To the Graduate Council:

I am submitting herewith a dissertation written by Patrick Douglas Gillespie entitled "Topics in Applied Algebraic Topology with a Measure-Theoretic Perspective." I have examined the final paper copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

Vasileios Maroulas, Major Professor

We have read this dissertation
and recommend its acceptance:

Conrad Plaut

Vyron Vellis

Henry Adams

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

To the Graduate Council:

I am submitting herewith a dissertation written by Patrick Douglas Gillespie entitled "Topics in Applied Algebraic Topology with a Measure-Theoretic Perspective." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

Vasileios Maroulas, Major Professor

We have read this dissertation
and recommend its acceptance:

Conrad Plaut

Vyron Vellis

Henry Adams

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

# Topics in Applied Algebraic Topology with a Measure-Theoretic Perspective

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Patrick Douglas Gillespie

December 2024

*To Rose and Violet*

# Acknowledgements

I would like to thank my advisor, Vasileios Maroulas, for introducing me to topological data analysis and applications of topology to machine learning, and for pushing me to do my best. I would also like to thank Henry Adams, Conrad Plaut, and Vyron Vellis for serving on my doctoral committee.

Thank you to Rose, for all the love and support you have given me throughout my PhD. I am grateful for my family for their consistent encouragement. I would also like to thank Eddie, Jared, and Brittany for much-needed camaraderie.

The work in this dissertation was partially sponsored by US Army Research Office Grant No. W911NF2120186.

# Abstract

There have been a growing number of applications of algebraic topology in recent years to data analysis and machine learning. Persistent homology, one of the main such applications, constructs a filtered simplicial complex from a finite dataset, often using the Vietoris–Rips complex, and then computes algebraic invariants of the filtered simplicial complex which can yield insight on the dataset. Beyond the application of persistent homology in data analysis, ideas from topology have also been applied to machine learning. Neural network architectures, such as simplicial convolutional networks, have been designed to use a topology aware message-passing operation. Building on this, cellular sheaves have been used to define a sheaf convolution operation which combats the problem of over-smoothing encountered in graph and simplicial convolutional networks.

In this dissertation, we study Vietoris complexes and cellular sheaves from a measure-theoretic perspective. First, we strengthen the known relationship between Vietoris complexes and Vietoris metric thickenings: a metric analogue of the Vietoris complex which is equipped with a Wasserstein metric obtained by regarding the points of the Vietoris complex as probability measures. This allows Vietoris complexes to be studied through their metric thickening counterparts, and as an example of this, we are able to prove a Hausmann-like result for Vietoris–Rips complexes of Euclidean submanifolds. We also investigate Vietoris complexes and metric thickenings of absolute neighborhood retracts.

Next, we turn our attention to cellular sheaves and their use within sheaf neural networks. Though it has been demonstrated that sheaf neural networks have advantages for certain types of data, they are sensitive to the choice of cellular sheaf used within their architecture. To mitigate issues arising from this fact, we replace the cellular sheaf in the network with a *distribution* over a space of cellular sheaves, which is then sampled from during inference. We train the resulting network using variational Bayesian inference, yielding what we refer to as a Bayesian sheaf neural network. As part of this work, we strengthen a result on the linear separation power of sheaf diffusion processes and define a novel family of probability distributions on special orthogonal groups via the Cayley transform.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Persistent homology is one of the central tools of topological data analysis and in recent years has found a number of applications to a diverse range of fields such as neuroscience [24, 9], biology and biochemistry [22, 67], materials science [82, 58, 77], and the study of sensor networks [26]. In order to apply persistent homology to a finite and discrete data set $X$, one must be able to convert $X$ into a filtered simplicial complex for which the persistent homology groups can then be computed. A common method of doing so is through the Vietoris–Rips construction, which associates a simplicial complex $\mathrm{VR}(X;r)$ to a metric space $X$ for each threshold value $r > 0$. The simplices of $\mathrm{VR}(X;r)$ are the finite subsets of $X$ with diameter less than $r$, so that as $r$ varies from 0 to $\mathrm{diam}(X)$, the simplicial complex $\mathrm{VR}(X;r)$ evolves from a discrete set to a fully connected complex—with the intermediate stages containing information relevant to $X$.

Because of the role of Vietoris–Rips complexes in applications of persistent homology, there has been growing interest in understanding the homotopy type of $|\mathrm{VR}(X;r)|$ (the geometric realization of $\mathrm{VR}(X;r)$) for not just a finite discrete set but instead the limiting case of a Riemannian manifold or Euclidean submanifold $X$ for arbitrary threshold value $r > 0$ [2, 4, 63]. However, when the Vietoris–Rips complex is considered for a metric space $X$ which is not discrete, the topology of

$|\mathrm{VR}(X;r)|$ becomes cumbersome to work with. For one, $\mathrm{VR}(X;r)$ in this case is not locally finite as a simplicial complex, hence $|\mathrm{VR}(X;r)|$ is not metrizable. Even if one equips $|\mathrm{VR}(X;r)|$ with the metric topology for simplicial complexes, its metric has no natural relationship with the metric of $X$.

To remedy some of these shortcomings, the Vietoris–Rips metric thickening $\mathrm{VR}^m(X;r)$ was introduced by Adamaszek, Adams, and Frick in [3]. The underlying set of $\mathrm{VR}^m(X;r)$ is the same as $|\mathrm{VR}(X;r)|$, yet $\mathrm{VR}^m(X;r)$ is equipped with a metric obtained by regarding the points of $|\mathrm{VR}(X;r)|$ as finitely supported probability measures and defining the distance between two points to be the Wasserstein distance between probability measures. With this metric, there exists an isometric embedding $X \to \mathrm{VR}^m(X;r)$. Similar to Hausmann's theorem [49] which provides the existence of a homotopy equivalence $|\mathrm{VR}(X;r)| \to X$ for sufficiently small values of $r$ when $X$ is a compact Riemannian manifold, it was shown in [3] that for a compact Riemannian manifold $X$ and $r$ sufficiently small, there exists a homotopy equivalence $\mathrm{VR}^m(X;r) \to X$. Moreover, the homotopy equivalence $\mathrm{VR}^m(X;r) \to X$ is naturally defined in terms of Fréchet means (which are unique in this case due to the choice of sufficiently small $r$) and has the embedding $X \to \mathrm{VR}^m(X;r)$ as its homotopy inverse, whereas the homotopy equivalence of Hausmann's theorem is highly non-canonical in the sense that it is defined in terms of a total ordering of $X$.

The relationship between $|\mathrm{VR}(X;r)|$ and $\mathrm{VR}^m(X;r)$ was studied in further detail in [5, 6]. Adams, Frick, and Virk showed in [5] that $|\mathrm{VR}(X;r)|$ and $\mathrm{VR}^m(X;r)$ have isomorphic homotopy groups for any separable metric space $X$ and arbitrary $r > 0$. More precisely, it was shown that the Vietoris complex $\mathcal{V}(\mathscr{U})$ and Vietoris metric thickening $\mathcal{V}^m(\mathscr{U})$, which generalize $\mathrm{VR}(X;r)$ and $\mathrm{VR}^m(X;r)$ respectively, have isomorphic homotopy groups whenever $\mathscr{U}$ is a uniformly bounded open cover of a separable metric space $X$. Their proof relies on comparing the nerve of a particular cover of $\mathcal{V}^m(\mathscr{U})$ with the nerve of $\mathscr{U}$, the latter of which is homotopy equivalent to $|\mathcal{V}(\mathscr{U})|$ by Dowker duality [27]. Critically, this proof does not produce a map between $|\mathcal{V}(\mathscr{U})|$ and $\mathcal{V}^m(\mathscr{U})$ which induces the isomorphisms of homotopy groups.

Ultimately, one would like to know when $|\mathcal{V}(\mathscr{U})|$ and $\mathcal{V}^m(\mathscr{U})$ have the same homotopy type, as this could allow the homotopy type of $|\mathcal{V}(\mathscr{U})|$ to be studied through metric techniques applied to $\mathcal{V}^m(\mathscr{U})$. In Chapter 3 we first strengthen the connection between $|\mathcal{V}(\mathscr{U})|$ and $\mathcal{V}^m(\mathscr{U})$ by proving a homological nerve theorem to show that they have isomorphic homology groups whenever $\mathscr{U}$ is uniformly bounded. In Chapter 4 we use a separate approach to improve upon this by showing that the natural bijection $|\mathcal{V}(\mathscr{U})| \to \mathcal{V}^m(\mathscr{U})$ is a weak homotopy equivalence. This work, as well as some of the preliminary material from Chapter 2 was published in [42]. As a consequence of the the weak homotopy equivalence, we are able to leverage existing results for the Vietoris metric thickening to show that for compact Euclidean submanifolds $M \subseteq \mathbb{R}^n$, the Vietoris–Rips complex $|\mathrm{VR}(M;r)|$ is homotopy equivalent to $M$ for all sufficiently small $r > 0$. This is an analogue of Hausmann's theorem [49] for Euclidean submanifolds, as Hausmann's theorem states that for any closed Riemannian manifold $M$ (with the metric space structure given by geodesic distance), $|\mathrm{VR}(M;r)|$ is homotopy equivalent to $M$ for all sufficiently small $r > 0$.

Given that the conclusion of Hausmann's theorem holds for both Riemannian manifolds and Euclidean submanifold, one might wonder whether it is possible to prove a version of Hausmann's theorem for an even more general class of spaces, encompassing both closed Riemannian manifolds and compact Euclidean submanifolds. It was conjectured in [63] that for a compact absolute neighborhood retract $X$, $|\mathrm{VR}(X;r)|$ is homotopy equivalent to $X$ for all sufficiently small $r > 0$. However in Chapter 5, we disprove this conjecture by producing a counter-example.

Next, we consider *cellular sheaves*, a topic in applied algebraic topology that has recently found application in machine learning. Geometric and topological machine learning are rapidly developing research areas which, broadly speaking, aim to use geometric priors about the data of interest to develop neural network architectures using methods and tools from differential geometry, topology, and graph theory [20, 72, 64]. Graph neural networks are a central example of geometric deep learning and have been applied to a wide range of tasks such as recommender systems [84], traffic modeling

[62, 85], and predicting properties of chemicals [43]. Building on the basic idea of graph neural networks, architectures such as simplicial convolutional neural networks and hypergraph neural networks extend the graph convolution operation to a higher order convolution operation, which has advantages when the dataset contains salient relationships between groups of nodes rather than just pairs such as citation networks [30, 34], chemical data [14, 13], and neural data [69]. However, such convolution operators suffer from over-smoothing, and thus tend to perform best on homophilic graph data, where nodes connected together by edges or higher dimensional simplices have a high likelihood of belonging to the same class [76, 12].

Several approaches have been taken to address the limited capabilities of graph convolutional networks and higher-order variants for heterophilic graph data (i.e. not homophilic) [83, 74, 87, 86]. One approach replaces the graph (or simplicial) convolution operation with a convolution operation defined in terms of a *cellular sheaf* [46, 12]. A cellular sheaf, briefly, is a data structure defined on a simplicial complex which associates a vector space to each simplex, and a linear transformation known as a *restriction map* to each pair of incident simplices, that is, a pair of simplices where one simplex is a face of the other. The restriction maps provide additional flexibility in the message passing operation. They allow neighboring nodes to maintain distinct features throughout the layers of the sheaf neural network: an advantageous quality when working with heterophilic graph data.

While sheaf neural networks were developed to address the issue of over-smoothing in graph neural networks, they are still susceptible to many of the issues common to all neural networks, such as overfitting to limited training data and lack of robustness to noisy data, weight initializations, and hyperparameter selections. One approach to mitigating such issues is to apply tools and ideas from Bayesian statistics to deep learning [50, 73]. Bayesian approaches to machine learning have been used to quantify uncertainty of neural network predictions [53, 40], reduce overfitting [39, 41], add robustness to adversarial attacks [16], as well as simply improve the predictions of machine learning algorithms [32, 44, 67, 71].

In Chapter 6, we apply a Bayesian framework to learning cellular sheaves within a sheaf neural network. Specifically, we regard the sheaf Laplacian—an object which generalizes the graph Laplacian to cellular sheaves—as a latent random variable of the sheaf neural network. In other words, instead of a network which learns a specific cellular sheaf, we consider a neural network which learns a probability distribution over a space of cellular sheaves, which is then sampled from during inference.

The main contributions are as follows. We introduce *Bayesian sheaf neural networks* (BSNN) in which the sheaf Laplacian operator of the sheaf neural network is a latent random variable whose distribution is parameterized by a function of the input data. Motivated by the evidence lower bound for variational Bayesian inference, we derive a loss function for our BSNN which contains a Kullback-Leibler (KL) divergence regularization term in order to efficiently train the BSNN. Additionally, we define a novel family of reparameterizable probability distributions on the rotation group $SO(n)$ that we refer to as *Cayley distributions*. The Cayley distributions have tractable density functions, and thus can be used to perform variational inference on the sheaf Laplacian within the BSNN when the sheaf restriction maps belong to $SO(n)$. We derive closed-form expressions for the KL divergence of the Cayley distributions from the uniform distribution on $SO(n)$ when $n = 2, 3$. Lastly, we evaluate BSNN models on several benchmark graph datasets, specifically, on node classification tasks with limited training data.

# Chapter 2

# Preliminaries

We begin by establishing notation and recalling preliminary concepts. We use $I$ to denote the unit interval, and $I^n$ to denote the unit $n$-cube $I^n = [0,1]^n$. Let $\Delta^n$ denote the standard $n$-simplex, i.e. $\Delta^n = \{(x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} : \sum_i x_i = 1 \text{ and } x_i \geq 0 \text{ for all } i\}$. We write $\partial I^n$ and $\partial \Delta^n$ for the boundaries of $I^n$ and $\Delta^n$ respectively.

## 2.1 Simplicial Complexes

An abstract simplicial complex $K$ on a vertex set $V$ is a collection of finite subsets of $V$ that is closed under taking subsets, that is, if $\sigma \in K$ and $\tau \subseteq \sigma$, then $\tau \in K$. If $\sigma \in K$ and $\operatorname{card}(\sigma) = n+1$, then $\sigma$ is an $n$-dimensional simplex, or $n$-simplex. We also refer to a 0-simplex as a vertex of $K$. Let $K_n$ denote the set of $n$-simplices of $K$. We use $K^{(n)}$ to denote the $n$-skeleton of $K$, that is, the set of all simplices of dimension not greater than $n$. The geometric realization $|K|$ is the set of all functions $\alpha : K_0 \to I$ which satisfy

(1) $\{v \in K_0 : \alpha(v) \neq 0\}$ is a simplex of $K$,

(2) $\sum_{v \in K_0} \alpha(v) = 1$.

If $\sigma \in K$ is a simplex of $K$, let $|\sigma| \subseteq |K|$ denote the set of all $\alpha \in |K|$ with support in $\sigma$. The topology of $|K|$ is the CW topology, in which a set $U$ is open in $|K|$ if $U \cap |\sigma|$

is open in $|\sigma| \cong \Delta^n$ for every $\sigma \in K$. If $\sigma \in K$ is an $n$-simplex of $K$, let $\partial\sigma$ denote the set of all $(n-1)$-simplices contained in $\sigma$. A triangulation $T$ of a space $X$ is a homeomorphism $T : |K| \to X$ for some simplicial complex $K$.

### 2.1.1 Complexes with the Metric Topology

For a vertex $v \in K$, the *barycentric coordinate* $\psi_v : |K| \to I$ is the function defined by $\psi_v(\alpha) = \alpha(v)$. For each vertex $v$, $\psi_v$ is continuous [66, Appendix 1, Corollary 2]. The barycentric coordinates can be used to define a metric $d_m$ on the set $|K|$ by setting $d_m(\alpha, \beta) = \sum_{v \in K_0} |\psi_v(\alpha) - \psi_v(\beta)|$. We denote by $|K|_m$ the space whose underlying set is $|K|$ and has the metric topology inherited from the metric $d_m$. The space $|K|$ has a finer topology than $|K|_m$, hence the identity map $|K| \to |K|_m$ is continuous, and is a homeomorphism if and only if $K$ is locally finite. Though $\mathrm{id} : |K| \to |K|_m$ is not a homeomorphism if $K$ is not locally finite, it is always a homotopy equivalence [66, Appendix 1, Theorem 10]. The following is a useful characterization of continuous maps into $|K|_m$.

**Lemma 2.0.1.** *[66, Appendix 1, Theorem 8] If $K$ is a simplicial complex and $Z$ is a topological space, a function $f : Z \to |K|_m$ is continuous if and only if $\psi_v \circ f : Z \to I$ is continuous for all vertices $v \in K_0$.*

## 2.2 Homotopy and Homology Groups

Two maps $f, g : X \to Y$ are homotopic if there exists a map $H : X \times I \to Y$ such that $H(\cdot, 0) = f$ and $H(\cdot, 1) = g$, and in this case we write $f \simeq g$. Given $A \subseteq X$ and $B \subseteq Y$ we write $f : (X, A) \to (Y, B)$ to denote a map $f : X \to Y$ satisfying $f(A) \subseteq B$. Given $A \subseteq X$, we say that $H : X \times I \to Y$ is a homotopy relative to $A$ (or rel. $A$) if $H(a, t) = H(a, 0)$ for all $t \in I$ and $a \in A$. If $H, G : X \times I \to Y$ are homotopies such that $H(x, 1) = G(x, 0)$ for all $x \in X$, the concatenation $H \cdot G : X \times I \to Y$ is the homotopy defined by

$$(H \cdot G)(x,t) = \begin{cases} H(x,2t) & t \in [0,1/2] \\ G(x,2t-1) & t \in [1/2,1]. \end{cases}$$

Let $X$ be a topological space and let $x_0 \in X$. The $n$th homotopy group of $(X, x_0)$ is denoted $\pi_n(X, x_0)$ and is the set of homotopy classes of maps $f : (I^n, \partial I^n) \to (X, x_0)$. We denote the homotopy class of a map $f$ by $[f]$. For a map $g : (X, x_0) \to (Y, y_0)$, the induced map on homotopy groups is denoted $g_* : \pi_n(X, x_0) \to \pi_n(Y, y_0)$. The set $\pi_0(X, x_0)$ is the set of path components of $X$ and does not depend on the basepoint, hence may simply be denoted $\pi_0(X)$. While $\pi_n(X, x_0)$ relies on a choice of basepoint $x_0 \in X$, when $X$ is path-connected, for any two basepoints $x_1, x_2 \in X$, there is an isomorphism $\pi_n(X, x_1) \cong \pi_n(X, x_2)$ and so in this case we may simply write $\pi_n(X)$.

A pair of spaces $X$ and $Y$ are homotopy equivalent if there exist maps $f : X \to Y$ and $g : Y \to X$ such that $gf \simeq \mathrm{id}_X$ and $fg \simeq \mathrm{id}_Y$, in which case we write $X \simeq Y$. A weak homotopy equivalence between $X$ and $Y$ is a map $f : X \to Y$ which induces a bijection $f_* : \pi_0(X) \to \pi_0(Y)$ on path components, and induces isomorphisms $f_* : \pi_n(X, x) \to \pi_n(Y, f(x))$ for all $x \in X$ and $n \geq 1$.

A chain complex is a collection of abelian groups $\{C_n\}_{n \in \mathbb{N}}$ and homomorphisms $\partial_n : C_n \to C_{n-1}$ (referred to as boundary maps or differentials) such that $\partial_n \partial_{n+1} = 0$. The $n$th homology of the chain complex $C_\bullet$ is $H_n(C_\bullet) = \ker \partial_n / \operatorname{im} \partial_{n+1}$.

Given a simplicial complex $K$, the group of $n$-chains in $K$, denoted $C_n(K)$, is the free abelian group generated by the $n$-simplices of $K$, that is, $C_n(K) = \{\sum_{i=1}^k a_i \sigma_i : a_i \in \mathbb{Z}, k \in \mathbb{N}, \sigma_i \in K_n\}$. Place a total order on the vertex set of $K$ so that each simplex $\sigma \in K$ may be uniquely written as $\sigma = [v_0, v_1, \ldots, v_n]$ for $v_0 < v_1 < \cdots < v_n$. A boundary map $\partial_n : C_n(K) \to C_{n-1}(K)$ is defined on a basis element $\sigma = [v_0, v_1, \ldots, v_n]$ by setting $\partial_n(\sigma) = \sum_{i=0}^n (-1)^i [v_0, \ldots, \hat{v}_i, \ldots, v_n]$ where $\hat{v}_i$ denotes the omission of vertex $v_i$. The $n$th homology group of $K$, denoted $H_n(K)$, is the $n$th homology group of the chain complex

$$\cdots \to C_{n+1}(K) \to C_n(K) \to C_{n-1}(K) \to \cdots .$$

For a topological space $X$, the group of singular $n$-chains in $X$, denoted $S_n(X)$, is the free abelian group generated by the collection of all maps $\Delta^n \to X$, that is, $S_n(X) = \{\sum_{i=1}^k a_i\sigma_i : a_i \in \mathbb{Z}, k \in \mathbb{N}, \sigma_i : \Delta^n \to X\}$. The boundary map $\partial_n : S_n(X) \to S_{n-1}(X)$ is defined analogous to the simplicial case by setting $\partial_n(\sigma) = \sum_{i=0}^n (-1)^i \sigma|_{[v_0,\ldots,\hat{v}_i,\ldots,v_n]}$ where $\sigma|_{[v_0,\ldots,\hat{v}_i,\ldots,v_n]}$ denotes the restriction of $\sigma : \Delta^n \to X$ to the face of $\Delta^n$ obtained by omitting the vertex $v_i$ of $\Delta^n$. The singular chain groups and boundary maps form a chain complex

$$\cdots \to S_{n+1}(X) \to S_n(X) \to S_{n-1}(X) \to \cdots ,$$

and the $n$th homology group of $X$ is $H_n(X) = \ker \partial_n / \operatorname{im} \partial_{n+1}$.

## 2.3 Vietoris Complexes

If $(X, d_X)$ is a metric space and $r \in \mathbb{R}$, the (open) Vietoris–Rips complex $\mathrm{VR}(X; r)$ is the simplicial complex whose vertex set consists of the points of $X$ and whose simplices are the finite subsets of $X$ with diameter less than $r$. A similar construction is the Čech complex $\check{\mathrm{C}}(X; r)$ whose simplices are finite subsets $\{x_1, \ldots, x_n\} \subseteq X$ such that there exists $z \in X$ satisfying $d_X(z, x_i) < r$ for all $i \leq n$.

Both of these constructions are generalized by the Vietoris complex. Given a cover $\mathscr{U}$ of $X$, the *Vietoris complex* $\mathcal{V}(\mathscr{U})$ is the simplicial complex whose vertex set is $X$ and contains a simplex $\sigma = \{x_1, x_2, \ldots, x_n\} \subseteq X$ if $\sigma \subseteq U$ for some $U \in \mathscr{U}$. If $\mathscr{U}$ is the cover of $X$ by open sets of diameter less than $r$, then $\mathcal{V}(\mathscr{U}) = \mathrm{VR}(X; r)$. Alternatively, if $\mathscr{U}$ is the open cover of $X$ by open balls of radius $r$, then $\mathcal{V}(\mathscr{U}) = \check{\mathrm{C}}(X; r)$.

**Remark 2.1.** The open Vietoris–Rips complex is sometimes denoted $\mathrm{VR}_<(X; r)$ in order to distinguish it from the closed Vietoris–Rips complex $\mathrm{VR}_\leq(X; r)$: the complex whose simplices are the finite subsets of $X$ with diameter less than or equal to $r$. While work has been done on understanding the homotopy type of Vietoris–Rips complexes $|\mathrm{VR}_{\leq}(X;r)|$ with the $\leq$ convention and their corresponding metric thickenings $\mathrm{VR}^m_{\leq}(X;r)$ [3, 70], in this dissertation we will only be concerned with open Vietoris–Rips complexes and their corresponding metric thickenings. Hence the notation $\mathrm{VR}(X;r)$ in the present work will always refer to the open Vietoris–Rips complex.

A cover $\mathscr{U}$ of $X$ is said to be *uniformly bounded* if there exists $D < \infty$ such that $\mathrm{diam}(U) \leq D$ for all $U \in \mathscr{U}$. The covers defining $\mathrm{VR}(X;r)$ and $\check{\mathrm{C}}(X;r)$ for any $0 < r < \infty$ are examples of uniformly bounded covers. If $\mathscr{U}$ is a cover of $X$ and $f : Y \to X$ is a map, let $f^{-1}\mathscr{U}$ denote the cover of $Y$ given by $f^{-1}\mathscr{U} = \{f^{-1}(U) : U \in \mathscr{U}\}$.

## 2.4 Vietoris Metric Thickenings

Let $(X, d_X)$ be a metric space. The support of a measure $\mu$ on the Borel sets of $X$ is the set consisting of points $x \in X$ where $\mu(U) > 0$ for every open neighborhood $U$ of $x$. The support of $\mu$ is denoted $\mathrm{supp}(\mu)$. Let $\mathcal{P}^{\mathrm{fin}}(X)$ denote the set of all probability measures on $X$ with finite support. If $\mu, \nu \in \mathcal{P}^{\mathrm{fin}}(X)$, a *coupling between* $\mu$ *and* $\nu$ is a probability measure $\gamma$ on $X \times X$ whose marginals on the first and second factors of $X \times X$ are $\mu$ and $\nu$ respectively, that is, $\gamma$ satisfies $\gamma(A \times X) = \mu(A)$ and $\gamma(X \times A)) = \nu(A)$ for all Borel subsets $A \subseteq X$. If $\mathscr{C}$ denotes the set of couplings between $\mu, \nu \in \mathcal{P}^{\mathrm{fin}}(X)$, the 1-Wasserstein distance between $\mu$ and $\nu$ is

$$d_W(\mu, \nu) = \inf_{\gamma \in \mathscr{C}} \int_{X \times X} d_X(x, y) \gamma(dx \times dy).$$

Since we will only consider the 1-Wasserstein distance between finitely supported probability measures, we can alternatively define $d_W(\mu, \nu)$ as an infimum of finite sums in the following manner. For $x \in X$, let $\delta_x$ denote the Dirac probability measure at $x$. Each measure $\mu \in \mathcal{P}^{\mathrm{fin}}(X)$ has a unique representation $\mu = \sum_{i \in \mathcal{I}} a_i \delta_{x_i}$ in which $\text{supp}(\mu) = \{x_i\}_{i\in\mathcal{I}} \subseteq X$ is a finite set indexed by $\mathcal{I}$, $a_i > 0$ for all $i \in \mathcal{I}$, and $\sum_{i\in\mathcal{I}} a_i = 1$. If it will not cause confusion, we may omit the finite index set $\mathcal{I}$ and simply write $\mu = \sum_i a_i\delta_{x_i}$. Now if $\mu = \sum_i a_i\delta_{x_i}$ and $\nu = \sum_j b_j\delta_{y_j}$, a coupling between $\mu$ and $\nu$ can be represented as a sum $\gamma = \sum_{i,j} \gamma_{i,j}\delta_{(x_i,y_j)}$ such that $\gamma_{i,j} \geq 0$ for all $i$ and $j$, $\sum_i \gamma_{i,j} = b_j$ for all $j$, and $\sum_j \gamma_{i,j} = a_i$ for all $i$. With this in mind, the 1-Wasserstein distance is then

$$d_W(\mu,\nu) = \inf_{\gamma\in\mathscr{C}} \sum_{i,j} \gamma_{i,j} d_X(x_i, y_j).$$

The Wasserstein metric is sometimes known as the optimal transport metric. A coupling $\gamma$ between $\mu = \sum_i a_i\delta_{x_i}$ and $\nu = \sum_j b_j\delta_{y_j}$ can be regarded as a transport plan between $\mu$ and $\nu$, with the constants $\gamma_{i,j}$ specifying the amount of mass moved from $x_i$ to $y_j$, and $\gamma_{i,j}d_X(x_i, y_i)$ is the associated cost.

If $\mathscr{U}$ is a cover of $X$, the *Vietoris metric thickening* $\mathcal{V}^m(\mathscr{U})$ is the metric space of all $\mu \in \mathcal{P}^{\text{fin}}(X)$ such that $\text{supp}(\mu)$ is contained in an element of $\mathscr{U}$, equipped with the 1-Wasserstein metric. Note that the map $X \to \mathcal{V}^m(\mathscr{U})$ defined by $x \mapsto \delta_x$ is an isometric embedding.

The underlying set of $\mathcal{V}^m(\mathscr{U})$ is the same as $|\mathcal{V}(\mathscr{U})|$ since any $\alpha \in |\mathcal{V}(\mathscr{U})|$ is a function $\alpha : X \to I$ with finite support contained in an element of $\mathscr{U}$ satisfying $\sum_{x\in X} \alpha(x) = 1$, hence can be equally regarded as a finitely supported probability measure $\mu \in \mathcal{V}^m(\mathscr{U})$. However, the topology of $\mathcal{V}^m(\mathscr{U})$ is coarser than that of $|\mathcal{V}(\mathscr{U})|$, so that while the natural bijection $\text{id} : |\mathcal{V}(\mathscr{U})| \to \mathcal{V}^m(\mathscr{U})$ is continuous [3, Proposition 6.1], it is not in general a homeomorphism.

