wasserstein distances, we can also compute their summary statistics, such as total persistence and the p-norm.

However, for certain applications, such summaries may lose information and more expressive measurements of the persistent homology features may be desired.

Persistence Diagrams



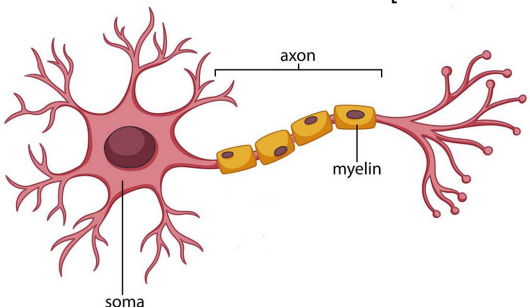
Some common linear representations of persistence diagrams. From left to right: A persistence diagram. Its persistence surface, which is a persistence measure. The corresponding persistence silhouette. The corresponding Betti Curve.

Computations are made using the sklearn-tda library: https://github.com/MathieuCarriere/sklearntda.

Image source: Divol and Lacombe [2019]

Biological Neural Networks

The human brain is estimated to have between 10 to 20 billion neurons in the cerebral cortex and 55 to 70 billion neurons in the cerebellum [von Bartheld et al., 2016].



Gray matter is primarily composed of neuron somas, and white matter is mostly made of axons wrapped in myelin *source: Johns Hopkins Medicine*. Not labeled are **dendrites** branching out of the cell body to receive transmissions from other neurons and the **terminal buttons** at the opposite end serve to transmit neuronal signals.

Biological Neural Networks

- Neurons receive electrochemical inputs from other neurons at the dendrites.

Biological Neural Networks

- Neurons receive electrochemical inputs from other neurons at the dendrites.
- If the sum of all the inputs is above its activation threshold, the neuron will activate.

Biological Neural Networks

- Neurons receive electrochemical inputs from other neurons at the dendrites.
- If the sum of all the inputs is above its activation threshold, the neuron will activate.
- It will transmit an electrochemical signal along its axon in order to pass that signal on to other neurons through its terminal buttons.

Biological Neural Networks

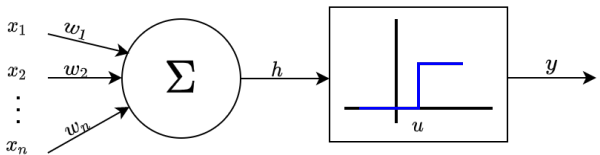
- Neurons receive electrochemical inputs from other neurons at the dendrites.
- If the sum of all the inputs is above its activation threshold, the neuron will activate.
- It will transmit an electrochemical signal along its axon in order to pass that signal on to other neurons through its terminal buttons.
- Bressler and Menon [2010] argue that large-scale brain networks provide a coherent framework for understanding cognition, allowing for a principled exploration for how cognitive functions emerge from such organization.

Artificial Neural Networks

About 60 years ago, McCulloch and Pitts [1943] proposed a computational model for an artificial neuron. It computes a weighted sum of its n inputs x_j and it outputs a 1 if the sum is above a certain threshold u, and a 0 otherwise. That is,

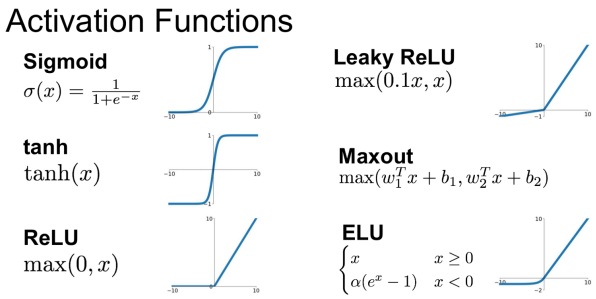
$$y = \theta\left(\sum_{j=1}^{n} w_j x_j - u\right),$$

where $\theta(\cdot)$, the activation function, is a unit step function at zero and w_j is the weight associated with the jth input.



