Eurographics Symposium on Geometry Processing 2015
Mirela Ben-Chen and Ligang Liu
(Guest Editors)

Volume 34 (2015), Number 5

This is a modified version of the paper published in Computer Graphics Forum

A one-dimensional homologically persistent skeleton of an unstructured point cloud in any metric space

V. Kurlin[1]

[1] Microsoft Research Cambridge, 21 Station Road, Cambridge CB1 2FB, UK and
Department of Mathematical Sciences, Durham University, Durham DH1 3LE, UK

Abstract

Real data are often given as a noisy unstructured point cloud, which is hard to visualize. The important problem is to represent topological structures hidden in a cloud by using skeletons with cycles. All past skeletonization methods require extra parameters such as a scale, a noise bound or weights of various quality criteria.

We define a homologically persistent skeleton, which depends only on a cloud of points and contains optimal subgraphs representing 1-dimensional cycles in the cloud across all scales. The full skeleton is a universal structure encoding topological persistence of cycles directly on the cloud. Hence a 1-dimensional shape of a cloud can be now easily predicted by visualizing our skeleton instead of guessing a scale for the original unstructured cloud.

We derive more subgraphs to reconstruct provably close approximations to an unknown graph given only by a noisy sample in any metric space. For a cloud of n points in the plane, the full skeleton and all its important subgraphs can be computed in time $O(n \log n)$. A general approximate algorithm has a time at most $O(n^3)$.

Categories and Subject Descriptors (according to ACM CCS): I.5.1 [Pattern Recognition]: Models—Structural

1. Introduction: motivations and our contributions

A *point cloud* C is a finite metric space, namely a finite set of points with only pairwise distances. The traditional way to visualize a shape of C is to select a scale α and join all points of C at a distance not more than 2α. If a cloud C is high-dimensional, choosing a suitable scale can be hard.

Topological Data Analysis resolves this difficulty by studying data across all scales. The output is a persistence diagram of pairs $(\text{birth}, \text{death})$ encoding life spans of topological features such as closed cycles from their birth to death. In the persistence diagram we can select pairs corresponding to closed cycles with a high persistence death $-$ birth.

However, the pairs $(\text{birth}, \text{death})$ alone are insufficient to actually locate cycles in a given cloud C. That is why, to visualize the persistence of 1-dimensional cycles hidden in C across all scales, we introduce a Homologically Persistent Skeleton $\text{HoPeS}(C)$ whose vertices are all points of C.

Now all 1-dimensional cycles that persist in a given cloud C over many scales are visualized directly on C. The skeleton $\text{HoPeS}(C)$ depends only on C, but contains shortest subgraphs that provably represent the 1-dimensional topology of C at every continuous scale, see Optimality Theorem 9.

$\text{HoPeS}(C)$ is obtained from a Minimum Spanning Tree $\text{MST}(C)$ by adding all critical edges giving birth to (homology classes of) 1D cycles in the persistence diagram, see two red critical edges with labels $(\text{birth}, \text{death})$ in Fig. 1.

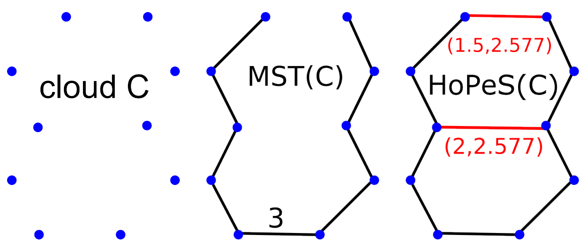

Figure 1: *Left: cloud $C \subset \mathbb{R}^2$. Middle:* $\text{MST}(C)$. *Right: a Homologically Persistent Skeleton with 2 red critical edges.*

Assuming that a cloud C is a noisy sample of an unknown graph G, we find another natural hierarchy of derived subgraphs $\mathrm{HoPeS}_{k,l}(C)$ that provably approximate G, see Theorem 15. Starting from any unstructured cloud C, we can visualize the full skeleton and then study its subgraphs that contain most persistent cycles depending on integers $k,l \geq 1$, which are easier to choose than a continuous scale α.

Here is a summary of our motivations for $\mathrm{HoPeS}(C)$.

• Visualize in one universal skeleton 1-dimensional cycles hidden in a high-dimensional cloud C across all scales α.

• Extend a classical Minimum Spanning Tree $\mathrm{MST}(C)$ of a cloud C to an optimal graph $\mathrm{HoPeS}(C)$ with cycles.

• Solve the skeletonization problem for unstructured clouds with guarantees and without using any extra parameters.

A high-level description of our contributions is below.

• Definition 6 introduces a Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$ of a point cloud C, which is the first universal structure that visualizes 1-dimensional cycles across all scales directly on the given cloud C in any metric space.

• For any $\alpha > 0$, Theorem 9 proves that $\mathrm{HoPeS}(C)$ contains a reduced skeleton with a minimum length over all graphs G having the homology of 'thickened' C at the scale α.

• The new Gap Search method in Propositions 11 and 13 strengthens the seminal stability of persistence diagrams by providing bijections between natural finite subdiagrams.