Recalling the barycentric coordinate functions $\psi_v$, we may on occasion find it useful to represent $\mu \in \mathcal{V}^m(\mathscr{U})$ as the sum $\mu = \sum_{x\in X} \psi_x(\mu)\delta_x$, which, while indexed by $X$, has only finitely many non-zero terms.

# Chapter 3

# Homology Groups of Vietoris Metric Thickenings

In this chapter, we show that the Vietoris complex $\mathcal{V}(\mathscr{U})$ and Vietoris metric thickening $\mathcal{V}^m(\mathscr{U})$ have isomorphic homology groups. In [5], for any any $n \in \mathbb{N}$, the authors construct an open cover $\widetilde{M}_{\mathscr{U}}$ of $\mathcal{V}^m(\mathscr{U})$ that is *good up to level* $n$, that is, the intersection of any collection of at most $n$ sets from $\widetilde{M}_{\mathscr{U}}$ is either empty or contractible. The authors of [5] remarked that these covers could potentially be used in a Mayer-Vietoris spectral sequence to show that $\mathcal{V}(\mathscr{U})$ and $\mathcal{V}^m(\mathscr{U})$ have isomorphic homology groups. The argument we present does just this, as we will prove a homological nerve theorem—which is ultimately an application of the Mayer-Vietoris spectral sequence—in order to show that $\mathcal{V}(\mathscr{U})$ and $\mathcal{V}^m(\mathscr{U})$ have isomorphic homology groups.

Given a cover $\mathscr{U} = \{U_i\}_{i\in\mathcal{I}}$ of a topological space $X$, the *nerve* of $\mathscr{U}$, which we denote $\mathcal{N}(\mathscr{U})$, is a simplicial complex whose vertex set is $\mathcal{I}$ and whose simplices are the finite subsets $\sigma \subset \mathcal{I}$ such that the intersection $\bigcap_{i\in\sigma} U_i$ is nonempty. There are many nerve theorems, each of which relate a space $X$ with $\mathcal{N}(\mathscr{U})$, but vary on the assumptions placed on $X$ and $\mathscr{U}$, as well as the conclusions drawn.

One of the earliest examples of a nerve theorem is due to Borsuk [15]. Borsuk proved that if $X$ is a finite-dimensional compact metric space and $\mathscr{A}$ is a finite cover of $X$ by closed subsets of $X$ such that intersection of any subset of $\mathscr{A}$ is an absolute retract, then $X$ and $\mathcal{N}(\mathscr{A})$ have the same homotopy type.

Another early example of a nerve theorem is contained in the work of Leray in [59] and [60]. His work implies that if $X$ is a finite simplicial complex and $\mathscr{A}$ is a finite cover of $X$ by subcomplexes such that the intersection of any subset of $\mathscr{A}$ has trivial cohomology, then $H^n(X) \cong H^n(\mathcal{N}(\mathscr{A}))$ for all $n$. The analogous nerve theorem for homology can be found in [21] for example, in which K. Brown notes that the theorem is "essentially due to Leray".

A sharper form of this homological nerve theorem is proven and used by R. Meshulam [68], which relaxes the condition that intersections of finite subsets of the cover $\mathscr{A}$ are homologically trivial, but only shows that $H_j(X) \cong H_j(\mathcal{N}(\mathscr{A}))$ for $j \leq n$ for a particular $n$. However, Meshulam assumes that $X$ is a finite simplicial complex, $\mathscr{U}$ is a finite cover of $X$ by subcomplexes, and takes homology to have coefficients in a field.

We show that the homological nerve theorem in [68] holds for the case where $X$ is an arbitrary topological space and $\mathscr{U}$ is an open cover of $X$.

**Theorem 3.1.** *Let $\mathscr{U} = \{U_i\}_{i\in\mathcal{I}}$ be an open cover of a topological space $X$, and let $N$ be the nerve of this cover. Fix an integer $k \in \mathbb{N}$. If $\widetilde{H}_j(\cap_{i\in\sigma}U_i) = 0$ for all $\sigma \in N^{(k)}$ and $j \in \{0, \ldots, k - \dim\sigma\}$, then*

(1) *$H_j(X) \cong H_j(N)$ for all $j \in \{0, \ldots, k\}$*

(2) *if $H_{k+1}(N) \neq 0$ then $H_{k+1}(X) \neq 0$.*

## 3.1 Mayer-Vietoris Spectral Sequence

We give a brief description of the spectral sequence of a cover, which is sometimes known as the Mayer-Vietoris spectral sequence. Fix an open cover $\mathscr{U} = \{U_i\}_{i\in\mathcal{I}}$ of $X$ and to simplify the notation, let $N = \mathcal{N}(\mathscr{U})$ be the nerve of $\mathscr{U}$. For a simplex $\sigma \subset \mathcal{I}$ in $N$, let $U_\sigma$ denote the intersection $U_\sigma = \cap_{i\in\sigma}U_i$.

Given the open cover $\mathscr{U}$ there is an associated double complex $(A, \partial', \partial'')$ (see Figure 3.1). A *double complex* is a collection of abelian groups $\{A_{p,q}\}_{p,q\in\mathbb{Z}}$ along with two collections of homomorphisms

$$\partial' : A_{p,q} \to A_{p-1,q} \qquad \partial'' : A_{p,q} \to A_{p,q-1}$$

which satisfy $\partial'\partial' = \partial''\partial'' = 0$ and $\partial'\partial'' = \partial''\partial'$. To define the double complex associated to $\mathscr{U}$, set $A_{p,q} = \bigoplus_{\sigma\in N_p} S_q(U_\sigma)$. The vertical differentials $\partial'' : A_{p,q} \to A_{p,q-1}$ are induced by the boundary maps $\partial : S_q(U_\sigma) \to S_{q-1}(U_\sigma)$, and the horizontal differentials $\partial' : A_{p,q} \to A_{p-1,q}$ are defined as follows.

Fix a total order on the vertices $\{v_i\}_{i\in\mathcal{I}}$ of $N$ so that each simplex of $N$ has a unique representation $\sigma = [v_0, \ldots, v_p]$ for which $v_0 < \cdots < v_p$. Then if $\sigma = [v_0, \ldots, v_p]$ is a $p$-simplex with $v_0 < \cdots < v_p$, define $f_j(\sigma)$ to be the $(p-1)$-simplex $f_j(\sigma) = [v_0, \ldots, \widehat{v_j}, \ldots, v_p]$ in which $\widehat{v_j}$ denotes that the vertex $v_j$ is omitted. Since $U_\sigma \subset U_{\partial_j\sigma}$, we have that $f_j$ defines an inclusion

$$S_q(U_\sigma) \to S_q(U_{f_j(\sigma)}) \to \bigoplus_{\tau\in N_{p-1}} S_q(U_\tau)$$

for each $\sigma$. These inclusions induce maps $\delta_j : \bigoplus_{\sigma\in N_p} S_q(U_\sigma) \to \bigoplus_{\tau\in N_{p-1}} S_q(U_\tau)$. We then define $\partial' : A_{p,q} \to A_{p-1,q}$ by setting $\partial' = \sum_{i=0}^{p}(-1)^i\delta_i$. One may check that $\partial'\partial' = 0$ by expanding out

$$\partial'\partial' = \sum_{k=0}^{p}\sum_{j=0}^{p-1}(-1)^{k+j}\delta_j\delta_k$$

and using the relation $\delta_j\delta_k = \delta_{k-1}\delta_j$ if $j < k$. Note that $\partial'\partial'' = \partial''\partial'$.

Given the double complex $A$, we may form the total complex $\operatorname{Tot} A$, whose degree $n$ term is $(\operatorname{Tot} A)_n = \bigoplus_{p+q=n} A_{p,q}$. The differential $\partial$ of $\operatorname{Tot} A$ is defined by setting $\partial(c) = \partial'(c) + (-1)^p\partial''(c)$ for $c \in A_{p,q}$, for all $p, q \in \mathbb{N}$.

$$
\begin{array}{ccccccccc}
 & & \vdots & & \vdots & & \vdots & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longleftarrow & \bigoplus_{\sigma\in N_0} S_2(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_1} S_2(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_2} S_2(U_\sigma) & \longleftarrow & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longleftarrow & \bigoplus_{\sigma\in N_0} S_1(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_1} S_1(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_2} S_1(U_\sigma) & \longleftarrow & \cdots \\
 & & \downarrow & & \downarrow{\scriptstyle \partial''} & & \downarrow & & \\
0 & \longleftarrow & \bigoplus_{\sigma\in N_0} S_0(U_\sigma) & \overset{\partial'}{\longleftarrow} & \bigoplus_{\sigma\in N_1} S_0(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_2} S_0(U_\sigma) & \longleftarrow & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}
$$

**Figure 3.1:** The double complex $(A, \partial', \partial'')$

There are two natural filtrations $F'$ and $F''$ of $\operatorname{Tot} A$, essentially by the columns and rows of $A$ respectively. The first filtration is given by $F'_k(\operatorname{Tot} A)_n = \bigoplus_{p\leq k} A_{p,n-p}$ and the second filtration is $F''_k(\operatorname{Tot} A)_n = \bigoplus_{q\leq k} A_{n-q,q}$. The filtrations $F'$ and $F''$ of $\operatorname{Tot} A$ give rise to *spectral sequences* $E'$ and $E''$ respectively. Very briefly, a spectral sequence is a sequence of objects $\{E^r, \partial_r\}_{r\in\mathbb{N}}$, where each "page" $E^r$ is a bigraded abelian group, i.e. $E^r = \{E^r_{p,q}\}_{p,q\in\mathbb{Z}}$ for abelian groups $E^r_{p,q}$, with differentials $\partial_r : E^r_{p,q} \to E^r_{p-1+r,q-r}$, and the terms of $E^{r+1}$ are obtained as the homology groups of the previous page $E^r$ with respect to $\partial_r$. Under mild condition which will apply in what follows, for each pair $p, q$, the terms $E^r_{p,q}$ stabilize for large enough $r$, that is $E^{r_1}_{p,q} \cong E^{r_0}_{p,q}$ for all $r_1 \geq r_0$ for some fixed $r_0$ and in this case we define $E^\infty_{p,q} := E^{r_0}_{p,q}$.

The terms $E^\infty_{p,q}$ from the spectral sequence associated to a filtration $F$ of $\operatorname{Tot} A$ are closely related to the homology groups of $\operatorname{Tot} A$. Specifically,

$$E^\infty_{p,q} \cong F_p(H_{p+q}(\operatorname{Tot} A))/F_{p-1}(H_{p+q}(\operatorname{Tot} A)),$$

where the filtration on $H_\bullet(\operatorname{Tot} A)$ is defined by setting $F_p(H_\bullet(\operatorname{Tot} A))$ as the image of $H_\bullet(F_p(\operatorname{Tot} A))$ under the map induced by the injection $F_p(\operatorname{Tot} A) \to \operatorname{Tot} A$. However, if $E^{\infty}_{p,q}$ is non-zero only when $p = 0$, we have the much more direct relationship $E^{\infty}_{0,q} \cong H_q(\operatorname{Tot} A)$.

We avoid giving a comprehensive treatment of spectral sequences, which can be found in [65], and instead state some facts relevant to our application. For the spectral sequence $E'$ associated with the filtration $F'$ of $\operatorname{Tot} A$ defined earlier, the 0th page $E^0$ is simply $A$ and the corresponding differential is $\partial''$. The $E^1$ page is obtained by taking homology of $A$ with respect to $\partial''$, and the differential $\partial'$ on $A$ induces a differential we also denote by $\partial'$ on $E^1$. Hence the $E^2$ page is obtained by taking the homology of $E^1$ with respect to $\partial'$. For the second spectral sequence $E''$, the $E^0$ page is again $A$, but now $E^1$ is obtained by taking the homology of $A$ with respect to $\partial'$, and then the $E^2$ page is obtained by next taking the homology of $E^1$ with respect to the differential induced by $\partial''$.

We will use the second spectral sequence to show that $H_\bullet(\operatorname{Tot} A) \cong H_\bullet(X)$, which we then compare with the first spectral sequence in Section 3.2 to prove Theorem 3.1.

Let $E = E''$ be the second spectral sequence of the double complex $A$. To describe the $E^1$ page explicitly, we need the following proposition.

**Proposition 3.1.** *Let $q \in \mathbb{N}$ and let $A_{\bullet,q}$ denote the chain complex*

$$\ldots \longrightarrow \bigoplus_{\sigma \in N_2} S_q(U_\sigma) \xrightarrow{\partial'} \bigoplus_{\sigma \in N_1} S_q(U_\sigma) \xrightarrow{\partial'} \bigoplus_{\sigma \in N_0} S_q(U_\sigma) \longrightarrow 0.$$

*Then $H_j(A_{\bullet,q}) \cong 0$ for all $j > 0$ and $H_0(A_{\bullet,q}) \cong S^{\mathscr{U}}_q(X)$.*

Here $S^{\mathscr{U}}_q(X)$ denotes the group of singular $q$-chains in $X$ whose elements $\sum_{i=0}^m n_i \sigma_i$ satisfy the condition that each singular simplex $\sigma_i$ has image in an element of $\mathscr{U}$.

To prove Proposition 3.1, we provide a generalization of the proof in [21, pg. 166], which assumes that $X$ is a $CW$-complex and $\mathscr{U}$ is a cover of $X$ by subcomplexes. A similar proposition, but for cohomology, is proved by Frigerio and Maffei in [37].

*Proof.* We prove the proposition by giving an alternative characterization of the groups $\bigoplus_{\sigma \in N_p} S_q(U_\sigma)$. We begin by noting that $\bigoplus_{\sigma \in N_p} S_q(U_\sigma)$ has a basis $B$ consisting of pairs $(\sigma, f)$ where $\sigma$ is a $p$-simplex of $N$ and $f$ is a map $f : \Delta^q \to U_\sigma$. For any map $f : \Delta^q \to X$, let $N^f$ be the subcomplex of $N$ consisting of simplices $\sigma$ such that $\mathrm{im}(f) \subset U_\sigma$. Then there is a bijection between $B$ and the set of pairs $(f, \sigma)$ where $f$ is an arbitrary map $f : \Delta^q \to X$ and $\sigma$ is a $p$-simplex of $N^f$. This is to say that for each $p$, there exists an isomorphism

$$\bigoplus_{\sigma \in N_p} S_q(U_\sigma) \cong \bigoplus_{f:\Delta^q \to X} C_p(N^f).$$

Moreover, these isomorphisms define an isomorphism of chain complexes

$$\begin{array}{ccccccccc} \cdots \longrightarrow & \bigoplus_{\sigma \in N_2} S_q(U_\sigma) & \longrightarrow & \bigoplus_{\sigma \in N_1} S_q(U_\sigma) & \longrightarrow & \bigoplus_{\sigma \in N_0} S_q(U_\sigma) & \longrightarrow 0 \\ & \downarrow \cong & & \downarrow \cong & & \downarrow \cong & \\ \cdots \to & \bigoplus_{f:\Delta^q \to X} C_2(N^f) & \to & \bigoplus_{f:\Delta^q \to X} C_1(N^f) & \to & \bigoplus_{f:\Delta^q \to X} C_0(N^f) & \to 0 \end{array}$$

where the differential in the bottom complex is induced by the boundary maps $C_p(N^f) \to C_{p-1}(N^f)$ on simplicial $p$-chains. Observe that for each $f : \Delta^q \to X$, the complex $N^f$ consists of all finite subsets of the set $\{i \in \mathcal{I} : \mathrm{im}(f) \subset U_i\}$. Hence $N^f$ is either empty or contractible. Hence $H_p(N^f) \cong 0$ for all $p > 0$ and the homology groups of the bottom chain complex (and hence the top as well) are zero at each position except $\bigoplus_{f:\Delta^q \to X} C_0(N^f)$. Here we note that $H_0(N_f)$ is either 0 or $\mathbb{Z}$, depending on whether $N_f$ is empty or not, which in turn depends on whether $f$ has image in some element of $\mathscr{U}$. Thus we can take the set of maps $f : \Delta^q \to X$ which have image in some element of $\mathscr{U}$ to be a basis for $\bigoplus_{f:\Delta^q \to X} H_0(N_f)$. This implies that $\bigoplus_{f:\Delta^q \to X} H_0(N_f) \cong S^{\mathscr{U}}(X)$, completing the proof. □

Using Proposition 3.1, we see that the $E^1$ page of the second spectral sequence is of the form

$$
\begin{array}{ccccc}
 & \vdots & \vdots & \vdots & \\
 & \downarrow & \downarrow & \downarrow & \\
0 & S_2^{\mathscr{U}}(X) & 0 & 0 & \cdots \\
 & \downarrow & \downarrow & \downarrow & \\
0 & S_1^{\mathscr{U}}(X) & 0 & 0 & \cdots \\
 & \downarrow & \downarrow & \downarrow & \\
0 & S_0^{\mathscr{U}}(X) & 0 & 0 & \cdots \\
 & \downarrow & \downarrow & \downarrow & \\
 & 0 & 0 & 0 &
\end{array}
$$

and where the differentials are induced by $\partial''$. Let $H_q^{\mathscr{U}}(X)$ denote the $q$-th homology group of the chain complex $S_\bullet^{\mathscr{U}}(X)$, which we note is isomorphic to $H_q(X)$ since $\mathscr{U}$ is an open cover [48, Proposition 2.21]. The $E^2$ page then consists of the homology groups $H_q^{\mathscr{U}}(X)$ along a single column and zeros elsewhere. Hence the second spectral sequence *collapses* at the $E^2$ page, i.e. $E^r \cong E^2$ for all $r \geq 2$. This is due to the fact that the differentials of all subsequent pages are maps between $E^r_{p,q}$ and $E^r_{p-1+r,q-r}$, and thus must either map to or from a trivial group, hence are the zero homomorphism. Then we see that $E^\infty_{0,q} \cong E^2_{0,q} \cong H_q(X)$ for all $q \in \mathbb{N}$ and $E^\infty_{p,q} \cong E^2_{p,q} \cong 0$ if $p > 0$. Hence the homology of the total complex of $A$ is isomorphic to the homology of $X$, i.e. $H_\bullet(\operatorname{Tot} A) \cong H_\bullet(X)$.

## 3.2 Homological Nerve Theorem

We are now ready to prove Theorem 3.1. We are able to generalize the result of [68] by using the homology spectral sequence of the cover $\mathscr{U}$, rather than the cohomology spectral sequence.

*Proof of Theorem 3.1.* Given an open cover $\mathscr{U}$ of a space $X$, let $N$ be the nerve of $\mathscr{U}$, let $A$ be the double complex associated to $\mathscr{U}$, and let $E = E'$ be the first spectral sequence of the double complex $A$. The terms of the $E^1$ page of the first spectral sequence are obtained by taking the homology of $A$ with respect to $\partial''$. Hence $E^1$ is

$$\begin{array}{ccccccc} & & \vdots & & \vdots & & \vdots \\ 0 & \longleftarrow & \bigoplus_{\sigma\in N_0} H_2(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_1} H_2(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_2} H_2(U_\sigma) \longleftarrow \dots \\ 0 & \longleftarrow & \bigoplus_{\sigma\in N_0} H_1(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_1} H_1(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_2} H_1(U_\sigma) \longleftarrow \dots \\ 0 & \longleftarrow & \bigoplus_{\sigma\in N_0} H_0(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_1} H_0(U_\sigma) & \longleftarrow & \bigoplus_{\sigma\in N_2} H_0(U_\sigma) \longleftarrow \dots \\ & & 0 & & 0 & & 0, \end{array}$$

where the differentials are induced by $\partial'$. For each $m \in \mathbb{N}$, there exists a surjective map $g_m : E^1_{m,0} \to C_m(N)$ defined as follows. For each $\sigma \in N$, let $P_\sigma$ denote the set of path components of $U_\sigma$, identify $H_0(U_\sigma)$ with $\bigoplus_{i\in P_\sigma} \mathbb{Z}$, and let $f_\sigma : H_0(U_\sigma) \to \mathbb{Z}$ be the map which sends $(n_i)_{i\in P_\sigma}$ to the sum $\sum_{i\in P_\sigma} n_i$. Then for each $m \in \mathbb{N}$, let $g_m : \bigoplus_{\sigma\in N_m} H_0(U_\sigma) \to \bigoplus_{\sigma\in N_m} \mathbb{Z}$ be the map induced by the collection $\{f_\sigma : \sigma \in N_m\}$. It is not too difficult to check that the collection $\{g_m : m \in \mathbb{N}\}$ defines a morphism of chain complexes, $g : E^1_{\bullet,0} \to C_\bullet(N)$.

Under the assumption that $\widetilde{H}_j(U_\sigma) \cong 0$ for all $\sigma \in N^{(k)}$ and $j \in \{0, \dots, k - \dim\sigma\}$, we see that for all $m \leq k$, the $m$-th antidiagonal of the $E^1$ page contains only one nontrivial term, $\bigoplus_{\sigma\in N_m} H_0(U_\sigma)$. Moreover, $g_m : \bigoplus_{\sigma\in N_m} H_0(U_\sigma) \to C_m(N)$ is an isomorphism for $m \leq k$. Then from the commutative diagram

$$\begin{array}{ccccccccc} E^1_{k+2,0} & \longrightarrow & E^1_{k+1,0} & \longrightarrow & E^1_{k,0} & \longrightarrow & E^1_{k-1,0} & \longrightarrow & \dots \\ \downarrow g_{k+2} & & \downarrow g_{k+1} & & \downarrow \cong & & \downarrow \cong & & \\ C_{k+2}(N) & \longrightarrow & C_{k+1}(N) & \longrightarrow & C_k(N) & \longrightarrow & C_{k-1}(N) & \longrightarrow & \dots \end{array}$$

we immediately see that $E^2_{m,0} \cong H_m(N)$ for all $m \leq k-1$. Using the fact that $g_{k+1}$ is surjective and $g_k, g_{k-1}$ are isomorphisms, it is also straightforward to see that $g_k$ induces an isomorphism $E^2_{k,0} \cong H_k(N)$ and $g_{k+1}$ induces a surjection $E^2_{k+1,0} \to H_{k+1}(N)$. Note that for $m \leq k$, the $m$-th antidiagonal of the $E^2$ page contains only one nontrivial term, $E^2_{m,0}$, and that $E^2_{p,q} \cong E^\infty_{p,q}$ for $p+q \le k$. Hence for $0 \le m \le k$, $H_m(\operatorname{Tot} A) \cong E^2_{m,0} \cong H_m(N)$. Consequently, $H_m(X) \cong H_m(\operatorname{Tot} A) \cong H_m(N)$ for all $m \le k$. Lastly since there is a surjection $E^2_{k+1,0} \to H_{k+1}(N)$, if $H_{k+1}(N) \neq 0$, then we must also have $E^2_{k+1,0} \neq 0$. Since the differentials entering and leaving the term $E^r_{k+1,0}$ are zero homomorphisms for all $r \ge 2$, we have $E^\infty_{k+1,0} \cong E^2_{k+1,0}$, which in turn implies that $H_{k+1}(X) \cong H_{k+1}(\operatorname{Tot} A) \neq 0$. □

**Remark 3.2.** The fact that $\mathscr{U}$ is an open cover is only used for the isomorphism $H^{\mathscr{U}}_\bullet(X) \cong H_\bullet(X)$ which is used to establish $H_\bullet(X) \cong H_\bullet(\operatorname{Tot} A)$. Hence Theorem 3.1 holds slightly more generally for any space $X$ and cover $\mathscr{U}$ such that $H^{\mathscr{U}}_\bullet(X) \cong H_\bullet(X)$, for example if $\mathscr{U}$ is a collection of sets whose interiors cover $X$.

## 3.3 Vietoris Complexes and Metric Thickenings have Isomorphic Homology Groups

Having proved Theorem 3.1, we are now equipped to show that $\mathcal{V}(\mathscr{U})$ and $\mathcal{V}^m(\mathscr{U})$ have isomorphic homology groups. Before doing so however, we recall the open covers of $\mathcal{V}^m(\mathscr{U})$ constructed in [5] that are good up to level $n$. Following the terminology of [5], if $U$ is a subset of $X$ and $0 < p < 1$, we say that a subset $A \subseteq \mathcal{V}^m(\mathscr{U})$ has the *mass concentration property* for the pair $(p, U)$ if $\mu(U) > p$ for all $\mu \in A$. In this case, we write that $A$ has $\mathrm{MCP}(p, U)$. If $U \subseteq X$, a subset $A \subseteq \mathcal{P}^{\mathrm{fin}}(X)$ is *$U$-pumping convex* if for any $\mu \in A$ and $\nu \in \mathcal{P}^{\mathrm{fin}}(X)$ satisfying $\operatorname{supp}(\nu) \subseteq \operatorname{supp}(\mu) \cap U$ we have that $(1-t)\mu + t\nu \in A$ for all $t \in I$.

If $U$ is a subset of $X$, let $M_U \subseteq \mathcal{P}^{\mathrm{fin}}(X)$ be the set of all $\mu \in \mathcal{P}^{\mathrm{fin}}(X)$ such that $\operatorname{supp}(\mu) \subseteq U$. If $\mathscr{U}$ is an open cover of $X$, then $M_{\mathscr{U}} = \{M_U : U \in \mathscr{U}\}$ is a cover of $\mathcal{V}^m(\mathscr{U})$. However, $M_{\mathscr{U}}$ is not necessarily an open cover. In [5] it was shown that the cover $M_{\mathscr{U}}$ can be "thickened" with respect to a choice of $0 < p < 1$ to construct an open cover $\widetilde{M_{\mathscr{U}}}$ of $\mathcal{V}^m(\mathscr{U})$ with several key properties.

**Proposition 3.2.** *Let $\mathscr{U}$ be an open cover of $X$. For every $0 < p < 1$, there exists an open cover $\widetilde{M}_{\mathscr{U}} = \{\widetilde{M}_U\}_{U \in \mathscr{U}}$ of $\mathcal{V}^m(\mathscr{U})$ such that for every $U \in \mathscr{U}$,*

(1) $M_U \subseteq \widetilde{M}_U$,

(2) $\widetilde{M}_U$ *is $U$-pumping convex,*

(3) $\widetilde{M}_U$ *has* $\mathrm{MCP}(p, U)$.

*Proof.* See [5, pg. 8]. □

**Remark 3.3.** The open covers in [5] are defined in such a way that the covers' elements are open and satisfy properties (1)–(3) by construction. However, a simpler way to verify Proposition 3.2 is the following. For any $0 < p < 1$, let $\widetilde{M}_{\mathscr{U}}$ be the collection of sets of the form $\widetilde{M}_U = \{\mu \in \mathcal{V}^m(\mathscr{U}) : \mu(U) > p\}$ for $U \in \mathscr{U}$. Then it is straightforward to check that conditions (1)–(3) hold. Lastly, the sets $\widetilde{M}_U$ are open in $\mathcal{V}^m(\mathscr{U})$ due to the fact that for any $U \in \mathscr{U}$ and $\mu \in \widetilde{M}_U$, we have $d_X(\mathrm{supp}(\mu) \cap U, U^C) > 0$ since $\mathrm{supp}(\mu)$ is a finite set. Hence if $\nu \notin \widetilde{M}_U$, then any coupling $\gamma$ between $\mu$ and $\nu$ must transport a mass of at least $\mu(U) - p$ a distance of at least $d_X(\mathrm{supp}(\mu) \cap U, U^C) > 0$, or more precisely, $\gamma$ must satisfy

$$\gamma\Big((\mathrm{supp}(\mu) \cap U) \times U^C\Big) \geq \mu(U) - p.$$

From this it follows that $d_W(\mu, \nu) > (\mu(U) - p) d_X(\mathrm{supp}(\mu) \cap U, U^C)$, and since this bound does not depend on $\nu$, it follows that $\mu$ is an interior point of $\widetilde{M}_U$. Hence the sets $\widetilde{M}_U$ are open in $\mathcal{V}^m(\mathscr{U})$.

