Topology and its Applications 279 (2020) 107254

Contents lists available at ScienceDirect

Topology and its Applications

www.elsevier.com/locate/topol

New families of stable simplicial filtration functors

Samir Chowdhury[a], Nathaniel Clause[a], Facundo Mémoli[a,*], Jose Ángel Sánchez[b], Zoe Wellner[c]

[a] The Ohio State University, Columbus, OH, United States of America
[b] University of North Carolina, Chapel Hill, Chapel Hill, NC, United States of America
[c] Carnegie Mellon University, Pittsburgh, PA, United States of America

ARTICLE INFO

Article history:
Received 8 May 2019
Received in revised form 5 April 2020
Accepted 6 May 2020
Available online 19 May 2020

Keywords:
Persistent homology
Filtrations
Barcodes
Gromov-Hausdorff

ABSTRACT

When quantifying the topological properties of a metric dataset using persistent homology, the first step is to produce a simplicial filtration on the data. The Čech and Vietoris-Rips filtrations are two of the workhorses of persistent homology, but over time, various other filtrations which capture different properties of data or have different computational burdens have been introduced. Towards a program of characterizing all the possible simplicial filtrations on a metric dataset, we introduce and develop the framework of valuation-induced stable filtration functors. This framework is based on the concept of curvature sets due to Gromov, and encapsulates the Vietoris-Rips and various other filtrations while simultaneously providing a model for generating families of novel filtration functors that capture diverse features present in datasets. We further extend this foundation by incorporating the notion of basepoint-dependent filtration functors and proving the associated functoriality and stability properties. This rich theoretical framework provides a unifying language for various extant simplicial filtrations, and is also a mechanism for generating arbitrarily large families of novel filtration functors with control over basepoint dependence/independence as well as the locality of the filtration. We exemplify our constructions on both toy datasets and on 3D shapes from a publicly available shape database. Our paper is accompanied by a Matlab software package incorporating an interactive platform for visualizing and testing new filtrations on datasets.

1. Introduction

Persistent homology (PH) is a data analysis technique for estimating the geometric and topological features of datasets at a range of scale parameters. In the last few decades, PH has become a widely recognized tool in computational topology and topological data analysis [1–5]. The PH pipeline is summarized below:

[*] Corresponding author.
E-mail addresses: chowdhury.57@osu.edu (S. Chowdhury), clause.15@osu.edu (N. Clause), memoli@math.osu.edu (F. Mémoli), jsgomez@live.unc.edu (J.A. Sánchez), zwellner@andrew.cmu.edu (Z. Wellner).

https://doi.org/10.1016/j.topol.2020.107254

Dataset (finite metric space) → Filtered simplicial complex → Persistence diagram/barcode.

Here the first arrow represents a *filtration map* that builds a simplicial structure on the finite metric space at a range of scale parameters. The second arrow represents the homology functor (with field coefficients), which tracks the topological features of the simplicial structure across a range of scales. These features are summarized as *persistence diagrams* or *barcodes* [4,5]. The Čech and Vietoris-Rips (VR) filtrations are the main methods for filtering datasets due to their interpretability [3]. These filtration maps are *functorial*, i.e. compatible metric spaces are mapped to compatible filtered spaces.

The notion of stability guarantees that if two datasets have similar geometric structure (independent of sample sizes), then their persistence diagrams should be similar. Stability is determined by quantifying differences between finite metric spaces ($\mathcal{M}$) and comparing them to differences between persistence diagrams ($\mathcal{D}$). The former is given by the Gromov-Hausdorff distance d_{GH}, a distance in the category of compact metric spaces [6]. The standard method for the latter is the bottleneck distance d_{B}, a distance between diagrams [3]. In particular, both the Vietoris-Rips and Čech filtration functors are stable [7,8]:

Theorem 1. *For X, Y finite metric spaces, all $k \in \mathbb{Z}_+$, and Φ either the Čech or the VR filtration,*

$$d_{\mathrm{B}}(\mathrm{dgm}_k^{\Phi}(X), \mathrm{dgm}_k^{\Phi}(Y)) \leq 2 \cdot d_{\mathrm{GH}}(X, Y).$$

This result carries two primary interpretations. On one hand, two datasets that have similar shape should induce similar barcodes, offering the possibility of analyzing shape through a careful interpretation of the diagrams. On the other hand, one can regard the distance between diagrams as an approximation from below of d_{GH}. Computing the Gromov Hausdorff distance is an NP-hard problem [9], while the bottleneck distance can be calculated in polynomial time [10]. Thus, instead of the computationally impractical Gromov-Hausdorff distance, one uses the approximation of d_{GH} from d_{B} as a classifier for applying clustering methods and machine learning [7].

The preceding theorem suggests the following question: does there exist a filtration functor Φ (of up to NP-hard complexity) and $k \in \{0, 1, 2, \ldots\}$ such that the map dgm_k^{Φ} is an isometric embedding of $\mathcal{M}$ into $\mathcal{D}$? Our first result (Proposition 16) shows the impossibility of having such a filtration functor. However, by relaxing our requirement of having a *single* filtration functor, we find (Theorem 17) a *family* $(\Phi_\alpha)_\alpha$ of stable filtration functors such that for every pair of spaces $X, Y \in \mathcal{M}$, there is a functor Φ_α attaining $d_{\mathrm{GH}}(X, Y)$ through the bottleneck distance between the corresponding 0-dimensional diagrams. This family is of course computationally intractable unless P=NP, but Theorem 17 suggests that to approximate d_{GH} between two metric spaces, one should generate "large" families of filtrations on the spaces. This approach follows the ideas espoused in [11–13]. In the setting of inverse problems, i.e. generating families of filtrations that can reconstruct a metric space up to isometry, there is a growing body of literature—see [14–20].

The main goal of this article is to create a framework for producing new filtration functors that satisfy a stability property such as the one seen in Theorem 1, that provide new approximations to d_{GH} for classification purposes, and that capture different geometric and topological features than the Čech and Vietoris-Rips filtrations. This is part of a larger program of characterizing *all* the possible filtration functors acting on metric spaces.

1.1. Overview of our results

We start by recalling the notion of curvature sets, which form a full invariant for a metric space [21, Theorem $3.27\frac{1}{2}$]. We then define a new concept: *valuations*, which are real-valued functions on sets of metric matrices (Definition 20). We prove that valuations induce filtration functors, which we call local filtration functors. This framework allows us to prove that by imposing a stability property on valuations, we guarantee the stability of the filtration functor induced by the valuation (Theorem 26). We further show that any Lipschitz function $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ generates a valuation which in turn induces a stable filtration functor (Proposition 27).

By analogy with the Vietoris-Rips filtration, which is tightly related to the metric notion of diameter, we define and study a filtration produced via ultrametricity [22], which measures how far a dataset is from having an ultrametric (i.e. tree-like) structure. In addition to proving and conjecturing various theoretical results on the ultrametricity barcodes of particular spaces, we show via computational experiments that the ultrametricity filtration may be more discriminative than the Vietoris-Rips filtration when comparing groups of phylogenetic datasets.

Filtrations that emphasize geometric properties of specific regions in a dataset are becoming increasingly important. For example, the authors of [23] considered the (superlevel set) filtrations of a metric graph (G, d_G) induced by the collection $\{d_G(p, \cdot) : p \in G\}$, proved stability, and demonstrated its use on datasets. This idea was further studied in [15] from the perspective of reconstructing metric graphs. In line with these ideas, we define *basepoint filtration functors* (Definition 40), which are filtrations depending on a choice of basepoint that provides a local perspective on the dataset. This framework generalizes standard filtration functors, since any filtration can be regarded as a constant basepoint filtration on the points of the space. We develop procedures for defining new basepoint filtration functors by adjusting the notion of valuation (Definition 47).

We prove that stable behavior of an *adjusted* valuation induces stability of the induced basepoint filtration (Theorem 49). This is the most general stability theorem proved in this article. We also prove in Proposition 50 that basepoint filtrations induced by adjusted valuations are stable with respect to slight changes in basepoint, that is, two nearby basepoints generate similar barcodes. We then develop a specific basepoint filtration functor called the eccentricity basepoint filtration in Example 51, and use this filtration for computational examples on particular datasets to provide both intuition on how basepoint filtrations operate, and topological interpretability of this specific filtration functor.

Software packages for both the ultrametricity and eccentricity basepoint filtrations are available on https://github.com/NateClause. In particular, the eccentricity filtration package provides an interactive user interface for easy experimentation on datasets.

1.2. Organization of the paper

Section 2 contains the necessary background. Section 3 contains two results motivating the perspective presented in this paper. In Section 4 we define valuations and present stability results, a method for generating valuations, and a particular example—the ultrametricity filtration functor. In Section 5 we further generalize our framework via the notion of basepoint filtration functors and present the eccentricity basepoint filtration functor. Theorem 49 in this section provides the most general stability result of the paper. Finally in Section 6 we describe our computational examples.

2. Preliminaries

By $\mathbb{R}_+, \mathbb{Z}_+$ we denote the non-negative reals/integers. For a set X, we let $\mathrm{pow}(X)$ denote the set of all finite, non-empty subsets of X. By $\mathcal{M}$ we denote the collection of all finite metric spaces. The Hausdorff distance between closed subsets of a metric space is denoted by d_{H}. It is given as follows: for A, B subsets of a finite metric space X (which are always closed),

$$d_{\mathrm{H}}(A, B) := \max\left(\max_{a\in A}\min_{b\in B} d_X(a, b),\ \max_{b\in B}\min_{a\in A} d_X(a, b)\right).$$

Given a finite metric space (X, d_X) we consider the map $\iota_X : \mathrm{pow}(X) \to \mathcal{M}$ given by $\sigma \mapsto (\sigma, d_X|_{\sigma\times\sigma})$, i.e. ι_X takes a subset of X and endows it with the restriction of the metric of X. By the *diameter* $\mathrm{diam}(X)$ of a finite metric space (X, d_X) we mean the number $\max_{x,x'\in X} d_X(x, x')$, and by the *eccentricity function associated to* (X, d_X) we mean the function $\mathrm{ecc}_X : X \to \mathbb{R}_+$ given by $x \mapsto \max_{x'\in X} d_X(x, x')$. For two metric spaces (X, d_X) and (Y, d_Y), we say that X embeds into Y isometrically, denoted $X \hookrightarrow Y$, if there is a subset $Z \subset Y$ and a function $h : X \to Z$ such that h is a distance-preserving bijection.

Throughout this article, we will use homology with field coefficients. We write H_k to denote the kth simplicial homology functor; refer to [24] for background on simplicial homology.

Given two finite metric spaces (X, d_X) and (Y, d_Y) and any non-empty relation $R \subset X \times Y$ we consider its *distortion* [6] given by

$$\mathrm{dis}(R) := \max_{(x,y),(x',y')\in R} |d_X(x, x') - d_Y(y, y')|.$$

A particular class of relations between sets X and Y is given by *correspondences*: these are relations R such that the canonical projection maps are surjective. We denote by $\mathcal{R}(X, Y)$ the set of all correspondences between X and Y. The *Gromov-Hausdorff distance* between $X, Y \in \mathcal{M}$ is defined as $d_{\mathrm{GH}}(X, Y) := \frac{1}{2} \min\limits_{R\in\mathcal{R}(X,Y)} \mathrm{dis}(R)$.

Definition 2 *([21]).* Given a metric space (X, d_X) and $n \in \mathbb{N}$ consider the map $D_X^{(n)} : X^n \to \mathbb{R}^{n\times n}$ given by $(x_1, \ldots, x_n) \mapsto \big(d_X(x_i, x_j)\big)_{i,j=1}^{n}$. The *n-th curvature set* of (X, d_X) is the collection of $n \times n$ matrices $\mathrm{K}_n(X) := \mathrm{im}(D_X^{(n)})$.

As an example, consider the three point space X with distances $d_X = \begin{pmatrix} 0&1&2\\ 1&0&1\\ 2&1&0 \end{pmatrix}$. Then the first three curvature sets are:

$$\mathrm{K}_1(X) = \left\{\left(\begin{smallmatrix}0\end{smallmatrix}\right)\right\}, \mathrm{K}_2(X) = \left\{\left(\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right), \left(\begin{smallmatrix}0&2\\2&0\end{smallmatrix}\right), \left(\begin{smallmatrix}0&0\\0&0\end{smallmatrix}\right)\right\},$$

$$\mathrm{K}_3(X) = \left\{\begin{pmatrix}0&1&2\\1&0&1\\2&1&0\end{pmatrix}, \begin{pmatrix}0&2&1\\2&0&1\\1&1&0\end{pmatrix}, \begin{pmatrix}0&1&1\\1&0&2\\1&2&0\end{pmatrix}, \begin{pmatrix}0&0&2\\0&0&2\\2&2&0\end{pmatrix}, \begin{pmatrix}0&2&0\\2&0&2\\0&2&0\end{pmatrix}, \begin{pmatrix}0&2&2\\2&0&0\\2&0&0\end{pmatrix},\right.$$
$$\left.\begin{pmatrix}0&0&1\\0&0&1\\1&1&0\end{pmatrix}, \begin{pmatrix}0&1&0\\1&0&1\\0&1&0\end{pmatrix}, \begin{pmatrix}0&1&1\\1&0&0\\1&0&0\end{pmatrix}, \begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix}\right\}.$$

Remark 3. Curvature sets enjoy the following type of functoriality: suppose $X \hookrightarrow Y$ isometrically, then, for any $n \in \mathbb{N}$ one has $\mathrm{K}_n(X) \subset \mathrm{K}_n(Y)$.

In $(\mathbb{R}^n, \ell_\infty)$, curvature sets are stable with respect to the Gromov-Hausdorff distance.

Theorem 4 *([25]). For any pair of compact metric spaces, and all $n \in \mathbb{N}$,*

$$d_{\mathrm{H}}^{(\mathbb{R}^{n\times n},\ell_\infty)}(\mathrm{K}_n(X), \mathrm{K}_n(Y)) \leq 2 \cdot d_{\mathrm{GH}}(X, Y).$$

2.1. Filtrations, filtration functors, and their stability

Filtrations on finite metric spaces are defined as below.

Definition 5. Let X be a finite set. A *filtration on X* is any map $\Phi_X : \mathrm{pow}(X) \to \mathbb{R}$ which satisfies the *monotonicity condition*: $\Phi_X(\tau) \leq \Phi_X(\sigma)$, for all $\tau \subset \sigma \subset X$.

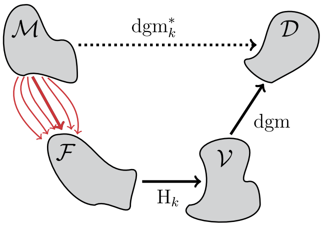

Fig. 1. A persistent homology method (shown here as a composite map dgm_k^*) maps finite metric spaces $\mathcal{M}$ to the space of all filtered spaces $\mathcal{F}$, then to persistent vector spaces $\mathcal{V}$, and finally to the space of diagrams/barcodes $\mathcal{D}$. Our goal is to understand the structure of the maps shown in red, i.e. the filtration functors $\mathcal{M} \to \mathcal{F}$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

Any pair (X, Φ_X) where X is a finite set and Φ_X is a filtration over X will be called a *filtered space*. An element σ of $\mathrm{pow}(X)$ will be referred to as a *simplex* and the value of $\Phi_X(\sigma)$ will be its *filtration value*. The collection of all such pairs will be denoted by $\mathcal{F}$. A filtered space (X, Φ_X) is naturally associated to a simplicial filtration. For each $t \in \mathbb{R}$, consider the (sublevel) set $\Phi_X[t] = \{\sigma \in \mathrm{pow}(X) : \Phi_X(\sigma) \leq t\}$. The monotonicity condition guarantees that $\Phi_X[t]$ is a simplicial complex at any $t \in \mathbb{R}$. As $t \to \infty$, this complex $\Phi_X[t]$ grows until it becomes the full simplicial complex over X. Applying the kth homology functor $\mathrm{H}_k : \mathcal{F} \to \mathcal{V}$ transforms this filtration $\{\Phi_X[t] \subset \Phi_X[t']\}_{t \leq t'}$ into a sequence of vector spaces with linear maps, i.e. a *persistent vector space*.

More generally, a persistent vector space is a family of vector spaces $\mathbb{U} = \{U^\delta \xrightarrow{\mu_{\delta,\delta'}} U^{\delta'}\}_{\delta \leq \delta'}$ such that: (1) $\mu_{\delta,\delta}$ is the identity for each $\delta \in \mathbb{R}$, and (2) $\mu_{\delta,\delta''} = \mu_{\delta',\delta''} \circ \mu_{\delta,\delta'}$ for each $\delta \leq \delta' \leq \delta''$. Furthermore, we make the additional assumption that all vector spaces are finite dimensional: (3) $\dim(U^\delta) < \infty$ for all $\delta \in \mathbb{R}$. The latter assumption is often referred to as *tameness* [26]. To each persistent vector space, it is then possible to associate a full invariant called a *persistence diagram* or *persistence barcode* [26]. We let $\mathrm{dgm} : \mathcal{V} \to \mathcal{D}$ denote the *diagram map* that maps a persistent vector space to its barcode. Given a filtered space (X, Φ_X) we define $\mathrm{dgm}_k^{\Phi_X}(X)$ to be the k-th dimensional persistence diagram of (X, Φ_X).

The space $\mathcal{D}$ of all persistence diagrams comes equipped with the so called *bottleneck distance* $d_\mathrm{B} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}_+$. We refer the reader to [3] for its definition.