• For any ε-sample C of an unknown graph G, Theorem 15 gives conditions on G when subskeletons $\mathrm{HoPeS}_{k,l}(C)$ have the homotopy type of G and are within the 2ε-offset of G.

• Corollary 16 proves that $\mathrm{HoPeS}_{k,l}(C)$ is globally stable for any small perturbations of noisy samples C of graphs G.

2. Comparison with related past skeletonization work

Morse-Smale complex $\mathrm{MS}(f,M)$ is defined for a function f on a manifold M (or for a discrete gradient field on a complex). To compare $\mathrm{MS}(f,M)$ with $\mathrm{HoPeS}(C)$, which depends on a cloud C, we need more structure on C. In practice f is a density depending on close neighbors of a point in C. Let $d_C(p)$ be the distance from $p \in \mathbb{R}^d$ to a closest point of C. Then $\mathrm{MS}(d_C,\mathbb{R}^d)$ is a subdivision of a Delaunay triangulation and has the 1D skeleton larger than $\mathrm{HoPeS}(C)$.

Forman's discrete Morse theory for a cell complex with a discrete gradient field builds a smaller homotopy equivalent complex whose number of critical cells is minimized by the algorithm in [LLT04]. Optimality Theorem 9 minimizes the total length of a skeleton, not the number of critical edges. Removing low persistent edges from the full skeleton $\mathrm{HoPeS}(C)$ to get smaller skeletons $\mathrm{HoPeS}_{k,l}(C)$ is similar to the simplification [EHZ03] of Morse-Smale complexes, where critical points with close heights are cancelled.

All known skeletonization algorithms for clouds seem to require extra input parameters such as a scale α or a noise bound ε. Hence all these algorithms can not accept our minimal input, which is only a cloud C. That is why any experimental comparison of solutions to 2 different problems will be unfair and we compare only theoretical aspects below.

Delaunay-based skeletons. R. Singh et al. [SCP00] approximated a skeleton of a shape by a subgraph of a Delaunay triangulation using 3 thresholds: K for the minimum number of edges in a cycle and $\delta_{min}, \delta_{max}$ for inserting/merging Voronoi regions. Similar parameters are used in [KK02].

Skeletonization via Reeb graphs. Starting from a noisy sample C of an unknown graph G with a scale parameter, X. Ge et al. [GSBW11] considered the Reeb graph of the Vietoris-Rips complex on a cloud C at a given scale α. The Reeb graph is not intrinsically embedded into any space even if $C \subset \mathbb{R}^2$. The reconstruction in [GSBW11, Theorem 3.1] outputs a graph with a correct homotopy type, while all our derived skeletons $\mathrm{HoPeS}_{k,l}(C)$ also give close geometric approximations in the 2ε-offset of an unknown graph G.

Metric graph reconstruction. M. Aanjaneya et al. [ACC*12] studied a related problem approximating a metric on a large input graph Y by a metric on a small output graph $\hat{X}$. If Y is a good ε-approximation to an unknown graph X, then [ACC*12, Theorem 2] guarantees the existence of a homeomorphism $X \to \hat{X}$ that distorts the metrics on X and $\hat{X}$ with a multiplicative factor $1+c\varepsilon$ for $c > \frac{30}{b}$, where $b > 14.5\varepsilon$ is the length of a shortest edge of X. According to [GSBW11], the algorithm may not run on all inputs, but only for carefully chosen parameters. The skeletons $\mathrm{HoPeS}_{k,l}(C)$ are well-defined for any cloud C and $k,l \geq 1$.

Filamentary structures using Reeb-type graphs. F. Chazal et al. [CHS15] defined the α-Reeb graph G of a metric space X at a user-defined scale α. If X is ε-close to an unknown graph with edges of minimum length 8ε, then the output G is $34(\beta(G)+1)\varepsilon$-close to the input X, where $\beta(G)$ is the first Betti number of G, see [CHS15, Theorem 4.9]. The algorithm has the very fast time $O(n\log n)$ for n points in X and needs the scale α.

Graph Induced Complex GIC. T. Dey et al. [DFW13] built GIC depending on a scale α and a user-defined graph that spans a cloud C. If C is an ε-sample of a good manifold, then GIC has the same homology H_1 as the Vietoris-Rips complex on C at scales $\alpha \geq 4\varepsilon$. Theorem 15 describes graphs G that can be geometrically and topologically approximated from any ε-sample C without extra input parameters.

Skeleton for α-offsets in $\mathbb{R}^2$. This work extends [CK13, Kur14b, Kur14a] from locating holes in 2D clouds to a full skeleton. The Gap Search method in section 6 vastly improves [Kur15a, Theorem 7], which was stated for one subskeleton under stronger assumptions on a graph $G \subset \mathbb{R}^2$.

The key advantage of a Homologically Persistent Skeleton HoPeS(C) is its universal scale-independent structure. In comparison with the persistence diagram of isolated dots (homology classes), HoPeS(C) shows all persistent cycles directly on a cloud C. In comparison with all algorithms that require a scale α, the skeleton HoPeS(C) contains a hierarchy of derived skeletons $\mathrm{HoPeS}_{k,l}(C)$ independent of α.