**Corollary 3.3.1.** *If $\mathscr{U}$ is a uniformly bounded open cover of a metric space $X$, then $\mathcal{V}^m(\mathscr{U})$ and $\mathcal{V}(\mathscr{U})$ have isomorphic homology groups.*

*Proof.* Write $\mathscr{U} = \{U_i\}_{i \in \mathcal{I}}$. Let $n \in \mathbb{N}$, let $1 - 1/n < p < 1$, and let $\widetilde{M}_{\mathscr{U}}$ be the open cover of $\mathcal{V}^m(\mathscr{U})$ from Proposition 3.2 corresponding to the value $p$. Then Proposition 4.4 of [5] ensures that the cover $\widetilde{M}_{\mathscr{U}}$ is good up to level $n$. Then $\widetilde{H}_j(\cap_{i \in \sigma} \widetilde{M}_{U_i}) \cong 0$ for all $\sigma \in N^{(n-1)}$ and $j \in \mathbb{N}$, in which case Theorem 3.1 then implies that $H_j(\mathcal{V}^m(\mathscr{U})) \cong H_j(\mathcal{N}(\widetilde{M}_{\mathscr{U}}))$ for all $j \leq n-1$. It was shown in [5] that the $n$-skeleton of $\mathcal{N}(\widetilde{M}_{\mathscr{U}})$ coincides with that of $\mathcal{N}(\mathscr{U})$, hence $H_j(\mathcal{N}(\widetilde{M}_{\mathscr{U}})) \cong H_j(\mathcal{N}(\mathscr{U})) \cong H_j(\mathcal{V}(\mathscr{U}))$ for all $j \leq n-1$, where the second isomorphism follows from Dowker duality [27]. In total, we have that $H_j(\mathcal{V}^m(\mathscr{U})) \cong H_j(\mathcal{V}(\mathscr{U}))$ for all $j \leq n-1$. Since $n$ was arbitrary, we conclude that $\mathcal{V}^m(\mathscr{U})$ and $\mathcal{V}(\mathscr{U})$ have isomorphic homology groups. □

Since the Vietoris–Rips complex $\mathrm{VR}(X;r)$ and the Čech complex $\check{\mathrm{C}}(X;r)$ are special cases of Vietoris complexes $\mathcal{V}(\mathscr{U})$, the following corollary is immediate.

**Corollary 3.3.2.** *For all $r > 0$, we have isomorphisms $H_n(\mathrm{VR}^m(X;r)) \cong H_n(\mathrm{VR}(X;r)$ and $H_n(\check{\mathrm{C}}^m(X;r)) \cong H_n(\check{\mathrm{C}}(X;r)$ for every $n \in \mathbb{N}$.*

# Chapter 4

# Vietoris Metric Thickenings and Complexes are Weakly Homotopy Equivalent

In this chapter, we further strengthen the connection between the Vietoris complex $\mathcal{V}(\mathscr{U})$ and the Vietoris metric thickening $\mathcal{V}^m(\mathscr{U})$. Our main result is to show that the natural bijection $|\mathcal{V}(\mathscr{U})| \to \mathcal{V}^m(\mathscr{U})$ is a weak homotopy equivalence.

**Theorem 4.1.** *For any uniformly bounded open cover $\mathscr{U}$ of a metric space $X$, the natural bijection* $\mathrm{id} : |\mathcal{V}(\mathscr{U})| \to \mathcal{V}^m(\mathscr{U})$ *is a weak homotopy equivalence.*

Theorem 4.1 can be seen as a significant step towards a proof that $|\mathcal{V}(\mathscr{U})|$ and $\mathcal{V}^m(\mathscr{U})$ are homotopy equivalent. Whitehead's theorem states that a weak homotopy equivalence $f : X \to Y$ is a homotopy equivalence if $X$ and $Y$ have the homotopy types of CW complexes. Thus, in light of Theorem 4.1, to show that $|\mathcal{V}(\mathscr{U})|$ and $\mathcal{V}^m(\mathscr{U})$ are homotopy equivalent, it suffices to show that $\mathcal{V}^m(\mathscr{U})$ has the homotopy type of a CW complex.

The structure of this chapter is as follows. In Section 4.1, we discuss $|\mathcal{V}(\mathscr{U})|_m$, the Vietoris complex with the metric topology for simplicial complexes and its relationship with $\mathcal{V}^m(\mathscr{U})$. In Section 4.2, we prove a technical lemma about the existence of homotopies for certain maps of simplices $\Delta^n \to \mathcal{V}^m(\mathscr{U})$. In Section 4.3, we recall the Freudenthal–Kuhn triangulation of $\mathbb{R}^n$, which we use to construct a class of triangulations of $I^n$. In Section 4.4 we prove Theorem 4.1 by showing that if $f : I^n \to \mathcal{V}^m(\mathscr{U})$ is a map such that the restriction $f|_{\partial I^n}$ is continuous as a map $f|_{\partial I^n} : \partial I^n \to |\mathcal{V}(\mathscr{U})|_m$, then $f$ can be homotoped rel. $\partial I^n$ to a continuous map $g : I^n \to |\mathcal{V}(\mathscr{U})|_m$. The main step for constructing such a homotopy involves triangulating $I^n$ using the Freudenthal-Kuhn triangulation, which allows us to then iteratively apply the homotopy of Section 4.2. Finally, in Section 4.5, we explore some of the novel consequences of Theorem 4.1.

## 4.1 The Vietoris Complex with the Metric Topology

Let $\mathscr{U}$ be an open cover of a metric space $X$. Though we can equip $|\mathcal{V}(\mathscr{U})|$ with the metric topology for simplicial complexes to obtain the space $|\mathcal{V}(\mathscr{U})|_m$, note that this space is distinct from $\mathcal{V}^m(\mathscr{U})$. For one, the vertex set $V \subseteq |\mathcal{V}(\mathscr{U})|_m$ is discrete, whereas the same set with the subspace topology from $\mathcal{V}^m(\mathscr{U})$ is homeomorphic to $X$.

Our reason for considering $|\mathcal{V}(\mathscr{U})|_m$ is the following. Because $\mathrm{id} : |K| \to |K|_m$ is a homotopy equivalence for any simplicial complex $K$, we will prove Theorem 4.1 by showing that $\mathrm{id} : |\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ is a weak homotopy equivalence. We make this replacement because the continuity of maps $f : Z \to |\mathcal{V}(\mathscr{U})|_m$ affords a convenient characterization via Lemma 2.0.1.

First, we show that the bijection $|\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ is continuous so long as $\mathscr{U}$ is uniformly bounded.

**Proposition 4.1.** *If $\mathscr{U}$ is a uniformly bounded open cover of $X$, then the identity map $|\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ is continuous.*

*Proof.* Since $\mathscr{U}$ is uniformly bounded, let $D = \sup_{U \in \mathscr{U}} \mathrm{diam}(U)$. Let $\mu \in \mathcal{V}^m(\mathscr{U})$ and let $0 < r < 2D$. Let $B = B_{\mathcal{V}^m(\mathscr{U})}(\mu, r)$ be the open ball in $\mathcal{V}^m(\mathscr{U})$ centered at $\mu$ with radius $r$, and let $B' = B_{|\mathcal{V}(\mathscr{U})|_m}(\mu, r/D)$ be the open ball in $|\mathcal{V}(\mathscr{U})|_m$

centered at $\mu$ of radius $r/D$. We claim that $B' \subseteq B$, which will show that the identity $|\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ is continuous. So let $\nu \in B'$. Then

$$\sum_{x\in X} |\psi_x(\mu) - \psi_x(\nu)| = \sum_{x\in X} |\mu(x) - \nu(x)| < \frac{r}{D}. \tag{4.1}$$

Note that if $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\nu)$ were disjoint, we would have $\sum_{x\in X} |\mu(x)-\nu(x)| = 2$. Since $r/D < 2$, it follows that $\mathrm{supp}(\mu) \cap \mathrm{supp}(\nu) \neq \emptyset$.

Let $\gamma$ be any coupling between $\mu$ and $\nu$ which satisfies

$$\gamma(x, x) = \min\big(\mu(x), \nu(x)\big)$$

for all $x \in X$. In other words, $\gamma$ is a transport plan between $\mu$ and $\nu$ which keeps fixed any mass they already have in common. Note that (4.1) implies that $\gamma$ has at most $r/(2D)$ of its mass off the diagonal of $X \times X$. Hence

$$\int_{X\times X} d_X(x, y)\gamma(dx \times dy) < \mathrm{diam}\Big(\mathrm{supp}(\mu) \cup \mathrm{supp}(\nu)\Big) \cdot \frac{r}{2D}.$$

Finally, we note that the diameters of $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\nu)$ are each less than $D$, and since $\mathrm{supp}(\mu) \cap \mathrm{supp}(\nu) \neq \emptyset$, it follows that $\mathrm{diam}(\mathrm{supp}(\mu) \cup \mathrm{supp}(\nu)) < 2D$. Thus

$$d_W(\mu, \nu) \le \int_{X\times X} d_X(x, y)\gamma(dx \times dy) < r,$$

which shows that $B' \subseteq B$. □

We recall a lemma which implies that if $f, g : Z \to \mathcal{V}^m(\mathscr{U})$ are continuous, then whenever the linear homotopy $H : Z \times I \to \mathcal{V}^m(\mathscr{U})$ between $f$ and $g$ is well-defined, it is also continuous.

**Lemma 4.1.1.** *[5, Lemma 2.1] Let $X$ and $Z$ be metric spaces, and let $f, g : Z \to \mathcal{P}^{\mathrm{fin}}(X)$ be continuous maps. Then the linear homotopy $H : Z \times I \to \mathcal{P}^{\mathrm{fin}}(X)$ given by $H(z, t) = (1 - t)f(z) + tg(z)$ is continuous.*

We now show that the linear homotopy of Lemma 4.1.1 is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$ under similar conditions.

**Lemma 4.1.2.** *Let $X$ and $Z$ be metric spaces, let $\mathscr{U}$ be an open cover of $X$, let $f, g : Z \to \mathcal{V}^m(\mathscr{U})$ be functions, and suppose the linear homotopy $H : Z \times I \to \mathcal{V}^m(\mathscr{U})$ given by $H(z,t) = (1-t)f(z) + tg(z)$ is well-defined. If both $f$ and $g$ are continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$, then so is $H$.*

*Proof.* If both $f$ and $g$ are continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$, then $\psi_v \circ f$ and $\psi_v \circ g$ are continuous for every vertex $v \in X$ of $\mathcal{V}(\mathscr{U})$. Note that

$$\begin{aligned}(\psi_v \circ H)(z,t) &= \psi_v\Big((1-t)f(z) + tg(z)\Big)\\ &= (1-t)(\psi_v \circ f)(z) + t(\psi_v \circ g)(z).\end{aligned}$$

Hence $\psi_v \circ H$ is continuous for each vertex $v$, in which case $H$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$ by Lemma 2.0.1. □

We will also recall the following lemma, which describes how under certain conditions a subset $A$ with $\mathrm{MCP}(p,U)$ can be mapped continuously into $M_U = \{\mu \in \mathcal{V}^m(\mathscr{U}) : \mathrm{supp}(\mu) \subseteq U\}$ (recall the notation from Section 3.3), and then show in the lemma directly after how this mapping is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$.

**Lemma 4.1.3.** *[5, Lemma 4.3] Let $X$ be a metric space, and let the open set $U \subseteq X$ have finite diameter. Suppose $A \subseteq \mathcal{P}^{\mathrm{fin}}(X)$ has $\mathrm{MCP}(p,U)$ for some $0 < p < 1$, and let $\phi : X \to [0,1]$ be an $L$-Lipschitz map with $\phi^{-1}(0) = U^C$. Then $f : A \to M_U \subseteq \mathcal{P}^{\mathrm{fin}}(X)$ defined as*

$$f\Big(\sum_{i\in\mathcal{I}} a_i\delta_{x_i}\Big) = \sum_{i\in\mathcal{I}} \Big(\frac{a_i\phi(x_i)}{\sum_{j\in\mathcal{I}} a_j\phi(x_j)}\Big)\delta_{x_i}$$

*is continuous. If $A$ is furthermore $U$-pumping convex, then the homotopy $H : A \times [0,1] \to A$ defined by $H(\mu,t) = (1-t)\mu + tf(\mu)$ is well-defined and continuous.*

**Lemma 4.1.4.** *Suppose $A \subseteq \mathcal{V}^m(\mathscr{U})$ has* $\mathrm{MCP}(p, U)$ *for some $p > 0$ and open set $U \subseteq X$. Let $\phi : X \to I$ be a continuous map satisfying $\phi^{-1}(0) = U^C$. The map $\Phi : A \to M_U$ defined by*

$$\Phi\Big(\sum_{i\in\mathcal{I}} a_i\delta_{x_i}\Big) = \sum_{i\in\mathcal{I}} \Big(\frac{a_i\phi(x_i)}{\sum_{j\in\mathcal{I}} a_j\phi(x_j)}\Big)\delta_{x_i}.$$

*is continuous with respect to the topologies of $A$ and $M_U$ regarded as subspaces of $|\mathcal{V}(\mathscr{U})|_m$.*

*Proof.* Let $A$ and $M_U$ have the subspace topology with respect to $|\mathcal{V}(\mathscr{U})|_m$. Note that a point $\mu = \sum_i a_i\delta_{x_i} \in \mathcal{V}^m(\mathscr{U})$ can alternatively be written $\mu = \sum_{x\in X} \psi_x(\mu)\delta_x$. Then if $v \in \mathcal{V}(\mathscr{U})$ is a vertex, and $\mu = \sum_{x\in X} \psi_x(\mu)\delta_x \in A$, we have

$$\begin{aligned}\psi_v\Big(\Phi\Big(\sum_{x\in X} \psi_x(\mu)\delta_x\Big)\Big) &= \psi_v\Big(\sum_{x\in X}\Big(\frac{\psi_x(\mu)\phi(x)}{\sum_{y\in X}\psi_y(\mu)\phi(y)}\Big)\delta_x\Big)\\ &= \frac{\psi_v(\mu)\phi(v)}{\sum_{y\in X}\psi_y(\mu)\phi(y)}.\end{aligned} \tag{4.2}$$

The fact that $A$ has $\mathrm{MCP}(p, U)$ implies that there exists $y \in U$ such that $\psi_y(\mu) > 0$. Then since $\phi(y) > 0$ for all $y \in U$, we have that

$$\sum_{y\in X} \psi_y(\mu)\phi(y) > 0$$

for all $\mu \in A$. Hence $\psi_v \circ \Phi$ is well-defined. Additionally, the function $A \to I$ given by $\mu \mapsto \sum_{y\in X} \psi_y(\mu)\phi(y)$ is continuous due to the fact that

$$\Big|\sum_{y\in X}\psi_y(\mu)\phi(y) - \sum_{y\in X}\psi_y(\nu)\phi(y)\Big| \le \sum_{y\in X}\phi(y)\big|\psi_y(\mu) - \psi_y(\nu)\big| \le d_m(\mu,\nu),$$

where $d_m$ is the metric on $|\mathcal{V}(\mathscr{U})|_m$. Since $\psi_v : A \to I$ is continuous, (4.2) implies that $\psi_v \circ \Phi$ is continuous. Since $v$ was arbitrary, Lemma 2.0.1 implies that $\Phi : A \to M_U$ is continuous with respect to the topologies of $A$ and $M_U$ as subspaces of $|\mathcal{V}(\mathscr{U})|_m$. ☐

## 4.2 Deforming Maps of Simplices into Vietoris Metric Thickenings

To show that $|\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ is a weak homotopy equivalence, our goal is to show that given a map $f : I^n \to \mathcal{V}^m(\mathscr{U})$ such that $f|_{\partial I^n}$ is continuous as a map $f|_{\partial I^n} : \partial I^n \to |\mathcal{V}(\mathscr{U})|_m$, there exists a homotopy $H : I^n \times I \to \mathcal{V}^m(\mathscr{U})$ rel. $\partial I^n$ between $f$ and a continuous map $g : I^n \to |\mathcal{V}(\mathscr{U})|_m$. To do this, we will construct a homotopy inductively on the $k$-skeleton of a triangulation of $I^n$. In this section, we establish a lemma which will serve as an important tool in constructing this homotopy.

First, we start with a lemma concerning the mass concentration property for compact sets.

**Lemma 4.1.5.** *Suppose $A \subseteq \mathcal{P}_X^{\text{fin}}$ is compact and has* $\mathrm{MCP}(p, U)$ *for some* $0 < p < 1$ *and open set $U \subseteq X$. Then there exists an open set $V \subseteq U$ such that $d(V, U^C) > 0$ and $A$ has* $\mathrm{MCP}(p, V)$.

*Proof.* For each $i \in \mathbb{N}$, let $V_i = \{x \in U : d(x, U^C) > 1/i\}$ and let $A_i = \{\mu \in A : \mu(V_i) > p\}$. Note that for each $i \in \mathbb{N}$, $V_i$ is open in $X$. We first show that $A_i$ is open in $A$ for all $i \in \mathbb{N}$. Let $i \in \mathbb{N}$, let $\mu \in A_i$, and choose $r > 0$ satisfying

$$0 < r < (\mu(V_i) - p) \cdot d\Big(\operatorname{supp}(\mu) \cap V_i, V_i^C\Big).$$

This is possible since $\operatorname{supp}(\mu) \cap V_i$ is a finite set disjoint from the closed set $(V_i)^C$, hence $d(\operatorname{supp}(\mu) \cap V_i, V_i^C) > 0$. If $\nu \in A \setminus A_i$, then $\nu(V_i) \leq p$. In this case, we have that $d_W(\mu, \nu) > r$ as any coupling $\gamma$ between $\mu$ and $\nu$ must move a mass of at least $\mu(V_i) - p$ a distance of at least $d(\operatorname{supp}(\mu) \cap V_i, V_i^C)$, or more precisely, $\gamma$ must satisfy

$$\gamma\Big((\operatorname{supp}(\mu) \cap V_i) \times V_i^C\Big) \geq \mu(V_i) - p.$$

Hence $B(\mu; r) \cap A \subseteq A_i$, showing that $A_i$ is open in $A$.

Next we show that $\{A_i\}_{i\in\mathbb{N}}$ is an open cover of $A$. So let $\mu \in A$. Because $d(\mathrm{supp}(\mu) \cap U, U^C) > 0$, if $i \in \mathbb{N}$ is chosen large enough so that $1/i < d(\mathrm{supp}(\mu) \cap U, U^C)$, we have that $\mathrm{supp}(\mu) \cap U \subseteq V_i$ and thus $\mu(V_i) = \mu(U) > p$. Thus $\{A_i\}_{i\in\mathbb{N}}$ is an open cover of $A$. Since $A$ is compact, we may find a finite subcover $\{A_{i_1}, A_{i_2}, \dots, A_{i_N}\}$. We may assume that $i_1 < i_2 < \cdots < i_N$, in which case $A_{i_1} \subseteq A_{i_2} \subseteq \cdots \subseteq A_{i_N}$, and thus $A = A_{i_N}$. This implies that for all $\mu \in A$, $\mu(V_{i_N}) > p$, that is, $A$ has MCP$(p, V_{i_N})$. By construction, $V_{i_N}$ satisfies $d(V_{i_N}, U^C) > 0$. $\square$

We are now ready to state and prove the main lemma of this section. Its statement and proof are inspired by [5, Prop. 4.4], though with alterations specific to our approach.

**Lemma 4.1.6.** *Let $\mathscr{U}$ be a uniformly bounded open cover of a metric space $X$. Let $N \in \mathbb{N}$ with $N \geq 2$, and let $\widetilde{M}_{\mathscr{U}}$ be the open cover of $\mathcal{V}^m(\mathscr{U})$ constructed with respect to a choice of $p$ satisfying $1 - 1/N < p < 1$. Let $U_1, \dots, U_n$ be a collection of elements of $\mathscr{U}$ with $n \leq N$, and set $U = \bigcap_{i\leq n} U_i$. Suppose $V \subseteq U$ is an open set which satisfies $d(V, U^C) > 0$. Then given a map*

$$g : (\Delta^k, \partial\Delta^k) \to (\bigcap_{i\leq n} \widetilde{M}_{U_i}, M_V),$$

*there exists a homotopy $H : \Delta^k \times I \to \bigcap_{i\leq n} \widetilde{M}_{U_i}$ rel. $\partial\Delta^k$ between $g$ and a map $g' : \Delta^k \to M_{V'}$, where $V' \subseteq U$ is an open set which satisfies $d(V', U^C) > 0$. Moreover, if $g$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$, then we may assume $H$ is as well.*

*Proof.* Since $1 - 1/N < p < 1$, we have that $0 < N(1 - p) < 1$. Since each $\widetilde{M}_{U_i}$ has MCP$(p, U_i)$, if $\mu \in \widetilde{M}_{U_i}$, then $\mu(U_i^C) < 1 - p$. Now if $\mu \in \bigcap_{i\leq n} \widetilde{M}_{U_i}$, it follows that

$$\mu(\bigcup_{i\leq n} U_i^C) < n(1 - p) \leq N(1 - p) < 1.$$

This implies that $\mu(\bigcap_{i\le n} U_i) > 1-N(1-p) > 0$, showing that $\bigcap_{i\le n} \widetilde{M}_{U_i}$ has MCP$(q,U)$ for $q = 1-N(1-p)$. Let $A = g(\Delta^k) \subseteq \bigcap_{i\le n} \widetilde{M}_{U_i}$. Then $A$ is compact and has MCP$(q,U)$, hence by Lemma 4.1.5, there exists an open set $U' \subseteq U$ such that $d(U', U^C) > 0$ and $A$ has MCP$(q,U')$. Let $\epsilon = d(V, U^C)/2$ and let $V' = U' \cup \bigcup_{x\in V} B(x;\epsilon)$. Then $A$ has MCP$(q,V')$, $d(V,(V')^C) > 0$, and $d(V,U^C) > 0$. Then we may find an $L$-Lipschitz function $\phi : X \to I$ for some $L > 0$ such that $V \subseteq \phi^{-1}(1)$ and $(V')^C = \phi^{-1}(0)$. Lemma 4.1.3 implies that the map $\Phi : A \to M_{V'}$ defined by

$$\Phi\Big(\sum_i a_i \delta_{x_i}\Big) = \sum_i \Big(\frac{a_i \phi(x_i)}{\sum_j a_j \phi(x_j)}\Big)\delta_{x_i}$$

is continuous. Set $g' = \Phi \circ g$. Define $H : \Delta^k \times I \to \bigcap_{i\le n} \widetilde{M}_{U_i}$ by $H(v,t) = (1-t)g(v) + tg'(v)$. Note that since $\widetilde{M}_{U_i}$ is $U_i$-pumping convex for every $i \le n$, $\bigcap_{i\le n} \widetilde{M}_{U_i}$ is $U$-pumping convex. Then $H$ is well-defined since $\bigcap_{i\le n} \widetilde{M}_{U_i}$ is $U$-pumping convex and $\mathrm{supp}(g'(v)) \subseteq \mathrm{supp}(g(v)) \cap U$. Consequently, $H$ is continuous by Lemma 4.1.1.

Hence $g'$ has image in $M_{V'}$. Moreover, since $V \subseteq \phi^{-1}(1)$, we have that $\Phi(\mu) = \mu$ for all $\mu \in A \cap M_V$, hence $g'|_{\partial\Delta^k} = g|_{\partial\Delta^k}$. Then by the definition of $H$, we see that $H(z,t) = g(z)$ for all $z \in \partial\Delta^k$ and $t \in I$.

Lastly, if $g$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$, then so is $g'$ by Lemma 4.1.4. Then from Lemma 4.1.2, we also have that $H$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$. □

## 4.3 The Freudenthal–Kuhn Triangulation

**Definition 4.1** (Freudenthal–Kuhn triangulation of $\mathbb{R}^n$)**.** *For $i \in \{1,2,\dots,n\}$, let $e_i$ denote the canonical $i$-th basis vector of $\mathbb{R}^n$. Let $\pi \in S_n$ be a permutation of the set $\{1,2,\dots,n\}$. For $x \in \mathbb{Z}^n \subseteq \mathbb{R}^n$, let $\sigma(x,\pi)$ denote the convex hull of the points $v_0, v_1, \dots, v_n$ which are given by the equations $v_0 = x$ and $v_i = v_{i-1} + e_{\pi(i)}$. The collection $\{\sigma(x,\pi) : x \in \mathbb{Z}^n, \pi \in S_n\}$ defines a triangulation of $\mathbb{R}^n$ [79, Lemma 3.2],* *sometimes known as the Freudenthal–Kuhn triangulation of $\mathbb{R}^n$ due to its use by H. Freudenthal [35] as well as H. Kuhn [56].*

Less formally, the Freudenthal–Kuhn triangulation is constructed by first subdividing $\mathbb{R}^n$ into unit $n$-cubes of the form

$$Q_x = [x_1, x_1 + 1] \times \cdots \times [x_n, x_n + 1]$$

for $x = (x_1, \ldots, x_n) \in \mathbb{Z}^n$. Each $n$-cube $Q_x$ is then triangulated by $n!$ many $n$-simplices in a canonical way so that the triangulations of each $Q_x$ fit together to define a triangulation of $\mathbb{R}^n$.

**Proposition 4.2.** *Let $n \in \mathbb{N}$. There exists $\alpha(n) \in \mathbb{N}$ such that for every $\epsilon > 0$ there exists a triangulation $T : |K| \to I^n$ which satisfies*

(1) $\operatorname{diam} T(|\sigma|) \leq \epsilon$ *for all $\sigma \in K$,*

(2) *every vertex $v \in K_0$ is contained in at most $\alpha(n)$ distinct $n$-simplices of $K$.*

*Proof.* Set $\alpha(n) = 2^n \cdot n!$. Let $\epsilon > 0$ and choose $p \in \mathbb{N}$ such that $\sqrt{n}/p < \epsilon$. Let $T : |K| \to \mathbb{R}^n$ be the Freudenthal–Kuhn triangulation of $\mathbb{R}^n$. Note that each $x \in \mathbb{Z}^n$ is contained in $2^n$ distinct cubes $Q_x$. Moreover, for each $x \in \mathbb{Z}^n$, the triangulation $T$ identifies $Q_x$ with a subcomplex of $K$ consisting of $n!$ many $n$-simplices. Since the vertex set of $K$ is identified with $\mathbb{Z}^n$ under $T$, each vertex of $K$ is contained in at most $\alpha(n) = 2^n \cdot n!$ distinct $n$-simplices of $K$. Observe that the diameter of each simplex in the triangulation is bounded by the diameter of a unit $n$-cube, which has diameter $\sqrt{n}$. Let $L \subseteq K$ be the subcomplex given by $|L| = T^{-1}([0, p]^n)$. If we compose the restriction of $T$ to $|L|$ with the linear map $[0, p]^n \to I^n$ which scales each coordinate by a factor of $1/p$, we obtain a triangulation $T' : |L| \to I^n$ such that $\operatorname{diam} T'(|\sigma|) \leq \sqrt{n}/p < \epsilon$ for all $\sigma \in L$ and such that each vertex of $L$ is contained in at most $\alpha(n) = 2^n \cdot n!$ distinct $n$-simplices of $L$. □

**Remark 4.2.** Barycentric subdivisions can be used to produce triangulations of $I^n$ which satisfy condition (1) of Proposition 4.2 for arbitrarily small $\epsilon$. However, since the upper bound on the number of distinct $n$-simplices which intersect at a point would depend on the number of barycentric subdivisions performed, this approach would not allow one to produce a value $\alpha(n)$ which is independent of $\epsilon$ and satisfies condition (2).

## 4.4 Vietoris Metric Thickenings and Complexes are Weakly Homotopy Equivalent

In this section, $\mathscr{U}$ will always denote a uniformly bounded open cover of a metric space $X$, so that $\mathrm{id} : |\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ is continuous by Proposition 4.1. To show that $\mathrm{id} : |\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ is a weak homotopy equivalence, we use the following characterization of weak homotopy equivalences.

**Lemma 4.2.1.** *A map $f : X \to Y$ between spaces $X$ and $Y$ is a weak homotopy equivalence if and only if $f$ satisfies the condition that for every commuting diagram of the form*

$$\begin{array}{ccc} \partial I^n & \xrightarrow{\alpha} & X \\ \big\downarrow{\scriptstyle i} & & \big\downarrow{\scriptstyle f} \\ I^n & \xrightarrow{\beta} & Y \end{array}$$

*there exists a map $\phi : I^n \to X$ such that $\alpha = \phi \circ i$ and $\beta \simeq f \circ \phi$ rel. $\partial I^n$.*

The statement, in a slightly more general form, can be found in [18, Theorem 11.12]. The proof amounts to first noting that $f : X \to Y$ is a weak homotopy equivalence if and only if the inclusion $i : X \to M_f$ is a weak homotopy equivalence, where $M_f$ denotes the mapping cylinder of $f$. Then since the above condition on $f$ is equivalent to the condition that $\pi_n(M_f, X) = 0$ for all $n \geq 1$ and $\pi_0(X) \to \pi_0(M_f)$ is onto, the lemma follows.

If $L \subseteq K$ is a subcomplex of the simplicial complex $K$, then $|K| \times \{0\} \cup |L| \times I$ is a retract of $|K| \times I$. The following lemma is a slightly sharper statement of this fact, where the condition placed on the retraction follows directly from standard proofs, for example [48, Proposition 0.16].