Throughout this work, we will restrict our scope to filtrations with non-negative filtration values. To be precise, for $(X, \Phi_X) \in \mathcal{F}$ and $\sigma \subset X$, $\Phi_X(\sigma) \geq 0$.

It is possible to define a distance $d_\mathcal{F}$ between filtered spaces as follows [27]: given (X, Φ_X) and (Y, Φ_Y) in $\mathcal{F}$, let

$$d_\mathcal{F}((X, \Phi_X), (Y, \Phi_Y)) := \inf_{Z, \pi_X, \pi_Y} \max_{\sigma \subset Z} |\Phi_X(\pi_X(\sigma)) - \Phi_Y(\pi_Y(\sigma))|, \tag{1}$$

where Z is any finite set where π_X and π_Y are surjective maps from Z to X and Y, respectively. The following is a technical result that will be useful in this paper:

Theorem 6 *([27]). For all finite filtered spaces (X, Φ_X), $(Y, \Phi_Y) \in \mathcal{F}$, and all $k \in \mathbb{Z}_+$,*

$$d_\mathrm{B}(\mathrm{dgm}_k^{\Phi_X}(X), \mathrm{dgm}_k^{\Phi_Y}(Y)) \leq d_\mathcal{F}((X, \Phi_X), (Y, \Phi_Y)).$$

Theorem 6 establishes the stability of the map $\mathrm{dgm} \circ \mathrm{H}_k : \mathcal{F} \to \mathcal{D}$ for all $k \in \mathbb{Z}_+$. Observe from Fig. 1 that the complementary portion of the persistent homology pipeline consists of the filtration maps from finite metric spaces to filtered spaces. We want a process for generating stable maps from the category of finite metric spaces to the category of filtered spaces. Towards this end, we start by describing category structures on $\mathcal{M}$ and $\mathcal{F}$.

Definition 7. By $\mathcal{M}^{\mathrm{iso}}$ we denote the category with $\mathcal{M}$ as objects, and isometric embeddings as morphisms. We impose a category structure on $\mathcal{F}$ by specifying the arrows: these are the maps $h : (X, \Phi_X) \to (Y, \Phi_Y)$ such that for any $\sigma \subset X$, we have $\Phi_X(\sigma) \geq \Phi_Y(h(\sigma))$.

Definition 8. We say that a filtration map $\Phi : \mathcal{M} \to \mathcal{F}$ is *functorial on* $\mathcal{M}^{\mathrm{iso}}$ if for all pairs of spaces $X, Y \in \mathcal{M}$ and any isometric embedding $h : X \to Y$, $\Phi_X(\sigma) \geq \Phi_Y(h(\sigma))$ for all $\sigma \subset X$. We also call Φ an $\mathcal{M}^{\mathrm{iso}}$*-filtration functor.*

Remark 9. By defining morphisms to be 1-Lipschitz functions instead of isometries, one obtains another interesting category structure on $\mathcal{M}$, denoted $\mathcal{M}^{\mathrm{Lip}}$. Interestingly, the "valuation-induced filtrations" we consider in this work are not all $\mathcal{M}^{\mathrm{Lip}}$-functorial. However, under certain assumptions, we can obtain $\mathcal{M}^{\mathrm{Lip}}$-functoriality, and it turns out that these "valuation-induced filtrations" differ from the VR-filtration by a 1-Lipschitz map. We do not include these results in the current paper, as they are in a somewhat different direction.

Remark 10. One may ask if the inequalities in the preceding definitions may be reversed in a meaningful way. We say that a filtration map $\Phi : \mathcal{M} \to \mathcal{F}$ is a *contravariant* $\mathcal{M}^{\mathrm{iso}}$ filtration functor if for the isometric embedding $h : X \to Y$, we have $\Phi_X(\sigma) \leq \Phi_Y(h(\sigma))$ for $\sigma \subset X$. When we do not specifically mention "contravariant", it is understood that we are referring to the "covariant" case in the definitions above.

The filtration maps we consider in this work will be functorial on $\mathcal{M}^{\mathrm{iso}}$, so we also make the following definition for convenience:

Definition 11. We say that $\Phi : \mathcal{M} \to \mathcal{F}$ is a *filtration functor* if it is $\mathcal{M}^{\mathrm{iso}}$-functorial.

Example 12. The Čech filtration functor $\Phi^{\check{\mathrm{C}}} : \mathcal{M} \to \mathcal{F}$ is defined as the map $(X, d_X) \mapsto (X, \Phi^{\check{\mathrm{C}}}_X)$, where $\Phi^{\check{\mathrm{C}}}_X : \mathrm{pow}(X) \to \mathbb{R}$ is given by

$$\Phi^{\check{\mathrm{C}}}_X(\sigma) := \min_{p \in X} \max_{x \in \sigma} d_X(p, x), \qquad \forall\, X \in \mathcal{M},\ \sigma \subset X.$$

To see functoriality, let $h : X \to Y$ be an isometric embedding. Then we have:

$$\Phi^{\check{\mathrm{C}}}_X(\sigma) = \min_{p \in X} \max_{x \in \sigma} d_X(p, x) \geq \min_{p \in Y} \max_{x \in h(\sigma)} d_Y(p, x) = \Phi^{\check{\mathrm{C}}}_Y(h(\sigma)).$$

Example 13. The Vietoris-Rips filtration functor $\Phi^{\mathrm{VR}} : \mathcal{M} \to \mathcal{F}$ is the map $(X, d_X) \mapsto (X, \Phi^{\mathrm{VR}}_X)$, where

$$\Phi^{\mathrm{VR}}_X(\sigma) := \mathrm{diam}(\iota_X(\sigma)), \qquad \forall\, X \in \mathcal{M},\ \sigma \subset X.$$

Functoriality for Φ^{VR} is clear, as the filtration value of a simplex does not depend on the ambient space for the Vietoris-Rips filtration. This property will be generalized in Section 4.

We now define stability of filtration functors.

Definition 14. We say that a filtration functor $\Phi : \mathcal{M} \to \mathcal{F}$ is *stable* if there exists $L \geq 0$ such that for all $X, Y \in \mathcal{M}$,

$$d_{\mathcal{F}}((X, \Phi_X), (Y, \Phi_Y)) \leq L \cdot d_{\mathrm{GH}}(X, Y).$$

If the above condition holds for a given $L \geq 0$ we call Φ an *L-stable filtration functor.*

Given any filtration functor Φ and $k \in \mathbb{Z}_+$, by dgm_k^Φ we denote the composite map

$$\mathrm{dgm} \circ \mathrm{H}_k \circ \Phi : \mathcal{M} \to \mathcal{D}.$$

Combining the stability of filtration functors with Theorem 6, one has the following:

Corollary 15 *([27]). Let $L \geq 0$ and $\Phi : \mathcal{M} \to \mathcal{F}$ be any L-stable filtration functor. Then for all $k \in \mathbb{N}$ and all $X, Y \in \mathcal{M}$,*

$$d_\mathrm{B}\big(\mathrm{dgm}_k^\Phi(X), \mathrm{dgm}_k^\Phi(Y)\big) \leq L \cdot d_\mathrm{GH}(X, Y).$$

3. An impossibility and an existence theorem

Our first result shows that there does not exist a single filtration functor that can achieve the Gromov-Hausdorff distance between any two metric spaces. Here we restrict attention to filtration functors sharing the property of the Vietoris-Rips and Čech filtration functors that all 0-simplices have filtration value zero.

Proposition 16. *Let Φ be any 1-stable filtration functor such that $\Phi_X(\{x\}) = 0$ for all $X \in \mathcal{M}$ and $x \in X$. Let $k \in \{0, 1, 2, \ldots\}$. Then there exist two different finite metric spaces X and Y such that $d_\mathrm{B}(\mathrm{dgm}_k^\Phi(X), \mathrm{dgm}_k^\Phi(Y)) < d_\mathrm{GH}(X, Y)$.*

However, it is possible to find a *family* of filtrations achieving d_GH.

Theorem 17. *There exists a family of 1-stable filtration functors $\mathfrak{F}$ such that*

$$d_\mathrm{GH}(X, Y) = \sup_{\Phi \in \mathfrak{F}} d_\mathrm{B}\big(\mathrm{dgm}_0^\Phi(X), \mathrm{dgm}_0^\Phi(Y)\big), \text{ for all } X, Y \in \mathcal{M}.$$

Remark 18. The theorem indicates that in order to fully capture geometric dissimilarity between finite metric spaces it is enough to only consider the case of 0-dimensional barcodes. However, this is merely an existence result: it does not offer guidance in terms of how the family $\mathfrak{F}$ should be chosen for applications.

Prior to proving these statements, we review the notion of *homothetic spaces.* Two spaces (X, d_X), (X, d'_X) are said to be *homothetic* if $d'_X = \lambda \cdot d_X$ for some $\lambda \geq 0$. Such spaces are related by the following equality [25, Example 3.3]:

$$d_\mathrm{GH}((X, d_X), (X, \lambda \cdot d_X)) = \tfrac{|\lambda - 1|}{2} \mathrm{diam}(X, d_X).$$

We now proceed to the proofs.

Proof of Proposition 16. Suppose first that $k \geq 1$. Towards a contradiction, let $\Phi : \mathcal{M} \to \mathcal{F}$ be a 1-stable filtration functor such that:

$$d_\mathrm{GH}(X, Y) = d_\mathrm{B}(\mathrm{dgm}_k^\Phi(X), \mathrm{dgm}_k^\Phi(Y)), \qquad \forall\, X, Y \in \mathcal{M}.$$

Let $X \in \mathcal{M}$ be the discrete 2-point space with unit distance, and let $* \in \mathcal{M}$ be the 1-point space. The only correspondence between X and $*$ is the product $X \times *$. Then,

$$d_\mathrm{GH}(X, *) = \tfrac{1}{2}\mathrm{dis}(X \times *) = \tfrac{1}{2} \max_{(x,*),(x',*) \in X \times *} |d_X(x, x') - d_*(*, *)| = \tfrac{1}{2}\mathrm{diam}(X) = \tfrac{1}{2}.$$

Both X and $*$ have trivial persistent homology in dimensions greater than 0. Thus we have:

$$\tfrac{1}{2} = d_{\mathrm{GH}}(*, X) = d_{\mathrm{B}}(\mathrm{dgm}_k^{\Phi}(*), \mathrm{dgm}_k^{\Phi}(X)) = d_{\mathrm{B}}(\emptyset, \emptyset) = 0.$$

Here the second equality holds by assumption. This is a contradiction, so such a Φ does not exist.

Next we consider the case $k = 0$. Towards a contradiction, let $\Phi : \mathcal{M} \to \mathcal{F}$ be a 1-stable filtration functor such that:

$$d_{\mathrm{GH}}(X, Y) = d_{\mathrm{B}}(\mathrm{dgm}_k^{\Phi}(X), \mathrm{dgm}_k^{\Phi}(Y)), \qquad \forall\, X, Y \in \mathcal{M}.$$

We proceed by first characterizing the persistence diagram of a two-point space for different interpoint distances. For each $r > 0$, let $X_r := \{0, r\} \subset \mathbb{R}$ be the space having two points at distance r from each other. By assumption, $\Phi_{X_r}(\{0\}) = \Phi_{X_r}(\{r\}) = 0$. Then there exists $f(r) > 0$ such that,

$$\mathrm{dgm}_0^{\Phi}(X_r) = \{[0, \infty), [0, f(r))\}.$$

Also it is clear that $\mathrm{dgm}_0^{\Phi}(*) = \{[0, \infty)\}$. From this we see that,

$$d_{\mathrm{GH}}(X_r, *) = d_{\mathrm{B}}(\mathrm{dgm}_0^{\Phi}(X), \mathrm{dgm}_0^{\Phi}(*)) = d_{\mathrm{B}}(\{[0, \infty), [0, f(r))\}, \{[0, \infty)\}) = \tfrac{f(r)}{2}.$$

Here the first equality holds by assumption. We know that $d_{\mathrm{GH}}(X, *) = \frac{\mathrm{diam}(X)}{2}$ for all spaces $X \in \mathcal{M}$. Thus we have $d_{\mathrm{GH}}(X_r, *) = \frac{r}{2}$. It follows from the preceding computation that $f(r) = r$.

Now let $0 < \varepsilon < 1$. Because X_1 and $X_{1+\varepsilon}$ are homothetical spaces [25, Example 3.3], $d_{\mathrm{GH}}(X_1, X_{1+\varepsilon}) = \frac{\varepsilon}{2}$. But the previous work shows that $d_{\mathrm{B}}(\mathrm{dgm}_0^{\Phi}(X_1), \mathrm{dgm}_0^{\Phi}(X_{1+\varepsilon})) = d_{\mathrm{B}}(\{[0, \infty), [0, 1)\}, \{[0, \infty), [0, 1 + \varepsilon)\}) = \varepsilon \neq \frac{\varepsilon}{2} = d_{\mathrm{GH}}(X_1, X_{1+\varepsilon})$. We conclude from this contradiction that such a filtration functor Φ does not exist. □

Proof of Theorem 17. For each $Z \in \mathcal{M}$ we define the filtration functor $\Upsilon^{(Z)} : \mathcal{M} \to \mathcal{F}$ given by $X \mapsto \Upsilon_X^{(Z)}$, where

$$\Upsilon_X^{(Z)}(\sigma) := d_{\mathrm{GH}}(X, Z) \qquad \forall\, \sigma \subset X \in \mathcal{M}.$$

This filtration is constant on simplices. At the simplicial level, it yields an empty simplicial complex on the interval $[0, d_{\mathrm{GH}}(X, Z))$, and the complete simplicial complex, i.e. the power set of X, on the interval $[d_{\mathrm{GH}}(X, Z), \infty)$. Thus $\mathrm{dgm}_0^{\Upsilon^{(Z)}}(X) = \{[d_{\mathrm{GH}}(X, Z), \infty)\}$.

By the definition of the bottleneck distance [3] and the triangle inequality,

$$d_{\mathrm{B}}\left(\mathrm{dgm}_0^{\Upsilon^{(Z)}}(X), \mathrm{dgm}_0^{\Upsilon^{(Z)}}(Y)\right) = |d_{\mathrm{GH}}(X, Z) - d_{\mathrm{GH}}(Y, Z)| \leq d_{\mathrm{GH}}(X, Y).$$

This holds for all $X, Y, Z \in \mathcal{M}$, thus all functors in $\mathfrak{F}$ are 1-stable.

Since $d_{\mathrm{GH}}(Y, Y) = 0$, it follows that

$$d_{\mathrm{B}}\left(\mathrm{dgm}_0^{\Upsilon^{(Y)}}(X), \mathrm{dgm}_0^{\Upsilon^{(Y)}}(Y)\right) = d_{\mathrm{GH}}(X, Y) \qquad \forall X, Y \in \mathcal{M}.$$

We conclude that,

$$d_{\mathrm{GH}}(X, Y) = \sup_{Z \in \mathcal{M}} d_{\mathrm{B}}\left(\mathrm{dgm}_0^{\Upsilon^{(Z)}}(X), \mathrm{dgm}_0^{\Upsilon^{(Z)}}(Y)\right) \qquad \forall X, Y \in \mathcal{M}.$$

Setting $\mathfrak{F} := \left\{\Upsilon^{(Z)} : \mathcal{M} \to \mathcal{F} :\ Z \in \mathcal{M}\right\}$ gives us the required family. □

Remark 19. See [11, Theorem 15] for a related result in the category of topological spaces.

The proof of this theorem exhibits a family $\mathfrak{F}$ which recovers the Gromov-Hausdorff distance through the bottleneck distances induced by the filtration functors in $\mathfrak{F}$. The family $\mathfrak{F}$ is not suitable for applications, since for any $X \in \mathcal{M}$ and $\Upsilon_Z \in \mathfrak{F}$, computing $\mathrm{dgm}_0^{\Upsilon_Z}(X)$ requires computing $d_{\mathrm{GH}}(X, Z)$. Thus the filtration functors in $\mathfrak{F}$ do not reduce the computational problem of estimating d_{GH}. The theorem does, however, guarantee the existence of families of filtration functors that recover the Gromov-Hausdorff distance. Such a theorem motivates identifying families of stable filtration functors that are sufficiently rich in terms of the metric information they capture.

4. Valuations

In this section we introduce the notion of a *valuation*, which is our main tool for defining families of filtration functors.

4.1. Definition and stability

The n-th curvature set $\mathrm{K}_n(X)$ of a metric space X contains all the metric information about n-tuples of points in X. One possibility for inducing a filtration on X is to use the information in $\mathrm{K}_n(X)$. This requires us to specify a rule that assigns a real number to elements of $\mathrm{pow}(\mathbb{R}^{n\times n})$. This is done via the notion of valuations.

Definition 20 *(Valuation).* Given $n \in \mathbb{N}$, an *n-valuation* is any map $\nu_n : \mathrm{pow}(\mathbb{R}^{n\times n}) \to \mathbb{R}$ such that ν_n is *monotonic*, meaning that $\nu_n(A) \le \nu_n(B)$ for all $A \subset B \in \mathrm{pow}(\mathbb{R}^{n\times n})$. We will denote by $\mathfrak{V}_n$ the set of all n-valuations.[1]

We are interested in defining filtrations through valuations. Monotonicity is imposed on valuations so that they induce filtrations.

Definition 21 *(Filtration functor induced by a valuation).* Given $n \in \mathbb{N}$ and any $\nu_n \in \mathfrak{V}_n$ we induce the filtration functor $\Phi^{\nu_n} : \mathcal{M} \to \mathcal{F}$ by writing:

$$\Phi_X^{\nu_n}(\sigma) := (\nu_n \circ \mathrm{K}_n \circ \iota_X)(\sigma), \qquad \forall X \in \mathcal{M}, \forall \sigma \subset X.$$

We write Φ^{ν_n} to denote the *filtration functor induced by ν_n*.