Artificial Neural Networks

Since its introduction, the McCulloch-Pitts neuron has been generalized and different activation functions such as sigmoid (e.g. logistic), Gaussian, and ReLU - rectified linear unit (i.e. the max between 0 and the linear unit) - have been used with the latter being the most popular for training deep networks.



source: *medium.com*

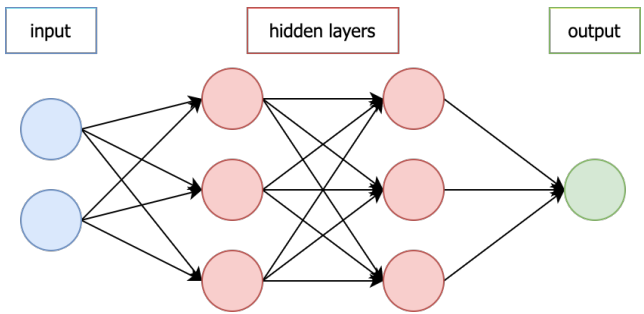
Artificial Neural Networks

ANNs are weighted directed graphs with artificial neurons as nodes and directed edges between neurons representing outputs and inputs. These networks can be constructed with different characteristics depending on the problem they should attempt to solve.

Artificial Neural Networks

ANNs are weighted directed graphs with artificial neurons as nodes and directed edges between neurons representing outputs and inputs. These networks can be constructed with different characteristics depending on the problem they should attempt to solve.

For example, a *feed-forward neural network*:



For more architectures, here is a zoo of them.

BNNs Research Questions

Can network structure from a persistent homology perspective answer the following questions:

Can network structure from a persistent homology perspective answer the following questions:

1. How does cognitive function emerge from large-scale brain networks?

BNNs Research Questions

Can network structure from a persistent homology perspective answer the following questions:

1. How does cognitive function emerge from large-scale brain networks?
2. What mechanisms underlie the dynamic engagement and disengagement of brain areas responsible for the formation and dissolution of large-scale functional networks?

BNNs Research Questions

Can network structure from a persistent homology perspective answer the following questions:

1. How does cognitive function emerge from large-scale brain networks?
2. What mechanisms underlie the dynamic engagement and disengagement of brain areas responsible for the formation and dissolution of large-scale functional networks?
3. How do different large-scale functional networks cooperate, compete and coordinate their activity during complex cognitive behavior?

BNNs Research Questions

Can network structure from a persistent homology perspective answer the following questions:

1. How does cognitive function emerge from large-scale brain networks?
2. What mechanisms underlie the dynamic engagement and disengagement of brain areas responsible for the formation and dissolution of large-scale functional networks?
3. How do different large-scale functional networks cooperate, compete and coordinate their activity during complex cognitive behavior?
4. Is knowledge constructed dynamically by large-scale functional networks?

BNNs Research Questions

One interesting area of research in neuroscience is concerned with how neuronal signaling generates time-varying functional connectivity (TVFC) throughout multiple brain regions. TVFC, or chronnectomics as introduced by Calhoun et al. [2014], describes a dynamic view of connectivity where different regions of the brain (which all may be evolving spatially in time) are coupled with strength measured as functions of time.

One interesting area of research in neuroscience is concerned with how neuronal signaling generates time-varying functional connectivity (TVFC) throughout multiple brain regions. TVFC, or chronnectomics as introduced by Calhoun et al. [2014], describes a dynamic view of connectivity where different regions of the brain (which all may be evolving spatially in time) are coupled with strength measured as functions of time.

Calhoun et al. [2014] show through examples that this perspective has been important in assessing the impact of mental illnesses in the brain, while others such as Saggar et al. [2018] applied this perspective to study how the brain dynamically adapts for efficient functioning through multiple tasks.