**Lemma 4.2.2.** *Let $K$ be a simplicial complex and let $L \subseteq K$ be a subcomplex. Let $i : |K| \times \{0\} \cup |L| \times I \to |K| \times I$ denote the inclusion. There exists a retraction $r : |K| \times I \to |K| \times \{0\} \cup |L| \times I$ such that $i \circ r(|\sigma| \times I) \subseteq |\sigma| \times I$ for all $\sigma \in K$.*

**Definition 4.2.** *If $K$ is a simplicial complex and $\mathscr{A}$ is a cover of $|K|$, we say that $K$ is subordinate to $\mathscr{A}$ if for every simplex $\sigma \in K$, $|\sigma|$ is contained in some element of $\mathscr{A}$. Additionally, given a cover $\mathscr{A}$ of a space $X$, we say that a triangulation $T : |K| \to X$ is subordinate to $\mathscr{A}$ if $T(|\sigma|)$ is contained in some element of $\mathscr{A}$ for every simplex $\sigma \in K$.*

To prove that the bijection $|\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ satisfies the condition of Lemma 4.2.1, we will show that for any map $f : I^n \to \mathcal{V}^m(\mathscr{U})$ such that $f|_{\partial I^n}$ is continuous as a map $f|_{\partial I^n} : \partial I^n \to |\mathcal{V}(\mathscr{U})|_m$, there exists a (free) homotopy $H : I^n \times I \to \mathcal{V}^m(\mathscr{U})$ between $f$ and a map $f' : I^n \to |\mathcal{V}(\mathscr{U})|_m$, and where $H$ satisfies the condition that $H|_{\partial I^n \times I} : \partial I^n \times I \to |\mathcal{V}(\mathscr{U})|_m$ is continuous. Then, as we will describe in the proof of Lemma 4.2.5, $H$ will induce a homotopy rel. $\partial I^n$ between $f$ and a map $f'' : I^n \to |\mathcal{V}(\mathscr{U})|_m$.

The construction of the homotopy $H$ consists of two main steps. For the first step, we homotope $f$ to a map $g : I^n \to \mathcal{V}^m(\mathscr{U})$ such that there exists a triangulation of $I^n$ subordinate to $g^{-1}M_{\mathscr{U}}$. This would be unnecessary if $M_{\mathscr{U}}$ were an open cover of $\mathcal{V}^m(\mathscr{U})$, however, we remind the reader that this is not the case in general. Thus we instead start with a triangulation of $I^n$ subordinate to $f^{-1}\widetilde{M}_{\mathscr{U}}$ (since $\widetilde{M}_{\mathscr{U}}$ *is* an open cover of $\mathcal{V}^m(\mathscr{U})$) and carefully deform $f$ to a map $g$ which carries each simplex of the triangulation into an element of $M_{\mathscr{U}}$.

**Lemma 4.2.3.** *Let $n \in \mathbb{N}$ and let $f : I^n \to \mathcal{V}^m(\mathscr{U})$ be a map such that $f|_{\partial I^n}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$. There exists a homotopy $H : I^n \times I \to \mathcal{V}^m(\mathscr{U})$ between $f$ and a map $g$ such that*

(1) *$H|_{\partial I^n \times I}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$,*

(2) *$I^n$ admits a triangulation subordinate to $g^{-1}M_{\mathscr{U}}$.*

*Proof.* Let $\alpha(n)$ be the integer guaranteed to exist by Proposition 4.2 and let $1 - 1/\alpha(n) < p < 1$. Let $\widetilde{M}_{\mathscr{U}}$ be the open cover of $\mathcal{V}^m(\mathscr{U})$ from Proposition 3.2 obtained with respect to the value $p$. By the Lebesgue number lemma, there exists $\epsilon > 0$ such that if $A$ is a subset of $I^n$ with $\operatorname{diam} A < \epsilon$, then $f(A)$ is contained in some element of $\widetilde{M}_{\mathscr{U}}$. By the construction of the integer $\alpha(n)$, there exists a triangulation $T : |K| \to I^n$ such $\operatorname{diam} T(|\sigma|) < \epsilon$ for all $\sigma \in K$, hence $T$ is subordinate to $f^{-1}\widetilde{M}_{\mathscr{U}}$, and such that each vertex of $K$ is contained in at most $\alpha(n)$ many $n$-simplices of $K$. Note that this implies that each simplex of $K$ is contained in at most $\alpha(n)$ many $n$-simplices of $K$. For convenience, we identify $I^n$ with $|K|$ so that $f$ is a map $f : |K| \to \mathcal{V}^m(\mathscr{U})$.

Recall that $K_n$ is the set of all $n$-simplices of $K$, and fix a function $\ell : K_n \to \mathscr{U}$ which assigns to each $n$-simplex $\sigma$ an open set $U \in \mathscr{U}$ such that $f(\sigma) \subseteq \widetilde{M}_U$. For each simplex $\sigma \in K$, let $\mathcal{S}_\sigma$ be the set of $n$-simplices in $K$ which contain $\sigma$. By construction, $\operatorname{card}(\mathcal{S}_\sigma) \leq \alpha(n)$ for all $\sigma \in K$ with $\sigma \neq \emptyset$.

We produce a homotopy $H$ between $f$ and a map $g$ by induction on the $k$-skeleton of $K$. So let $k < n$ and suppose there exists a homotopy $H_k$ between $f$ and a map $f_k : |K| \to \mathcal{V}^m(\mathscr{U})$ such that

(a) $H_k|_{\partial I^n \times I}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$,

(b) for every $\sigma \in K$, $f_k(|\sigma|) \subseteq \bigcap_{U \in \ell(\mathcal{S}_\sigma)} \widetilde{M}_U$,

(c) $f_k$ satisfies the condition that for each $\sigma \in K^{(k)}$, if $U_\sigma$ denotes the open set $U_\sigma = \bigcap \ell(\mathcal{S}_\sigma)$, there exists an open set $V_\sigma \subseteq U_\sigma$ such that $d(V_\sigma, U_\sigma^C) > 0$ and $f_k(|\sigma|) \subseteq M_{V_\sigma}$.

Note that condition (c) states that the $k$-skeleton of $K$ is subordinate to the cover $f_k^{-1}M_{\mathscr{U}}$, but in a slightly stronger form which allows Lemma 4.1.6 to be applied to the $(k+1)$-simplices of $K$ in the induction step. We also remark that the sets $U_\sigma$ are non-empty for all $\sigma \neq \emptyset$. This is a consequence of the fact that for $\sigma \neq \emptyset$, $\bigcap_{U\in\ell(\mathcal{S}_\sigma)} \widetilde{M}_U$ is non-empty and each $\widetilde{M}_U$ has MCP$(p,U)$ for $p > 1 - 1/\alpha(n)$, hence (recalling the beginning of the proof of Lemma 4.1.6) it follows that $\bigcap_{U\in\ell(\mathcal{S}_\sigma)} \widetilde{M}_U$ has MCP$(q, U_\sigma)$ for some $q > 0$.

To begin the induction step, let $\sigma \in K_{k+1}$ and let $V_{\partial\sigma}$ denote the open set $\bigcup_{\tau\in\partial\sigma} V_\tau$ where each $V_\tau$ is the open set guaranteed to exist by (c) for the $k$-simplex $\tau \in \partial\sigma$. Then $d(V_{\partial\sigma}, U_\sigma^C) > 0$. Since $\widetilde{M}_{\mathscr{U}}$ was constructed with respect to the value $p > 1 - 1/\alpha(n)$ and $\ell(\mathcal{S}_\sigma)$ is a subset of $\mathscr{U}$ with at most $\alpha(n)$ elements, we may apply Lemma 4.1.6 to the map

$$f_k|_{|\sigma|} : (\Delta^{k+1}, \partial\Delta^{k+1}) \to \Big( \bigcap_{U\in\ell(\mathcal{S}_\sigma)} \widetilde{M}_U, M_{V_{\partial\sigma}} \Big).$$

Hence $f_k|_{|\sigma|}$ is homotopic rel. $\partial\Delta^{k+1}$ by a homotopy $H'_\sigma : \Delta^{k+1} \times I \to \bigcap_{U\in\ell(\mathcal{S}_\sigma)} \widetilde{M}_U$ to a map $f'_\sigma$ with image in $M_{V_\sigma}$ where $V_\sigma \subseteq U_\sigma$ satisfies $d(V_\sigma, U_\sigma) > 0$. Moreover, if $\sigma \subseteq \partial I^n$, then since $f_k|_{|\sigma|}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$, we may assume $H'_\sigma$ is as well. Note that condition (b) of the induction hypothesis is needed to ensure that $f_k|_{|\sigma|}$ has codomain $\bigcap_{U\in\ell(\mathcal{S}_\sigma)} \widetilde{M}_U$ so that Lemma 4.1.6 can be applied. Let $f' : |K^{(k+1)}| \to \mathcal{V}^m(\mathscr{U})$ be the map defined by $f'|_{|\sigma|} = f'_\sigma$ for all $\sigma \in K_{k+1}$, and let $H'$ be the homotopy between $f_k|_{|K^{(k+1)}|}$ and $f'$ given by $H'|_{|\sigma|\times I} = H'_\sigma$. By Lemma 4.2.2 there exists a retraction $r : |K| \times I \to |K| \times \{0\} \cup |K^{(k+1)}| \times I$ such that $i \circ r(|\sigma| \times I) \subseteq |\sigma| \times I$ for all $\sigma \in K$. Let $f_k \cup H'$ denote the map $|K| \times \{0\} \cup |K^{(k+1)}| \times I \to \mathcal{V}^m(\mathscr{U})$ such that $f_k \cup H'|_{|K|\times\{0\}} = f_k$ and $f_k \cup H'|_{|K^{(k+1)}|\times I} = H'$. Then $G = (f_k \cup H') \circ r : |K| \times I \to \mathcal{V}^m(\mathscr{U})$ defines a homotopy between $f_k$ and the map $f_{k+1}$ defined by $f_{k+1} = G(\cdot, 1)$. Let $H_{k+1} = H_k \cdot G$ be the concatenation of $H_k$ and $G$. Then $H_{k+1}$ is a homotopy between $f$ and $f_{k+1}$ and we now verify that

(a) $H_{k+1}|_{\partial I^n \times I}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$,

(b) for every $\sigma \in K$, $f_{k+1}(|\sigma|) \subseteq \bigcap_{U \in \ell(\mathcal{S}_\sigma)} \widetilde{M}_U$,

(c) $f_{k+1}$ satisfies the condition that for each $\sigma \in K^{(k+1)}$, if $U_\sigma$ denotes the open set $U_\sigma = \bigcap \ell(\mathcal{S}_\sigma)$, there exists an open set $V_\sigma \subseteq U_\sigma$ such that $d(V_\sigma, U_\sigma^C) > 0$ and $f_{k+1}(|\sigma|) \subseteq M_{V_\sigma}$.

First, recall that $f_k|_{|\sigma|}$ is continuous relative to $|\mathcal{V}(\mathscr{U})|_m$ for all $\sigma \in K$ such that $|\sigma| \subseteq \partial I^n$, and $H'_\sigma$ is continuous relative to $|\mathcal{V}(\mathscr{U})|_m$ for all $k+1$ simplices $\sigma$ such that $|\sigma| \subseteq \partial I^n$. Then $f_k \cup H'$ restricted to the set

$$A = \partial I^n \times I \cap \Big( |K| \times \{0\} \cup |K^{(k+1)}| \times I \Big)$$

is continuous relative to $|\mathcal{V}(\mathscr{U})|_m$. Since $r(\partial I^n \times I) \subseteq A$ holds due to the fact that $i \circ r(|\sigma| \times I) \subseteq |\sigma| \times I$ for all $\sigma \in K$, we see that $G|_{\partial I^n \times I}$ is continuous relative to $|\mathcal{V}(\mathscr{U})|_m$. This implies (a).

To see that (b) holds, let $\sigma \in K$. If $\sigma \in K^{(k)}$, then $f_{k+1}|_{|\sigma|} = f_k|_{|\sigma|}$ and we are done by the induction hypothesis. If $\sigma \in K^{(k+1)}$, then

$$f_{k+1}(|\sigma|) \subseteq H'_\sigma(|\sigma| \times I) \subseteq \bigcap_{U \in \ell(\mathcal{S}_\sigma)} \widetilde{M}_U.$$

Lastly, suppose $\sigma \in K \setminus K^{(k+1)}$. Then $f_{k+1}(|\sigma|) \subseteq G(|\sigma| \times I)$, which is in turn contained in

$$f_k(|\sigma|) \cup H'\big(|\sigma^{(k+1)}| \times I\big)$$

by the fact that $G = (f \cup H') \circ r$ and $i \circ r(|\sigma| \times I) \subseteq |\sigma| \times I$ for all $\sigma \in K$. Observe that $f_k(|\sigma|) \subseteq \bigcap_{U \in \ell(\mathcal{S}_\sigma)} \widetilde{M}_U$ and

$$H'\big(|\sigma^{(k+1)}| \times I\big) \subseteq \bigcup_{\tau \in \sigma_{k+1}} \bigcap_{U \in \ell(\mathcal{S}_\tau)} \widetilde{M}_U \subseteq \bigcap_{U \in \ell(\mathcal{S}_\sigma)} \widetilde{M}_U$$

where $\tau \in \sigma_{k+1}$ denotes that $\tau$ is a $(k+1)$-simplex of $\sigma$, and the last subset inclusion follows from the fact that if $\tau \subseteq \sigma$, then $\ell(\mathcal{S}_\sigma) \subseteq \ell(\mathcal{S}_\tau)$, and so $\bigcap_{U \in \ell(\mathcal{S}_\tau)} \widetilde{M}_U \subseteq \bigcap_{U \in \ell(\mathcal{S}_\sigma)} \widetilde{M}_U$. Hence (b) follows.

Lastly, (c) follows directly by the construction of $f_{k+1}$. Thus we have completed the induction step.

The base case is satisfied trivially by starting induction at $k = -1$, setting $f_{-1} = f$, and letting $H_{-1}$ be the trivial homotopy $H_{-1}(x,t) = f(x)$. Here we use the convention that $K^{(-1)} = \{\emptyset\}$ and $d(\emptyset, A) = \infty$ for any subset $A \subseteq X$ so that condition (c) is satisfied by setting $V_\sigma = \emptyset$ for $\sigma = \emptyset \in K^{(-1)}$.

Hence by induction there exists a homotopy $H = H_n$ between $f$ and a map $g = f_n$ which satisfies conditions (1) and (2) in the statement of the lemma. □

**Lemma 4.2.4.** *Let $n \in \mathbb{N}$ and let $f : I^n \to \mathcal{V}^m(\mathscr{U})$ be a map such that $f|_{\partial I^n}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$. If $I^n$ admits a triangulation subordinate to $f^{-1}M_{\mathscr{U}}$, then there exists a homotopy $H : I^n \times I \to \mathcal{V}^m(\mathscr{U})$ between $f$ and a map $g : I^n \to |\mathcal{V}(\mathscr{U})|_m$ such that $H|_{\partial I^n \times I}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$.*

*Proof.* Suppose $T : |K| \to I^n$ is a triangulation subordinate to $f^{-1}M_{\mathscr{U}}$. To simplify notation, for any simplex $\sigma \in K$, we identify $|\sigma|$ with its image in $I^n$ under $T$. Given an $n$-simplex $\sigma \in K$, write $\sigma = [v_0, \dots, v_n]$. Express each point of $x \in |\sigma|$ in barycentric coordinates by writing $x = \sum_{i=0}^n x_i v_i$. Let $g_\sigma : \Delta^n \to \mathcal{V}^m(\mathscr{U})$ be the map defined by $g_\sigma(\sum_{i=0}^n x_i v_i) = \sum x_i f(v_i)$. For each $n$-simplex $\sigma \in K$, $g_\sigma$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$ because $g_\sigma$ is an affine mapping of $\Delta^n$ into the simplex of $|\mathcal{V}(\mathscr{U})|_m$ with vertex set $\bigcup_i \operatorname{supp} f(v_i)$. Note that $\bigcup_i \operatorname{supp} f(v_i)$ does in fact define a simplex of $\mathcal{V}(\mathscr{U})$ due to the fact that $T$ is subordinate to $f^{-1}M_{\mathscr{U}}$, which implies that $f(|\sigma|) \subseteq M_U$ for some $U \in \mathscr{U}$, and so $\bigcup_i \operatorname{supp} f(v_i) \subseteq U$. The continuity of $\mathrm{id} : |\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ implies that $g_\sigma$ is also continuous with respect to the topology of $\mathcal{V}^m(\mathscr{U})$.

For all $\sigma \in K_n$, let $H_\sigma : \Delta^n \times I \to \mathcal{V}^m(\mathscr{U})$ be the linear homotopy between $f|_{|\sigma|}$ and $g_\sigma$. Since $T$ is subordinate to $f^{-1}M_{\mathscr{U}}$, for each $\sigma \in K$, there exists $U_\sigma \in \mathscr{U}$ such that $f(x) \in M_{U_\sigma}$ for all $x \in |\sigma|$. Then for any $t \in I$, we have that

$$H_\sigma(x,t) = (1-t)f(x) + tg(x) = (1-t)f(x) + t\sum_{i=0}^{n} x_i f(v_i) \in M_{U_\sigma}.$$

Hence $H_\sigma$ is well-defined, and is thus continuous by Lemma 4.1.1. It is straightforward to see that if $\sigma \cap \sigma' = \tau$, then $H_\sigma|_{|\tau|} = H_{\sigma'}|_{|\tau|}$. So let $H : I^n \times I \to \mathcal{V}^m(\mathscr{U})$ be the map defined by $H|_{|\sigma|} = H_\sigma$ for all $\sigma \in K_n$. Then $H$ is a homotopy between $f$ and the map $g : I^n \to \mathcal{V}^m(\mathscr{U})$ satisfying $g|_{|\sigma|} = g_\sigma$. Moreover, if $\tau \in K$ satisfies $|\tau| \subseteq \partial I^n$, then $f|_{|\tau|}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$. Since $g|_{|\tau|}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$, so too is $H|_{|\tau| \times I}$ by Lemma 4.1.2. Hence $H|_{\partial I^n \times I}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$. □

**Lemma 4.2.5.** *Let $f : I^n \to \mathcal{V}^m(\mathscr{U})$ be a map such that $f|_{\partial I^n}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$. Then $f$ is homotopic rel. $\partial I^n$ to a map $g : I^n \to |\mathcal{V}(\mathscr{U})|_m$.*

*Proof.* Suppose $f : I^n \to \mathcal{V}^m(\mathscr{U})$ is a map such that $f|_{\partial I^n}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$. We may then apply Lemma 4.2.3 to find a homotopy $H : I^n \times I \to \mathcal{V}^m(\mathscr{U})$ between $f$ and a map $f' : I^n \to \mathcal{V}^m(\mathscr{U})$ such that $H|_{\partial I^n \times I}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$ and $I^n$ admits a triangulation subordinate to $f'^{-1}M_{\mathscr{U}}$.

We have that $f'|_{\partial I^{n+1}}$ is continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$ since $H|_{\partial I^n \times I}$ is. Hence we may apply Lemma 4.2.4 to obtain a homotopy $H' : I^n \times I \to \mathcal{V}^m(\mathscr{U})$ between $f'$ and a map $f'' : I^n \to |\mathcal{V}(\mathscr{U})|_m$ such that $H'|_{\partial I^n \times I}$ is continuous relative to $|\mathcal{V}(\mathscr{U})|_m$. Let $G : I^n \times I \to \mathcal{V}^m(\mathscr{U})$ be the concatenation of $G = H \cdot H'$. Let $A = \partial I^n \times I \cup I^n \times \{1\}$ and let $\phi : A \to I^n \times \{0\}$ be a homeomorphism which fixes $\partial I^n \times \{0\}$. Finally, let $g = G|_A \circ \phi^{-1}$. Then $G$ induces a homotopy rel. $\partial I^n$ between

$f$ and $g$. Because $G|_{\partial I^n \times I}$ and $f''$ are both continuous with respect to the topology of $|\mathcal{V}(\mathscr{U})|_m$, so too is $g$. ☐

We now prove our main result.

**Theorem 1.** *For any uniformly bounded open cover $\mathscr{U}$ of a metric space $X$, the natural bijection* $\mathrm{id} : |\mathcal{V}(\mathscr{U})| \to \mathcal{V}^m(\mathscr{U})$ *is a weak homotopy equivalence.*

*Proof.* Combining Lemma 4.2.1 and Lemma 4.2.5, we see that $\mathrm{id} : |\mathcal{V}(\mathscr{U})|_m \to \mathcal{V}^m(\mathscr{U})$ is a weak homotopy equivalence. Since $\mathrm{id} : |K| \to |K|_m$ is a homotopy equivalence for any simplicial complex $K$ [66, Appendix 1, Theorem 10], in particular for the case $K = \mathcal{V}(\mathscr{U})$, the theorem follows. ☐

To strengthen the weak homotopy equivalence of Theorem 4.1 to a homotopy equivalence, it suffices to show that $\mathcal{V}^m(\mathscr{U})$ has the homotopy type of a CW complex. We reiterate a question posed in [5].

**Question.** If $\mathscr{U}$ is a uniformly bounded open cover of a metric space $X$, does $\mathcal{V}^m(\mathscr{U})$ have the homotopy type of a CW complex? More specifically, is $\mathcal{V}^m(\mathscr{U})$ an absolute neighborhood retract?

For the definition of absolute neighborhood retract, see Section 5.1.

## 4.5 Consequences of the Weak Homotopy Equivalence

One immediate consequence of Theorem 4.1 is that a weak homotopy equivalence induces isomorphisms of homology and cohomology groups [48, Proposition 4.21]. Hence Theorem 4.1 yields a separate proof of Corollary 3.3.1 from the previous chapter.

**Corollary 4.2.1.** *For any uniformly bounded open cover $\mathscr{U}$ of a metric space $X$, the natural bijection* $\mathrm{id} : |\mathcal{V}(\mathscr{U})| \to \mathcal{V}^m(\mathscr{U})$ *induces isomorphisms of (co)homology groups.*

However, this corollary is in fact slightly stronger than Corollary 3.3.1 as the isomorphisms between the homology groups of $|\mathcal{V}(\mathscr{U})|$ and $\mathcal{V}^m(\mathscr{U})$ are now induced by $\mathrm{id} : |\mathcal{V}(\mathscr{U})| \to \mathcal{V}^m(\mathscr{U})$. This has new implications for the relationship between the persistent homology of $\mathrm{VR}(X;\bullet)$ and $\mathrm{VR}^m(X;\bullet)$, as well as the persistent homology of the Čech complex $\check{\mathrm{C}}(X;\bullet)$ and Čech metric thickening $\check{\mathrm{C}}^m(X;\bullet)$.

A *persistence module* is a family of $\mathbb{K}$-vector spaces $\{V_r\}_{r\in\mathbb{R}}$, where $\mathbb{K}$ is any field, along with a collection of linear maps $\phi_{r,s} : V_r \to V_s$ for every choice of $r \le s$ such that $\phi_{r,r} = \mathrm{id}$ and $\phi_{r,t} = \phi_{s,t} \circ \phi_{r,s}$ whenever $r \le s \le t$. More succinctly, a persistence module is a functor $(\mathbb{R}, \le) \to \mathbf{Vect}_{\mathbb{K}}$ from the poset $(\mathbb{R}, \le)$ to the category of $\mathbb{K}$-vector spaces.

For a fixed metric space $X$, note that $|\mathrm{VR}(X;\bullet)|$ and $\mathrm{VR}^m(X;\bullet)$ define functors from $(\mathbb{R}, \le)$ to **Top**, the category of topological spaces. Let $H_n : \mathbf{Top} \to \mathbf{Vect}_{\mathbb{K}}$ denote the $n$-th singular homology functor with coefficients in $\mathbb{K}$. Then for any metric space $X$ and $n \in \mathbb{N}$, $H_n \circ |\mathrm{VR}(X;\bullet)|$ and $H_n \circ \mathrm{VR}^m(X;\bullet)$ are both examples of persistence modules.

Note that the following diagram, in which the horizontal arrows are the inclusions, commutes for any $r_1 \le r_2$.

$$\begin{array}{ccc} |\mathrm{VR}(X;r_1)| & \hookrightarrow & |\mathrm{VR}(X;r_2)| \\ {\scriptstyle \mathrm{id}}\downarrow & & \downarrow{\scriptstyle \mathrm{id}} \\ \mathrm{VR}^m(X;r_1) & \hookrightarrow & \mathrm{VR}^m(X;r_2) \end{array}$$

Hence Theorem 4.1 and the above observation imply that the functors $H_n \circ |\mathrm{VR}(X;\bullet)|$ and $H_n \circ \mathrm{VR}^m(X;\bullet)$ are naturally isomorphic. The same is true if we replace $\mathrm{VR}(X;\bullet)$ and $\mathrm{VR}^m(X;\bullet)$ with $\check{\mathrm{C}}(X;\bullet)$ and $\check{\mathrm{C}}^m(X;\bullet)$ respectively. Hence we have the following.

**Corollary 4.2.2.** *For any metric space $X$ and $n \in \mathbb{N}$, the persistence modules $H_n \circ |\mathrm{VR}(X;\bullet)|$ and $H_n \circ \mathrm{VR}^m(X;\bullet)$ (respectively, $H_n \circ |\check{\mathrm{C}}(X;\bullet)|$ and $H_n \circ \check{\mathrm{C}}^m(X;\bullet)$) are isomorphic.*

A similar result was proved through different methods by Adams, Memoli, Moy, and Wang [6, Corollary 5.10], who have shown that for a totally bounded metric space $X$, the interleaving distance between the persistence modules $H_n \circ |\mathrm{VR}(X;\bullet)|$ and $H_n \circ \mathrm{VR}^m(X;\bullet)$ (respectively, $H_n \circ |\check{\mathrm{C}}(X;\bullet)|$ and $H_n \circ \check{\mathrm{C}}^m(X;\bullet)$) is zero. While this is enough to conclude that $\mathrm{VR}(X;\bullet)$ and $\mathrm{VR}^m(X;\bullet)$ have identical (undecorated) persistence diagrams, their result does not imply that the persistence modules are isomorphic.

To state a final application of Theorem 4.1, we recall the *reach* of a subset $X \subseteq \mathbb{R}^n$. First, given $X \subseteq \mathbb{R}^n$, if

$$Y = \{y \in \mathbb{R}^n : \exists x_1 \neq x_2 \in X, d(y,x_1) = d(y,x_2) = d(y,X))\},$$

then the *medial axis* of $X$ is $\overline{Y}$, the closure of $Y$. Then if $\tau = d(X,\overline{Y})$, we say that $X$ has reach $\tau$. In other words, if $X$ has reach $\tau$, then for every point $x \in \mathbb{R}^n$ whose distance to $X$ is less than $\tau$, $x$ has a unique closest point in $X$. It was shown in [7] that for a subset $X \subseteq \mathbb{R}^n$ with positive reach $\tau > 0$, $\mathrm{VR}^m(X;r)$ is homotopy equivalent to $X$ for all $0 < r < \tau$. We are able to use Theorem 4.1 to conclude that the same is true for $\mathrm{VR}(X;r)$.

**Corollary 4.2.3.** *If $X \subseteq \mathbb{R}^n$ has reach $\tau > 0$, then $|\mathrm{VR}(X;r)|$ is homotopy equivalent to $X$ for all $0 < r < \tau$.*

*Proof.* Assuming that $X \subseteq \mathbb{R}^n$ has reach $\tau > 0$, the Vietoris–Rips metric thickening $\mathrm{VR}^m(X;r)$ is homotopy equivalent to $X$ for all $r < \tau$ [7]. Since the natural map $|\mathrm{VR}(X;r)| \to \mathrm{VR}^m(X;r)$ is a weak homotopy equivalence for any $r > 0$ by Theorem 4.1, we have that $|\mathrm{VR}(X;r)|$ is weakly homotopy equivalent to $X$ for all $r < \tau$. Since $X$ is a set of positive reach, it is a retract of its tubular neighborhood $\{y \in \mathbb{R}^n : d(y,X) < \tau\}$ which has the homotopy type of a CW complex, and thus $X$ itself has the homotopy type of a CW complex, see [48, Prop. A.11]. Lastly, Whitehead's theorem implies that $|\mathrm{VR}(X;r)|$ and $X$ are in fact homotopy equivalent for all $r < \tau$. □

Note that any closed $C^2$ submanifold $M \subseteq \mathbb{R}^n$ has positive reach [78], hence Corollary 4.2.3 applies to such spaces.

# Chapter 5

# Vietoris Complexes of Absolute Neighborhood Retracts

Hausmann's theorem [49] states that for a closed Riemannian manifold $M$ with the geodesic metric, there exists a quantity $r(M) > 0$ such that $|\mathrm{VR}(M;r)|$ is homotopy equivalent to $M$ for all $0 < r < r(M)$. At the end of the previous chapter, we showed that the conclusion of Hausmann's theorem also holds for subsets $X \subseteq \mathbb{R}^n$ of positive reach with the Euclidean metric.

Though sets of positive reach form a quite general class of spaces, there exist many simple examples of spaces that are not of positive reach, yet the conclusion of Hausmann's theorem still holds. For example, it is possible to show that if $X$ the boundary of the square $[-1,1]^2$, that is $X = \partial[-1,1]^2$, then $|\mathrm{VR}(X;r)|$ and $X$ are homotopy equivalent for all $0 < r < 2$.