Lemma 22. *Φ^{ν_n} is a well-defined filtration map that is functorial on $\mathcal{M}^{\mathrm{iso}}$.*

Proof. Let $\nu_n \in \mathfrak{V}_n$ be a valuation. Let $X \in \mathcal{M}$ and $\tau \subset \sigma \subset X$. By Remark 3, $\tau \subset \sigma$ implies $\mathrm{K}_n(\iota_X(\tau)) \subset \mathrm{K}_n(\iota_X(\sigma))$. Since ν_n is monotonic,

$$\Phi_X^{\nu_n}(\tau) = \nu_n(\mathrm{K}_n(\iota_X(\tau))) \le \nu_n(\mathrm{K}_n(\iota_X(\sigma))) = \Phi_X^{\nu_n}(\sigma).$$

Thus, $\Phi_X^{\nu_n}$ is a filtration on X and Φ^{ν_n} is a filtration map. To see $\mathcal{M}^{\mathrm{iso}}$-functoriality, let $h : X \to Y$ be an isometric embedding. Then in fact we have the following equality:

$$\Phi_X^{\nu_n}(\sigma) = \nu_n(\mathrm{K}_n(\iota_X(\sigma))) = \nu_n(\mathrm{K}_n(\iota_Y(h(\sigma)))) = \Phi_Y^{\nu_n}(h(\sigma)). \quad \square$$

[1] Our use of the term valuation deviates from the usual meaning: we do not assume modularity [28].

Definition 23. A filtration functor $\Phi : \mathcal{M} \to \mathcal{F}$ is *local* if for some $n \in \mathbb{N}$ there exists a valuation $\nu_n \in \mathfrak{V}_n$ such that $\Phi = \Phi^{\nu_n}$, in which case we say that Φ is a *n-local filtration functor*. If no such n exists then we say that Φ is *global*.

Remark 24. We make the following remarks:

1. Notice that if the filtration functor Φ is n-local then it is n'-local for all $n' \geq n$. By convention, whenever we say a filtration functor is n-local, we are referring to the minimal n for which this is true.
2. *The Vietoris-Rips filtration functor is* 2*-local.* Note that for any $X \in \mathcal{M}$ and $\sigma \subset X$:

$$\Phi_X^{\mathrm{VR}}(\sigma) = \mathrm{diam}(\sigma) = \max_{\{x,x'\} \subset \sigma} d_X(x,x') = \max_{\alpha \in \mathrm{K}_2(\iota_X(\sigma))} \max_{i,j} \alpha_{ij}.$$

   In words, the filtration value of a simplex under the Vietoris-Rips filtration functor depends only on the pairwise distances between points in the simplex, which is information contained in $\mathrm{K}_2(\sigma)$. A 2-valuation ν_2 generates the Vietoris-Rips filtration functor: $\nu_2(A) := \max_{\alpha \in A} \|\alpha\|_\infty$ for all $A \in \mathrm{pow}(\mathbb{R}^{2\times 2})$.
3. *The Čech filtration functor is not local.* To show this, we use a counterexample: Let (P, d_P) be the metric space consisting of three equidistant points $P = \{p_1, p_2, p_3\}$ at distance 2 from each other. Let (Q, d_Q) be the metric space consisting of four points $Q = \{q_0, q_1, q_2, q_3\}$ where the metric is given by the matrix

$$d_Q = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 \\ 1 & 2 & 0 & 2 \\ 1 & 2 & 2 & 0 \end{pmatrix}.$$

   Since the subset $S = \{q_1, q_2, q_3\} \subset Q$ and P are isometric, $\mathrm{K}_n(\iota_Q(S)) = \mathrm{K}_n(P)$ for all $n \in \mathbb{N}$. If the Čech functor were n-local for some $n \in \mathbb{N}$, then there would exist $\nu_n \in \mathfrak{V}_n$ such that $1 = \Phi_Q^{\mathrm{C}}(S) = \nu_n(\mathrm{K}_n(\iota_Q(S))) = \nu_n(\mathrm{K}_n(P)) = \Phi_P^{\mathrm{C}}(P) = 2$, which yields a contradiction.
4. *All local filtration maps are* $\mathcal{M}^{\mathrm{iso}}$*-functorial.* This follows from Lemma 22.

We now restrict to a suitable class of valuations that yield stable filtration functors.

Definition 25. Let $n \in \mathbb{N}$ and let $\nu_n \in \mathfrak{V}_n$. Given $L \geq 0$, we say that ν_n is *L-stable* if:

$$|\nu_n(A) - \nu_n(B)| \leq L \cdot d_{\mathrm{H}}^{(\mathbb{R}^{n\times n}, \ell_\infty)}(A, B), \qquad \text{for all nonempty } A, B \subset \mathbb{R}^{n\times n}.$$

We denote by $\mathfrak{V}_n^L$ the subset of $\mathfrak{V}_n$ consisting of all L-stable n-valuations.

By invoking Theorem 6 and the stability of the K_n map (cf. Theorem 4), we obtain:

Theorem 26. *Let $n \in \mathbb{N}$ and $L > 0$. Then for any $\nu_n \in \mathfrak{V}_n^L$, one has:*

$$d_{\mathrm{B}}\left(\mathrm{dgm}_k^{\Phi^{\nu_n}}(X), \mathrm{dgm}_k^{\Phi^{\nu_n}}(Y)\right) \leq 2L \cdot d_{\mathrm{GH}}(X,Y), \qquad \forall X, Y \in \mathcal{M}, \forall k \in \mathbb{N}.$$

We omit the proof, as it will follow from a more general result (Theorem 49).

4.2. Max-induced valuations

Thus far we have discussed the definition and properties of valuations, and here we provide a simple way of *generating* valuations. As we show below, any Lipschitz function can be used to generate valuations.

Given a function $f : \mathbb{R}^{n\times n} \to \mathbb{R}$, we define its *max-induced valuation* or *max-valuation*, denoted ν^f, to be:

$$\nu_n^f(A) := \max_{\alpha\in A} f(\alpha), \qquad \forall\, A \in \mathrm{pow}(\mathbb{R}^{n\times n}).$$

The valuations in this particular class are well behaved when imposing minimal conditions on the function that induces them. In particular, if we restrict our scope to functions that are Lipschitz with respect to the sup norm, we obtain a family of stable valuations.

Proposition 27. *Let $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ be L-Lipschitz with respect to the ℓ_∞ norm on $\mathbb{R}^{n\times n}$. Then the max-induced valuation ν_n^f is L-stable.*

Proof. Let $A, B \in \mathrm{pow}(\mathbb{R}^{n\times n})$ and $\delta = d_{\mathrm{H}}^{(\mathbb{R}^{n\times n},\ell_\infty)}(A, B)$. Let $\alpha_0 \in A$ such that $f(\alpha_0) = \nu_n^f(A)$. There exists $\beta_0 \in B$ such that $\|\alpha_0 - \beta_0\|_\infty \le \delta$. From the Lipschitz continuity,

$$\nu_n^f(A) - \nu_n^f(B) \le f(\alpha_0) - f(\beta_0) \le L\,\delta.$$

Following the symmetrical argument choosing $\beta_0 \in B$,

$$|\nu_n^f(A) - \nu_n^f(B)| \le L \cdot d_{\mathrm{H}}^{(\mathbb{R}^{n\times n},\ell_\infty)}(A, B). \quad \square$$

In the next subsection we study a particular example of a filtration functor generated by a max-induced valuation: the ultrametricity filtration functor Φ^{ult}.

4.3. The ultrametricity filtration functor

We now define a filtration based on ultrametricity, which is a measure of the defect of a metric from being ultrametric [29].

Definition 28. Let $(X, u_X) \in \mathcal{M}$ be a metric space. Then u_X is an *ultrametric* if:

$$u_X(x_1, x_3) \le \max\{u_X(x_1, x_2), u_X(x_2, x_3)\} \quad \forall\, x_1, x_2, x_3 \in X. \qquad (*)$$

One significant property of ultrametric spaces is the following: if (X, u_X) is ultrametric, then for all 2-simplices $\sigma \subset X$, one can write $\sigma = \{x_1, x_2, x_3\}$ such that $u_X(x_1, x_2) = u_X(x_2, x_3) \ge u_X(x_1, x_3)$. In other words, every triangle in an ultrametric space is isosceles, with two longest sides equal. This property actually characterizes ultrametric spaces, and is used in the following definition [22].

Definition 29 *([22]).* Let $X \in \mathcal{M}$. The *ultrametricity of* X is defined as

$$\mathrm{ult}(X) := \max_{x_1,x_2,x_3\in X} \left(d_X(x_1, x_3) - \max\{d_X(x_1, x_2), d_X(x_2, x_3)\}\right).$$

We call $\mathrm{ult} : \mathcal{M} \to \mathbb{R}$ the *ultrametricity map*. This naturally induces a filtration functor.

Definition 30 *(Ultrametricity filtration functor).* The *ultrametricity filtration functor* is the map $\Phi^{\mathrm{ult}} : \mathcal{M} \to \mathcal{F}$ given by writing $\Phi_X^{\mathrm{ult}}(\sigma) := \mathrm{ult}(\iota_X(\sigma))$ for each $\sigma \subset X$.

The map Φ^{ult} is a well-defined filtration functor that satisfies 4-stability. All of this can be proved by showing first that it is 3-local.

Proposition 31. Φ^{ult} *is a* 3*-local and* 4*-stable filtration functor, induced by the* 2*-stable valuation* $\nu_{\mathrm{ult}} \in \mathfrak{V}_3$ *given by writing* $\nu_{\mathrm{ult}}(A) := \max_{\alpha\in A}(\alpha_{13} - \max\{\alpha_{12}, \alpha_{23}\})$*, for* $A \subset \mathbb{R}^{n\times n}$*.*

Proof. Let us prove that Φ^{ult} is a filtration generated by ν_{ult}. For all $X \in \mathcal{M}$ and $\sigma \subset X$,

$$\begin{aligned}\Phi^{\mathrm{ult}}_X(\sigma) &= \max_{x_1,x_2,x_3\in\sigma} (d_X(x_1,x_3) - \max\{d_X(x_1,x_2), d_X(x_2,x_3)\})\\ &= \max_{\alpha\in \mathrm{K}_3(\iota_X(\sigma))} (\alpha_{13} - \max\{\alpha_{12},\alpha_{23}\}) = \nu_{\mathrm{ult}}(\mathrm{K}_3(\iota_X(\sigma))).\end{aligned}$$

Given $A \subset B \in \mathrm{pow}(\mathbb{R}^{n\times n})$, we have $\nu_{\mathrm{ult}}(A) \le \nu_{\mathrm{ult}}(B)$ because we take a max over a larger set. From this, Φ^{ult} is a well-defined 3-local filtration. This valuation is max-induced by the function $f : \mathbb{R}^{3\times 3} \to \mathbb{R}$ given by $f(\alpha) = \alpha_{13} - \max\{\alpha_{12},\alpha_{23}\}$. Since f is 2-Lipschitz, by Proposition 27, ν_{ult} is 2-stable. Let us denote the k-th dimensional persistence diagram map induced by Φ^{ult} as $\mathrm{dgm}^{\mathrm{ult}}_k$. By applying Theorem 26, Φ^{ult} is 4-stable: for all $k \in \mathbb{N}$ and $X, Y \in \mathcal{M}$,

$$d_{\mathrm{B}}\left(\mathrm{dgm}^{\mathrm{ult}}_k(X), \mathrm{dgm}^{\mathrm{ult}}_k(Y)\right) \le 4 \cdot d_{\mathrm{GH}}(X,Y). \quad \square$$

Remark 32. There are other interesting features of this filtration that make it essentially different from the Vietoris-Rips or the Čech filtration functors. We list some in the following remarks:

1. For any two-point metric space $P = \{p_1, p_2\}$ with distance d_P, one has $\mathrm{ult}(P) = 0$. Then, for any $X \in \mathcal{M}$, every 1-simplex of $\mathrm{pow}(X)$ has filtration value 0. This implies that $\mathrm{dgm}^{\mathrm{ult}}_0(X) = \{[0,\infty)\}$ for any $X \in \mathcal{M}$.
2. For a space $X \in \mathcal{M}$ and a simplex $\sigma \subset X$ the filtration value of σ with respect to Φ^{ult} equals $\max_{x_1,x_2,x_3\in\sigma} \mathrm{ult}(\iota_X(\{x_1,x_2,x_3\}))$. From this, for all $t \in \mathbb{R}_+$, $\Phi^{\mathrm{ult}}[t]$ is the flag complex generated by $\{\tau \subset X : \tau \text{ a 2-simplex}, \Phi^{\mathrm{ult}}_X(\tau) \le t\}$. This means that at all time $\Phi^{\mathrm{ult}}_X[t]$ is the flag complex of the 2-skeleton of $\Phi^{\mathrm{ult}}_X[t]$. This draws a parallel between Φ^{ult} and Φ^{VR}, since for all $t \ge 0$ $\Phi^{\mathrm{VR}}_X[t]$ is the flag complex generated by the 1-skeleton of $\Phi^{\mathrm{VR}}_X[t]$.
3. Despite this parallel structure, Φ^{ult} appears to be more informative than Φ^{VR} for certain datasets arising in phylogenetics, as we show in Section 6.1.2.

Example 33. Using Remark 32, we now construct examples of metric spaces with nontrivial 1 and 2-dimensional persistence with respect to Φ^{ult}. Illustrations are provided in Fig. 2. First let $X' = \{p_1, p_2, p_3\}$ be a three point metric space corresponding to an isosceles triangle with two shortest sides equal, each having length α. Let the length of the longest side be β. Then $\mathrm{ult}(X') = \beta - \alpha > 0$, and we can force this quantity to be arbitrarily large. Thus $\mathrm{dgm}^{\mathrm{ult}}_1(X')$ consists of a single bar $[0, \mathrm{ult}(X'))$. Next we obtain X from X' by "gluing in" three metric spaces X_1, X_2, X_3, each having diameter $\ll \alpha$, to the three points of X'. The metric on X is given as follows: for any $x_i \in X_i$ and $x_j \in X_j$, define $d_X(x_i,x_j) := d_{X_i}(x_i,p_i) + d_{X'}(p_i,p_j) + d_{X_j}(p_j,x_j)$. Then X is a metric space of arbitrary cardinality having arbitrarily long bars in $\mathrm{dgm}^{\mathrm{ult}}_1$.

Next let $Y' = \{q_1, q_2, q_3\}$ be an isosceles right triangle in the plane, with q_1, q_3 the endpoints of the hypotenuse. Let m denote the midpoint of the hypotenuse; then m is equidistant to each q_i, q_j pair. Next let Y be the suspension of Y' obtained by adjoining a, b above and below m at sufficiently large distance so that $\mathrm{ult}(\iota_Y(a, q_i, q_j)) = 0$ for each $i, j \in \{1,2,3\}$, and likewise for b. Then in the ultrametricity filtration, the 2-simplices $[a, q_i, q_j], [b, q_i, q_j]$ enter at time 0, and they form a 2-cycle that remains alive until the entry of $[a, q_i, q_j, q_k], [b, q_i, q_j, q_k]$. This happens when $[q_i, q_j, q_k]$ enters, which is controlled by $\mathrm{ult}(Y')$. Thus $\mathrm{dgm}^{\mathrm{ult}}_2(Y)$ is nontrivial. A similar gluing procedure as in the case of X above can be carried out to extend this example to metric spaces with arbitrary cardinality having arbitrarily long bars in $\mathrm{dgm}^{\mathrm{ult}}_2$.

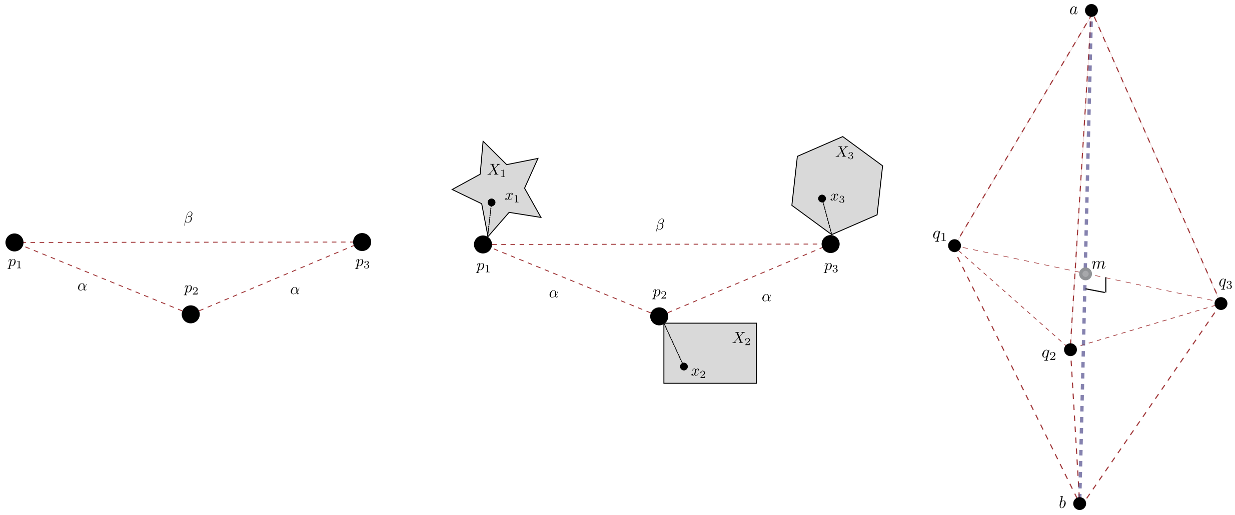

Fig. 2. Illustrations of the metric spaces in Example 33 with nontrivial $\mathrm{dgm}_k^{\mathrm{ult}}$, for $k = 1, 2$.