BNNs Research Questions

One interesting area of research in neuroscience is concerned with how neuronal signaling generates time-varying functional connectivity (TVFC) throughout multiple brain regions. TVFC, or chronnectomics as introduced by Calhoun et al. [2014], describes a dynamic view of connectivity where different regions of the brain (which all may be evolving spatially in time) are coupled with strength measured as functions of time.

Calhoun et al. [2014] show through examples that this perspective has been important in assessing the impact of mental illnesses in the brain, while others such as Saggar et al. [2018] applied this perspective to study how the brain dynamically adapts for efficient functioning through multiple tasks.

My particular interest in this area is how the chronnectome itself changes overtime depending on the age of the individual.

Deep neural networks are powerful learning models for many problems, however one can cause these networks to misclassify their inputs simply by adding a little bit of humanly imperceptible noise to the example, making it an *adversarial example* [Szegedy et al., 2013].

Deep neural networks are powerful learning models for many problems, however one can cause these networks to misclassify their inputs simply by adding a little bit of humanly imperceptible noise to the example, making it an *adversarial example* [Szegedy et al., 2013].

Adversarial examples in ANNs indicate that the internal representations the network is using to classify inputs is not straightforward nor intuitive.

ANNs Research Questions

Deep neural networks are powerful learning models for many problems, however one can cause these networks to misclassify their inputs simply by adding a little bit of humanly imperceptible noise to the example, making it an *adversarial example* [Szegedy et al., 2013].

Adversarial examples in ANNs indicate that the internal representations the network is using to classify inputs is not straightforward nor intuitive.

Using TDA and statistical methods, I plan to investigate the short-comings in the current research related to the detection of adversarial examples in various ANNs architectures.

ANNs Research Questions

TDA can be used on the input data or to understand how the activation network changes given a particular input, we can also use TDA to inform our machine learning models by incorporating topological information directly into the design of the model - for the purpose of regularizing or analyzing the model.

TDA can be used on the input data or to understand how the activation network changes given a particular input, we can also use TDA to inform our machine learning models by incorporating topological information directly into the design of the model - for the purpose of regularizing or analyzing the model.

For the later situation, Rieck et al. [2018] treat a feed-forward neural network as a stack of bipartite graphs and propose a complexity measure arising from topological features arising in filtrations (using network parameters) of the neural network graph. Using this approach, they were able to distinguish between well trained and badly trained networks.

ANNs Research Questions

TDA can be used on the input data or to understand how the activation network changes given a particular input, we can also use TDA to inform our machine learning models by incorporating topological information directly into the design of the model - for the purpose of regularizing or analyzing the model.

For the later situation, Rieck et al. [2018] treat a feed-forward neural network as a stack of bipartite graphs and propose a complexity measure arising from topological features arising in filtrations (using network parameters) of the neural network graph. Using this approach, they were able to distinguish between well trained and badly trained networks.

My interest is to use a similar approach for model analysis in various ANNs architectures.

ANNs + BNNs Research Questions

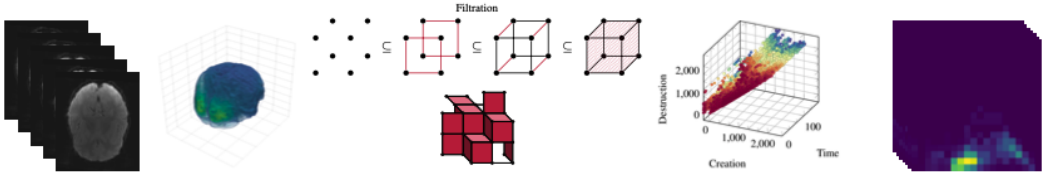
What topological features are revealed from ANNs applied to fMRI data?, and what connections exist between the topological features of the activation network and the ones of the fMRI data itself?

TDA+BNNs: state-of-the-art

Rieck et al. [2020] analyze time varying fMRI data using *cubical complexes*, which does not require the creation of a correlation graph and uses the 'raw' voxel activations themselves.