**Proposition 5.1.** *If $X = \partial[-1,1]^2$, then $|\mathrm{VR}(X;r)|$ and $X$ are homotopy equivalent for all $0 < r < 2$.*

For a proof see Appendix A.

Thus it is reasonable to ask whether there is a more general class of spaces for which the consequence of Hausmann's theorem holds. One well-behaved class of metric spaces is that of *absolute neighborhood retracts* (ANRs). It was conjectured in [63] that when $X$ is a compact ANR, $|\mathrm{VR}(X;r)|$ is homotopy equivalent to $X$ for all sufficiently small $r < 0$, i.e. the conclusion of Hausmann's theorem holds for all compact ANRs.

In this chapter, we study Vietoris complexes of compact ANRs. We begin by showing that for any ANR $X$, there exists an open cover $\mathscr{U}$ of $X$ such that for any open refinement $\mathscr{V} \leq \mathscr{U}$, $|\mathcal{V}(\mathscr{V})|$ *dominates* $X$, that is, there exists maps $r : |\mathcal{V}(\mathscr{V})| \to X$ and $i : X \to |\mathcal{V}(\mathscr{V})|$ such that $r \circ i \simeq \mathrm{id}$. This result implies that both the Vietoris–Rips complex $|\mathrm{VR}(X;r)|$ as well as the Čech complex $|\check{\mathrm{C}}(X;r)|$ dominate $X$ for all sufficiently small $r$ when $X$ is a compact ANR. Second, we provide a method of constructing compact ANRs for which $|\mathrm{VR}(X;r)|$ is not homotopy equivalent to $X$ for all sufficiently small $r < 0$, which disproves Conjecture 4 of [63].

## 5.1 ANRs are Dominated by Vietoris Complexes for Sufficiently Fine Covers

We begin with some preliminary definitions.

**Definition 5.1.** *A metric space $X$ is an absolute neighborhood retract for metric spaces (*ANR*) if whenever $X$ is embedded as a closed subspace of a metric space $Y$, there exists an open neighborhood $U$ of $X$ in $Y$ and a retraction $r : U \to X$.*

For a cover $\mathscr{U}$ of a space $X$ and $x \in X$, the star of $x$ with respect to $\mathscr{U}$ is $\mathrm{st}(x, \mathscr{U}) = \bigcup\{U \in \mathscr{U} : x \in U\}$. Given a cover $\mathscr{U}$ of a space $X$, a *barycentric refinement* of $\mathscr{U}$ is a cover $\mathscr{V}$ such for every $x \in X$, there exists $U \in \mathscr{U}$ such that $\mathrm{st}(x, \mathscr{V}) \subseteq U$.

Let $\mathscr{U}$ be a cover of $Y$. Two maps $f, g : X \to Y$ are $\mathscr{U}$*-near* if for every $x \in X$, there exists $U \in \mathscr{U}$ such that $f(x), g(x) \in U$. The maps $f$ and $g$ are said to be $\mathscr{U}$*-homotopic* if there exists a homotopy $H : X \times I \to Y$ between $f$ and $g$ such that for every $x \in X$, there exists $U \in \mathscr{U}$ with $H(\{x\} \times I) \subseteq U$. If $Y$ is an ANR, and $\mathscr{U}$ is an open cover of $Y$, there exists an open cover $\mathscr{V}$ of $Y$ such that whenever two maps $f, g : X \to Y$ are $\mathscr{V}$-near, they are $\mathscr{U}$-homotopic [28, Lemma 7.2]. Given a simplicial complex $K$, a subcomplex $L \subseteq K$, and a cover $\mathscr{U}$ of a space $X$, we say that $f : |L| \to X$ is a *partial realization with respect to* $\mathscr{U}$ if for any $\sigma \in K$, $f(|\sigma| \cap |L|) \subseteq U$ for some $U \in \mathscr{U}$. A map $g : |K| \to X$ is a *full realization with respect to* $\mathscr{U}$ if for every $\sigma \in K$, $g(|\sigma|) \subset U$ for some $U \in \mathscr{U}$. Another property of ANRs is the following. For any ANR $X$ and an open cover $\mathscr{U}$ of $X$, there exists an open refinement $\mathscr{V} \leq \mathscr{U}$ such that for any simplicial complex $K$ and subcomplex $L \subseteq K$ containing all the vertices of $K$, any partial realization $f : |L| \to X$ with respect to $\mathscr{V}$ extends to a full realization $|K| \to X$ with respect to $\mathscr{U}$ [29].

**Definition 5.2.** *A space $A$ is dominated by $X$ if there exist maps $i : A \to X$ and $r : X \to A$ such that $ri \simeq \mathrm{id}_A$.*

It is implicitly shown in [63, Proposition 9.8] that a for compact ANR $X$, $|\mathrm{VR}(X; r)|$ dominates $X$ for all sufficiently small $r > 0$. The proof can be summarized as follows. Any metric space $X$ can be isometrically embedding into an injective metric space $E$, and $B_{r/2}(X, E) = \{y \in E : d_E(y, X) < r/2\}$ is homotopy equivalent to $|\mathrm{VR}(X; r)|$. Then since $X$ is a compact ANR, there exists $r$ small enough such that $B_{r/2}(X, E)$ retracts onto $X$.

We use a different approach to prove a more general fact.

**Theorem 5.1.** *Suppose $X$ is an* ANR*. There exists an open cover $\mathscr{U}$ of $X$ such that for any open refinement $\mathscr{V} \leq \mathscr{U}$, $|\mathcal{V}(\mathscr{V})|$ dominates $X$.*

*Proof.* Since $X$ is an ANR, there exists a cover $\mathscr{W}$ of $X$ such that whenever two maps $f, g : Y \to X$ are $\mathscr{W}$-near, they are homotopic. Let $\mathscr{W}'$ be a barycentric refinement of $\mathscr{W}$. Applying again the fact that $X$ is an ANR, there exists a refinement $\mathscr{U} \leq \mathscr{W}'$ such that for any subcomplex $L$ of an arbitrary simplicial complex $K$, any partial realization $|L| \to X$ with respect to $\mathscr{U}$ extends to a full realization $|K| \to X$ with respect to $\mathscr{W}'$.

For this cover $\mathscr{U}$, we now show that for any open refinement $\mathscr{V} \leq \mathscr{U}$, $|\mathcal{V}(\mathscr{V})|$ dominates $X$. So let $\mathscr{V} \leq \mathscr{U}$ be a open refinement of $\mathscr{U}$ and let $\mathscr{V}'$ be a barycentric refinement of $\mathscr{V}$. Let $\{\phi_x\}_{x\in X}$ be a partition of unity subordinate to $\mathscr{V}'$. Let $f : X \to |\mathcal{V}(\mathscr{V})|$ be the map which sends a point $y \in X$ to the point of $|\mathcal{V}(\mathscr{V})|$ defined by the barycentric coordinates $\phi_x(y)$. Note that $\{x \in X : \phi_x(y) \neq 0\} \subseteq \mathrm{st}(y, \mathscr{V}') \subseteq V$ for some $V \in \mathscr{V}$, hence $f$ is well-defined.

Define $g : |\mathcal{V}(\mathscr{V})| \to X$ by setting $g(x) = x$ for each vertex of $\mathcal{V}(\mathscr{V})$ (a partial realization on the 0-skeleton of $\mathcal{V}(\mathscr{V})$ with respect to $\mathscr{V} \leq \mathscr{U}$) and extending this to a full realization $g : |\mathcal{V}(\mathscr{V})| \to X$ with respect to $\mathscr{W}'$. If $v$ is a vertex of the unique open simplex containing $f(x)$, there exists $W_1 \in \mathscr{W}'$ such that $x, g(v) \in W_1$ simply by the fact that $x, g(v) \in V'$ for some $V' \in \mathscr{V}'$ and $\mathscr{V}'$ refines $\mathscr{W}'$. Additionally, there exists $W_2 \in \mathscr{W}'$ such that $g(v), gf(x) \in W_2$ by the fact that $g : |\mathcal{V}(\mathscr{V})| \to X$ is a realization with respect to $\mathscr{W}'$. Then

$$x, gf(x) \in W_1 \cup W_2 \subseteq \mathrm{st}(g(v), \mathscr{W}') \subseteq W$$

for some $W \in \mathscr{W}$. Hence $gf$ and $\mathrm{id}_X$ are $\mathscr{W}$-near, and thus are homotopic. □

Theorem 5.1 can be applied to not only show that $|\mathrm{VR}(X;r)|$ dominates an ANR $X$ for sufficiently small $r$, but also that the Čech complex $|\check{\mathrm{C}}(X;r)|$ dominates $X$ for sufficiently small $r > 0$, a fact which does not follow from the proof of [63, Proposition 9.8].

**Corollary 5.1.1.** *Suppose $X$ is a compact* ANR*. There exists $r > 0$ such that $|\mathrm{VR}(X;s)|$ (resp. $|\check{\mathrm{C}}(X;s/2)|$) dominates $X$ for any $0 < s < r$.*

*Proof.* By Theorem 5.1, there exists a cover $\mathscr{U}$ of $X$ such that for any refinement $\mathscr{V} \leq \mathscr{U}$, $|\mathcal{V}(\mathscr{V})|$ dominates $X$. Since $X$ is compact, there exists $r > 0$ such that the collection $\mathscr{D}_r$ of all open sets with diameter at most $r$ refines $\mathscr{U}$. Then if $0 < s < r$, $\mathscr{D}_s$ refines $\mathscr{D}_r$, hence $\mathscr{U}$ as well, which implies that $|\mathcal{V}(\mathscr{D}_s)| = |\mathrm{VR}(X;s)|$ dominates $X$. Likewise, for such a value of $r$, the collection $\mathscr{B}_{r/2}$ of all open balls of radius $r/2$ refines $\mathscr{U}$, hence $|\mathcal{V}(\mathscr{B}_{s/2})| = |\check{\mathrm{C}}(X;s/2)|$ dominates $X$ for any $0 < s < r$. □

**Remark 5.2.** Given a cover $\mathscr{V}$ of a space $X$, let $\overline{\mathscr{V}}$ denote the cover $\overline{\mathscr{V}} = \{\overline{V} : V \in \mathscr{V}\}$. In the statement of Theorem 5.1, if $\mathscr{V}$ is an open cover such that $\overline{\mathscr{V}}$ refines $\mathscr{U}$, then using the same proof, one can show that $|\mathcal{V}(\overline{\mathscr{V}})|$ dominates $X$. This is allows one to prove the analogue of Corollary 5.1.1 for closed Vietoris–Rips and Čech complexes, $\mathrm{VR}_{\leq}(X;r)$ and $\check{\mathrm{C}}_{\leq}(X;r)$.

We also have an analogous result to Theorem 5.1 for Vietoris metric thickenings, though now, $\mathcal{V}^m(\mathscr{V})$ not only dominates $X$, but retracts onto $X$.

**Proposition 5.2.** *Suppose $X$ is an* ANR. *There exists an open cover $\mathscr{U}$ of $X$ such that for any refinement $\mathscr{V} \leq \mathscr{U}$, $\mathcal{V}^m(\mathscr{V})$ retracts onto $X$.*

*Proof.* Let $\mathcal{V}^m(X)$ denote the Vietoris metric thickening with respect to the cover of $X$ by $\{X\}$, i.e. $\mathcal{V}^m(X)$ contains *all* finitely support probability measures on $X$. Recall that the map $x \mapsto \delta_x$ identifies $X$ with a closed subset of $\mathcal{V}^m(X)$. Since $X$ is an ANR, there exists an open neighborhood $W$ of $X$ in $\mathcal{V}^m(X)$ and a retraction $r : W \to X$. For each $x \in X$, choose $s_x > 0$ so that $B_{\mathcal{V}^m(X)}(\delta_x, s_x) \subseteq W$. Define $\mathscr{U} = \{B_X(x, s_x) : x \in X\}$. Then $\mathcal{V}^m(\mathscr{U}) \subseteq W$ since for any $\mu \in \mathcal{V}^m(\mathscr{U})$, $\mu$ has support in some $B_X(x_0, s_{x_0})$, in which case $d_W(\mu, \delta_{x_0}) < s_{x_0}$ showing that $\mu \in W$. Hence if $\mathscr{V} \leq \mathscr{U}$, the restriction of $r$ to $\mathcal{V}^m(\mathscr{V}) \subseteq W$ defines the desired retraction. □

In the case that the ANR $X$ is assumed to be compact, we have the metric thickening analogue to Corollary 5.1.1, with essentially the same proof.

**Corollary 5.2.1.** *If $X$ is a compact* ANR, *there exists $r > 0$ such that* $\mathrm{VR}^m(X;s)$ *(resp. $\check{\mathrm{C}}^m(X;s/2)$) retracts onto $X$ whenever $0 < s < r$.*

## 5.2 A Counter-example

In this section, we provide a counter-example to the following conjecture.

**Conjecture** ([63]). Let $(X, d_X)$ be a compact ANR metric space. Then, there exists $r(X) > 0$ such that $|\mathrm{VR}(X;r)|$ is homotopy equivalent to $X$ for any $r \in (0, r(X)]$.

To start, we consider the wedge sum of a collection of spaces with a particular metric. Recall that the wedge sum $\bigvee_{i\in\mathcal{I}}(X_i, b_i)$ of a collection of spaces $\{X_i\}_{i\in\mathcal{I}}$ where each $X_i$ has basepoint $b_i$ is the quotient of the disjoint union $\coprod_i X_i$ under the equivalence relation which identifies all of the basepoints $b_i$. If the basepoints $b_i$ are arbitrary or understood from context, we may simply write $\bigvee_{i\in\mathcal{I}} X_i$.

**Definition 5.3.** *Let $(X_i, d_i)$, $i \in \mathcal{I}$, be a collection of metric spaces, each with basepoint $b_i$. The* wedge metric $d_w$ *on the wedge sum $\bigvee_{i\in\mathcal{I}}(X_i, b_i)$ is defined by setting $d_w(x, y) = d_i(x, y)$ if $x, y \in X_i$, and $d_w(x, y) = d_i(x, b_i) + d_j(y, b_j)$ if $x \in X_i$ and $y \in X_j$ with $i \neq j$.*

**Proposition 5.3.** *Suppose $(X, d)$ is a finite-dimensional, locally contractible, contractible compact metric space and that there exists $r^* > 0$ such that $|\mathrm{VR}(X; r^*)|$ is not contractible. For each $i \in \mathbb{N}$, let $(X_i, d_i)$ be the metric space given by setting $X_i = X$ and $d_i = 2^{-i}d$. Then the space $Y = \bigvee_{i=1}^{\infty} X_i$ with the wedge metric $d_w$ is a compact ANR, yet for any $r > 0$, there exists $0 < s < r$ such that $|\mathrm{VR}(Y; s)|$ is not homotopy equivalent to $Y$.*

*Proof.* If $X$ is finite-dimensional, locally contractible, contractible, and compact, then $Y = \widetilde{\bigvee}_{i=1}^{\infty} X_i$ is as well. Hence $Y$ is an ANR [15]. For any $k \in \mathbb{N}$, [1, Proposition 3.7] implies that

$$|\mathrm{VR}(Y; r)| \simeq |\mathrm{VR}(\bigvee_{i=1}^{k-1} X_i; r)| \vee |\mathrm{VR}(X_k; r)| \vee |\mathrm{VR}(\bigvee_{i=k+1}^{\infty} X_i; r)|$$

for any $r > 0$, where each wedge sum is equipped with the wedge metric. Also, note that $\mathrm{VR}(X_k; r) = \mathrm{VR}(X; 2^k r)$. Hence $|\mathrm{VR}(X; 2^k r)|$ is a retract of $|\mathrm{VR}(Y; r)|$.

Now, for any $r > 0$, we may find $k \in \mathbb{N}$ such that $r^*/2^k < r$. Set $s = r^*/2^k$. Then $|\mathrm{VR}(Y; s)|$ retracts onto the non-contractible space $|\mathrm{VR}(X; r^*)|$, hence cannot be homotopy equivalent to the contractible space $Y$. □

A *correspondence* between sets $X$ and $Y$ is any relation $R \subseteq X \times Y$ such that the projections of $R$ onto each factor are surjective. The *distortion* of a non-empty relation $R \subseteq X \times Y$ is

$$\mathrm{dis}(R) = \sup_{(x_1,y_1),(x_2,y_2)\in R} |d_X(x_1,x_2) - d_Y(y_1,y_2)|.$$

The Gromov-Hausdorff distance $d_{\mathrm{GH}}$ between two compact metric spaces $X, Y$ can be defined as

$$d_{\mathrm{GH}}(X,Y) = \frac{1}{2} \cdot \inf_{R\in\mathscr{C}} \mathrm{dis}(R),$$

where $\mathscr{C}$ denotes the collection of all correspondences between $X$ and $Y$ [51].

In the following theorem we produce a counter-example to the conjecture from [63] stated at the beginning of the section.

**Theorem 5.3.** *There exists a compact* ANR $Y$ *such that for all* $r > 0$*, there exists* $0 < s < r$ *such that* $|\mathrm{VR}(Y;s)|$ *and* $Y$ *are not homotopy equivalent.*

*Proof.* By Proposition 5.3, it suffices to produce a compact metric space $X$ that is finite-dimensional, locally contractible, and contractible for which there exists $r^* > 0$ such that $|\mathrm{VR}(X;r^*)|$ is not contractible. This is done by setting $X$ to be a circle with a sufficiently small arc removed.

To give a precise argument, let $(S^1, d)$ denote the unit circle equipped with the geodesic metric. By Latschev's theorem [57], there exists $\epsilon > 0$ and $\delta > 0$ such that any metric space $X$ with $d_{\mathrm{GH}}(S^1, X) < \delta$, then $|\mathrm{VR}(X;\epsilon)| \simeq S^1$. Let $A \subseteq S^1$ be an open arc of length less than $\delta$, let $X = S^1 \setminus A$, and equip $X$ with the metric inherited at a subset of $S^1$. Let $b_1, b_2$ denote the two boundary points of $X$ in $S^1$. We claim that $d_{\mathrm{GH}}(S^1, X) < \delta$. To see this, let the relation $R$ be the graph of the (discontinuous) surjective function $f : S^1 \to X$ defined by setting $f(x) = x$ for $x \in X$, and $f(x) = b_1$ otherwise. It suffices to show that $\mathrm{dis}(R) \leq d(b_1,b_2) < \delta$. To show this, note that

(1) if $x_1, x_2 \in A$, then $f(x_1) = f(x_2) = b_1$ so $|d(x_1,x_2) - d(f(x_1), f(x_2))| = d(x_1,x_2) \leq d(b_1,b_2)$,

(2) if $x_1 \in A$ and $x_2 \in X$, then $|d(x_1,x_2) - d(f(x_1), f(x_2))| = |d(x_1,x_2) - d(b_1,x_2)| \leq d(x_1,b_1) \leq d(b_1,b_2)$,

(3) if $x_1, x_2 \in X$, then $|d(x_1, x_2) - d(f(x_1), f(x_2))| = |d(x_1, x_2) - d(x_1, x_2)| = 0$.

Thus, $\mathrm{dis}(R) \leq d(b_1, b_2) < \delta$, which implies that $d_{\mathrm{GH}}(S^1, X) < \delta$. Therefore $|\mathrm{VR}(X; \epsilon)| \simeq S^1$ and so $|\mathrm{VR}(X; \epsilon)|$ is not contractible. $\square$

See Figure 5.1 for a visual representation of the counterexample described by the proof of Theorem 5.3.

### 5.2.1 A Geodesic Counter-example

For a metric space $(X, d)$ and a path $\gamma : I \to X$, the length of $\gamma$ is $L(\gamma) = \sup\{\sum_{i=0}^{k} d(\gamma(t_i), \gamma(t_{i+1}))\}$ where the supremum is taken over all partitions $0 = t_0 < t_1 < \cdots < t_k = 1$ of the unit interval. A metric space $(X, d)$ is *geodesic* if for every pair of points $x, y \in X$, there exists a path $\gamma : I \to X$ between $x$ and $y$ such that $L(\gamma) = d(x, y)$. Hausmann's theorem applies not only to Riemannian manifolds $M$ such that the quantity $r(M)$ defined in [49] is greater than 0, but more generally to geodesic spaces $X$ such that $r(X) > 0$. While Theorem 5.3 shows that the conclusion of Hausmann's theorem does not hold for compact ANRs, one might ask whether it does for compact *geodesic* ANRs. However, a certain higher dimensional analogue of the counter-example in Theorem 5.3 shows that this is not the case.

**Proposition 5.4.** *There exists a compact geodesic* ANR $X$ *such that for all* $r > 0$, *there exists* $0 < s < r$ *such that* $|\mathrm{VR}(X; s)|$ *and* $X$ *are not homotopy equivalent.*

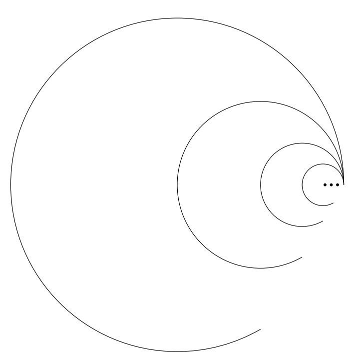

**Figure 5.1:** A visual representation of the space $Y$ entailed by the proof of Theorem 5.3. Note that the space is equipped with the wedge metric.

*Proof.* Note that if a collection of metric spaces $(X_i, d_i)$ are all geodesic, then the wedge sum $\bigvee_{i\in\mathcal{I}} X_i$ with the wedge metric $d_w$ is also geodesic. Thus by Proposition 5.3, it suffices to produce a finite-dimensional, locally contractible, contractible, compact *geodesic* metric space $X$ for which $|\mathrm{VR}(X, r)|$ is non-contractible for some value of $r > 0$.

Let $(S^2, d)$ denote the 2-sphere equipped with the geodesic metric with unit circumference. Again, by Latschev's theorem [57], there exists $\epsilon > 0$ and $\delta > 0$ such that any metric space $X$ with $d_{\mathrm{GH}}(S^2, X) < \delta$, then $|\mathrm{VR}(X;\epsilon)| \simeq S^2$. Let $A \subseteq S^2$ denote an open spherical cap whose bounding circle has circumference less than $2\delta$. Let $b_0$ be a point on the circle bounding $A$. Let $X = S^2 \setminus A$ and equip $X$ with the induced intrinsic metric $d_I$, that is $d_I(x, y) = \min\{L(\gamma) : \gamma : I \to X, \gamma(0) = x, \gamma(1) = y\}$. We claim that $d_{\mathrm{GH}}(X, S^2) < \delta$. To see this, let the relation $R$ be the graph of the surjective function $f : S^2 \to X$ given by $f(x) = b_0$ for all $x \in A$ and otherwise $f(x) = x$. It suffices to show that $\mathrm{dis}(R) < \delta$. First note that since the circumference of $A$ is $2\delta$, the geodesic distance between any two points in $A$ is less than $\delta$. We consider three cases.

(1) If $x_1, x_2 \in A$, then $f(x_1) = f(x_2) = b_0$ so $|d(x_1, x_2) - d_I(f(x_1), f(x_2))| = d(x_1, x_2) < \delta$.

(2) If $x_1 \in X$ and $x_2 \in A$, then $|d(x_1, x_2) - d_I(f(x_1), f(x_2))| = |d(x_1, x_2) - d_I(x_1, b_0)|$. Note that both $d(x_1, \partial A) \leq d(x_1, x_2) \leq d(x_1, \partial A) + \delta$ and $d(x_1, \partial A) \leq d_I(x_1, b_0) \leq d(x_1, \partial A) + \delta$. This implies that $|d(x_1, x_2) - d_I(x_1, b_0)| < \delta$.

(3) if $x_1, x_2 \in X$, then $|d(x_1, x_2) - d(f(x_1), f(x_2))| = |d(x_1, x_2) - d_I(x_1, x_2)|$. If there exists a geodesic $\gamma$ between $x_1$ and $x_2$ which does not pass through $A$, we have $d(x_1, x_2) = d_I(x_1, x_2)$, and thus $|d(x_1, x_2) - d_I(x_1, x_2)| < \delta$. Otherwise, the geodesic $\gamma$ intersects $A$ and in this case, we have that $d(x_1, \partial A) + d(x_2, \partial A) \leq d(x_1, x_2) \leq d_I(x_1, x_2) \leq d(x_1, \partial A) + d(x_2, \partial A) + \delta$, which again implies that $|d(x_1, x_2) - d_I(x_1, x_2)| < \delta$.

Therefore, $\mathrm{dis}(R) < \delta$, and hence $d_{\mathrm{GH}}(X, S^2) < \delta$. $\square$

# Chapter 6

# Bayesian Sheaf Neural Networks

We now turn our attention to an application of algebraic topology to machine learning. Given a simplicial complex $K$, consider the chain complex

$$\cdots \longrightarrow C_{n+1}(K;\mathbb{R}) \xrightarrow{\partial_{n+1}} C_n(K;\mathbb{R}) \xrightarrow{\partial_n} C_{n-1}(K;\mathbb{R}) \longrightarrow \cdots ,$$

where $C_n(K;\mathbb{R})$ denotes $n$-chains in $K$ with coefficients in $\mathbb{R}$, that is $C_n(K;\mathbb{R})$ consists of formal sums $\sum_{i=0}^{k} a_i\sigma_i$ where $a_i \in \mathbb{R}$ and $\sigma \in K_n$. The $n$th combinatorial Laplacian $L_n : C_n(K;\mathbb{R}) \to C_n(K;\mathbb{R})$ is $L_n = \partial_n^*\partial_n + \partial_{n+1}\partial_{n+1}^*$, where $\partial_n^*$ denotes the adjoint of $\partial_n$. Choosing a basis for $C_n(K;\mathbb{R})$, the $n$th combinatorial Laplacian $L_n$ can be represented by an $k \times k$ matrix, where $k$ denotes the number of $n$-simplices of $K$. Then given data in the form of "feature" vectors $x_\sigma \in \mathbb{R}^d$ associated to each $n$-simplex of $K$, one may aggregate these vectors in a $k \times d$ matrix $X$, and the $n$th combinatorial Laplacian $L_n$ can be used as a feature propogation mechanism which operates on $X$ by matrix multiplication $L_nX$. By incorporating a learnable weight matrix $W \in \mathbb{R}^{d\times d}$ and a non-linear activation function $\sigma$ such as $\sigma(x) = \max(0, x)$, the features $X$ can be updated by computing $Y = \sigma(L_nXW)$, where $\sigma$ is applied element-wise. This forms the basic idea behind the layers of a *simplicial convolutional neural network* [30]. Graph convolutional networks—which have found wide-ranging application—are a special case: where $K$ is a 1-dimensional simplicial complex, i.e. a graph. Building on this idea, sheaf neural networks replace the combinatorial Laplacian $L_n$ with a sheaf Laplacian $L^n_{\mathcal{F}}$ defined in terms of a *cellular sheaf* $\mathcal{F}$ over $K$. Briefly, a cellular sheaf is a structure which assigns a linear transformation to each incident pair of simplices $\sigma \subseteq \tau$. Sheaf neural networks were introduced to combat the problem of over-smoothing which can occur in graph and simplicial convolutional networks [12]. Over-smoothing, commonly discussed in the context of graph neural networks, is the phenomena in which the features of neighboring nodes converge to similar values as they are updated by successive graph convolutional layers. This is an issue for node classification tasks involving *heterophilic* graph data, where neighboring node have a low likelihood of belonging to the same class.

In applications to real-world data, there is often no a priori choice of cellular sheaf to use, hence it is necessary to learn a suitable cellular sheaf as part of the network. However, this leaves the network sensitive to the learned sheaf and creates additional potential for the network to overfit to training data, reducing how well the model generalizes to unseen data. In this chapter, we develop a Bayesian framework for learning cellular sheaves within sheaf neural networks in order to increase the robustness of such models, yielding what we refer to as Bayesian sheaf neural networks (BSNN). In Section 6.1, we review preliminary topics such as neural networks, cellular sheaves, and sheaf Laplacians. We review how the sheaf Laplacian can be used to define a sheaf neural network, and the connection to a sheaf diffusion process. We prove a result regarding the linear separation power of a sheaf diffusion process for sheaves consisting of special orthogonal maps. In Section 6.2, we describe our variational approach to learning cellular sheaves within a sheaf neural network. In Section 6.3, we define a reparameterizable family of distributions on $SO(n)$, which are used to define the variational distribution for cellular sheaves with special orthogonal restriction maps. In Section 6.4, we evaluate our BSNN on several node classification tasks for heterophilic web-page datasets. We end with a final discussion, as well as potential directions for future work in Section 6.5.

## 6.1 Preliminaries

### 6.1.1 Neural Networks

We give a very simplified review of neural networks. In very broad terms, a neural network can be described as a family of functions $f_\phi : X \to Y$ from a set of observations $X$ to a set of targets $Y$, that is parameterized by *weights* $\phi$. Given a dataset $\{(x_i, y_i)\}_i \subseteq X \times Y$, one wishes to find a value $\phi$ for which the predictions $f_\phi(x_i)$ closely match the true targets $y_i$. This is done by selecting a *loss function* $L : Y \times Y \to \mathbb{R}$, so that the objective becomes to minimize $\sum_i L(f_\phi(x_i), y_i)$. Neural networks and loss functions are typically differentiable, so that during training, the minimization of $\sum_i L(f_\phi(x_i), y_i)$ can be performed by gradient descent.