In Section 6.1.2, we demonstrate some practical applications of Φ^{ult}. These computational experiments demonstrate cases where the ultrametricity filtration provides more persistence information than the Vietoris-Rips filtration. In addition to empirical applications on finite metric spaces, we have observed that Φ^{ult} yields surprising results for path connected metric spaces such as the circle $\mathbb{S}^1$. Notably, we find that $\mathrm{dgm}_1^{\mathrm{ult}}$ is trivial for path connected metric spaces (this subsumes the $\mathbb{S}^1$ case). Computational experiments suggest that for a collection of standard path connected metric spaces $\mathrm{dgm}_2^{\mathrm{ult}}$ is trivial. In contrast, computational results suggest that $\mathrm{dgm}_3^{\mathrm{ult}}(\mathbb{S}^1)$ consists of a single bar $[0, \frac{2\pi}{7})$.

Proposition 34. *Let (X, d_X) be a path connected metric space. Then $\mathrm{dgm}_1^{\mathrm{ult}}(X)$ is trivial.*

To prove this proposition, we first provide the following lemma. The proof of the lemma requires two applications of the special property that all 1-simplices enter at time 0 in the ultrametricity filtration.

Lemma 35. *Let (X, d_X) be a path connected metric space, and consider the 1-cycle $\sigma := [v_0, v_1] + [v_1, v_2] + [v_2, v_0]$ where $v_0, v_1, v_2 \in X$. Let $\epsilon > 0$. Then σ becomes a 2-boundary by time ϵ under the ultrametricity filtration.*

Proof. Let $\epsilon > 0$ be given. Fix a continuous curve γ connecting v_1, v_2. We fix $n \in \mathbb{N}$ such that we can select consecutive points $v_1 = x_0, x_1, \ldots, x_n = v_2$ with x_i on γ for $0 \le i \le n$ and $d_X(x_i, x_{i+1}) < \epsilon$. Fix such a choice $x_1, x_2, \ldots, x_{n-1}$. Then consider the 2-chain $\tau := \sum_{i=0}^{n-1} [v_0, x_i, x_{i+1}]$. We first claim that τ will appear in the ultrametricity filtration by time ϵ. To see this, we consider the ultrametricity of each 2-simplex in the sum.

Let $0 \le i \le n-1$, and let $d_i := d_X(v_0, x_i)$. By the triangle inequality we know that $d_i < d_{i+1} + \epsilon$ and likewise $d_{i+1} < d_i + \epsilon$. This gives $|d_i - d_{i+1}| < \epsilon$. We know the ultrametricity of $\tau_i := [v_0, x_i, x_{i+1}]$ is given as the difference of the two largest distances between the vertices. If these two largest distances are d_i and d_{i+1}, then by the above, $\mathrm{ult}(\tau_i) < \epsilon$. The only other case is if one of the two largest distances in the 2-simplex is $d_X(x_i, x_{i+1})$. Without loss of generality, let d_i be one of the other two largest distances. But as one of the two largest distances is $d_X(x_i, x_{i+1})$, we know $d_{i+1} < d_X(x_i, x_{i+1}) < \epsilon$, and so by the triangle inequality and the preceding observations,

$$d_i < d_{i+1} + \epsilon \implies d_i < d_X(x_i, x_{i+1}) + \epsilon \implies d_i - d_X(x_i, x_{i+1}) < \epsilon$$

But then as $d_X(x_i, x_{i+1}) < \epsilon$ already, we also have $d_X(x_i, x_{i+1}) - d_i < \epsilon$, and so $|d_X(x_i, x_{i+1}) - d_i| < \epsilon$, meaning in either case we get $\mathrm{ult}(\tau_i) < \epsilon$.

So we have that each τ_i appears by time ϵ. Thus the entire 2-chain τ appears by time ϵ. Now we can compute $\sigma' := \partial\tau = [v_0, v_1] + [v_2, v_0] + \sum_{i=0}^{n-1}[x_i, x_{i+1}]$. This is not σ, but we claim that σ', σ are homologous. To see this, observe that σ, σ' differ by the boundary of the 2-chain $\sum_{i=1}^{n-1}[x_0, x_i, x_{i+1}]$. Then following the analogous case argument from above, we get that each 2-simplex $[x_0, x_i, x_{i+1}]$ appears in the ultrametricity filtration by time ϵ. Thus we have that σ is homologous, via a boundary that exists by time ϵ, to the 1-cycle σ' which is a 2-boundary by time ϵ. Hence the death time of σ as a homology class is $< \epsilon$. □

Now we use Lemma 35 to prove the proposition. Notice that we again use the property that all 1-simplices enter at time 0 in the ultrametricity filtration.

Proof of Proposition 34. Let $\epsilon > 0$ be given, and $\sigma = \sum_{i=0}^{n-1}[v_i, v_{i+1}] + [v_n, v_0]$ be a 1-cycle. We can rewrite σ as $\sigma = \sum_{i=0}^{n-1}[v_0, v_i] + [v_i, v_{i+1}] - [v_0, v_{i+1}]$. Then by Lemma 35, we have that each term $[v_0, v_i] + [v_i, v_{i+1}] - [v_0, v_{i+1}]$ in σ is a 2-boundary of some 2-chain τ_i by time ϵ. Thus we have $\sigma = \partial\left(\sum_{i=0}^{n-1}\tau_i\right)$, which is a 2-chain appearing by time ϵ, and so the claim is shown for the case of $\mathrm{dgm}_1^{\mathrm{ult}}(X)$. □

As $\mathbb{S}^1$ is a path connected space, Proposition 34 tell us that the 1-dimensional persistence of $\mathbb{S}^1$ is trivial. One wonders if this triviality holds into higher dimensions. We have computationally found that it is trivial in dimension 2 as well, but there is nontrivial persistence in dimension 3.

Remark 36 *(The case of* $\mathrm{dgm}_2^{\mathrm{ult}}$*).* We have computationally observed that for the interval $[0, 1]$, and for the geodesic and Euclidean versions of $\mathbb{S}^1$ and $\mathbb{S}^2$, $\mathrm{dgm}_2^{\mathrm{ult}}$ is negligible for a wide range of sampling densities of these spaces.

Conjecture 1. *For any compact geodesic space X, we have* $\mathrm{dgm}_2^{\mathrm{ult}}(X)$ *is trivial.*

Remark 37 *(The case of* $\mathrm{dgm}_3^{\mathrm{ult}}$*).* Let $\mathbb{S}_m^1$ for $m \in \mathbb{Z}_+$ be an equidistant sampling of m points on $\mathbb{S}^1$, with the metric induced by the geodesic metric on $\mathbb{S}^1$. We used Javaplex [30] and Dionysus [31] to compute the persistent homology of $\mathbb{S}_m^1$ for dimensions up to 3 under the ultrametricity filtration. For $m = 7, 8, \ldots, 50$, let $m = 7 \cdot q + r$ be the Euclidean algorithm factorization of m, with $0 \leq r < 7$. Then, we experimentally observed that, for each m in the range above, the 3-dimensional persistence will consist of some number repetitions of the interval $\left[\frac{2\pi}{m}, \frac{4\pi}{m}\right)$, alongside one other interval of significant length:

$$\mathrm{dgm}_3^{\mathrm{ult}}(\mathbb{S}^1) \ni [b, d) = \begin{cases} \left[\frac{2\pi}{m}, \frac{2\cdot q\pi}{m}\right) & r = 0 \\ \left[\frac{2\pi}{m}, \frac{2\cdot(q+1)\pi}{m}\right) & r > 0 \end{cases}$$

This computational experiment coupled with the stability result in Proposition 31 suggests the following conjecture:

Conjecture 2. *For the geodesic space $\mathbb{S}^1$, which we get as a metric space via taking $m \to \infty$, we have*

$$\mathrm{dgm}_3^{\mathrm{ult}}(\mathbb{S}^1) = \left[0, \frac{2\pi}{7}\right).$$

We want to point out that although the Vietoris-Rips barcodes of S1 are now known, it took significant effort to characterize them fully [32].

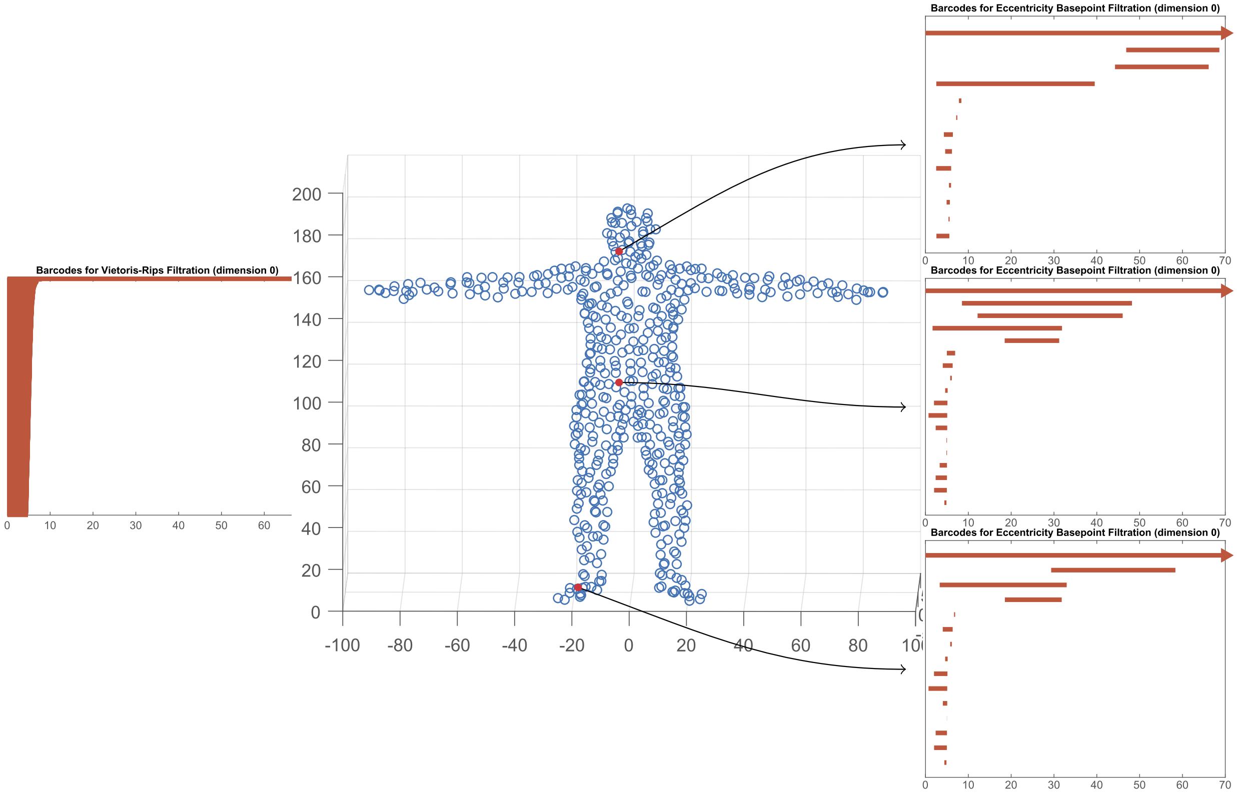

Fig. 3. Left: VR barcode. **Right:** barcodes from different basepoints using the eccentricity basepoint filtration (Section 6.2). Observe that the basepoint filtration is much more informative and is more sensitive to the presence of extremities in data.

5. Generalizing to basepoint filtration functors

The global and local filtration functors that we have considered so far can be generalized further by allowing these filtrations to depend on a basepoint (see e.g. [23,20]). We proceed to the definition and properties of basepoint and local basepoint filtrations, and end with a concrete example of a basepoint filtration.

5.1. Basepoint filtration functors

Continuing the ideas discussed in Section 1 and Theorem 17, we now consider an assignment of a filtration functor to each point of a space, which we refer to as a *basepoint filtration functor*. Instead of mapping each finite metric space to a single filtered space, this functor maps a metric space to a collection of filtrations, each depending on a choice of basepoint (see Fig. 3). This representation gives specific localized information about the space, and allows the user to pick out filtrations emphasizing the importance of a particular region of the space.

Definition 38. For any $X \in \mathcal{M}$, a *basepoint family of filtrations on* X is any collection

$$\{\Psi_{X,x_0} : \mathrm{pow}(X) \to \mathbb{R}\}_{x_0 \in X},$$

where for all $x_0 \in X$, Ψ_{X,x_0} is a filtration on X.

Example 39. Given a geodesic space (X, d_X), the sublevel set filtrations associated to $\{d_X(p, \cdot)\}_{p \in X}$ form a basepoint family of filtrations.

Definition 40. A *basepoint filtration map* is a map $\Psi : \mathcal{M} \to \mathrm{pow}(\mathcal{F})$ such that Ψ takes any $X \in \mathcal{M}$ to a basepoint family of filtrations $\Psi_X = \{\Psi_{X,x_0} : \mathrm{pow}(X) \to \mathbb{R}\}_{x_0 \in X}$. Basepoint filtration maps are said to be $(\mathcal{M}^{\mathrm{iso}})$ functorial if the following condition is satisfied:

Given $X, Y \in \mathcal{M}$ and $h : X \to Y$ an isometric embedding, one has $\Psi_{X,x_0}(\sigma) \geq \Psi_{Y,h(x_0)}(h(\sigma))$ for all $\sigma \subset X$ and $x_0 \in X$.

Definition 41. When we simply refer to a *basepoint filtration functor*, we mean an $\mathcal{M}^{\mathrm{iso}}$-functorial basepoint filtration map.

Remark 42. If instead we have $\Psi_{X,x_0}(\sigma) \leq \Psi_{Y,h(x_0)}(h(\sigma))$, then Ψ is said to be a *contravariant* basepoint filtration functor.

Note that any filtration as we have previously defined gives rise to a basepoint filtration:

Example 43 *(Constant basepoint filtration functors).* Let $\Phi : \mathcal{M} \to \mathcal{F}$ be any filtration functor. We define the constant basepoint filtration functor Ψ^{const} to be

$$\Psi^{\mathrm{const}}_X(x_0) = \Phi_X, \qquad \forall\, X \in \mathcal{M},\ x_0 \in X.$$

For example, for all $x_0 \in X$, define Ψ_{X,x_0} to be the Vietoris-Rips filtration. Then this basepoint filtration functor contains the same information as the Vietoris-Rips filtration.

As before, we have to impose stability conditions on these filtrations so that there is a relation not only among spaces, but also among filtrations of the same space based at different basepoints. We now define the cost function associated to a basepoint filtration functor and use it to define a notion of stability.

Definition 44. Given $X, Y \in \mathcal{M}$, $k \in \mathbb{Z}_+$, and a basepoint filtration functor Ψ, the *k-dimensional cost function induced by* Ψ is the map $\mathcal{C}_{\Psi,k} : X \times Y \to \mathbb{R}_+$ given by

$$(x, y) \mapsto d_{\mathrm{B}}\left(\mathrm{dgm}_k^{\Psi_{X,x}}(X), \mathrm{dgm}_k^{\Psi_{Y,y}}(Y)\right).$$

Definition 45. Let $L > 0$. A basepoint filtration functor Ψ is L-stable if

$$\min_{R \in \mathcal{R}(X,Y)} \max_{(x_0,y_0) \in R} \mathcal{C}_{\Psi,k}(x_0, y_0) \leq L \cdot d_{\mathrm{GH}}(X, Y), \text{ for all } X, Y \in \mathcal{M}, \text{ for all } k \in \mathbb{Z}_+.$$

Having defined the preliminaries of basepoint filtrations, we now merge this concept with that of valuations to obtain "local" basepoint filtrations. Here "local" refers to the locality of valuations described in Definition 23. Additionally we incorporate the notion of a *point descriptor* which encodes information about the particular position of a basepoint in the space.

5.2. Local basepoint filtration functors

Let $n \in \mathbb{N}$. For all $X \in \mathcal{M}$, $x_0 \in X$ and $\sigma \subset X$, we define the *basepoint n-th curvature set of* σ to be

$$\mathrm{K}_n(x_0, \sigma) := \left\{ D_X^{(n+1)}(x_0, x_1, \ldots, x_n) \in \mathbb{R}^{(n+1)\times(n+1)} : x_1, \ldots, x_n \in \sigma \right\}.$$

This set generates $\mathrm{K}_n(\iota_X(\sigma))$ and appends the distances between σ and the basepoint x_0 to the first row and column, i.e. the top-left corner of the matrix. This means that if we define the projection $\pi : \mathbb{R}^{(n+1)\times(n+1)} \to \mathbb{R}^{n\times n}$ to be $\pi(A) = (a_{ij})_{i,j=2}^{n+1}$, then $\pi(\mathrm{K}_n(x_0, \sigma)) = \mathrm{K}_n(\iota_X(\sigma))$.