TDA+BNNs: state-of-the-art

Rieck et al. [2020] analyze time varying fMRI data using *cubical complexes*, which does not require the creation of a correlation graph and uses the 'raw' voxel activations themselves.



(a) fMRI images (b) fMRI volume (c) Cubical complex (d) Persistence diagrams (e) Persistence images

They represent an fMRI stack as a volume from which they create a sequence of cubical complexes. The persistent homology of this sequence results in a set of time-varying persistence diagrams . They calculate summary statistics from the diagrams (not shown), and convert them to vector representations for analysis tasks.

TDA+BNNs: state-of-the-art

They empirically demonstrate that this representation of fMRI data captures age-related differences between individuals.

TDA+BNNs: state-of-the-art

They empirically demonstrate that this representation of fMRI data captures age-related differences between individuals.

Their approach opens future opportunities for the analysis of the geometry of brain state trajectories and link states back to events, as well as serves as a basis for the investigation of neurological pathologies from a new topological perspective.

TDA+ANNs: state-of-the-art

Oballe et al. [2021] introduce a new trainable ANN unit called "ToFU" (Topological Functional Unit)

TDA+ANNs: state-of-the-art

Oballe et al. [2021] introduce a new trainable ANN unit called "ToFU" (Topological Functional Unit)

1. ToFU uses a PD dissimilarity function as its activation.

TDA+ANNs: state-of-the-art

Oballe et al. [2021] introduce a new trainable ANN unit called "ToFU" (Topological Functional Unit)

1. ToFU uses a PD dissimilarity function as its activation.
2. They provide a signal classification example where ToFU learns features that outperform traditional topological vectorizations and remain competitive with those derived from Fourier analysis.

TDA+ANNs: state-of-the-art

Oballe et al. [2021] introduce a new trainable ANN unit called "ToFU" (Topological Functional Unit)

1. ToFU uses a PD dissimilarity function as its activation.
2. They provide a signal classification example where ToFU learns features that outperform traditional topological vectorizations and remain competitive with those derived from Fourier analysis.
3. They build a variational autoencoder architecture that demonstrates how ToFU learns pertinent topology present in the data itself.

References I

Steven L Bressler and Vinod Menon. Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences*, 14(6):277–290, 2010.

Vince D Calhoun, Robyn Miller, Godfrey Pearlson, and Tulay Adalı. The chronnectome: time-varying connectivity networks as the next frontier in fmri data discovery. *Neuron*, 84(2):262–274, 2014.

Vincent Divol and Théo Lacombe. Understanding the topology and the geometry of the persistence diagram space via optimal partial transport. *arXiv preprint arXiv:1901.03048*, 2019.

Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

Christopher Oballe, David Boothe, Piotr J Franaszczuk, and Vasileios Maroulas. Tofu: Topology functional units for deep learning. *Foundations of Data Science*, 2021.

References II

Bastian Rieck, Matteo Togninalli, Christian Bock, Michael Moor, Max Horn, Thomas Gumbsch, and Karsten Borgwardt. Neural persistence: A complexity measure for deep neural networks using algebraic topology. *arXiv preprint arXiv:1812.09764*, 2018.

Bastian Rieck, Tristan Yates, Christian Bock, Karsten Borgwardt, Guy Wolf, Nicholas Turk-Browne, and Smita Krishnaswamy. Uncovering the topology of time-varying fmri data using cubical persistence. *arXiv preprint arXiv:2006.07882*, 2020.

Manish Saggar, Olaf Sporns, Javier Gonzalez-Castillo, Peter A Bandettini, Gunnar Carlsson, Gary Glover, and Allan L Reiss. Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature communications*, 9(1):1–14, 2018.

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

References III

Christopher S von Bartheld, Jami Bahney, and Suzana Herculano-Houzel. The search for true numbers of neurons and glial cells in the human brain: a review of 150 years of cell counting. *Journal of Comparative Neurology*, 524(18):3865–3895, 2016.