A *multi-layer perceptron* (MLP) is one of the simplest types of neural network. An MLP is the composition of several feed-forward layers, each of which is a function $\mathbb{R}^n \to \mathbb{R}^k$ of the form $x \mapsto \sigma(Wx + b)$, where $W \in \mathbb{R}^{k\times n}$ is a *weight* matrix, $b \in \mathbb{R}^k$ is a vector of weights referred to as *biases*, and $\sigma : \mathbb{R} \to \mathbb{R}$ is a nonlinear function, for example $\sigma(x) = \max(0, x)$, which is applied element-wise to $Wx + b \in \mathbb{R}^k$. The collection of all matrices $W$ and biases $b$ within an MLP are its parameters, which are updated during training.

### 6.1.2 Cellular Sheaves

A *cellular sheaf* $\mathcal{F}$ on a simplicial complex $K$ is a structure which assigns a vector space $\mathcal{F}(\sigma)$ to each simplex $\sigma \in K$, and assigns to each incident pair $\tau \subseteq \sigma$ a linear transformation $\mathcal{F}_{\tau\subseteq\sigma} : \mathcal{F}(\tau) \to \mathcal{F}(\sigma)$ such that $\mathcal{F}_{\sigma\subseteq\sigma} : \mathcal{F}(\sigma) \to \mathcal{F}(\sigma)$ is the identity map for every $\sigma$, and with the compatibility condition that $\mathcal{F}_{\tau\subseteq\rho} = \mathcal{F}_{\sigma\subseteq\rho} \circ \mathcal{F}_{\tau\subseteq\sigma}$ whenever $\tau \subseteq \sigma \subseteq \rho$. In keeping with the terminology from sheaf theory [19], the vector space $\mathcal{F}(\sigma)$ is referred to as the *stalk* of $\mathcal{F}$ over $\sigma$, and the transformations $\mathcal{F}_{\tau\subseteq\sigma}$ are referred to as *restriction maps*. Cellular sheaves can be defined on cell complexes, a slight generalization of simplicial complexes, hence their name. For simplicity of presentation, we restrict our attention to simplicial complexes.

The motivation for considering cellular sheaves in the context of geometric deep learning is that given a cellular sheaf $\mathcal{F}$, there is a corresponding sheaf Laplacian $L_{\mathcal{F}}$. Just as how the message passing operation in many graph convolutional neural networks is defined in terms of the graph Laplacian $L = D - A$ [55], the sheaf Laplacian $L_{\mathcal{F}}$ can be used to define a message passing operation.

To begin defining the sheaf Laplacian for a cellular sheaf $\mathcal{F}$ over a finite simplicial complex $K$, assume a total order is placed on the vertices of $K$, so that each simplex $\sigma \in K$ may be written $\sigma = [v_0, v_1, \ldots, v_n]$ with $v_0 < v_1 < \cdots < v_n$. Given a $k$-simplex $\sigma = [v_0, v_1, \ldots, v_k]$, we denote its $(k-1)$-dimensional faces by $f_i(\sigma) := [v_0, \ldots, \hat{v}_i, \ldots, v_k]$ for $i \in \{0, \ldots, k\}$, where $\hat{v}_i$ denotes the omission of vertex $v_i$.

An $n$-chain of $(K, \mathcal{F})$ is an element $x \in C_n(K; \mathcal{F})$, where $C_n(K; \mathcal{F}) := \bigoplus_{\sigma \in K_n} \mathcal{F}(\sigma)$ denotes the direct sum of the vector spaces $\mathcal{F}(\sigma)$ over all $n$-dimensional simplices $\sigma \in K_n$. An $n$-chain $x \in C_n(K; \mathcal{F})$ can be represented as a formal sum $x = \sum_i y_i \sigma_i$ where $\sigma_i$ are $n$-simplices of $K$ and $y_i \in \mathcal{F}(\sigma_i)$. The $n$-th boundary map $\partial_n : C_n(K; \mathcal{F}) \to C_{n-1}(K; \mathcal{F})$ is the linear transformation given by defining $\partial_n$ on vectors $y \in \mathcal{F}(\sigma)$ within the stalk of an $n$-simplex $\sigma$ to be

$$\partial_n(y) = \sum_{i=0}^{n+1} (-1)^i \mathcal{F}^{\mathsf{T}}_{f_i(\sigma) \subseteq \sigma}(y) f_i(\sigma). \tag{6.1}$$

The $n$-th sheaf Laplacian of $(K, \mathcal{F})$ is the linear transformation $L^n_{\mathcal{F}} : C_n(K; \mathcal{F}) \to C_n(K; \mathcal{F})$ defined by

$$L^n_{\mathcal{F}} = \partial_n^* \partial_n + \partial_{n+1} \partial_{n+1}^*, \tag{6.2}$$

where $\partial_n^* : C_{n-1}(K; \mathcal{F}) \to C_n(K; \mathcal{F})$ denotes the adjoint of $\partial_n$. In the case that $\mathcal{F}(\sigma) = \mathbb{R}$ for all $\sigma \in K$ and $\mathcal{F}_{\tau \subseteq \sigma} = \mathrm{id}_{\mathbb{R}}$ for all pairs $\tau \subseteq \sigma$, the $n$-th sheaf Laplacian $L^n_{\mathcal{F}}$ is simply the combinatorial Laplacian of the simplicial complex $K$ [36]. For further information on cellular sheaves and sheaf Laplacians, we refer the interested reader to [45].

### Cellular Sheaves over Graphs

Although the sheaf Laplacians $L^n_{\mathcal{F}}$ for multiple dimensions $n$ could be used as message passing operations within a sheaf neural network, due to the prevalence of graph-based data, one is primarily concerned with cellular sheaves over graphs, which are 1-dimensional simplicial complexes, and the 0th sheaf Laplacian $L^0_{\mathcal{F}}$, denoted by $L_{\mathcal{F}}$.

Given a sheaf $\mathcal{F}$ over a graph $G$ with $n$ nodes, and assuming the stalk $\mathcal{F}(u)$ over each node $u \in V(G)$ has the same dimension $d$, the sheaf Laplacian, $L_{\mathcal{F}}$, may be represented as a block matrix of size $nd \times nd$. The off-diagonal blocks $L_{ij} \in \mathbb{R}^{d\times d}$ are $L_{ij} = -\mathcal{F}^{\mathsf{T}}_{u_j \subseteq e}\mathcal{F}_{u_i \subseteq e}$, whenever there exists an edge between the nodes $u_i$ and $u_j$, and $L_{ij} = 0$ otherwise. The diagonal blocks are $L_{ii} = \sum_{u_i \subseteq e} \mathcal{F}^{\mathsf{T}}_{u_i \subseteq e}\mathcal{F}_{u_i \subseteq e}$, where the sum is over all edges $e$ such that $u_i \subseteq e$. If we let $D$ denote the $nd \times nd$ block diagonal matrix whose diagonal blocks are $L_{ii}$, and all other blocks are zero matrices, then the normalized 0th sheaf Laplacian $\Delta_{\mathcal{F}}$ defined in [46] is $\Delta_{\mathcal{F}} := D^{-1/2}L_{\mathcal{F}}D^{-1/2}$. Note $D$ is symmetric positive semi-definite, hence there exists a matrix $D^{1/2}$ such that $D^{1/2}D^{1/2} = D$. Moreover, we only consider sheaves such that the restriction maps $\mathcal{F}_{u\subseteq e}$ are invertible, in which case $D$ is invertible, and so $D^{-1/2}$ is well-defined.

## 6.1.3 Sheaf Diffusion

Let $G$ be a graph with $n$ nodes and with $d$-dimensional node features, and let $\mathcal{F}$ be a cellular sheaf over $G$ with $d$-dimensional stalks. We represent the node features as an $nd$-dimensional vector $X \in \mathbb{R}^{nd}$ so that the normalized sheaf Laplacian $\Delta_{\mathcal{F}}$, as a $nd \times nd$ block matrix, updates $X$ by matrix-vector multiplication $\Delta_{\mathcal{F}}X$. Motivated by how the graph convolutional layers of [55] can be viewed as a discretization of a heat diffusion process $\frac{d}{dt}X(t) = -\Delta X(t)$ defined in terms of the normalized graph Laplacian $\Delta$, the authors of [12] consider a similar diffusion process which now uses the sheaf Laplacian:

$$X(0) = X, \quad \tfrac{d}{dt}X(t) = -\Delta_{\mathcal{F}}X(t). \tag{6.3}$$

Applying the Euler method with unit time-steps to approximate solutions to Eq. (6.3) yields

$$X(t+1) = X(t) - \Delta_{\mathcal{F}}X(t). \tag{6.4}$$

The authors of [12] use Eq. (6.4) to define sheaf diffusion layers as follows. Allowing for $f \in \mathbb{N}$ distinct feature channels, let $X \in \mathbb{R}^{nd\times f}$ be a feature matrix. Given a sheaf Laplacian operator $\Delta_{\mathcal{F}(t)}$ for layer $t$, a sheaf diffusion layer is defined by setting

$$X_{t+1} = X_t - \sigma\big(\Delta_{\mathcal{F}(t)}(I_n \otimes W_{t,1})X_t W_{t,2}\big). \tag{6.5}$$

Here $W_{t,1} \in \mathbb{R}^{d\times d}$ and $W_{t,2} \in \mathbb{R}^{f\times f}$ are weight matrices for layer $t$, $\sigma$ is a non-linear activation function, $\otimes$ denotes a Kronecker product.

As a measure of the expressive power of sheaf layers based on Eq. (6.4), Bodnar et al. [12] study the ability of the sheaf diffusion process in Eq. (6.3) to linearly separate the features of nodes belonging to different classes.

**Definition 6.1.** *A class of sheaves $\mathcal{H}^d$ with d-dimensional stalks is said to have* linear separation power *over a class of labeled graphs $\mathcal{G}$ if for every $G \in \mathcal{G}$, there exists a sheaf $\mathcal{F} \in \mathcal{H}^d$ such that the sheaf diffusion process determined by $\mathcal{F}$ on $G$ linearly separates the features $X(t)$ of the classes of $G$ in the time limit $t \to \infty$, for almost all initial conditions $X(0)$.*

As part of a sheaf learning mechanism within a sheaf neural network, a fixed class of sheaves $\mathcal{H}^d$ is parameterized, and the input data is mapped to a suitable sheaf $\mathcal{F} \in \mathcal{H}^d$ through this parameterization. For a node classification task with $C$ classes, the collection of sheaves $\mathcal{H}^d$ used would ideally have linear separation power over the collection of connected graphs with $C$ classes. While one could always take $\mathcal{H}^d$ to be as general as possible, it can be preferable to use a more restricted collection of sheaves, as this can reduce the number of model parameters involved in learning a cellular sheaf. For example, learning cellular sheaves whose restriction maps are diagonal or special orthogonal matrices involves fewer parameters than when the restriction maps are general linear maps. Recall that a special orthogonal matrix, i.e. element of $SO(n)$, is a matrix $Q$ such that $Q^T = Q^{-1}$ and $\det(Q) = 1$.

It was shown that the class of sheaves $\mathcal{H}^d_{\text{diag}}$ with invertible diagonal restriction maps has linear separation power over the collection $\mathcal{G}_C$ of connected graphs with $C$ classes whenever $C \leq d$ [12]. However, for $\mathcal{H}^d_{so}$, the class of sheaves with $d \times d$ special orthogonal restriction maps, [12, Proposition 13] only guarantees the existence of a $d$ such that $\mathcal{H}^d_{so}$ has linear separation power over $\mathcal{G}_C$ when $C \leq 8$.

**Remark 6.1.** In [12], sheaves with orthogonal restriction maps, rather than special orthogonal, are considered. However, since the sheaf learning mechanism in their SNN is a continuous parametrization of the orthogonal group $O(n)$ by the connected space $\mathbb{R}^m$, it must have its image contained in a connected component of $O(n)$. Given that this implies that the restriction maps $\mathcal{F}_{u \subseteq e}$ will either all have determinant 1 or all have determinant $-1$, the maps $\mathcal{F}^{\mathsf{T}}_{u \subseteq e}\mathcal{F}_{v \subseteq e}$ used to define the sheaf Laplacian must always have determinant equal to 1. Thus it is effectively equivalent to learning restriction maps belonging to $SO(n)$.

The class of sheaves with special orthogonal restriction maps is of particular interest for sheaf neural networks—including our Bayesian sheaf neural networks—due to both theoretical and practical properties. From the theoretical point of view, certain results concerning the kernel of $\Delta_{\mathcal{F}}$ can be proven for sheaves using rotations as restriction maps [12, Section 3.1]. Regarding the more practical aspects, rotations offer a nice middle-ground between diagonal and general linear map in terms of balancing the number of parameters with the expressivity of the resulting sheaf. Moreover, when using rotations, the blocks along the diagonal of the sheaf Laplacian $L_{\mathcal{F}}$ are scalar multiples of the identity matrix. This makes computing the normalized sheaf Laplacian $\Delta_{\mathcal{F}}$ easier. Because of this, we take a moment to improve existing results on the linear separation power of the class of sheaves $\mathcal{H}^d_{so}$ with $d \times d$ special orthogonal restriction maps. In particular, we prove that for any $C \in \mathbb{N}$, we may find $d$ large enough such that $\mathcal{H}^d_{so}$ has linear separation power over $\mathcal{G}_C$. In the following proposition, $\lfloor \cdot \rfloor$ denotes the floor function.

**Proposition 6.1.** *For any $d \geq 1$, if $\mathcal{G}$ denotes the class of connected graphs with $C \leq 7\lfloor \frac{d}{4} \rfloor$ classes, then $\mathcal{H}^d_{so}$ has linear separation power over $\mathcal{G}$.*

In what follows, given a cellular sheaf $\mathcal{F}$ over a graph $G = (V, E)$ and initial conditions $X(0) = X$, let $X(t)$ denote the solution to Eq. (6.3). The limit of $X(t)$ as $t \to \infty$ exists, and is equal to the orthogonal projection of $X(0)$ onto $\ker(\Delta_{\mathcal{F}})$ [47]. If $\mathcal{F}$ has $d$-dimensional stalks and $G$ has $n$ nodes, although $X(t) \in \mathbb{R}^{nd}$, it will be useful to equally regard $X(t)$ as an $n \times d$ matrix. With this in mind, let $X(t)_u$ denote the row of $X(t)$ corresponding to node $u$. We write $u \in V_l$ if node $u$ has label $l$. For a cellular sheaf $\mathcal{F}$ with $d$-dimensional stalks over a graph $G = (V, E)$ with labeling $\ell : V \to \mathcal{L}$ and initial conditions $X(0)$, we say that $\mathcal{F}$ *linearly separates the classes of* $(G, X(0), \ell)$ if for each label $l \in \mathcal{L}$, the sets $\{\lim_{t\to\infty} X(t)_u : u \in V_l\} \subseteq \mathbb{R}^d$ and $\{\lim_{t\to\infty} X(t)_u : u \notin V_l\} \subseteq \mathbb{R}^d$ are linearly separable. With this terminology, Definition 6.1 states that a collection of cellular sheaves $\mathcal{H}$ has linear separation power over a collection of labelled graphs $\mathcal{G}$, if for each graph $G = (V, E) \in \mathcal{G}$ with labelling $\ell : V \to \mathcal{L}$, there exists a cellular sheaf $\mathcal{F} \in \mathcal{H}$ over $G$ such that $\mathcal{F}$ linearly separates the classes of $(G, X(0), \ell)$ for almost all initial conditions $X(0)$, i.e. all but a measure zero subset of the initial conditions.

*Proof of Proposition 6.1.* Let $d \geq 1$ and let $d = 4k + m$ for $0 \leq m < 4$. That is, $k = \lfloor \frac{d}{4} \rfloor$. Since proving the proposition for $C = 7k$ also proves the cases $C < 7k$, we may assume that $C = 7k$. Let $\mathcal{L} = \{1, 2, \ldots, C\}$ denote the set of class labels for the nodes of a graph $G = (V, E) \in \mathcal{G}$ and let $\ell : V \to \mathcal{L}$ denote the labelling of the nodes of $G$. For each $1 \leq i \leq k$, let $J_i = \{7(i-1)+1, 7(i-1)+2, \ldots, 7(i-1)+7\}$, let $\mathcal{L}_i = \{j : j \in J_i\} \cup \{0\}$, and let $\ell_i : V \to \mathcal{L}_i$ denote a relabeling of $G$ where for each $u \in V$, $f_i(u) = f(u)$ whenever $f(u) \in J_i$, and $f_i(u) = 0$ otherwise. That is, the relabeling $\ell_i$ retains the seven labels in $J_i$, and replaces all other labels with a single distinct eighth label.

By [12, Proposition 13], for each $1 \leq i \leq k$ we may find a cellular sheaf $\mathcal{F}^i \in \mathcal{H}^4_{so}$ with 4-dimensional stalks which linearly separates the eight classes of $(G, X(0), \ell_i)$ for almost all initial conditions $X(0) \in \mathbb{R}^{4n}$. Let $\mathcal{F}$ denote the cellular sheaf with restriction maps $\mathcal{F}_{u \subseteq e} = \mathcal{F}^1_{u \subseteq e} \oplus \mathcal{F}^2_{u \subseteq e} \oplus \cdots \oplus \mathcal{F}^k_{u \subseteq e} \oplus I_m$. We claim that $\mathcal{F}$ linearly separates the classes of the original labelling $(G, X(0), \ell)$ for almost all initial conditions $X(0) \in \mathbb{R}^{n \times d}$.

Given initial conditions $X(0) \in \mathbb{R}^{n \times d}$, for each $1 \leq i \leq k$ let $X_i(0) \in \mathbb{R}^{n \times 4}$ be the matrix consisting of columns $4i - 3$ through $4i$ of $X(0)$. Note that for almost all initial conditions $X(0)$, it will be true that for all $1 \leq i \leq k$, $\mathcal{F}^i$ linearly separates the classes of $(G, X_i(0), \ell_i)$. Next, for any class $l \in \mathcal{L}$, find $i$ such that $l \in J_i$. Since $\mathcal{F}^i$ linearly separates the classes of $(G, X_i(0), \ell_i)$, we may find a hyperplane $A \subseteq \mathbb{R}^4$ which separates the class $l$ from all other classes in $\mathcal{L}_i$, and hence $\mathcal{L}$ as well. Then $\mathbb{R}^4 \oplus \cdots \oplus A \oplus \cdots \oplus \mathbb{R}^4 \oplus \mathbb{R}^m$ is a hyperplane in $\mathbb{R}^d$ separating class $l$ from all other classes. □

**Remark 6.2.** An extra consequence of Proposition 6.1 is that it also provides a new upper bound on the stalk dimension $d$ necessary for $\mathcal{H}^d_{gen}$, the collection of sheaves with $d \times d$ general linear restriction maps, to have linear separation power over the collection of connected graph $\mathcal{G}_C$ with $C$ classes. This is due to the simple fact that $\mathcal{H}^d_{so} \subseteq \mathcal{H}^d_{gen}$.

## 6.2 Bayesian Sheaf Neural Networks

We briefly establish some notation. Let $P$ be a distribution on a measurable space $\mathcal{X}$ with density $p$ with respect to some reference measure $\mu$. If $X$ is a random variable with distribution $P$, we write $X \sim P$. For a function $f : \mathcal{X} \to \mathbb{R}$, we denote by $\mathbb{E}_P[f(X)]$ the expectation of $f(X)$ with respect to the distribution $P$, that is, $\mathbb{E}_P[f(X)] = \int_{\mathcal{X}} p(x) f(x)\, \mu(dx)$. If $Q$ is another distribution over $\mathcal{X}$, and $Q$ has density

$q$ with respect to $\mu$, the Kullback-Leibler divergence of $P$ from $Q$ is

$$\mathrm{KL}(P \parallel Q) = \int_{\mathcal{X}} p(x) \log \left( \frac{p(x)}{q(x)} \right) \mu(dx).$$

To simplify notation, we will denote a distribution by its density function. For instance, an expression of the form $X \sim p(X|Y)$ denotes that the random variable $X$ is distributed according to the conditional distribution of $X$ given $Y$.

### 6.2.1 Problem Statement

Consider a semi-supervised node classification problem, in which we have a graph dataset $G = (V, E)$ together with a node feature matrix $X$, a collection of observed node labels, $y$, and a collection of unobserved node labels $y^*$, all of which are random samples. Motivated by the expressive power of the sheaf diffusion process, we model the conditional probability $p_\theta(y|X)$ using a sheaf neural network whose layers are given by Eq. 6.5, with network parameters $\theta$ and a cellular sheaf $\mathcal{F}$.

Rather than take the learned cellular sheaf $\mathcal{F}$ (or sheaves) to be a deterministic function of $X$ as in [12], we instead regard $\mathcal{F}$ as a latent random variable of our model. See Figure 6.1 for a depiction of the probability graphical model. Note that the posterior distribution $p_\theta(\mathcal{F}|X, y)$ will be intractable due to the fact that the labels $y$ are modeled as the output of a neural network. Hence, to do approximate posterior inference of $\mathcal{F}$ given the feature matrix, $X$, the observed node labels $y$, and model parameters $\theta$, we assume the intractable posterior distribution $p_\theta(\mathcal{F}|X, y)$ belongs to a predetermined family of variational distributions $q_\phi(\mathcal{F}|X, y)$ parameterized by weights $\phi$.

We begin by noting that the log-likelihood $\log p_\theta(X, y)$ can be written

$$\log p_\theta(X, y) = \mathrm{KL}(q_\phi(\mathcal{F}|X, y) \parallel p_\theta(\mathcal{F}|X, y)) + \mathcal{L}(\theta, \phi; X, y), \tag{6.6}$$

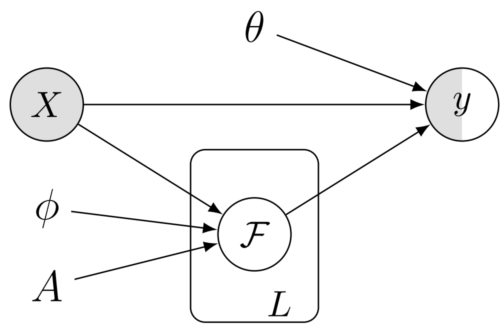


**Figure 6.1:** Probability graphical model of the Bayesian sheaf neural network for a semi-supervised node classification problem. The adjacency matrix of the graph is $A$, and $\theta$, $\phi$ are weights of the network.

where

$$\mathcal{L}(\theta, \phi; X, y) := \mathbb{E}_{q_\phi(\mathcal{F}|X,y)}\big[-\log q_\phi(\mathcal{F}|X, y) + \log p_\theta(\mathcal{F}, X, y)\big]. \tag{6.7}$$

The quantity $\mathcal{L}(\theta, \phi; X, y)$ is known as the *evidence lower bound* or ELBO, due to the fact that the KL divergence from one distribution to another is always non-negative, hence $\log p_\theta(X, y) \geq \mathcal{L}(\theta, \phi, X, y)$. Note from Eq. (6.6) that maximizing the ELBO with respect to the parameters $\theta$ and $\phi$ has the effect of maximizing the log-likelihood of the data, as well as minimizing the KL divergence of the variational distribution $q_\phi(\mathcal{F}|X, y)$ from the true posterior distribution $p_\theta(\mathcal{F}|X, y)$. For this reason, the ELBO is commonly used as quantity to be optimized within variation Bayesian inference. See [11] for an overview of variation inference. In order to maximize $\mathcal{L}(\theta, \phi; X, y)$ with respect to the parameters $\theta, \phi$, we use the following equivalent expression:

$$\mathcal{L}(\theta, \phi; X, y) = \mathbb{E}_{q_\phi(\mathcal{F}|X,y)}\big[\log p_\theta(X, y|\mathcal{F})\big] - \mathrm{KL}(q_\phi(\mathcal{F}|X, y) \parallel p_\theta(\mathcal{F})). \tag{6.8}$$

In many cases, a closed-form expression can be derived for the KL-divergence term in the right hand side of Eq. (6.8), leaving only the gradient of the expectation term with respect to $\theta$ and $\phi$ to be estimated. See Appendix B for calculations of the KL divergence in certain cases. Similar to [54], we estimate $\nabla_\phi \mathbb{E}_{q_\phi(\mathcal{F}|X,y)}[\log p_\theta(X, y|\mathcal{F})]$ by making use of the so-called *reparameterization trick*. In our case, for a random cellular sheaf $\mathcal{F}$ distributed as $\mathcal{F} \sim q_\phi(\mathcal{F}|X, y)$, the reparameterization trick expresses

$\mathcal{F}$ as a deterministic function $\mathcal{F} = g_\phi(\epsilon, X)$ of random noise $\epsilon \sim p(\epsilon)$. This allows us to rewrite the expectation with respect to $q_\phi(\mathcal{F}|X, y)$ as an expectation with respect to $p(\epsilon)$ which in turn allows the gradient to be moved within the expectation. That is,

$$\begin{aligned}\nabla_\phi \mathbb{E}_{q_\phi(\mathcal{F}|X,y)}[\log p_\theta(X, y|\mathcal{F})] &= \nabla_\phi \mathbb{E}_{p(\epsilon)}[\log p_\theta(X, y|g_\phi(\epsilon, X))] \\ &= \mathbb{E}_{p(\epsilon)}[\nabla_\phi \log p_\theta(X, y|g_\phi(\epsilon, X))] \\ &\approx \frac{1}{K}\sum_{k=1}^{K} \nabla_\phi \log p_\theta(X, y|g_\phi(\epsilon^{(k)}, X)),\end{aligned}$$

where $\epsilon^{(k)}$ are independent and identically distributed samples of $p(\epsilon)$ for $k \in \{1, 2, \ldots, K\}$.

Hence, when the KL divergence term of Eq. (6.8) can be computed analytically, if $\mathcal{F}^{(k)} = g_\phi(\epsilon^{(k)}, X)$ for $\epsilon^{(k)} \sim p(\epsilon)$, our estimator is

$$\widetilde{\mathcal{L}}_\lambda(\theta, \phi; X, y) = \frac{1}{K}\sum_{k=1}^{K} \log p_\theta(X, y|\mathcal{F}^{(k)}) - \lambda\, \mathrm{KL}(q_\phi(\mathcal{F}|X, y) \parallel p_\theta(\mathcal{F})). \tag{6.9}$$

Here, the weight $\lambda$ of the KL divergence term is included to combat issues with the KL divergence term vanishing during training, which we cyclically anneal during training in a manner similar to [38]. When the KL divergence term of Eq. (6.8) cannot be computed analytically, we may estimate it in a similar manner by noting that $\mathrm{KL}(q_\phi(\mathcal{F}|X, y) \parallel p_\theta(\mathcal{F})) = \mathbb{E}_{q_\phi(\mathcal{F}|X,y)}[\log(q_\phi(\mathcal{F}|X, y)/p_\theta(\mathcal{F}))]$. In this case, the estimator, also denoted by $\widetilde{\mathcal{L}}_\lambda$, is given by

$$\widetilde{\mathcal{L}}_\lambda(\theta, \phi; X, y) = \frac{1}{K}\sum_{k=1}^{K} \Big( \log p_\theta(X, y|\mathcal{F}^{(k)}) - \lambda \log \frac{q_\phi(\mathcal{F}^{(k)}|X, y)}{p_\theta(\mathcal{F}^{(k)})} \Big), \tag{6.10}$$

where again, $\mathcal{F}^{(k)} = g_\phi(\epsilon^{(k)}, X)$ for $\epsilon^{(k)} \sim p(\epsilon)$.