Given a natural number n and any stable $(n+1)$-valuation ν_{n+1}, one can construct the basepoint filtration functor $\Psi^{\nu_{n+1}}$ defined as follows: each $X \in \mathcal{M}$ induces the basepoint map $\Psi_X^{\nu_{n+1}} : X \to \mathcal{F}$ given by

$$x \mapsto \Psi_{X,x}^{\nu_{n+1}},$$

where the filtration value of a simplex $\sigma \subset X$ is given by

$$\Psi_{X,x}^{\nu_{n+1}}(\sigma) := \nu_{n+1}(\mathrm{K}_n(x, \sigma)).$$

The family $\Psi^{\nu_{n+1}}$ is a basepoint family of filtration functors induced by ν_{n+1}. This is a possible path through which we can generate basepoint filtration functors, but it has a disadvantage. Although we are considering information about the basepoint and the simplex, we are not taking into account the particular position of the basepoint in the space. This could be measured by other extrinsic quantities that cannot be computed using only the basepoint curvature sets, which are intrinsic. For this, we define first the notion of functions that describe the point with respect to the whole space.

To do so, fix $\ell \in \mathbb{N}$ and let $\mathcal{M}_\ell$ be the collection of all triplets (X, d_X, f_X) where (X, d_X) is a finite metric space and $f_X : X \to \mathbb{R}^\ell$ is a function. We will think of a map from $\mathcal{M}$ to $\mathcal{M}_\ell$ to be a relation that, for each $X \in \mathcal{M}$ and a point $x_0 \in X$, assigns a set of quantities that describe the position of the point with respect to the space.

Definition 46. A *point descriptor* $\rho : \mathcal{M} \to \mathcal{M}_\ell$ is a map $(X, d_X) \mapsto (X, d_X, \rho_X : X \to \mathbb{R}^\ell)$ such that there exists a constant $K > 0$ with the property that for all $X, Y \in \mathcal{M}$ and any correspondence $R \subset X \times Y$,

$$\max_{(x_0,y_0)\in R} \|\rho_X(x_0) - \rho_Y(y_0)\|_\infty \leq K \cdot \mathrm{dis}(R).$$

We say that ρ is *K-stable* if this condition is satisfied for some $K \geq 0$.

By incorporating point descriptors in valuations, we consider extrinsic features of the basepoint in the filtration value of simplices. We refer to these objects as *adjusted valuations*.

Definition 47 *(Adjusted valuation).* Let $n, \ell \in \mathbb{N}$. An (n, ℓ)-*adjusted valuation* is a map

$$\nu_{n,\ell} : \mathrm{pow}(\mathbb{R}^{n\times n}) \times \mathbb{R}^\ell \to \mathbb{R}$$

with an altered version of monotonicity: for any fixed $v \in \mathbb{R}^\ell$, $\nu_{n,\ell}(A, v) \geq \nu_{n,\ell}(B, v)$ for all $B \subset A \in \mathrm{pow}(\mathbb{R}^{n\times n})$. We say that an adjusted valuation $\nu_{n,\ell}$ is *L-stable* if for all $A, B \in \mathrm{pow}(\mathbb{R}^{n\times n})$ and $v, w \in \mathbb{R}^\ell$,

$$|\nu_{n,\ell}(A, v) - \nu_{n,\ell}(B, w)| \leq L \cdot \max\{d_{\mathrm{H}}^{(\mathbb{R}^{n\times n},\ell_\infty)}(A, B), \|v - w\|_\infty\}.$$

Combining the notion of a point descriptor and an adjusted valuation produces a more general and informative class of filtration functors than those obtained simply by applying valuations to the basepoint curvature set.

Definition 48 *(Local basepoint filtration functor).* Let $n, \ell \in \mathbb{N}$. A basepoint filtration functor Ψ is (n, ℓ)-*local* if there exists a point descriptor ρ with image in $\mathbb{R}^\ell$ and an adjusted valuation $\nu_{n,\ell} : \mathrm{pow}(\mathbb{R}^{n\times n}) \times \mathbb{R}^\ell \to \mathbb{R}$ such that for any $X \in \mathcal{M}$ and $x_0 \in X$:

$$\Psi_{X,x_0}(\sigma) = \nu_{n,\ell}(\mathrm{K}_{n-1}(x_0,\sigma),\rho(x_0)) \qquad \forall\,\sigma \subset X.$$

We now move on to proving the stability of these constructions.

5.3. Stability results for local basepoint filtrations

Recall the notion of “covariance” and “contravariance” from Remarks 10 and 42.

Theorem 49. *Let Ψ be an (n,ℓ)-local basepoint filtration functor for some $n,\ell \in \mathbb{N}$. Let $\nu_{n,\ell}$ be an L-stable adjusted valuation and ρ a K-stable point descriptor with image in $\mathbb{R}^\ell$ such that Ψ is generated by $\nu_{n,\ell}$ and ρ. Then, for all $k \geq 0$, and all finite metric spaces X, Y,*

$$\min_{R\in\mathcal{R}(X,Y)} \max_{(x_0,y_0)\in R} \mathcal{C}_{\Psi,k}(x_0,y_0) \leq 2L\cdot\max\{1,K\}\cdot d_{\mathrm{GH}}(X,Y).$$

Moreover, the theorem also holds if Ψ is contravariant.

Proof. The proof does not rely on the covariant or contravariant structure of Ψ, so we only need to show the inequality. Let $X, Y \in \mathcal{M}$ and $R_0 \in \mathcal{R}(X,Y)$ such that $\frac{1}{2}\mathrm{dis}(R_0) = d_{\mathrm{GH}}(X,Y)$. Now, notice that if π_X, π_Y are the canonical projections from R_0 to X and Y respectively, then (R_0, π_X, π_Y) is a triplet with a set and surjective maps to X and Y, respectively.

Now let $\sigma \subset R_0$ and $(x_0,y_0) \in R_0$. Notice that $(\sigma, \pi_X|_\sigma, \pi_Y|_\sigma)$ is a triplet of a set and two surjective maps to $\pi_X(\sigma)$ and $\pi_Y(\sigma)$. Let $\sigma_X = \pi_X(\sigma)$ and $\sigma_Y = \pi_Y(\sigma)$ be the metric spaces generated by the projections. We observe that

$$\begin{aligned}|\Psi_{X,x_0}(\pi_X(\sigma)) - \Psi_{Y,y_0}(\pi_Y(\sigma))| &= |\nu_{n,\ell}(\mathrm{K}_{n-1}(x_0,\sigma_X),\rho_X(x)) - \nu_{n,\ell}(\mathrm{K}_{n-1}(y_0,\sigma_Y),\rho_Y(y))| \\ &\leq L\cdot\max\{d_{\mathrm{H}}(\mathrm{K}_{n-1}(x_0,\sigma_X),\mathrm{K}_{n-1}(y_0,\sigma_Y)), \|\rho_X(x_0)-\rho_Y(y_0)\|_\infty\}.\end{aligned}$$

We would like to bound the two terms inside the max. From the definition of stability of point descriptors, we see that,

$$\|\rho_X(x_0)-\rho_Y(y_0)\|_\infty \leq K\cdot\mathrm{dis}(R_0) = 2K\cdot d_{\mathrm{GH}}(X,Y).$$

This bounds the second term. For the first term, let $\alpha = D_X^{(n)}(x_0,\ldots,x_{n-1}) \in \mathrm{K}_{n-1}(x_0,\sigma_X)$. There are $y_1,\ldots,y_{n-1}$ such that $(x_i,y_i) \in \sigma$ for all $i \in \{1,\ldots,n-1\}$. We also had $(x_0,y_0) \in R$ by assumption. Set $\beta = D_Y^{(n)}(y_0,y_1,\ldots,y_{n-1})$. Then we have

$$\|\alpha-\beta\|_\infty = \max_{0\leq i,j\leq n}\{|d_X(x_i,x_j) - d_Y(y_i,y_j)|\} \leq \mathrm{dis}(R_0) = 2\cdot d_{\mathrm{GH}}(X,Y).$$

Putting these observations together, we have

$$\begin{aligned}|\Psi_{X,x_0}(\pi_X(\sigma)) - \Psi_{Y,y_0}(\pi_Y(\sigma))| &\leq L\cdot\max\{2\cdot d_{\mathrm{GH}}(X,Y), 2K\cdot d_{\mathrm{GH}}(X,Y)\} \\ &= 2L\cdot\max\{1,K\}\cdot d_{\mathrm{GH}}(X,Y).\end{aligned}$$

Recalling the definition of $d_{\mathcal{F}}$ from Equation (1), it follows that for all $(x_0,y_0) \in R_0$,

$$d_{\mathcal{F}}((X,\Psi_{X,x_0}),(Y,\Psi_{Y,y_0})) \leq 2L\cdot\max\{1,K\}\cdot d_{\mathrm{GH}}(X,Y).$$

From this, we conclude

$$\begin{aligned}\min_{R\in\mathcal{R}(X,Y)}\max_{(x_0,y_0)\in R} d_{\mathcal{F}}((X,\Psi_{X,x_0}),(Y,\Psi_{Y,y_0})) &\leq \max_{(x_0,y_0)\in R_0} d_{\mathcal{F}}((X,\Psi_{X,x_0}),(Y,\Psi_{Y,y_0}))\\ &\leq 2L\cdot\max\{1,K\}\cdot d_{\mathrm{GH}}(X,Y).\end{aligned}$$

Finally, from Theorem 6, we conclude that for any $k\geq 0$,

$$\min_{R\in\mathcal{R}(X,Y)}\max_{(x_0,y_0)\in R}\mathcal{C}_{\Psi,k}(x_0,y_0)\leq 2L\cdot\max\{1,K\}\cdot d_{\mathrm{GH}}(X,Y).\quad\square$$

This result is the most general stability theorem we prove in this work. It is a generalization of Theorem 26, in light of the fact that all previously considered filtration functors can be viewed as basepoint filtration functors, as seen in Example 43.

We know that local filtration functors are well behaved when generated by stable adjusted valuations and point descriptors. Now we must consider a different type of stability. It should be true that changing the basepoint transforms the diagrams in a continuous way. This would align with the idea that the filtration depends on the perspective of the point; if the perspective changes by a small distance, the induced diagrams should incur small changes.

Proposition 50. *Let $X\in\mathcal{M}$ and $x,x'\in X$. Let Ψ be a local basepoint filtration functor. Let $\nu_{n,\ell}$ be an L-stable adjusted valuation and ρ a K-stable point descriptor with image in $\mathbb{R}^{\ell}$. Let us assume that $\nu_{n,\ell}$ and ρ generate the basepoint filtration functor Ψ. Then, for all $k\in\mathbb{N}$,*

$$d_{\mathrm{B}}(\mathrm{dgm}_k^{\Psi_{X,x}}(X),\mathrm{dgm}_k^{\Psi_{X,x'}}(X))\leq L\cdot\max\{1,K\}\cdot d_X(x,x').$$

Proof. Recalling the definition of $d_{\mathcal{F}}$ and Theorem 6, it is sufficient to show that there is a triplet (Z,π_X,π'_X) of a set Z and two surjective maps $\pi_X,\pi'_X:Z\to X$ such that

$$\max_{\sigma\subset Z}|\Psi_{X,x}(\pi_X(\sigma))-\Psi_{X,x'}(\pi'_X(\sigma))|\leq L\cdot\max\{1,K\}\cdot d_X(x,x').$$

Let $X\in\mathcal{M}$ and $x,x'\in X$. Then $(X,\mathrm{id}_X,\mathrm{id}_X)$ is a triplet of a set and two surjections to the set X. Let $\sigma\subset X$. Then,

$$\begin{aligned}|\Psi_{X,x}(\mathrm{id}_X(\sigma))-\Psi_{X,x'}(\mathrm{id}_X(\sigma))| &= |\Psi_{X,x}(\sigma)-\Psi_{X,x'}(\sigma)|\\ &= |\nu_{n,l}(\mathrm{K}_{n-1}(x,\sigma),\rho_X(x))-\nu_{n,l}(\mathrm{K}_{n-1}(x,\sigma),\rho_X(x'))|\\ &\leq L\cdot\max\{d_{\mathrm{H}}(\mathrm{K}_{n-1}(x,\sigma),\mathrm{K}_{n-1}(x',\sigma)),\|\rho_X(x)-\rho_X(x')\|_\infty\}.\end{aligned}$$

We need to bound the terms inside the max. Fix a correspondence $R=\{(x,x)\}_{x\in X}\cup\{(x,x')\}\in\mathcal{R}(X,X)$. It follows that

$$\mathrm{dis}(R)=\max_{z\in X}|d_X(z,x)-d_X(z,x')|\leq d_X(x,x'),$$

and so

$$\|\rho_X(x)-\rho_X(x')\|_\infty\leq K\cdot d_X(x,x').$$

Now let $\alpha\in\mathrm{K}_{n-1}(x,\sigma)$. There exist $x_1,\ldots,x_{n-1}\in\sigma$ such that $\alpha=D_X^{(n)}(x,x_1,\ldots,x_{n-1})$. It follows that $\beta=D_X^{(n)}(x',x_1,\ldots,x_{n-1})$ is an element of $\mathrm{K}_{n-1}(x',\sigma)$, and so

$$\|\alpha-\beta\|_\infty=\max_{1\leq i\leq n-1}|d_X(x,x_i)-d_X(x',x_i)|\leq d_X(x,x').$$

Similarly, we can prove that if we choose $\beta \in \mathrm{K}_{n-1}(x', \sigma)$, there is an element $\alpha \in \mathrm{K}_{n-1}(x, \sigma)$ such that $\|\alpha - \beta\|_\infty \leq d_X(x, x')$. This implies that $d_{\mathrm{H}}(\mathrm{K}_{n-1}(x, \sigma), \mathrm{K}_{n-1}(x', \sigma)) \leq d_X(x, x')$. Then,

$$|\Psi_{X,x}(\sigma) - \Psi_{X,x'}(\sigma)| \leq L \cdot \max\{d_X(x, x'), K \cdot d_X(x, x')\} = L \cdot \max\{1, K\} \cdot d_X(x, x'),$$

where the inequality comes from the K-stability of ρ. This concludes the proof. □

Having established the theoretical framework for basepoint filtration functors, we now proceed to a concrete example of such a functor.

5.4. Eccentricity basepoint filtration functor

Given a compact metric space (X, d_X), we recall the *eccentricity function* (see [25]) $\mathrm{ecc}_X : X \to \mathbb{R}_+$ defined by $x \mapsto \max_{x' \in X} d_X(x, x')$.

Example 51 *(Eccentricity basepoint filtration).* An interesting example of a *contravariant* basepoint filtration functor is the map Ψ^{ecc} defined so that for each $X \in \mathcal{M}$,

$$\Psi^{\mathrm{ecc}}_X = \{\Psi^{\mathrm{ecc}}_{X,x_0} : \mathrm{pow}(X) \to \mathbb{R}\}_{x_0 \in X},$$

where for $x_0 \in X$ and $\sigma \subset X$,

$$\Psi^{\mathrm{ecc}}_{X,x_0}(\sigma) := \max\left\{\mathrm{diam}(\iota_X(\sigma)), \frac{1}{2}\left(\mathrm{ecc}_X(x_0) - \min_{x' \in \sigma} d_X(x_0, x')\right)\right\}.$$

Remark 52. This definition is motivated by a construction that becomes intuitive when understood in the context of manifolds. For the purposes of this remark, consider the case of a surface X in Euclidean space. Let $x \in X$, and let $x' \in X$ be such that $d_X(x, x')$ achieves $\mathrm{ecc}_X(x)$. Let σ denote a closed ball containing x'. The filtration value of σ is then roughly the diameter of σ. Then as we “slide” σ towards x, say along a geodesic, the filtration value of σ approaches $\mathrm{ecc}_X(x)$. The preceding definition is the discrete analog of this idea.

The constant $\frac{1}{2}$ controls the dominance of the diameter term. Using 0 instead would recover the Vietoris-Rips filtration. In general, the smaller the constant, the greater the portion of the exterior (relative to the basepoint) of the space that is filtered similarly to the Vietoris-Rips filtration. This is illustrated with the experiment on the 3D scan of a cat in Section 6.2.

Lemma 53. *Ψ^{ecc} is a well-defined, contravariant basepoint filtration functor.*

Proof. Let $X \in \mathcal{M}$ and $x_0 \in X$. First we show that $\Psi^{\mathrm{ecc}}_{X,x_0}$ is a filtration on X. By definition, we already have that $\Psi^{\mathrm{ecc}}_{X,x_0}$ is a map from $\mathrm{pow}(X)$ to $\mathbb{R}_+$, so we only need to show that it satisfies the monotonicity condition. Let $\tau \subset \sigma \subset X$. Then we have $\mathrm{diam}(\iota_X(\tau)) \leq \mathrm{diam}(\iota_X(\sigma))$, and $\min\limits_{x' \in \sigma} d_X(x_0, x') \leq \min\limits_{x' \in \tau} d_X(x_0, x')$ since $\tau \subset \sigma$. This implies that

$$\frac{1}{2}\left(\mathrm{ecc}_X(x_0) - \min_{x' \in \sigma} d_X(x_0, x')\right) \geq \frac{1}{2}\left(\mathrm{ecc}_X(x_0) - \min_{x' \in \tau} d_X(x_0, x')\right).$$

Putting these two together gives that $\Psi^{\mathrm{ecc}}_{X,x_0}(\tau) \leq \Psi^{\mathrm{ecc}}_{X,x_0}(\sigma)$, so the monotonicity condition holds, and $\Psi^{\mathrm{ecc}}_{X,x_0}$ is a filtration on X.