### 6.2.2 Variational Sheaf Learner

We consider a variational sheaf learner in three different cases: where the learned sheaf $\mathcal{F}$ has restriction maps which are assumed to be (1) general linear, (2) special orthogonal, or (3) invertible diagonal. In all cases, we make the simplifying assumption that the variational distribution $q_\phi(\mathcal{F}|X, y)$ from the previous section factors as the product $\prod_{u\subseteq e} q_\phi(\mathcal{F}_{u\subseteq e}|X, y)$ over all incident node-edge pairs $u \subseteq e$. While this assumption somewhat restricts the family of variational distributions, it makes computing the KL divergence term in the ELBO tractable. For further details, see Appendix B. We take each factor $q_\phi(\mathcal{F}_{u\subseteq e}|X, y)$ to be a distribution $Q(\mu_{u\subseteq e}, \sigma_{u\subseteq e})$ which depends on the type of restriction map used within the sheaf, where the distributional parameters $\mu_{u\subseteq e}, \sigma_{u\subseteq e}$ are a function of the node features $X$ and $\phi$ as follows. For each incident node-edge pair $u \subseteq e$, if $u' \subseteq e$ denotes the other node incident with $e$, a multi-layer perceptron with learned weights $\phi$, denoted $\mathrm{MLP}_\phi$, maps the concatenated vector $[x_u||x_{u'}]$ to distributional parameters $\mu_{u\subseteq e}$ and $\sigma_{u\subseteq e}$:

$$[\, \mu_{u\subseteq e} \,||\, \sigma_{u\subseteq e} \,] = \mathrm{MLP}_\phi([x_u||x_{u'}]). \tag{6.11}$$

**Case 1: General Linear.** In the case where the sheaf restriction maps $\mathcal{F}_{u\subseteq e} : \mathcal{F}(u) \to \mathcal{F}(e)$ are general linear maps, $\mu_{u\subseteq e}$ and $\sigma_{u\subseteq e}$ are $d^2$-dimensional vectors. We take $Q(\mu_{u\subseteq e}, \sigma_{u\subseteq e})$ to be normal with diagonal covariance so that for $z \sim \mathcal{N}_{d^2}(\mu_{u\subseteq e}, \mathrm{diag}(\sigma_{u\subseteq e}))$ we set $\mathcal{F}_{u\subseteq e}$ to be the vector $z$ rearranged into a $d \times d$ matrix. The reparameterization trick in this case writes $z = \mu_{u\subseteq e} + \mathrm{diag}(\sigma_{u\subseteq e})\epsilon$ for $\epsilon \sim \mathcal{N}(\mathbf{0}, I_{d^2})$ so that $\mathcal{F}_{u\subseteq e}$, being the matrix obtained by reshaping the vector $z$, is a deterministic function of $\mu_{u\subseteq e}$, $\sigma_{u\subseteq e}$, and $\epsilon$. We use a standard normal prior where $p_\theta(\mathcal{F}) = \prod_{u\subseteq e} p_\theta(\mathcal{F}_{u\subseteq e})$ and $p_\theta(\mathcal{F}_{u\subseteq e}) = \mathcal{N}_{d^2}(\mathbf{0}, I_{d^2})$.

**Case 2: Special Orthogonal.** The case where sheaf restriction maps are assumed to be special orthogonal presents challenges. Typical probability distribution on $SO(d)$, such as the matrix Langevin distribution, are not easily reparameterizable. This makes incorporating such distributions into neural networks, that are trained via back-propagation, difficult. On the other hand, while reparameterizable distributions on $SO(d)$ have been constructed in [31] using the exponential map, the density functions of such distributions can be difficult to compute exactly, which obstructs one from computing, or even efficiently estimating, the KL divergence term in Eq. (6.8). In Section 6.3 we define what we refer to as Cayley distributions, a family of reparameterizable distributions $\mathcal{C}_d(M, \kappa)$ on $SO(d)$ with closed form expressions for their probability density functions. We use $\mathcal{C}_d(M, \kappa)$ as the variational distributions for sheaves with special orthogonal restriction maps. Hence when $\mathcal{F}_{u \subseteq e} \in SO(d)$, the distributional parameters are a mean rotation $\mu_{u \subseteq e} \in SO(d)$ and scalar concentration parameter $\sigma_{u \subseteq e} \in [0, 1)$, and we have $Q(\mu_{u \subseteq e}, \sigma_{u \subseteq e}) = \mathcal{C}_d(\mu_{u \subseteq e}, \sigma_{u \subseteq e})$. The reparameterization of the Cayley distributions is described in Remark 6.4. We use a prior distribution $p_\theta(\mathcal{F}) = \prod_{u \subseteq e} p_\theta(\mathcal{F}_{u \subseteq e})$ where each $p_\theta(\mathcal{F}_{u \subseteq e})$ is a uniform distribution on $SO(d)$.

**Case 3: Invertible Diagonal.** In the case where the sheaf restriction maps $\mathcal{F}_{u \subseteq e} : \mathcal{F}(u) \to \mathcal{F}(e)$ are assumed to be invertible diagonal matrices with respect to the standard basis, $\mu_{u \subseteq e}$ and $\sigma_{u \subseteq e}$ are $d$-dimensional vectors and we take $q_\phi(\mathcal{F}_{u \subseteq e} | X, y) = Q(\mu_{u \subseteq e}, \sigma_{u \subseteq e})$ to be a normal distribution with diagonal covariance, so that $\mathcal{F}_{u \subseteq e} = \mathrm{diag}(z)$ for $z \sim \mathcal{N}_d(\mu_{u \subseteq e}, \mathrm{diag}(\sigma_{u \subseteq e}))$. The reparameterization $\mathcal{F} = g_\phi(X, \epsilon)$ in this case is given by setting $\mathcal{F}_{u \subseteq e} = \mathrm{diag}(\mu_{u \subseteq e} + \sigma_{u \subseteq e}\epsilon)$ for $\epsilon \sim \mathcal{N}(\mathbf{0}, I_d)$. Lastly, we use a standard normal prior, where $p_\theta(\mathcal{F}) = \prod_{u \subseteq e} p_\theta(\mathcal{F}_{u \subseteq e})$ and $p_\theta(\mathcal{F}_{u \subseteq e}) = \mathcal{N}_d(\mathbf{0}, I_d)$.

### 6.2.3 Bayesian Sheaf Neural Network Overview

We give an overview of the Bayesian sheaf neural network architecture, which is also summarized in Figure 6.2. The variational sheaf learning mechanism maps the input node features to the sheaf distributional parameters as described in Section 6.2.2, though we briefly describe here as well. A linear layer first resizes the input node features to dimension $k$. Specifically, we use $k = df$ where $d$ and $f$ are the stalk dimensions and number of feature channels respectively. Then, for each incident

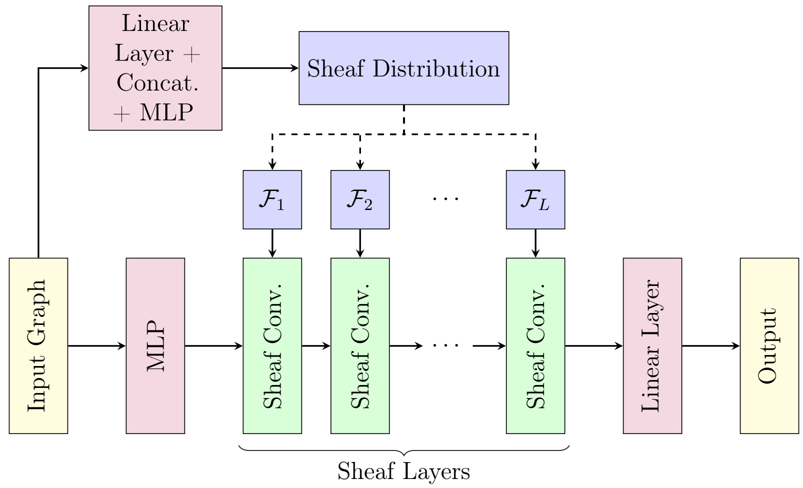


**Figure 6.2:** Visual depiction of BSNN Architecture. The variational sheaf learning mechanism is depicted in the upper half of the diagram. Dashed arrows represent sampling from the sheaf distribution.

node-edge pair $u \subseteq e$, the features of the pair of nodes corresponding to the edge $e$ are concatenated. Hence a node feature matrix of size $n \times k$ is transformed by this concatenation operation into a matrix of size $2e \times 2k$, where $n$ denotes the number of nodes, and $e$ denotes the number of edges in the graph. Lastly, an MLP maps these features indexed by the incident node-edge pairs $u \subseteq e$ to distributional parameters $\mu_{u \subseteq e}$ and $\sigma_{u \subseteq e}$ as described by Eq. (6.11), which define the distribution over the cellular sheaves. From this distribution, a collection of independent and identically distributed sheaves $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_L$ are sampled: one for each layer of the network.

Separately, a multi-layer perceptron maps the input node features $X \in \mathbb{R}^{n \times m}$ to features of dimension $X' \in \mathbb{R}^{n \times df}$ where $d$ is the stalk dimension and $f$ is the number of hidden channels. The features $X'$ are reshaped to size $nd \times f$. Sheaf layers described by Eq. (6.5) update the transformed features $X'$ using the sampled sheaves $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_L$. Finally, a linear layer maps the output of the final sheaf layer to a vector of class probabilities.

During training, a one-sample Monte Carlo estimate of Eq. (6.9) or Eq. (6.10) is computed, and the loss is back-propagated to update the network weights. During testing, ensembling is performed to obtain point estimates for the unobserved node labels $y^*$.

## 6.3 Reparameterizable Distributions on $SO(n)$

In order to perform variational inference for our BSNN when cellular sheaves with special orthogonal restriction maps are used, we require a family of distributions on $SO(n)$ that are both reparameterizable and have tractable density functions. We define a novel family of such distributions via the Cayley transform [23].

### 6.3.1 Cayley Distributions on $SO(n)$

We denote the set of all $n \times n$ skew-symmetric matrices by $\mathfrak{so}(n)$. The *Cayley transform* is a function $C : \mathfrak{so}(n) \to SO(n)$ defined by

$$C(A) = (I - A)^{-1}(I + A) = 2(I - A)^{-1} - I. \tag{6.12}$$

If $A$ is an $n \times n$ skew-symmetric matrix, then $x^\top Ax = 0$ for all $x \in \mathbb{R}^n$ due to the fact that $x^\top Ax = (x^\top Ax)^\top = -x^\top Ax$. Hence if $(I - A)x = 0$, then $x^\top(I - A)x = x^\top x = 0$ implying that $x = 0$. Thus $I - A$ is invertible, showing that the Cayley transform is well-defined. If $P \in SO(n)$ does not have $-1$ as an eigenvalue, $C^{-1}(P)$ exists and is given by

$$C^{-1}(P) = (P - I)(I + P)^{-1} = I - 2(I + P)^{-1}. \tag{6.13}$$

Hence $C$ is injective, and all but a measure zero subset of $SO(n)$ is contained in the image of $C$.

**Remark 6.3.** The Cayley transform is sometimes alternatively defined as $C(A) = (I + A)^{-1}(I - A) = (I - A)(I + A)^{-1}$. Since the two definitions differ only by the precomposition of $C$ with the reflection $A \mapsto -A$, either choice of definition could be used without meaningful change.

For a vector $\phi \in \mathbb{R}^{n(n-1)/2}$, let $X_\phi$ denote the skew symmetric matrix containing the entries of $\phi$ as its lower triangular entries. Explicitly, if $\phi = (\phi_1, \phi_2, \ldots, \phi_{n(n-1)/2})$ and if $x_{ij}$ denotes the $i,j$-th entry of $X_\phi$, then for $i > j$, set $x_{ij} = -x_{ji} = \phi_{ij-j(j+1)/2}$ and set $x_{ii} = 0$. We may vectorize the Cayley transform and regard it as a map $\widetilde{C} : \mathbb{R}^{n(n-1)/2} \to \mathbb{R}^{n^2}$ by setting $\widetilde{C}(\phi) = \operatorname{vec} C(X_\phi)$. The Jacobian (determinant) of $\widetilde{C}$ is $J\widetilde{C}(\phi) = \det(D\widetilde{C}(\phi)^\mathsf{T} D\widetilde{C}(\phi))^{1/2}$, where $D\widetilde{C}(\phi)$ denotes the total derivative of $\widetilde{C}$ at the point $\phi$. The Jacobian $J\widetilde{C}(\phi)$ can be expressed [61] in terms of $X_\phi$ by

$$J\widetilde{C}(\phi) = \frac{2^{3n(n-1)/4}}{\det(I + X_\phi)^{n-1}}. \tag{6.14}$$

A random variable $X$ taking values in $SO(n)$ is uniformly distributed if it is distributed according to the normalized Haar measure on $SO(n)$. Recall that a *left Haar measure* on a Lie group $G$, such as $SO(n)$, is a non-trivial measure $\mu$ on the Borel subsets of $G$ that is

(1) left translation invariant: $\mu(gS) = \mu(S)$ for all $g \in G$ and Borel $S \subset G$,

(2) finite on compact sets: $\mu(K) < \infty$ for compact $K \subseteq G$,

(3) outer regular on Borel sets: $\mu(S) = \inf\{\mu(U) : U \supseteq S,\ U \text{ open}\}$ for Borel $S \subseteq G$,

(4) inner regular on open sets: $\mu(U) = \sup\{\mu(K) : K \subseteq U,\ K \text{ compact}\}$ for open $U \subseteq G$.

Haar's theorem implies that there is a unique normalized left Haar measure, that is, a unique left Haar measure $\mu$ such that $\mu(G) = 1$. Right Haar measures are defined similarly but are right translation invariant, i.e. $\mu(Sg) = \mu(S)$ for all $g \in G$ and Borel $S \subset G$, rather than left translation invariant. Though in general left and right Haar measures need not be the same, for compact Lie groups, such as $SO(n)$, they coincide. Hence we may talk of simply a Haar measure on $SO(n)$.

**Definition 6.2.** *Let $C : \mathfrak{so}(n) \to SO(n)$ denote the Cayley transform and let $X$ be a uniformly distributed random variable on $SO(n)$. Given parameters $M \in SO(n)$ and*

$0 \le \kappa < 1$, *we define the* Cayley distribution $\mathcal{C}_n(M, \kappa)$ *on* $SO(n)$ *to be the distribution of* $Y = C(\frac{1-\kappa}{1+\kappa} C^{-1}(X))M$.

**Remark 6.4.** Note that by the definition of the Cayley distribution, a random variable $Y \sim \mathcal{C}_n(M, \kappa)$ can be expressed as a deterministic function of $M$, $\kappa$, and $X \sim U_{SO(n)}$. Hence these distributions are reparameterizable. For details on sampling from the uniform distribution on $SO(n)$, see for example [8].

In order to compute or efficiently estimate the KL divergence term in Eq. (6.9), an expression for the density of $\mathcal{C}_n(M, \kappa)$ is necessary. Using Eq. (6.14) for the Jacobian of the Cayley transform, we compute the density of $\mathcal{C}_n(M, \kappa)$ in the following theorem.

**Theorem 6.5.** *The Cayley distribution* $\mathcal{C}_n(M, \kappa)$ *has density*

$$f_n(P; M, \kappa) = (1 - \kappa^2)^{\frac{n(n-1)}{2}} \det\left(PM^\mathsf{T} - \kappa I\right)^{1-n}, \tag{6.15}$$

*with respect to the normalized Haar measure on* $SO(n)$.

Before proving Theorem 6.5, we first recall some preliminaries. For a subset $S$ of a metric space $M$ and for $\delta > 0$, let

$$\mathcal{H}^d_\delta(S) = \inf\Big\{\sum_{i=1}^\infty \alpha_d(\tfrac{1}{2}\operatorname{diam}(S))^d : \bigcup_{i=1}^\infty U_i \supseteq S, \operatorname{diam}(U_i) < \delta\Big\}$$

where $\alpha_d$ denotes the volume of a unit $d$-ball. The $d$-dimensional Hausdorff measure on the Borel subsets of $M$ is then given by $\mathcal{H}^d(S) = \lim_{\delta\to 0} H^d_\delta(S)$.

Let $\lambda^n$ denote the $n$-dimensional Lebesgue measure. The area formula [33] computes the Hausdorff measure of the image of a 1-to-1 and continuously differentiable function $f$ in terms of the Jacobian of $f$. We state a version presented in [80].

**Proposition 6.2** (Area formula)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}^m$ *be* 1*-to-*1 *and continuously differentiable. For any Borel subset* $A \subseteq \mathbb{R}^n$,

$$\int_A Jf(x)\lambda^n(dx) = \int_{f(A)} \mathcal{H}^n(dy).$$

*Moreover,*

$$\int_A g(f(x)) Jf(x) \lambda^n(dx) = \int_{f(A)} g(y) \mathcal{H}^n(dy), \tag{6.16}$$

*for all Borel measurable $g : \mathbb{R}^m \to \mathbb{R}$ for which one side of (6.16) exists.*

Equip $SO(n)$ with the metric induced by the Frobenius norm $\|A\| = \operatorname{tr}(A^\mathsf{T} A)$. Since left (or right) multiplication on $SO(n)$ is an isometry, it follows that the Hausdorff measure is a left (and right) translation invariant measure on $SO(n)$. Moreover, if $d = n(n-1)/2$ denotes the dimension of $SO(n)$, then $\mathcal{H}^d$ is a locally finite measure on $SO(n)$, and thus is a Haar measure by Haar's theorem.

Hence, Proposition 6.2 yields a change of variables formula for they Cayley transform. For example, if $X$ is a uniformly distributed random variable on $SO(n)$, then $C^{-1}(X)$ has density $p(x) = J\widetilde{C}(x)$ with respect to the Lebesgue measure on $\mathbb{R}^d$.

*Proof of Theorem 6.5.* Let $\mu$ denote the unique Haar measure on $SO(n)$ such that $\int_{SO(n)} d\mu = 1$. Let $X$ be a uniformly distributed random variable on $SO(n)$. Let $Y = C(\frac{1-\kappa}{1+\kappa} C^{-1}(X))$, and let $f_n(P; I, \kappa)$ denote the density of $Y$ with respect to $\mu$. For convenience, set $\gamma = \frac{1+\kappa}{1-\kappa}$. Then by Proposition 6.2 and the chain rule,

$$f_n(P; I, \kappa) = \gamma^{\frac{n(n-1)}{2}} J\widetilde{C}\Big(\gamma \widetilde{C}^{-1}(\operatorname{vec} P)\Big) J\widetilde{C}^{-1}(\operatorname{vec} P).$$

Using Eq. (6.14), we obtain

$$f_n(P; I, \kappa) = \gamma^{\frac{n(n-1)}{2}} \Big(\frac{\det(I + C^{-1}(P))}{\det(I + \gamma C^{-1}(P))}\Big)^{n-1}. \tag{6.17}$$

Using the equality $\det(I + C^{-1}(P)) = \frac{2^n}{\det(I+P)}$,

$$f_n(P; I, \kappa) = 2^{n(n-1)} \gamma^{\frac{n(n-1)}{2}} \det\Big((I + P)(I + \gamma C^{-1}(P))\Big)^{1-n}.$$

Next, from Eq. (6.13) we have that $I + \gamma C^{-1}(P) = (1+\gamma)I - 2\gamma(I+P)^{-1}$. By substituting and simplifying,

$$f_n(P; I, \kappa) = 2^{n(n-1)}\gamma^{\frac{n(n-1)}{2}} \det\Big((1+\gamma)(I+P) - 2\gamma I\Big)^{1-n}.$$

Factoring $1+\gamma$ out of the expression within the determinant, we have

$$f_n(P; I, \kappa) = 2^{n(n-1)}\Big(\frac{\gamma}{(1+\gamma)^2}\Big)^{\frac{n(n-1)}{2}} \det\Big(P - \frac{\gamma-1}{\gamma+1}I\Big)^{1-n},$$

or in terms of $\kappa$,

$$f_n(P; I, \kappa) = (1-\kappa^2)^{\frac{n(n-1)}{2}} \det\Big(P - \kappa I\Big)^{1-n}.$$

□

**Remark 6.6.** The family of Cayley distributions defined in [61] have density proportional to $\det(PM^\mathsf{T} + I)^\kappa$ of Eq. (6.15). Note that the concentration parameter $\kappa$ appears in the exponent of the determinant expression, whereas in our proposed family of Cayley distributions, $\kappa$ instead scales the identity matrix term within the determinant.

### 6.3.2 Special Cases of the Cayley Distribution

We now examine the distributions $\mathcal{C}_n(M, \kappa)$ for small values of $n$.

**Case n = 2.** If $P, M \in SO(2)$ are rotations by $\theta, \mu$ radians respectively, then $\mathcal{C}_2(M, \kappa)$ has density

$$f_2(P; M, \kappa) = \frac{1}{2\pi} \cdot \frac{1-\kappa^2}{1+\kappa^2 - 2\kappa\cos(\theta-\mu)}.$$

This is the density of a wrapped Cauchy distribution on $SO(2) \cong S^1$, see[52]. Hence for the case $n = 2$, the Cayley distributions coincide with known distributions.

**Case $\mathbf{n = 3}$.** Note that $SO(3)$ is homeomorphic to 3-dimensional real projective space $RP^3$, which is the quotient of $S^3$ under the identification of antipodal points. We show that $\mathcal{C}_3(M, \kappa)$ is related to an *angular central Gaussian* distribution [81] on $S^3$. Specifically, we show that $\mathcal{C}_3(M, \kappa)$ is the pushforward of an angular central Gaussian distribution on $S^3$ under the quotient map $\Phi : S^3 \to SO(3)$. If $x \in \mathbb{R}^{n+1}$ is a unit vector, $\Lambda \in \mathbb{R}^{(n+1)\times(n+1)}$ is a symmetric positive definite matrix, and $\alpha_n$ denotes the volume of the $n$-dimensional sphere $S^n$, the angular central Gaussian distribution $ACG_n(\Lambda)$ on $S^n$ has density

$$f(x; \Lambda) = \alpha_n^{-1} \det(\Lambda)^{-\frac{1}{2}} (x^\top \Lambda^{-1} x)^{-2}$$

with respect to the uniform measure on $S^n$.

**Proposition 6.3.** *Let $\Phi : S^3 \to \mathbb{RP}^3 \cong SO(3)$ denote the map which identifies antipodal points of $S^3$. For any $M \in SO(3)$ and $\kappa \in [0, 1)$, there exists a symmetric positive definite matrix $\Lambda_{M,\kappa} \in \mathbb{R}^{4\times 4}$ such that if $X \sim ACG_3(\Lambda_{M,\kappa})$, then $\Phi(X) \sim \mathcal{C}_3(M, \kappa)$.*

For a proof see Appendix A.

When $n = 2$ or $n = 3$, Eq. (6.14) for the Jacobian of the Cayley transform simplifies to $2^{3n(n-1)/4}(1 + \|\phi\|^2)^{1-n}$. One consequence of this fact is that $\mathcal{C}_n(I_n, \kappa)$ is the pushforward of a Cauchy distribution (multivariate in the case $n = 3$) under the Cayley transform. This fact makes it possible to compute the KL divergence of $\mathcal{C}_n(M, \kappa)$ from the uniform distribution on $SO(n)$ in these cases.

**Proposition 6.4.** *For $n = 2, 3$ the KL divergence of $\mathcal{C}_n(M, \kappa)$ from the uniform distribution is given by*

$$\begin{aligned}
\mathrm{KL}(\mathcal{C}_2(M, \kappa) \parallel U_{SO(2)}) &= -\log(1 - \kappa^2) \\
\mathrm{KL}(\mathcal{C}_3(M, \kappa) \parallel U_{SO(3)}) &= -\log(1 - \kappa^2) - 2\log(1 - \kappa) - 2\kappa.
\end{aligned}$$

For a proof see Appendix B.

Thus when learning a distribution of cellular sheaves with special orthogonal restriction maps in our BSNN, for the cases where the stalk dimension $d$ is 2 or 3, we are able to use Eq. (6.9) to estimate the ELBO by plugging in the expressions for the KL divergence from Proposition 6.4. When $d \geq 4$, we estimate the ELBO by simply using Eq. (6.10).

## 6.4 Node Classification Experiments

We compare our Bayesian sheaf neural networks (BSNN) against the sheaf neural networks (NSD) from [12] on the WebKB datasets, a set of real-world web-page datasets. These datasets have low edge-homophily coefficients, that is, there are relatively few edges which connect nodes belonging to the same class [87]. This is a type of graph data where sheaf neural networks are expected to outperform traditional graph neural networks. See Appendix C for further details on the datasets. To study the performance of our Bayesian sheaf models when training data is limited, we use 10 fixed splits of the datasets, each of which contains 32%/20%/48% of the nodes for training, validation, and testing respectively. For a given hyperparameter configuration, each model is trained and evaluated on the 10 fixed splits of the data and we compute the average validation accuracy. Identical hyperparameter sweeps are performed for each model. For the hyperparameter configuration that yields the best validation accuracy for a model, we train and evaluate the model on the 10 splits for 30 different random seeds. We report the mean test accuracies along with the standard deviation of the test accuracies across the 30 random seeds in Table 6.1. For each type of restriction map (diagonal, special orthogonal, and general linear), we compare the results of the BSNN model against the NSD model. The higher test accuracy is bolded, and further, a blue entry denotes that a Wilcoxon signed rank test indicates a statistically significant difference for a $p$-value threshold of $p = 0.05$. We also include results for a graph convolutional network (GCN) as a baseline comparison.

**Table 6.1:** Average test accuracies and standard deviations across 30 random seeds for the hyperparameter configuration yielding the highest validation accuracy for each model and dataset.

| | **Texas** | **Wisconsin** | **Cornell** |
|---|---|---|---|
| **Diag-BSNN** | 75.35±1.68 | **83.04**±0.87 | 72.23±1.76 |
| Diag-NSD | **75.49**±2.40 | 82.77±1.15 | **73.70**±1.45 |
| **SO(d)-BSNN** | **76.65**±1.48 | **82.02**±1.06 | **74.31**±1.10 |
| $O(d)$-NSD | 76.09±1.79 | 81.79±1.63 | 74.07±1.58 |
| **Gen-BSNN** | **76.32**±1.65 | **82.03**±0.94 | **72.29**±1.21 |
| Gen-NSD | 72.97±1.81 | 77.25±1.67 | 70.16±2.08 |
| GCN | 56.50±0.81 | 55.63±1.13 | 51.97±0.49 |

Additionally, to examine the sensitivity of the Bayesian and deterministic sheaf models to the hyperparameter selections, we report the mean test accuracy across all hyperparameter configurations in Table 6.2.

### 6.4.1 Results

Though the results are mixed in the case of diagonal restriction maps, the $SO(d)$-BSNN outperforms the corresponding $O(d)$-NSD model on all three datasets, and the Gen-BSNN significantly outperforms the corresponding Gen-NSD on all three dataset. This suggests that the comparative advantage of our Bayesian approach to learning cellular sheaves is more pronounced for sheaves with a greater number of parameters. We also note that on each of the three WebKB datasets, the best test accuracy out of all the models is attained by one of our BSNNs. Moreover, the standard deviation of the reported test accuracies for each BSNN is lower than the corresponding NSD model with the same type of restriction map in all but one case, showing that our BSNN models are less sensitive to the effects of random weight initializations.

From Table 6.2, the advantage of each BSNN model over the corresponding NSD model when considering the entire range of hyperparameters is definitive. For every type of restriction map and dataset, a Wilcoxon signed-rank test for a $p$-value threshold of $p = 0.05$ indicates a statistically significant difference between the results of the

**Table 6.2:** Average test accuracy across all hyperparameter configurations and corresponding standard deviations.

| | **Texas** | **Wisconsin** | **Cornell** |
|---|---|---|---|
| **Diag-BSNN** | **73.41** $\pm$2.84 | **77.78** $\pm$3.32 | **64.77** $\pm$5.51 |
| Diag-NSD | 71.48 $\pm$3.54 | 74.85 $\pm$4.57 | 62.18 $\pm$7.10 |
| **SO(d)-BSNN** | **72.88** $\pm$3.63 | **76.84** $\pm$4.68 | **65.21** $\pm$6.84 |
| $O(d)$-NSD | 71.98 $\pm$3.28 | 74.55 $\pm$5.03 | 61.73 $\pm$6.70 |
| **Gen-BSNN** | **73.12** $\pm$2.41 | **77.92** $\pm$3.23 | **65.77** $\pm$4.87 |
| Gen-NSD | 71.34 $\pm$2.70 | 74.64 $\pm$3.92 | 62.20 $\pm$5.69 |

BSNN model and the corresponding NSD model. Thus we observe that the BSNN models maintain good performance over a much wider range of hyperparameters when compared with the NSD with the same type of restriction maps.

## 6.5 Conclusion

In this chapter, we have introduced Bayesian sheaf neural networks: novel sheaf neural networks in which the sheaf Laplacian is a random latent variable, and where the loss function used in training contains a Kullback-Leibler divergence regularization term. Not only does this allow one to quantify uncertainty of the model arising from the learned sheaf, but we show that for node classifications tasks with limited training data, our BSNN models outperform deterministic sheaf models. Additionally, our experiments suggest that the BSNN models are more robust to different choices of hyperparameters and weight initializations than the corresponding deterministic sheaf models.