Next we check contravariance. Let $(Y, d_Y) \in \mathcal{M}$, and let $h : X \to Y$ be an isometric embedding. Let $x \in X$, $\sigma \subset X$. Note that $\mathrm{diam}(\iota_X(\sigma)) = \mathrm{diam}(\iota_Y(h(\sigma)))$, and $\min_{x' \in \sigma} d_X(x, x') = \min_{y \in h(\sigma)} d_Y(h(x), y)$. However, $\mathrm{ecc}_X(x) = \mathrm{ecc}_{\iota(h(X))}(h(x)) \leq \mathrm{ecc}_Y(h(x))$. Thus $\Psi^{\mathrm{ecc}}_{X,x}(\sigma) \leq \Psi^{\mathrm{ecc}}_{Y,h(x)}(h(\sigma))$. □

Thus we see that Ψ^{ecc} is a (contravariant) basepoint filtration functor. For this functor, $\mathrm{ecc}_X(x)$ is the point descriptor. Then $\forall A \in \mathrm{K}_n(x_0, X)$, the adjusted valuation is given by

$$\nu_{n+1,1}(A \times v) := \max\left\{ \max(a_{ij})_{2\le i,j\le n+1}, \frac{1}{2}\left(v - \min(a_{1j})_{2\le j\le n+1}\right) \right\}.$$

To prove stability of the eccentricity filtration, we just need that the adjusted valuation and ecc function are stable, which is an immediate consequence of the following:

Lemma 54 *([25]). Let $X, Y \in \mathcal{M}$ and let $R \in \mathcal{R}(X, Y)$. Then,*

1. $|\mathrm{diam}(X) - \mathrm{diam}(Y)| \le \mathrm{dis}(R)$,
2. *For all $(x, y) \in R$, $|\mathrm{ecc}_X(x) - \mathrm{ecc}_Y(y)| \le \mathrm{dis}(R)$.*

In Section 6.2 we provide a computational example illustrating the use of the eccentricity functor.

Remark 55. The structure of the eccentricity functor is similar to that of a filtration based on mean curvature that appeared in the master's thesis [33]. The approach taken in [33] was to look at the cluster tree of connected components instead of persistence diagrams obtained after applying the homology functor. A similar approach of tracking the connected components of the eccentricity functor may also be taken, although we do not pursue this direction in the current work. The resulting object will be a rooted tree with leaf nodes appearing at possibly different filtration values. This method of producing a family of rooted trees can also be shown to be stable using the techniques described above.

6. Computational examples

Throughout this section, we write FPS to mean the *farthest point sampling* procedure [34].[2] Persistence computations were carried out using Javaplex [30] and Ripser [35]. Our code for the experiments below is written in Matlab and can be found on https://github.com/NateClause/Ultrametricity_Filtration and https://github.com/NateClause/Basepoint-Filtration. We will use both persistence diagrams and barcodes to visualize the final results of running our code on data.

6.1. The Φ^{ult} filtration functor

We start by describing experiments carried out using the Φ^{ult} filtration functor. In this filtration, all 1-simplices appear at time 0, and so the 0-dimensional barcode is not informative. In our experiments, higher dimensional ultrametricity barcodes for datasets arising in phylogenetics appear to be more informative than those produced by the Vietoris-Rips filtration.

6.1.1. Implementation details

We produced code for computing the ultrametricity filtration and for producing the corresponding barcodes using Javaplex. The software can be found at https://github.com/NateClause/Ultrametricity_Filtration. Two programs were developed. The first one computes the barcodes of a finite metric space given the distance matrix representation. As the implementation is not optimized, we recommend starting with a small dataset (∼50 points). The second function displays a window in which the user chooses up to thirty points in the plane. After selecting the points, the algorithm computes the ultrametricity persistence

[2] This is sometimes referred to as *sequential max-min sampling* as well.

diagram of the point cloud endowed with the Euclidean subspace metric. This serves the purpose of giving a simple platform for experimenting with this filtration.

The following question plays an important role on the computational limitations of our algorithms: Is it possible to find a transformation of a distance matrix so that the Vietoris-Rips barcodes of the transformed matrix agree with the ultrametricity barcodes of the original matrix? If the answer is positive, then we can use optimized programs such as Ripser on the transformed matrix for efficient computation of the ultrametricity barcodes. Unfortunately, while we provide a positive answer to the question in the setting of Ψ^{ecc} (see Section 6.2), it is still open in the setting of Φ^{ult}. Nonetheless, our software still enables experimentation on different datasets using the ultrametricity filtration. We explore some examples below.

6.1.2. Φ^{ult} on phylogenetics datasets

The construction of Φ^{ult} naturally suggests its use on datasets arising in phylogenetics. In the next example, we used a dataset of mitochondrial DNA sequences for different primates curated through the NCBI GenBank genetic sequence database [36–40]. This dataset consisted of genetic subsequences—each comprising ~ 350 bases—coming from 12 different primates. The data was preprocessed using the Jukes-Cantor metric for comparing genetic sequences. This preprocessed data comprised a 12×12 distance matrix which we used for subsequent computations. We denote this matrix as D_P in the rest of this discussion.

As a first look at the data, we computed the single linkage dendrogram from D_P. This is visualized in Fig. 4. The four genera of the Hominidae family are easily distinguishable. Next we plotted the persistence barcodes of D_P using both the Vietoris-Rips filtration and the ultrametricity filtration. The ultrametricity filtration yielded a more complex barcode in dimension 1 than the VR filtration, and also yielded numerous bars in dimension 2 whereas the VR filtration produced none. To interpret the bars, we studied representative generators as returned by Javaplex. Because these representatives are not necessarily unique, we restricted our attention to intervals of the form $[b, d)$ where we were able to: (1) directly identify unique 2-simplices that were born at b, d respectively, and (2) verify that the vertices of these 2-simplices were involved in the representative returned for $[b, d)$. The direct verifications were performed using D_P.

The 2-cycle for one of these intervals forms a triangular bipyramid, as illustrated in Fig. 4. This 2-cycle persists on the interval $[0.029, 0.033)$, and is filled in when the 2-simplex comprising the European human, Western chimpanzee, and Puti orangutan is filled in. Notably, these three species come from three distinct genera of the Hominidiae family.

As an extra validation step for the ultrametricity filtration, we present results from performing the same analysis on a bird flu dataset [41] in Fig. 5. In this case, the Vietoris-Rips filtration yields trivial barcodes in dimensions greater than 0, and thus is no more informative than the single linkage dendrogram. However, the ultrametricity filtration continues to produce barcodes with complex structure in both dimensions 1 and 2.

These examples suggest that when comparing groups of phylogenetic datasets, the ultrametricity filtration provides stronger discrimination than the Vietoris-Rips filtration through the use of the bottleneck distance to compare barcodes in dimensions 1 and 2. We performed a meta-analysis on an HIV dataset to validate this claim. Following [42], we took a dataset of 1175 samples of the HIV-1 env gene released by the Los Alamos National Laboratory [43]. Each sample comprises a sequence of ~ 3400 nucleotides. This particular gene encodes the enveloping protein used by the virus to bind to a host cell. Next we performed a bootstrapping procedure where we took batches of 25 gene sequences at a time, computed their pairwise distance using the Jukes-Cantor metric (using Matlab's `seqpdist` function), and computed their persistence diagrams in dimensions 0-2 using both the Vietoris-Rips and ultrametricity filtrations. This procedure was repeated for 1000 iterations. Boxplots of the number of bars and the length of the longest bar in dimensions 1 and 2 for these 1000 iterations are provided in Fig. 6. These results support the hypothesis that when comparing groups of gene sequences, e.g. samples of HIV strains grouped by country or year, the ultrametricity filtration is more informative than the Vietoris-Rips filtration.

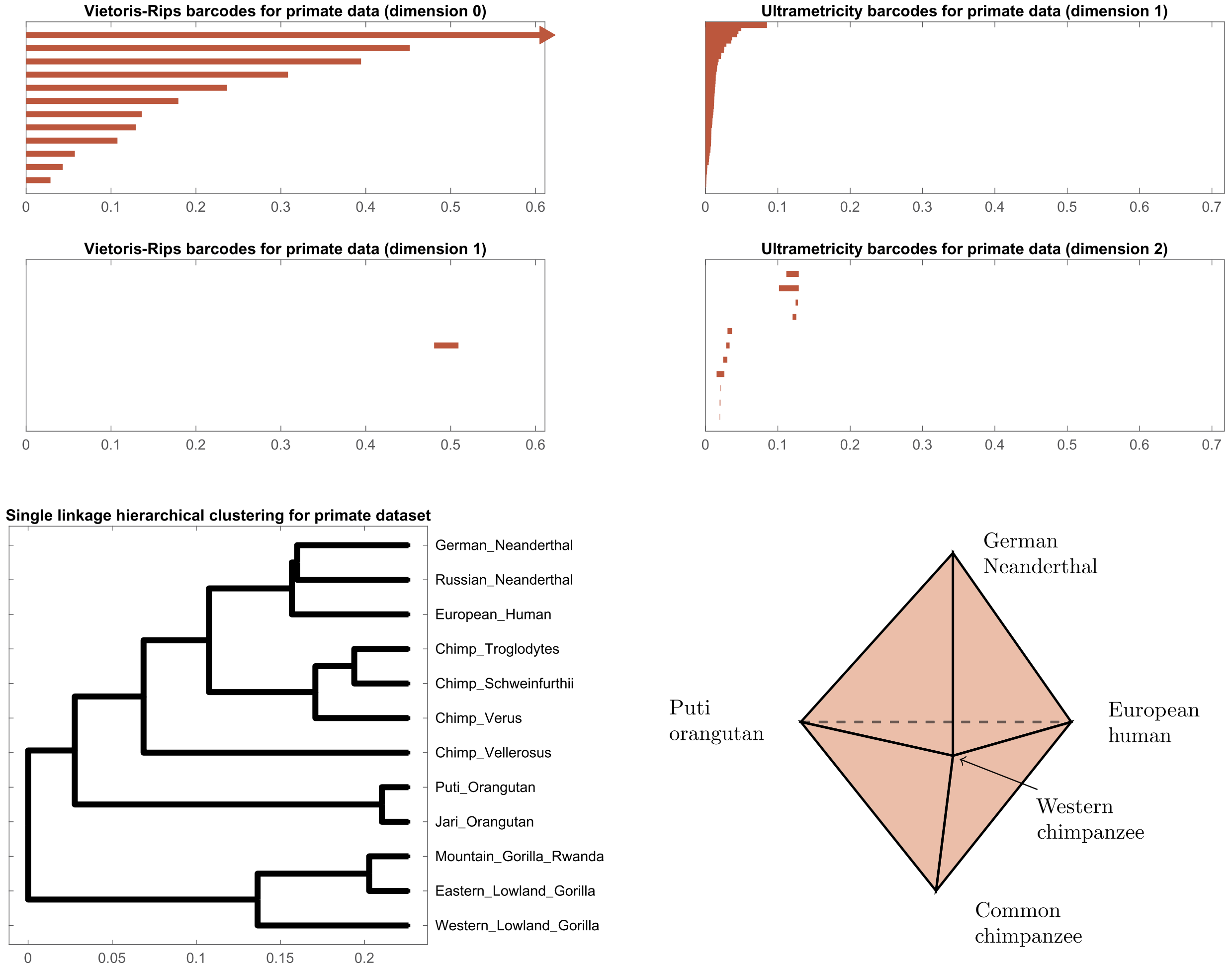

Fig. 4. Results from the primate dataset. **Top row:** The Vietoris-Rips barcode is less informative than the ultrametricity barcode in dimensions 1 and higher. The additional bars in the ultrametricity barcode suggest that ultrametricity-based lower bounds would be more effective than VR-based lower bounds for discriminating between such datasets. **Bottom left:** Single linkage dendrogram obtained from the 12 × 12 distance matrix obtained using the Jukes-Cantor metric for gene sequence comparison. The entries "Puti_Orangutan" and "Jari_Orangutan" do not refer to distinct species, but are simply the names of two Sumatran orangutans. **Bottom right:** Triangular bipyramid configuration of a set of generators for a nontrivial 2-dimensional homology class comprising the interval $[0.029, 0.033)$. The common chimpanzee and the western chimpanzee correspond to Chimp_Troglodytes and Chimp_Verus, respectively. This 2-cycle becomes a boundary when the European human-Western Chimpanzee-Puti Orangutan simplex is filled in. Observe from the dendrogram that these species come from three distinct genera.

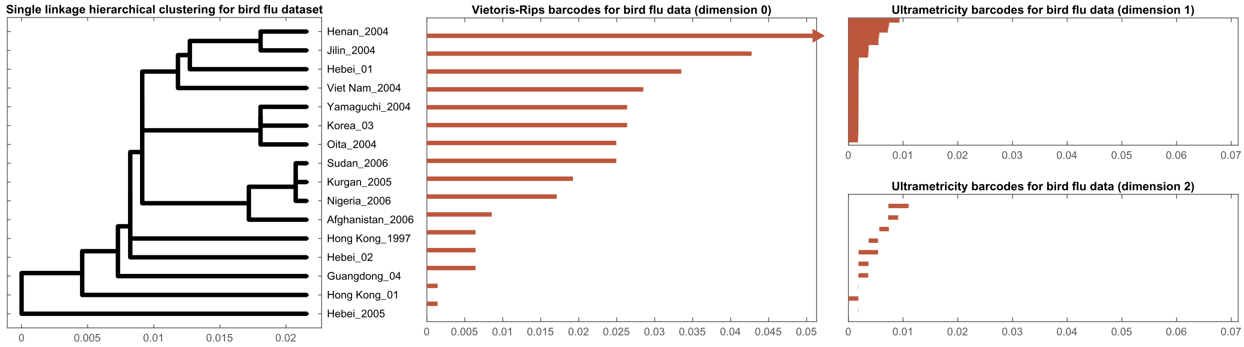

Fig. 5. Results from carrying out our analysis pipeline on a bird flu dataset. Dendrogram labels correspond to locations and times when each virus strain was sampled. The Vietoris-Rips filtration has no persistence in dimensions greater than 0, and is thus no more informative than the single linkage dendrogram. In comparison, the ultrametricity filtration produces complex barcodes in dimensions 1 and 2.

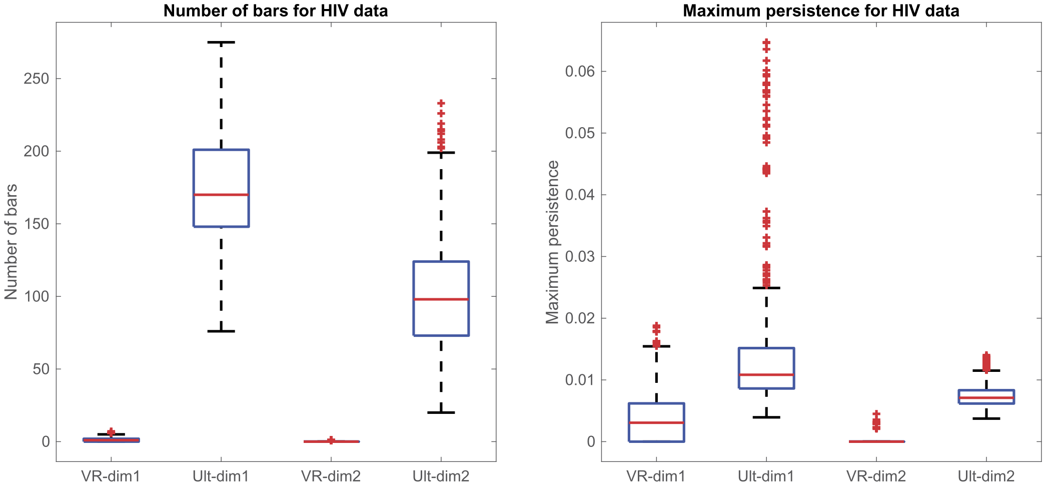

Fig. 6. Boxplots corresponding to the HIV dataset example. Across 1000 iterations, we find that the ultrametricity filtration produces more and longer bars in dimensions 1 and 2 than the Vietoris-Rips filtration.

Connections to literature on genetic recombination Genetic recombination refers to the event that genomes from different organisms combine to form a new genome with shared traits, and is a fundamental process in evolution. This is a multiscale phenomenon that ranges from the individual level (recombination between two organisms of the same species) to the population level (recombination between distantly related subspecies). Such events are often modeled by adding links between nodes in a phylogenetic tree. These links introduce cycles, and the multiscale nature naturally suggests the use of persistent homology for studying recombination. Such uses were demonstrated in [42], and subsequent work [44–46] developed the theoretical foundations of this approach from multiple perspectives.

While theoretical results supporting our empirical observations regarding ultrametricity barcodes for phylogenetic datasets are beyond the scope of the current work, we outline a plausibility argument for studying genetic recombination of populations with distantly related ancestors through ultrametricity. A nontrivial homology class in dimension 2, such as that generated by the five-point triangular bipyramid illustrated in Fig. 4, is obtained when the set of five samples $\{r_1, r_2, r_3, a_1, a_2\}$ are such that $\{\{a_1, r_i, r_j\} : 1 \leq i \neq j \leq 3\}$ and $\{\{a_2, r_i, r_j\} : 1 \leq i \neq j \leq 3\}$ form triangles with low ultrametricity, and $\{r_1, r_2, r_3\}$ forms a triangle with higher ultrametricity. This occurs when a_1, a_2 are ancestors of $\{r_1, r_2, r_3\}$ from a distant generation that are themselves not closely related, and $\{r_1, r_2, r_3\}$ are from nearby generations and are closely related by recombination. It may be the case that $\{r_1, r_2, r_3\}$ do not always form a triangle with high ultrametricity, but we do expect to observe triples with high ultrametricity given sufficient samples.

6.2. The Ψ^{ecc} basepoint functor

We now consider the Ψ^{ecc} basepoint functor. For a fixed filtration with basepoint x_0, the first simplices to enter the complex occur in the part of the space furthest away from the selected basepoint x_0. As the filtration value increases, simplices located closer to x_0 enter sequentially. To give more concrete intuition, we developed an interactive program for studying the eccentricity filtration on point cloud data and used it to study a publicly available shape dataset.

6.2.1. Implementation details

This interactive code can be found at https://github.com/NateClause/Basepoint-Filtration. The code utilizes both Javaplex [30] and Ripser [35] for the persistent homology calculations. Note that all higher dimensional simplices in an eccentricity filtration are determined by the 1-skeleton of the corresponding simplicial complex. This allows us to employ the following computational trick for computing persistence

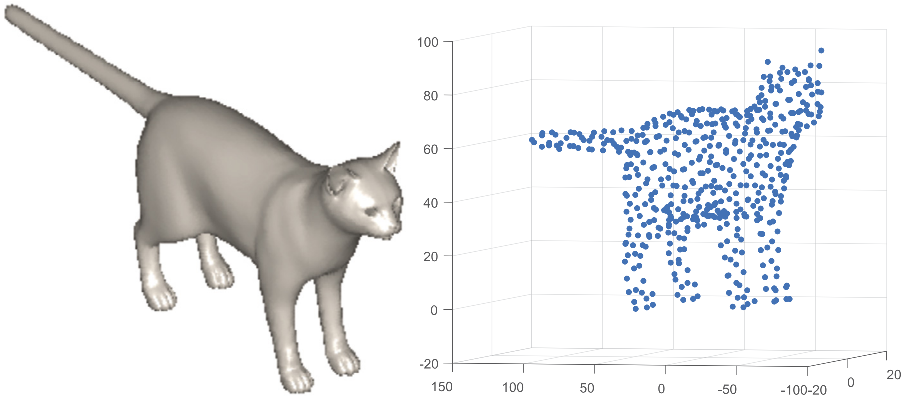

Fig. 7. Original scan of cat and point cloud approximation.

via Ripser. Instead of passing the standard distance matrix to Ripser, we precompute the filtration values for all pairs of points in the space, and pass this matrix of filtration values to Ripser as a distance matrix. When Ripser computes the Vietoris-Rips persistent homology of the space, the 1-dimensional simplices arrive at the time which is the “distance” between the two vertices as per the distance matrix. This is not the actual distance, but instead the filtration value of the simplex. Note that Ripser adds all 0-simplices at time 0, but since we are only using Ripser for computing persistent homology in dimensions 1 and higher, this does not affect the computations. Utilizing Ripser in this way is advantageous because Ripser is more computationally efficient than Javaplex and enables us to work with larger datasets. Since Ripser adds in all vertices at time 0, and this is not how the eccentricity filtration works, Javaplex is used to compute 0-dimensional persistent homology. We also make use of an option in Javaplex to return a representative cycle of a homology class. This gives useful information, as we note in the example below.

To run the code, the user inputs the point cloud for a finite metric space. The program then plots the dataset in 3D. Note that when working with higher-dimensional datasets, the user must preprocess the data to embed it into three dimensions. This can be done via standard techniques such as principal component analysis. The user then clicks on a point within the space. Upon doing so, the following actions occur: this point will be selected as the basepoint for the corresponding basepoint eccentricity filtration, the persistent homology of the space using this filtration will be computed, and the persistence barcodes will be plotted on the screen. For visualization help, the plot of the metric space never disappears, and after a basepoint is selected, it is highlighted and all the other points are colored by how close they are to the basepoint. This helps in understanding the behavior of the eccentricity term in the filtration. After observing the persistence barcodes, the user may then click on a new point on the original plot to select a new basepoint and the process will then be repeated. The code is modularized in a way to enable easy, quick changes to the filtration while maintaining other functionalities. If one wanted to computationally test a different basepoint filtration functor, one would only need to alter the portion of the code where the eccentricity functor is currently defined.

6.2.2. Ψ^{ecc} *on a 3D shape*

The data in the next example is from a high resolution scan of the surface of a cat from a public repository [47] which is modeled as a finite metric space with geodesic distances. The original scan contained over 27000 points in 3D, so in order to make it computationally practical we used the built-in FPS sampling tool in Javaplex to select 500 points to approximate the data. Fig. 7 shows the original image of the scan, as well as a plot of our 500 selected points.

The eccentricity filtration can give useful information about the “protrusions” of a space, via the 0-dimensional persistence intervals. This is demonstrated in Fig. 8.

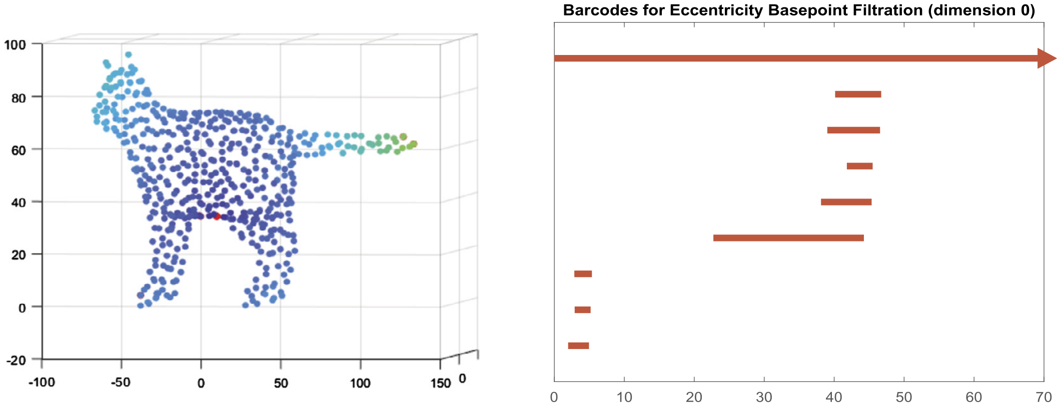

Fig. 8. The selected basepoint is on the belly of the cat. The longest bar corresponds to the tip of the tail, the four bars of similar length correspond to the legs, and the single medium-sized bar corresponds to the face.

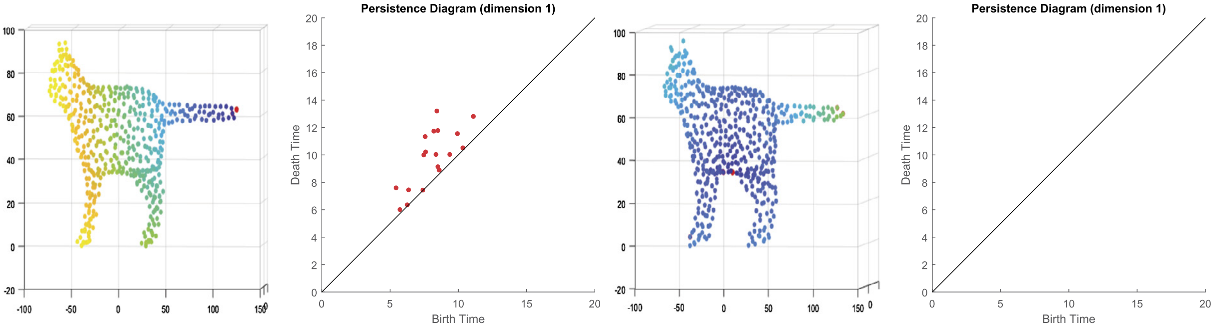

Fig. 9. Left: The red, highlighted basepoint selected for eccentricity filtration is close to the boundary of the metric space (on the tail). **Mid-right:** The red, highlighted basepoint selected for eccentricity filtration is central in the metric space (on the belly).

A central point on the figure occurs on the belly of the cat. When selecting this (or any other relatively central) point as the basepoint for the eccentricity filtration, one can see all of the "protrusions" of the space as 0-dimensional persistence intervals. The infinitely persisting class always starts at a point which realizes the eccentricity of the basepoint; in this case the tip of the tail. We use the homology class representative feature of Javaplex to see that one short interval in the bottom left of the diagram represents a class from the tail, which makes sense as at the edge of the tail the diameter term in the filtration is dominant, and thus the filtration behaves similarly to the Vietoris-Rips filtration in this region. Then the next two shorter intervals correspond to classes starting at the tips of the ears, until they merge with the larger class right above them, which corresponds to a class which originates from a point on the face of the cat. Lastly, the four remaining intervals of similar persistence correspond to the four classes starting at the end of each leg of the cat.

Next, we provide two plots of 1-dimensional persistence diagrams resulting from different selections of basepoint (Fig. 9).

In the left of Fig. 9, the tip of the tail is selected as a basepoint. Since this is at one edge of the space, a large portion of the space opposite this basepoint will behave similarly to rips. The longest persistent interval corresponds to the loop going around the main body of the cat. There are also lots of short persistent intervals, which mostly correspond to noise as 500 points cannot completely represent a space originating from 27000 points. In the right of Fig. 9, we see the 1-dimensional barcodes when a central point is chosen as the basepoint. As noted, the largest natural loop from this space is the one going around the main body. However, the basepoint lies on this loop, so it will not be realized as a loop in the eccentricity filtration. While choosing a central basepoint is useful for considering 0-dimensional persistence, such a choice of basepoint

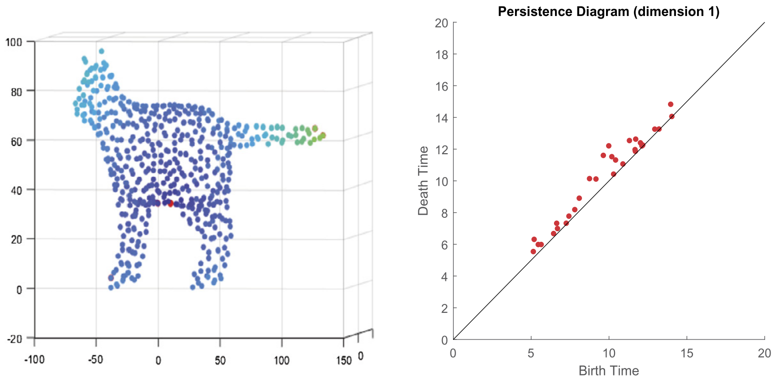

Fig. 10. Diagram with eccentricity constant $\frac{1}{4}$.

will more often than not generate significantly fewer and shorter 1-dimensional persistent intervals than with a choice of basepoint on the extremities of the space.

Next, we note that the eccentricity filtration can be adjusted by changing the constant of $\frac{1}{2}$ in front of the eccentricity term in Example 51. In Fig. 10, we provide the 1-dimensional barcode with the same central basepoint, but a constant of $\frac{1}{4}$ on the eccentricity term instead.

Note that there are many more, and longer, persistence intervals than the same filtration with a constant of $\frac{1}{2}$ on the eccentricity term. What this constant regulates is effectively how far away from the basepoint to filter the space as if it were the Vietoris-Rips filtration. The smaller the constant, the greater the portion on the exterior (relative to the basepoint) of the space is treated similarly to Vietoris-Rips. In fact, if this constant is 0 we see that the eccentricity filtration is equivalent to the Vietoris-Rips filtration.

7. Discussion

In this paper, we defined a framework for generating new filtrations (and including existing filtrations) on finite metric spaces using Gromov's curvature sets and provided easy-to-use practical implementations. Curvature sets are used because they comprise a full invariant of a metric space, and our computational examples substantiate the theoretical expectation that large families of filtrations capture multifaceted information about a dataset.

As discussed in Section 4.2, new filtrations can be easily generated using Lipschitz functions on $n \times n$ matrices. This leads to constructions such as the eccentricity and ultrametricity filtration. We have created an interactive platform for exploratory data analysis using the eccentricity filtration and provided a thorough example application on a 3D shape. We have also presented examples of phylogenetics datasets showing that the ultrametricity filtration outperforms the Vietoris-Rips filtration, which is the standard workhorse of persistent homology. Looking forward, we remark that a related quantity called *hyperbolicity* [29] can be used to produce a 4-local filtration functor, and applications of this functor to datasets remain open.

Acknowledgements

The material in this report emerged from a research project carried out while the authors were in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the Summer@ICERM 2017 REU [48]. The effort was supported by the National Science Foundation under Grants Nos. DMS-1439786, IIS-1422400, and CCF-1526513. Further development was carried out during a visit of J.Á. Sánchez to The Ohio State University in November 2017. The visit was supported by Grant CONACyT CB No. 252996 and The Ohio State University.

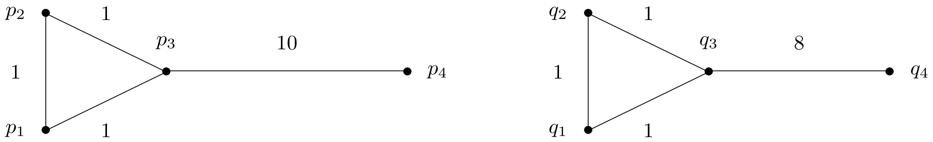

Fig. A.11. Two four-point metric spaces.

Appendix A. Supplementary information

A.1. The role of minimizing correspondences for the eccentricity filtration

Example 56. The definition of stability for basepoint filtration functors is based on a minimizing correspondence. In this example, we demonstrate why such a restriction must be made.

Consider the two four-vertex graphs $P = \{p_1, p_2, p_3, p_4\}$ and $Q = \{q_1, q_2, q_3, q_4\}$ as seen in Fig. A.11. We now refer to P and Q as the finite metric spaces of these graphs equipped with the graph metric.

The natural correspondence between these two metric spaces is

$$R_1 = \{(p_1, q_1), (p_2, q_2), (p_3, q_3), (p_4, q_4)\}.$$

This correspondence also is one whose distortion generates the Gromov-Hausdorff distance between P and Q of 1. To demonstrate the importance of using a minimal correspondence in stability, we compute the 0-dimensional cost function induced by Ψ^{ecc} for pairs of basepoints in both this correspondence and a different one.

We start by computing the values of $\mathcal{C}_{\Psi^{\mathrm{ecc}},0}(p_i, q_i)$ for pairs of basepoints in R_1. With p_1 as the selected basepoint, the filtration value of p_4 will be 0, the filtration value of p_2 and p_3 will be 5, and the filtration value of p_1 will be $\frac{11}{2}$. The edge $[p_2, p_3]$ will have filtration value 5, the edges $[p_1, p_2]$ and $[p_1, p_3]$ will both have filtration value $\frac{11}{2}$, and the edge $[p_3, p_4]$ will have filtration value 10. The connected component generated by p_4 appears at time 0, and the connected component generated by p_2/p_3 appears at time 5. p_1 is connected via edges to p_2 and p_3 immediately after it appears at time $\frac{11}{2}$ and thus does not generate a connected component. Lastly, the connected component generated by p_2/p_3 disappears at time 10 when it is connected to p_4 by edge $[p_3, p_4]$. Thus, the 0-dimensional persistence intervals with p_1 as the selected basepoint will be $[0, \infty)$ and $[5, 10)$. Note p_1 and p_2 are interchangeable, so these will also be the 0-dimensional persistence intervals with p_2 as the selected basepoint.

With p_3 as the selected basepoint, the filtration value of p_4 will be 0, the filtration value of p_1 and p_2 will be $\frac{9}{2}$, and the filtration value of p_3 will be 5. The edge $[p_1, p_2]$ will have filtration value $\frac{9}{2}$, the edges $[p_1, p_3]$ and $[p_2, p_3]$ will have filtration value 5, and the edge $[p_3, p_4]$ will have filtration value 10. The connected component generated by p_4 again appears at time 0, and the connected component generated by p_1, p_2 appears at time $\frac{9}{2}$. p_3 is connected via edges to p_1 and p_2 immediately after it appears at time 5 and thus does not generate a connected component. Lastly, the connected component generated by p_1/p_2 disappears at time 10 when it is connected to p_4 by edge $[p_3, p_4]$. Thus, the 0-dimensional persistence intervals with p_3 as the selected basepoint will be $[0, \infty)$ and $[\frac{9}{2}, 10)$.

Lastly, with p_4 as the selected basepoint, the filtration value of p_4 will be $\frac{11}{2}$, the filtration value of p_1 and p_2 will be 0, and the filtration value of p_3 will be $\frac{1}{2}$. The edges $[p_1, p_2]$, $[p_2, p_3]$ and $[p_1, p_3]$ will all have filtration value 1, and the edge $[p_3, p_4]$ will have filtration value 10. The connected components generated by p_1 and p_2 appear at time 0, and then are connected by $[p_1, p_2]$ at time 1. The connected component generated by p_3 appears at time $\frac{1}{2}$ and similarly is connected to become a single connected component with p_1 and p_2 at time 1. Lastly, the connected component generated by p_4 appears at time $\frac{11}{2}$ and is joined with the other component by edge $[p_3, p_4]$ at time 10. Thus, the 0-dimensional persistence intervals with p_4 as the selected basepoint will be $[0, \infty)$, $[0, 1)$, $[\frac{1}{2}, 1)$, and $[\frac{11}{2}, 10)$.

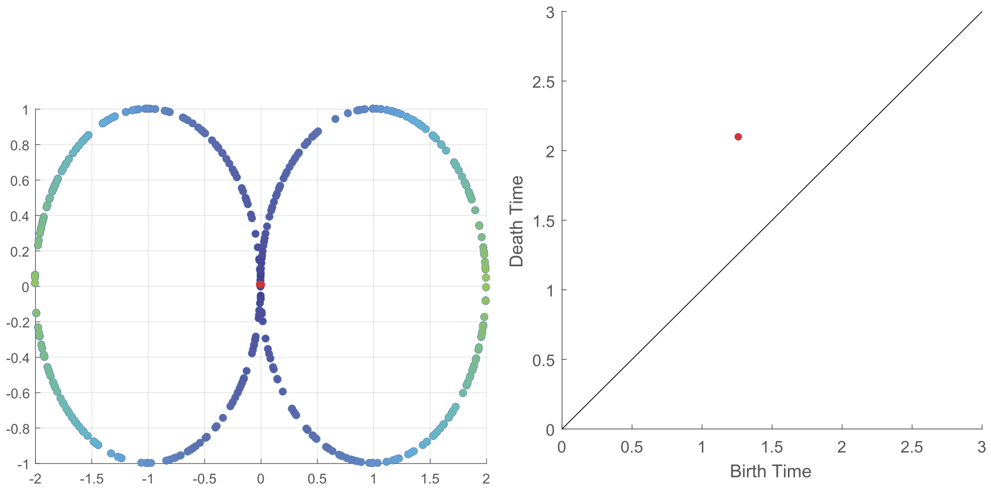

Fig. A.12. Fig. 8 persistence diagram with central basepoint.