# Bibliography


[1] Michal Adamaszek, Henry Adams, Ellen Gasparovic, Maria Gommel, Emilie Purvine, Radmila Sazdanovic, Bei Wang, Yusu Wang, and Lori Ziegelmeier. Vietoris-Rips and Čech complexes of metric gluings. In *34th International Symposium on Computational Geometry (SoCG 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018. 48

[2] Michał Adamaszek and Henry Adams. The Vietoris–Rips complexes of a circle. *Pacific Journal of Mathematics*, 290:1–40, 2017. 1

[3] Michał Adamaszek, Henry Adams, and Florian Frick. Metric reconstruction via optimal transport. *SIAM Journal on Applied Algebra and Geometry*, 2(4):597–619, 2018. 2, 10, 11

[4] Michał Adamaszek, Henry Adams, and S Reddy. On Vietoris–Rips complexes of ellipses. *Journal of Topology and Analysis*, 11(3):661–690, 2019. 1

[5] Henry Adams, Florian Frick, and Žiga Virk. Vietoris thickenings and complexes have isomorphic homotopy groups. *Journal of Applied and Computational Topology*, 20XX. 2, 12, 20, 21, 22, 25, 26, 29, 39

[6] Henry Adams, Facundo Mémoli, Michael Moy, and Qingsong Wang. The persistent topology of optimal transport based metric thickenings. *Algebraic & Geometric Topology*, 24(1):393–447, 2024. 2, 41

[7] Henry Adams and Joshua Mirth. Metric thickenings of Euclidean submanifolds. *Topology and its Applications*, 254:69–84, 2019. 41

[8] T. W. Anderson, I. Olkin, and L. G. Underhill. Generation of random orthogonal matrices. *SIAM Journal on Scientific and Statistical Computing*, 8(4):625–629, 1987. 70

[9] P Bendich, J Marron, E Miller, A Pieloch, and S Skwerer. Persistent homology analysis of brain artery trees. *The Annals of Applied Statistics*, 10(1):198–218, 2016. 1

[10] Lukas Biewald. Experiment tracking with weights and biases, 2020. Software available from wandb.com. 94

[11] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017. 63

[12] Cristian Bodnar, Francesco Di Giovanni, Benjamin Chamberlain, Pietro Liò, and Michael Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs. *Advances in Neural Information Processing Systems*, 35:18527–18541, 2022. 4, 54, 57, 58, 59, 61, 62, 74, 94

[13] Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Liò, Guido F Montúfar, and Michael Bronstein. Weisfeiler and Lehman go cellular: CW networks. *Advances in Neural Information Processing Systems*, 34:2625–2640, 2021. 4

[14] Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F Montúfar, Pietro Liò, and Michael Bronstein. Weisfeiler and Lehman go topological: Message passing simplicial networks. In *International Conference on Machine Learning*, pages 1026–1037. PMLR, 2021. 4

[15] Karol Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fundamenta Mathematicae*, 35(1):217–234, 1948. 12, 48

[16] Luca Bortolussi, Ginevra Carbone, Luca Laurenti, Andrea Patane, Guido Sanguinetti, and Matthew Wicker. On the robustness of Bayesian neural networks to adversarial attacks. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. 4

[17] Nizar Bouhlel and David Rousseau. A generic formula and some special cases for the Kullback–Leibler divergence between central multivariate Cauchy distributions. *Entropy*, 24(6), 2022. 92

[18] G Bredon. *Topology and geometry.* Springer-Verlag, New York, NY, 1993. 32

[19] Glen E Bredon. *Sheaf theory*, volume 170. Springer Science & Business Media, 2012. 55

[20] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021. 3

[21] K Brown. *Cohomology of Groups.* Springer-Verlag, New York, NY, 1982. 13, 16

[22] Z Cang, L Mu, J Ye, and G.-W Wei. Representability of algebraic topology for biomolecules in machine learning based scoring and virtual screening. *PLOS Computational Biology*, 14(2), 2018. 1

[23] Arthur Cayley. Sur quelques propriétés des déterminants gauches. *Journal für die reine und angewandte Mathematik*, 32:119–123, 1846. 68

[24] M Chung, J Hanson, J Ye, R Davidson, and S Pollak. Persistent homology in sparse regression and its application to brain morphometry. *IEEE Transactions on Medical Imaging*, 34(9):1928–1939, 2015. 1

[25] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012. 91

[26] V de Silva and R Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7(1):339–358, 2007. 1

[27] C Dowker. Homology groups of relations. *Annals of Mathematics*, 56(1):84–95, 1952. 2, 22

[28] James Dugundji. An extension of Tietze's theorem. *Pacific Journal of Math*, 1(3):353–367, 1951. 44

[29] James Dugundji. Absolute neighborhood retracts and local connectedness in arbitrary metric spaces. *Compositio Mathematica*, 13:229–246, 1958. 45

[30] Stefania Ebli, Michaël Defferrard, and Gard Spreemann. Simplicial neural networks. In *NeurIPS Workshop on Topological Data Analysis and Beyond*, 2020. 4, 53

[31] Luca Falorsi, Pim de Haan, Tim R Davidson, and Patrick Forré. Reparameterizing distributions on Lie groups. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3244–3253. PMLR, 2019. 66

[32] Xinjie Fan, Shujian Zhang, Bo Chen, and Mingyuan Zhou. Bayesian attention modules. *Advances in Neural Information Processing Systems*, 33:16362–16376, 2020. 4

[33] H Federer. *Geometric Measure Theory*. Springer-Verlag, 1969. 70

[34] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3558–3565, 2019. 4

[35] H Freudenthal. Simplizialzerlegungen von beschränkter Flachheit. *Annals of Mathematics*, 43:580–582, 1942. 31

[36] Joel Friedman. Computing Betti numbers via combinatorial Laplacians. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of Computing*, pages 386–391, 1996. 56

[37] R Frigerio and A Maffei. A remark on the Mayer-Vietoris double complex for singular cohomology. *Homology, Homotopy, and Applications*, 23(2):59–68, 2021. 16

[38] Hao Fu, Chunyuan Li, Xiaodong Liu, Jianfeng Gao, Asli Celikyilmaz, and Lawrence Carin. Cyclical annealing schedule: A simple approach to mitigating KL vanishing. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 240–250, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. 64

[39] Yarin Gal and Zoubin Ghahramani. Bayesian convolutional neural networks with Bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158*, 2015. 4

[40] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48*, pages 1050–1059, 2016. 4

[41] Yarin Gal and Zoubin Ghahramani. A theoretically grounded application of dropout in recurrent neural networks. *Advances in Neural Information Processing Systems*, 29, 2016. 4

[42] Patrick Gillespie. Vietoris thickenings and complexes are weakly homotopy equivalent. *Journal of Applied and Computational Topology*, 8(1):1–19, 2023. 3

[43] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning*, pages 1263–1272. PMLR, 2017. 4

[44] Ehsan Hajiramezanali, Arman Hasanzadeh, Krishna Narayanan, Nick Duffield, Mingyuan Zhou, and Xiaoning Qian. Variational graph recurrent neural networks. *Advances in Neural Information Processing Systems*, 32, 2019. 4

[45] J. Hansen and R. Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3:315–358, 2019. 57

[46] Jakob Hansen and Thomas Gebhart. Sheaf neural networks. In *NeurIPS 2020 Workshop on Topological Data Analysis and Beyond*, 2020. 4, 57

[47] Jakob Hansen and Robert Ghrist. Opinion dynamics on discourse sheaves. *SIAM Journal on Applied Mathematics*, 81(5):2033–2060, 2021. 60

[48] A. Hatcher. *Algebraic Topology*. Cambridge University Press, Cambridge, United Kingdom, 2002. 18, 33, 39, 41

[49] J Hausmann. On the Vietoris–Rips complexes and a cohomology theory for metric spaces. *Annals of Mathematics Studies*, 138:175–188, 1995. 2, 3, 43, 50

[50] Laurent Valentin Jospin, Hamid Laga, Farid Boussaid, Wray Buntine, and Mohammed Bennamoun. Hands-on Bayesian neural networks—a tutorial for deep learning users. *IEEE Computational Intelligence Magazine*, 17(2):29–48, 2022. 4

[51] Nigel J Kalton and Mikhail I Ostrovskii. Distances between Banach spaces. In *Forum Math*, volume 2, pages 17–48. Springer, 1999. 49

[52] Shogo Kato and M.C. Jones. An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli*, 19(1):154–171, 2013. 72

[53] Alex Kendall and Yarin Gal. What uncertainties do we need in Bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, 30, 2017. 4

[54] Diederik P Kingma and Max Welling. Auto-encoding variational Bayes. *arXiv preprint arXiv:1312.6114*, 2013. 63, 91

[55] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017. 56, 57

[56] H Kuhn. Some combinatorial lemmas in topology. *IBM Journal of Research and Development*, 4(5):518–524, 1960. 31

[57] Janko Latschev. Vietoris-Rips complexes of metric spaces near a closed Riemannian manifold. *Archiv der Mathematik*, 77(6):522–528, 2001. 49, 51

[58] Y Lee, S Barthel, P Dłotko, S Moosavi, K Hess, and B Smit. Quantifying similarity of pore-geometry in nanoporous materials. *Nature Communications*, 8(1):1–8, 2017. 1

[59] Jean Leray. Sur la forme des espaces topologiques et sur les points fixes des représentations. *J. Math. Pures. Appl.*, 24:95–167, 1945. 13

[60] Jean Leray. L'anneau spectral et l'anneau filtré d'homologie d'un espace localement compact et d'un application continue. *J. Math. Pures. Appl.*, 29(9):1–139, 1950. 13

[61] Carlos A. León, Jean-Claude Massé, and Louis-Paul Rivest. A statistical model for random rotations. *Journal of Multivariate Analysis*, 97(2):412–430, 2006. 69, 72

[62] Yaguang Li, Rose Yu, Cyrus Shahabi, and Yan Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR '18)*, 2018. 4

[63] Sunhyuk Lim, Facundo Mémoli, and Osman Berat Okutan. Vietoris–Rips persistent homology, injective metric spaces, and the filling radius. *Algebraic & Geometric Topology*, 24(2):1019–1100, 2024. 1, 3, 43, 44, 45, 46, 47, 49

[64] Ephy R Love, Benjamin Filippenko, Vasileios Maroulas, and Gunnar Carlsson. Topological convolutional layers for deep learning. *Journal of Machine Learning Research*, 24(59):1–35, 2023. 3

[65] S MacLane. *Homology*. Springer-Verlag, New York, NY, 1963. 16

[66] S. Mardešić and J. Segal. *Shape Theory*. North-Holland Publishing Company, Amsterdam, Netherlands, 1982. 7, 39

[67] Vasileios Maroulas, Cassie Putman Micucci, and Farzana Nasrin. Bayesian topological learning for classifying the structure of biological networks. *Bayesian Analysis*, 17(3):711–736, 2022. 1, 4

[68] Roy Meshulam. The clique complex and hypergraph matching. *Combinatorica*, 21:89–94, 2001. 13, 18

[69] Edward C Mitchell, Brittany Story, David Boothe, Piotr J Franaszczuk, and Vasileios Maroulas. A topological deep learning framework for neural spike decoding. *Biophysical Journal*, 123(17):2781–2789, September 2024. 4

[70] Michael Moy. Vietoris–Rips metric thickenings of the circle. *Journal of Applied and Computational Topology*, 7(4):831–877, 2023. 10

[71] Farzana Nasrin, Christopher Oballe, David Boothe, and Vasileios Maroulas. Bayesian topological learning for brain state classification. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 1247–1252. IEEE, 2019. 4

[72] Theodore Papamarkou, Tolga Birdal, Michael M Bronstein, Gunnar E Carlsson, Justin Curry, Yue Gao, Mustafa Hajij, Roland Kwitt, Pietro Liò, Paolo Di Lorenzo,

Vasileios Maroulas, Nina Miolane, Farzana Nasrin, Karthikeyan Natesan Ramamurthy, Bastian Rieck, Simone Scardapane, Michael T. Schaub, Petar Veličković, Bei Wang, Yusu Wang, Guo-Wei Wei, and Ghada Zamzmi. Position: Topological deep learning is the new frontier for relational learning. In *Forty-first International Conference on Machine Learning*, 2024. 3

[73] Theodore Papamarkou, Maria Skoularidou, Konstantina Palla, Laurence Aitchison, Julyan Arbel, David Dunson, Maurizio Filippone, Vincent Fortuin, Philipp Hennig, José Miguel Hernández-Lobato, Aliaksandr Hubin, Alexander Immer, Theofanis Karaletsos, Mohammad Emtiyaz Khan, Agustinus Kristiadi, Yingzhen Li, Stephan Mandt, Christopher Nemeth, Michael A Osborne, Tim G. J. Rudner, David Rügamer, Yee Whye Teh, Max Welling, Andrew Gordon Wilson, and Ruqi Zhang. Position: Bayesian deep learning is needed in the age of large-scale AI. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 39556–39586. PMLR, 21–27 Jul 2024. 4

[74] Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-GCN: Geometric graph convolutional networks. In *International Conference on Learning Representations*, 2020. 4

[75] F. W. Ponting and H. S. A. Potter. The volume of orthogonal and unitary space. *The Quarterly Journal of Mathematics*, os-20(1):146–154, 01 1949. 92

[76] T Konstantin Rusch, Michael M Bronstein, and Siddhartha Mishra. A survey on oversmoothing in graph neural networks. *SAM Research Report*, 2023, 2023. 4

[77] A Spannaus, K Law, P Luszczek, F Nasrin, C Micucci, P Liaw, L Santodonato, D Keffer, and V Maroulas. Materials fingerprinting classification. *Computer Physics Communications*, 266, 2021. 1

[78] Christoph Thäle. 50 years sets with positive reach–a survey. *Surveys in Mathematics and its Applications*, 3:123–165, 2008. 42

[79] M Todd. *The computation of fixed points and applications*. Springer Berlin Heidelberg, Berlin, Germany, 1976. 30

[80] Tim Traynor. Change of variable for Hausdorff measure (from the beginning). *Rendiconti dell'Istituto di Matematica dell'Università di Trieste*, 26:327–347, 1994. 70

[81] David E. Tyler. Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika*, 74(3):579–589, 1987. 73

[82] K Xia, X Feng, and G.-W. Tong, Y ad Wei. Persistent homology for the quantitative prediction of fullerene stability. *Journal of Computational Chemistry*, 36(6):408–422, 2014. 1

[83] Yujun Yan, Milad Hashemi, Kevin Swersky, Yaoqing Yang, and Danai Koutra. Two sides of the same coin: Heterophily and oversmoothing in graph convolutional neural networks. In *2022 IEEE International Conference on Data Mining (ICDM)*, pages 1287–1292. IEEE, 2022. 4

[84] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 974–983, 2018. 3

[85] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018. 4

[86] Lingxiao Zhao and Leman Akoglu. PairNorm: Tackling oversmoothing in GNNs. In *International Conference on Learning Representations*, 2020. 4

[87] Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in Neural Information Processing Systems*, 33:7793–7804, 2020. 4, 74

# Appendix

## A Additional Proofs

*Proof of Proposition 5.1.* We show that $\mathrm{VR}^m(X;r)$ and $X$ are homotopy equivalent for $0 < r < 2$, in which case the proposition follows from the weak homotopy equivalence $\mathrm{VR}(X;r) \to \mathrm{VR}^m(X;r)$ along with the fact that $X$ has the homotopy type of a CW complex, hence Whitehead's theorem may be applied.

Let $p : \mathbb{R}^2 \setminus (0,0) \to X$ be the radial projection map. Let $0 < r < 2$, and define a map $f : \mathrm{VR}^m(X;r) \to X$ as follows. For $x = \sum_i a_i \delta_{x_i} \in \mathrm{VR}^m(X;r)$, set $f(x) = p(\sum_i a_i x_i)$. Note this is well-defined due to the fact that $r < 2$, hence the convex hull of the points $\{x_i\}_i$ does not contain $(0,0)$. Let $i : X \to \mathrm{VR}^m(X;r)$ denote the inclusion $i(x) = \delta_x$. Note that $fi = \mathrm{id}_X$. It suffices to show that $if \simeq \mathrm{id}$. We claim that for any point $x = \sum_i a_i \delta_{x_i} \in \mathrm{VR}^m(X;r)$, the subset $\{x_i\}_i \cup \{f(x)\} \subseteq X$ has diameter less than $r$. If this is true, then the linear homotopy $H : \mathrm{VR}^m(X;r) \times I \to \mathrm{VR}^m(X;r)$ defined by $H(x,t) = (1-t)x + t \cdot if(x)$ is well-defined and thus is continuous (recall Lemma 4.1.1), which shows that $if \simeq \mathrm{id}$.

To prove the claim, let $x = \sum_i a_i \delta_{x_i} \in \mathrm{VR}^m(X;r)$. If the points $\{x_i\}_i$ are contained in a single side of $X$, then the claim is trivial. Otherwise, since $r < 2$ the points $\{x_i\}_i$ must be contained in two intersecting sides of the square $X$. By symmetry, we may assume that these two sides are $A = [-1,1] \times \{1\}$ and $B = \{1\} \times [-1,1]$, and let $q$ denote their intersection point $q = (1,1)$. Let $a, b \in (-1,1]$ be the largest values such that the points $\{x_i\}_i$ are contained in $[a,1] \times \{1\} \cup \{1\} \times [b,1] \subseteq A \cup B$.

Let $C = [a,1] \times \{1\} \cup \{1\} \times [b,1]$, and note that the radial projection $p$ maps $\mathrm{hull}(C \cup \{(0,0)\}) \setminus (0,0)$ into $C$, where $\mathrm{hull}(C \cup \{(0,0)\})$ denotes the convex hull of $C \cup \{(0,0)\}$. Hence $f(x) \in f(\mathrm{hull}(\{x_i\}_i)) \subseteq C$. Lastly we recall from the definition of the values $a, b$ that $\mathrm{diam}(C) < r$, proving our claim that $\mathrm{diam}(\{x_i\}_i \cup \{f(x)\}) < r$. □

*Proof of Proposition 6.3.* Representing points of $S^3$ as unit vectors $x = (a,b,c,d) \in \mathbb{R}^4$, the 2-fold cover $\Phi : S^3 \to SO(3)$ has an explicit description as

$$\Phi(x) = \begin{pmatrix} a^2+b^2-c^2-d^2 & 2(bc-ad) & 2(bd+ac) \\ 2(bc+ad) & a^2-b^2+c^2-d^2 & 2(cd-ab) \\ 2(bd-ac) & 2(cd+ab) & a^2-b^2-c^2+d^2 \end{pmatrix}.$$

A straightforward calculation shows that

$$\det(\Phi(x) - \kappa I) = (1-\kappa)^3 a^2 + (1+\kappa)^2(1-\kappa)(b^2+c^2+d^2).$$

If $f(P; M, \kappa)$ denotes the density of $\mathcal{C}_3(M,\kappa)$ with respect to the unit Haar measure on $SO(3)$, $g(x;\Lambda)$ denotes the density of $ACG_3(\Lambda)$ with respect to the unit Haar measure on $S^3$, and $\Lambda_\kappa$ is the diagonal matrix $\Lambda_\kappa = \mathrm{diag}((\frac{1+\kappa}{1-\kappa})^2, 1, 1, 1)$, then we have

$$\begin{aligned} f(\Phi(x); I, \kappa) &\propto \det(\Phi(x) - \kappa I)^{-2} \\ &= \Big((1-\kappa)^3 a^2 + (1+\kappa)^2(1-\kappa)(b^2+c^2+d^2)\Big)^{-2} \\ &\propto \Big((\frac{1-\kappa}{1+\kappa})^2 a^2 + b^2 + c^2 + d^2\Big)^{-2} \\ &\propto g(x; \Lambda_\kappa), \end{aligned}$$

where the symbol $\propto$ denotes "proportional to".

Moreover, for $M \in SO(3)$, let $y_M \in \Phi^{-1}(M^\intercal)$. Regarding $y_M$ as a unit quaternion, let $Q \in SO(4)$ be the rotation $z \mapsto z y_M$ for $z \in \mathbb{H}$. Note that $\Phi(x) M^\intercal = \Phi(x y_M) =$

$\Phi(Qx)$ and $g(x; Q^{\mathsf{T}}\Lambda_\kappa Q) = g(Qx; \Lambda_\kappa)$. Hence

$$f(\Phi(x); M, \kappa) = f(\Phi(Qx); I, \kappa) \propto g(x; Q^{\mathsf{T}}\Lambda_\kappa Q).$$

Lastly, since $\Phi$ is a homomorphism, the pushforward of the unit Haar measure on $S^3$ yields the unit Haar measure on $SO(3)$. This combined with the fact that $f(\Phi(x); M, \kappa) \propto g(x; Q^{\mathsf{T}}\Lambda_\kappa Q)$ implies that $\Phi(X) \sim \mathcal{C}_3(M, \kappa)$ whenever $X \sim ACG_3(\Lambda_{M,\kappa})$ for $\Lambda_{M,\kappa} = Q^{\mathsf{T}}\Lambda_\kappa Q$. □

# B KL Divergence Calculations

Recall from Section 6.2.2 that we assume that the variational distribution $q_\phi(\mathcal{F}|X, y)$ factors as $q_\phi(\mathcal{F}|X, y) = \prod_{u\subseteq e} q_\phi(\mathcal{F}_{u\subseteq e}|X, y)$ over all incident node-edge pairs $u \subseteq e$, and that we use a prior $p_\theta(\mathcal{F})$ which also factors in the same manner $p_\theta(\mathcal{F}) = \prod_{u\subseteq e} p_\theta(\mathcal{F}_{u\subseteq e})$. As a consequence of the chain rule for KL divergence, see for example [25, Theorem 2.5.3], we have

$$\mathrm{KL}(q_\phi(\mathcal{F}|X, y) \parallel p_\theta(\mathcal{F})) = \sum_{u\subseteq e} \mathrm{KL}(q_\phi(\mathcal{F}_{u\subseteq e}|X, y) \parallel p_\theta(\mathcal{F}_{u\subseteq e})).$$

Hence it suffices to be able to compute $\mathrm{KL}(q_\phi(\mathcal{F}_{u\subseteq e}|X, y) \parallel p_\theta(\mathcal{F}_{u\subseteq e}))$ for all $u \subseteq e$.

In the cases where we learn diagonal or general linear sheaf restriction maps, $\mathrm{KL}(q_\phi(\mathcal{F}_{u\subseteq e}|X, y) \parallel p_\theta(\mathcal{F}_{u\subseteq e}))$ is a KL divergence of a normal distribution with diagonal covariance from a standard normal distribution. The KL divergence of a normal distribution $\mathcal{N}(\mu, \mathrm{diag}(\sigma^2))$ with mean $\mu \in \mathbb{R}^d$ and diagonal covariance $\sigma^2 \in \mathbb{R}^d$ from the standard normal distribution is

$$\mathrm{KL}(\mathcal{N}(\mu, \mathrm{diag}(\sigma^2)) \parallel \mathcal{N}(\mathbf{0}, I_d)) = -\frac{1}{2}\sum_{i=1}^{d} \left(1 + \log(\sigma_i^2) - \mu_i^2 - \sigma_i^2\right). \tag{18}$$

See [54, Appendix B] for a derivation of this calculation.

In the case where the restriction maps are special orthogonal, and the stalk dimension $d$ is 2 or 3, we compute $\mathrm{KL}(q_\phi(\mathcal{F}_{u\subseteq e}|X, y) \parallel p_\theta(\mathcal{F}_{u\subseteq e}))$ using Proposition 6.4, which we now prove.

*Proof of Proposition 6.4.* First, we note that the integral of a function $f(P)$ on $SO(n)$ with respect to a Haar measure is invariant under the change of variables $P \mapsto PM$ for any $M \in SO(n)$. Hence $\mathrm{KL}(\mathcal{C}_n(M, \kappa) \parallel U_{SO(n)})$ does not depend on $M$, that is,

$$\mathrm{KL}(\mathcal{C}_n(M, \kappa) \parallel U_{SO(n)}) = \mathrm{KL}(\mathcal{C}_n(I, \kappa) \parallel U_{SO(n)}).$$

To simplify notation, set $d = n(n-1)/2$, let $\gamma = \frac{1+\kappa}{1-\kappa}$, and let $V_n$ denote the volume of $SO(n)$. Recall Eq. (6.17), and note that the density of $\mathcal{C}_n(I, \kappa)$ with respect to $\mathcal{H}^d$ can be expressed as

$$V_n^{-1}\gamma^d\Big(\frac{\det(I + C^{-1}(P)}{\det(I+\gamma C^{-1}(P))}\Big)^{n-1}$$

where $V_n$ denotes the volume of $SO(n)$. Then the change of variables formula yields

$$\mathrm{KL}(\mathcal{C}_n(M,\kappa) \parallel U_{SO(n)}) = \int_{\mathbb{R}^d} \frac{2^{\frac{3n(n-1)}{4}} V_n^{-1}\gamma^d}{\det(I+\gamma X_\phi)^{n-1}} \log\Big(\frac{\gamma^d \det(I+X_\phi)^{n-1}}{\det(I+\gamma X_\phi)^{n-1}}\Big)\, \lambda^d(d\phi),$$

where recall from Section 6.3 that $X_\phi$ denotes the skew symmetric matrix whose entries below the diagonal are the entries of the vector $\phi$.

When $n = 2, 3$, we have $\det(I + X_\phi) = 1 + \|\phi\|^2$ and $\det(I + \gamma X_\phi) = 1 + \gamma^2\|\phi\|^2$, hence

$$\mathrm{KL}(\mathcal{C}_n(M,\kappa) \parallel U_{SO(n)}) = \int_{\mathbb{R}^d} \frac{2^{\frac{3n(n-1)}{4}} V_n^{-1}\gamma^d}{(1+\gamma^2\|\phi\|^2)^{n-1}} \log\Big(\frac{\gamma^d (1+\|\phi\|^2)^{n-1}}{(1+\gamma^2\|\phi\|^2)^{n-1}}\Big)\, \lambda^d(d\phi).$$

Recall that by [75],

$$V_n = 2^{\frac{3n(n-1)}{4}} \prod_{i=2}^{n} \frac{\pi^{(i-1)/2}\Gamma((i-1)/2)}{\Gamma(i-1)}.$$

Hence $\mathrm{KL}(\mathcal{C}_n(M,\kappa) \parallel U_{SO(n)})$ is precisely the KL divergence of $\mathrm{Cauchy}_d(0, \frac{1}{\gamma^2}I)$ from $\mathrm{Cauchy}_d(0, I)$ for $n = 2$ or $n = 3$. By the expressions for the KL divergence between Cauchy distributions in [17], we have

$$\mathrm{KL}(\mathrm{Cauchy}_1(0, \tfrac{1}{\gamma^2}I) \parallel \mathrm{Cauchy}_1(0, I)) = \log\left(\frac{(\gamma+1)^2}{4\gamma}\right)$$

and

$$\mathrm{KL}(\mathrm{Cauchy}_3(0, \tfrac{1}{\gamma^2}I_3) \parallel \mathrm{Cauchy}_3(0, I_3)) = \log\left(\frac{(\gamma+1)^4}{16\gamma}\right) + \frac{2(1-\gamma)}{1+\gamma}.$$

Since $\gamma = \frac{1+\kappa}{1-\kappa}$ we have

$$\mathrm{KL}(\mathcal{C}_2(M, \kappa) \parallel U_{SO(2)}) = -\log(1 - \kappa^2)$$

and

$$\mathrm{KL}(\mathcal{C}_3(M, \kappa) \parallel U_{SO(3)}) = -\log(1 - \kappa^2) - 2\log(1 - \kappa) - 2\kappa.$$

□

# C Experimental Details

All experiments were run on a NVIDIA Tesla K80 GPU with two Intel Xeon E5-2637 CPUs with 64GB of memory. Hyperparameter tuning was done using Weights & Biases [10] to perform a grid search over the hyperparameter configurations described in Table C.1. The neural sheaf diffusion (NSD) models from [12] were implemented using the code available at https://github.com/twitter-research/neural-sheaf-diffusion. Our code is available at https://github.com/patrick-gillespie/bsnn.

**Table C.1:** Hyperparameter configurations for WebKB experiments.

| **Hyperparameter** | **Values** |
| --- | --- |
| Hidden channels | {8, 32} |
| Stalk dim $d$ | {2, 3, 4, 5} |
| Layers | {2, 3, 4, 5} |
| Dropout | {0.0, 0.3, 0.6} |
| Learning rate | 0.01 |
| Activation | ELU |
| Weight decay | 5e-4 |
| Patience | 200 |
| Max training epochs | 500 |
| Optimiser | Adam |
| Ensemble # | 3 (BSNN), N/A (NSD) |

## C.1 WebKB Dataset Descriptions

The Texas, Cornell, and Wisconsin datasets are web page datasets for the computer science departments of the respective universities. Each dataset is a graph in which nodes represent web pages and edges represent hyperlinks between them. In all datasets the node features are 1703-dimensional bag-of-word representations of the web pages, and each node belongs to one of the five classes: student, project, course, staff, and faculty. The Texas dataset contains 183 nodes and 325 edges. The Cornell dataset contains 183 nodes and 298 edges. The Wisconsin dataset contains 251 nodes and 515 edges

# Vita

Patrick Gillespie was born in Allentown, Pennsylvania on 22 June 1998. He first developed an interest in mathematics while attending Lehigh Carbon Community College, where he earned an Associate of Science in Mathematics before transferring to West Chester University. At West Chester, Patrick developed a passion for topology, where he performed research under the mentorship of Dr. Jeremy Brazas. He graduated from West Chester in 2020 with a Bachelor of Science in Mathematics.

Patrick began his PhD at the University of Tennessee in 2020. In January 2022, he joined Dr. Vasileios Maroulas's research group and studied the applications of topology to machine learning and data analysis in collaboration with the Army Research Lab.