Next, we need to compute the persistence intervals for Q with the various choices of basepoint. This computation follows the previous persistence interval computation very closely, so we simply state the persistence intervals for Q. With q_1 or q_2 as the selected basepoint, the 0-dimensional persistence intervals for Q will be $[0, \infty)$ and $[4, 8)$. With q_3 as the selected basepoint, the intervals will be $[0, \infty)$ and $[\frac{7}{2}, 8)$. Lastly, with q_4 as the selected basepoint, the intervals will be $[0, \infty)$, $[0, 1)$, $[\frac{1}{2}, 1)$, and $[\frac{9}{2}, 8)$.

We now compute the bottleneck distances for the pairs of basepoints in R_1. For the pair (p_1, q_1), the two $[0, \infty)$ intervals are matched, and the interval $[5, 10)$ for P is matched with the interval $[4, 8)$ for Q, yielding $\mathcal{C}_{\Psi^{\mathrm{ecc}},0}(p_1, q_1) = 1$. The matching is the same to compute $\mathcal{C}_{\Psi^{\mathrm{ecc}},0}(p_2, q_2) = 1$. For the pair (p_3, q_3), the two $[0, \infty)$ intervals are matched, and the interval $[\frac{9}{2}, 10)$ for P is matched with the interval $[\frac{7}{2}, 8)$ for Q, yielding $\mathcal{C}_{\Psi^{\mathrm{ecc}},0}(p_3, q_3) = 1$. Lastly, for the pair (p_4, q_4), the intervals $[0, \infty)$, $[0, 1)$, and $[\frac{1}{2}, 1)$ appear once for both P and Q and are thus matched with their duplicate. The interval $[\frac{11}{2}, 10)$ for P is then matched with the interval $[\frac{9}{2}, 8)$ for Q and this matching then yields $\mathcal{C}_{\Psi^{\mathrm{ecc}},0}(p_4, q_4) = 1$. Thus, for R_1 we get:

$$\max_{(x_0,y_0)\in R_1} \mathcal{C}_{\Psi^{\mathrm{ecc}},0}(x_0, y_0) = 1$$

Now consider another correspondence, $R_2 = \{(p_1, q_3), (p_2, q_2), (p_3, q_1), (p_4, q_4)\}$. For the pair (p_3, q_1), we have the two $[0, \infty)$ intervals matching with each other. This leaves us with the interval $[\frac{9}{2}, 10)$ for P and $[4, 8)$ for Q. These intervals are matched together, yielding $\mathcal{C}_{\Psi^{\mathrm{ecc}},0}(p_3, q_1) = \frac{3}{2}$. However, this means that

$$\max_{(x_0,y_0)\in R_2} \mathcal{C}_{\Psi^{\mathrm{ecc}},0}(x_0, y_0) \geq \frac{3}{2} > 1 = \max_{(x_0,y_0)\in R_1} \mathcal{C}_{\Psi^{\mathrm{ecc}},0}(x_0, y_0)$$

This shows that only considering a minimal correspondence is important in the definition of stability for basepoint filtration functors. For another example to consider, take a finite metric space with three or more points. Then even though the Gromov-Hausdorff distance between this space and itself is 0, for a correspondence that does not match points with themselves, it is possible that $\mathcal{C}_{\Psi^{\mathrm{ecc}},0}(x_0, y_0) > 0$, which would mean stability is not possible.

A.2. Eccentricity on a Fig. 8

An elementary example which gives a good basis for understanding for the eccentricity basepoint filtration is a Fig. 8. The finite metric space used for this code is a discrete Fig. 8 with 400 points equipped with the restriction of the geodesic distance. In Figs. A.12, A.13, and A.14, on the left we see the plot of the metric

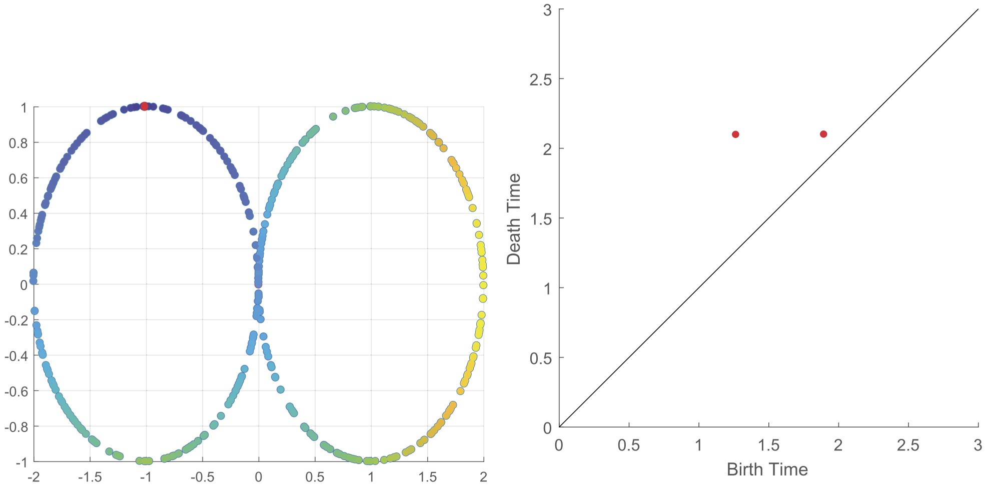

Fig. A.13. Persistence diagram with offset basepoint.

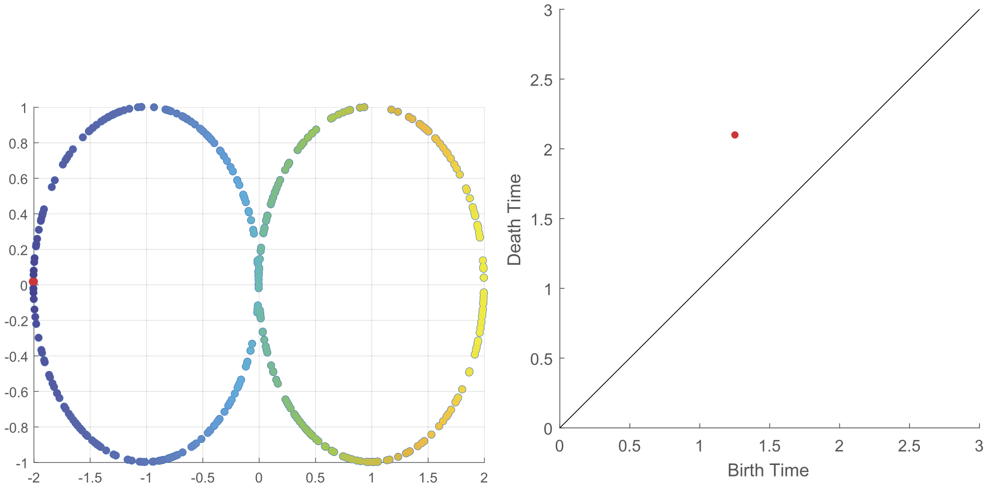

Fig. A.14. Fig. 8 persistence diagram with outer basepoint.

space, with the red point in the center indicating that this is the selected basepoint. The other points are then colored, with the darker blue points close to the basepoint, and the lighter yellow points being further away. On the right, we see the one-dimensional persistent barcodes of this space. It is not possible to tell visually, but there are two 1-dimensional persistent homology intervals; each corresponding to one of the loops from the circles.

This is what we would expect for the 1-dimensional persistent homology of a Fig. 8 through "standard" filtration methods such as Vietoris-Rips, since there are clearly two loops and both have equal size. However, as we move the basepoint around one of the circles, the persistent homology of the new filtration changes. In Fig. A.13, we see what happens when the selected basepoint (highlighted in red) is moved up along one of the circles.

We see in this image that one of the 1-dimensional persistence intervals is shrinking, while the other maintains the same length. The eccentricity filtration treats simplices far away from the basepoint like the Vietoris-Rips filtration, whereas simplices close to the basepoint are dominated by the eccentricity term. Thus the shrinking persistence interval corresponds to the loop on which the basepoint rests. The unchanging persistence interval then corresponds to the loop disjoint from the basepoint. As the basepoint continues further from the center, one of the persistence intervals disappears. This can be seen in Fig. A.14, where the basepoint is on the outer edge of the Fig. 8.

References

[1] P. Frosini, Measuring Shapes by Size Functions, Intelligent Robots and Computer Vision X: Algorithms and Techniques, vol. 1607, International Society for Optics and Photonics, 1992, pp. 122–134.

[2] V. Robins, Towards computing homology from finite approximations, in: Topology Proceedings, vol. 24, 1999, pp. 503–532.

[3] H. Edelsbrunner, J. Harer, Computational Topology: an Introduction, American Mathematical Soc., 2010.

[4] G. Carlsson, Topology and data, Bull. Am. Math. Soc. 46 (2) (2009) 255–308.

[5] H. Edelsbrunner, D. Morozov, Persistent homology: theory and practice, in: European Congress of Mathematics Kraków, 2–7 July, 2012, 2014, pp. 31–50.

[6] D. Burago, Y. Burago, S. Ivanov, A Course in Metric Geometry, vol. 33, American Mathematical Society, Providence, RI, 2001.

[7] F. Chazal, D. Cohen-Steiner, L.J. Guibas, F. Mémoli, S.Y. Oudot, Gromov-Hausdorff Stable Signatures for Shapes Using Persistence, Computer Graphics Forum, vol. 28, Wiley Online Library, 2009, pp. 1393–1403.

[8] F. Chazal, V. De Silva, S. Oudot, Persistence stability for geometric complexes, Geom. Dedic. 173 (1) (2014) 193–214.

[9] F. Schmiedl, Shape matching and mesh segmentation: mathematical analysis, algorithms and an application in automated manufacturing, PhD thesis, Technische Universität München, München, 2015.

[10] A. Efrat, A. Itai, M.J. Katz, Geometry helps in bottleneck matching and related problems, Algorithmica 31 (1) (2001) 1–28.

[11] P. Frosini, G. Jabłoński, Combining persistent homology and invariance groups for shape comparison, Discrete Comput. Geom. 55 (2) (2016) 373–409, https://doi.org/10.1007/s00454-016-9761-y.

[12] P. Frosini, Towards an observer-oriented theory of shape comparison: position paper, in: Proceedings of the Eurographics 2016 Workshop on 3D Object Retrieval, Eurographics Association, 2016, pp. 5–8.

[13] M.G. Bergomi, P. Frosini, D. Giorgi, N. Quercioli, Towards a topological–geometrical theory of group equivariant non-expansive operators for data analysis and machine learning, Nat. Mach. Intell. 1 (9) (2019) 423–433.

[14] K. Turner, S. Mukherjee, D.M. Boyer, Persistent homology transform for modeling shapes and surfaces, Inf. Inference 3 (4) (2014) 310–344.

[15] S. Oudot, E. Solomon, Barcode embeddings for metric graphs, arXiv preprint, arXiv:1712.03630.

[16] J. Curry, S. Mukherjee, K. Turner, How many directions determine a shape and other sufficiency results for two topological transforms, arXiv preprint, arXiv:1805.09782.

[17] R. Ghrist, R. Levanger, H. Mai, Persistent homology and Euler integral transforms, J. Appl. Comput. Topol. 2 (1–2) (2018) 55–60.

[18] R.L. Belton, B.T. Fasy, R. Mertz, S. Micka, D.L. Millman, D. Salinas, A. Schenfisch, J. Schupbach, L. Williams, Learning simplicial complexes from persistence diagrams, in: Canadian Conference on Computational Geometry, vol. 30, 2018.

[19] S. Oudot, E. Solomon, Inverse problems in topological persistence: a survey, in: Abel Symposia, 2019.

[20] A. Bittner, B.T. Fasy, M. Grudzien, S.G. Hajra, J. Huang, K. Pelatt, C. Thatcher, A. Tumurbaatar, C. Wenk, Comparing directed and weighted road maps, in: Research in Computational Topology, Springer, 2018, pp. 57–70.

[21] M. Gromov, Metric Structures for Riemannian and Non-Riemannian Spaces, Springer Science & Business Media, 2007.

[22] S. Chowdhury, F. Mémoli, Z.T. Smith, Improved error bounds for tree representations of metric spaces, in: Advances in Neural Information Processing Systems, 2016, pp. 2838–2846.

[23] T.K. Dey, D. Shi, Y. Wang, Comparing graphs via persistence distortion, in: 31st International Symposium on Computational Geometry (SoCG 2015), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

[24] J. Munkres, Elements of Algebraic Topology, CRC Press, 2018.

[25] F. Mémoli, Some properties of Gromov–Hausdorff distances, Discrete Comput. Geom. 48 (2) (2012) 416–440.

[26] F. Chazal, V. de Silva, M. Glisse, S. Oudot, The Structure and Stability of Persistence Modules, Springer, 2016.

[27] F. Mémoli, A distance between filtered spaces via tripods, arXiv preprint, arXiv:1704.03965.

[28] J.P. Kung, G.-C. Rota, C.H. Yan, Combinatorics: the Rota Way, Cambridge University Press, 2009.

[29] C. Semple, M.A. Steel, Phylogenetics, vol. 24, Oxford University Press on Demand, 2003.

[30] A. Tausz, M. Vejdemo-Johansson, H. Adams JavaPlex, A research software package for persistent (co)homology, in: H. Hong, C. Yap (Eds.), Proceedings of ICMS 2014, in: Lecture Notes in Computer Science, vol. 8592, 2014, pp. 129–136, software available at http://appliedtopology.github.io/javaplex/.

[31] D. Morozov, Dionysus 2, Software available at https://mrzv.org/software/dionysus2/.

[32] M. Adamaszek, H. Adams, The Vietoris–Rips complexes of a circle, Pac. J. Math. 290 (1) (2017) 1–40.

[33] Y. Yin, Shape classification via optimal transport and persistent homology, Master's thesis, The Ohio State University, 2019.

[34] T.F. Gonzalez, Clustering to minimize the maximum intercluster distance, Theor. Comput. Sci. 38 (1985) 293–306.

[35] U. Bauer, Ripser, 2016.

[36] I.V. Ovchinnikov, A. Götherström, G.P. Romanova, V.M. Kharitonov, K. Liden, W. Goodwin, Molecular analysis of Neanderthal DNA from the northern caucasus, Nature 404 (6777) (2000) 490–493.

[37] A. Sajantila, P. Lahermo, T. Anttinen, M. Lukka, P. Sistonen, M.-L. Savontaus, P. Aula, L. Beckman, L. Tranebjaerg, T. Gedde-Dahl, et al., Genes and languages in Europe: an analysis of mitochondrial lineages, Genome Res. 5 (1) (1995) 42–52.

[38] M. Krings, A. Stone, R.W. Schmitz, H. Krainitzki, M. Stoneking, S. Pääbo, Neandertal dna sequences and the origin of modern humans, Cell 90 (1) (1997) 19–30.

[39] M. Jensen-Seaman, K. Kidd, Mitochondrial DNA variation and biogeography of eastern gorillas, Mol. Ecol. 10 (9) (2001) 2241–2247.

[40] Mathworks, Building a phylogenetic tree for the hominidae species, https://www.mathworks.com/help/bioinfo/examples/building-a-phylogenetic-tree-for-the-hominidae-species.html. (Accessed 22 March 2020).

[41] Mathworks, Investigating the bird flu virus, https://www.mathworks.com/help/bioinfo/examples/investigating-the-bird-flu-virus.html. (Accessed 22 March 2020).

[42] J.M. Chan, G. Carlsson, R. Rabadan, Topology of viral evolution, Proc. Natl. Acad. Sci. USA 110 (46) (2013) 18566–18571.

[43] B. Foley, T. Leitner, C. Apetrei, B. Hahn, I. Mizrachi, J. Mullins, A. Rambaut, S. Wolinsky, B. Korber, HIV Sequence Database, Published by Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, NM, LA-UR-18-25673, 2018, available at http://www.hiv.lanl.gov/.

[44] L. Parida, F. Utro, D. Yorukoglu, A.P. Carrieri, D. Kuhn, S. Basu, Topological signatures for population admixture, in: International Conference on Research in Computational Molecular Biology, Springer, 2015, pp. 261–275.

[45] P.G. Cámara, A.J. Levine, R. Rabadan, Inference of ancestral recombination graphs through topological data analysis, PLoS Comput. Biol. 12 (8) (2016) e1005071.

[46] M. Lesnick, R. Rabadan, D.I. Rosenbloom, Quantifying genetic innovation: mathematical foundations for the topological study of reticulate evolution, SIAM J. Appl. Algebra Geom. 4 (1) (2020) 141–184.

[47] R.W. Sumner, J. Popović, Deformation Transfer for Triangle Meshes, ACM Transactions on Graphics (TOG), vol. 23, ACM, 2004, pp. 399–405.

[48] Summer@ICERM, Topological data analysis, 2017.