The derived skeletons are most persistent subgraphs of HoPeS(C) depending on integer indices $k,l \geq 1$, which are easier to choose a posteriori rather than a continuous scale α a priori. We may start from the 'simplest guess' $k = l = 1$ and then try $k = 2$, $l = 1$ without re-running the algorithm, but only selecting a different subgraph of HoPeS(C).

3. Filtrations of complexes and persistent homology

The shape of a cloud C in any metric space M is usually visualized by fixing a scale $\alpha > 0$ and considering the α-*offset* $C^\alpha \subset M$ that is the union of closed balls with the radius α and centers at all points of C. When the scale α is increasing, the α-offsets form the *filtration* $\{C^\alpha\}$ that is the nested sequence $C = C^0 \subset \ldots \subset C^\alpha \subset \ldots \subset C^{+\infty} = M$. A cloud C is an ε-*sample* of a set $G \subset M$ if $C \subset G^\varepsilon$ and $G \subset C^\varepsilon$.

Since C^α may have complicated shapes, but continuously deform to simpler Čech complexes [Kur15b, Appendix A]. A faster computable filtration on a cloud C is by the *Vietoris-Rips complexes*. For any scale α, the complex $\mathrm{VR}(C;\alpha)$ has a k-dimensional simplex on points $v_0,\ldots,v_k \in C$ whenever the distance $D(v_i,v_j) \leq 2\alpha$ for all $0 \leq i < j \leq k$.

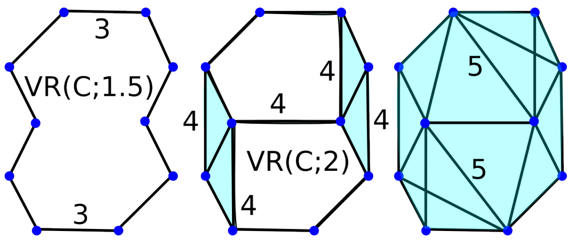

Figure 2: *The Vietoris-Rips complexes* $\mathrm{VR}(C;1.5)$, $\mathrm{VR}(C;2)$ *and* $\mathrm{VR}(C;2.5)$ *for the point cloud C in Fig. 1.*

$\mathrm{VR}(C;\alpha)$ can be high-dimensional even if $C \subset \mathbb{R}^2$. Any $\mathrm{VR}(C;\alpha)$ is uniquely determined by its 1-dimensional skeleton $\mathrm{VR}_1(C;\alpha)$ whose simplices are spanned by *cliques* (complete subgraphs) of $\mathrm{VR}_1(C;\alpha)$. The last picture in Fig. 2 shows a plane projection of $\mathrm{VR}(C;2.5)$ with 2 symmetric tetrahedra having longest edges of length 5.

The reader may consider only the Vietoris-Rips filtration $\{\mathrm{VR}(C;\alpha)\}$. However, our key results work for general complexes and require the concept of a length relative to an *ascending filtration* $\{Q(C;\alpha)\}$ of complexes on C, where $Q(C;0) = C$ and $Q(C;\alpha) \subset Q(C;\alpha')$ for any $\alpha \leq \alpha'$.

The *length* of the edge e between any two points of C *relative to* $\{Q(C;\alpha)\}$ is the doubled scale α when e enters $Q(\alpha)$, so $|e| = 2\min\{\alpha \,:\, e \subset Q(C;\alpha)\}$. If e doesn't enter $Q(C;\alpha)$ for any α, we set $|e| = +\infty$. The condition $Q(C;0) = C$ implies that the length of any edge between points of C is positive. For both filtrations $\{C^\alpha\}$ and $\{\mathrm{VR}(C;\alpha)\}$, this length $|e|$ coincides with the original metric on the cloud C.

The key idea of Topological Data Analysis is to study topological features such as homology that persist across many scales in a filtration $\{Q(C;\alpha)\}$. The 0-dimensional homology $H_0(Q)$ of a complex Q is the vector space formally generated by connected components of Q. The 1-dimensional homology $H_1(Q)$ is similarly formed by linear combinations of non-trivial 1-dimensional cycles, say with coefficients in $\mathbb{Z}_2 = \{0,1\}$, see [Kur15b, Appendix A].

For instance, $\mathrm{VR}(1.5)$ in Fig. 2 is one cycle, so $H_1 = \mathbb{Z}_2$ has dimension 1. The complex $\mathrm{VR}(C;2)$ looks like the character θ with 2 'independent' small cycles, whose 'sum' gives the big cycle, hence $H_1 = \mathbb{Z}_2 \oplus \mathbb{Z}_2$ has dimension 2. All cycles in the larger complex $\mathrm{VR}(C;2.5)$ are contractible, hence $H_1 = 0$. If a complex Q is disconnected, then its homology $H_1(Q)$ is considered as the direct sum of the 1-dimensional homology groups of all connected components of Q.

Increasing the scale α in a filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C, we will watch how (homology classes of) 1-dimensional cycles are born and die in $H_1(Q(C;\alpha))$. Any inclusion $f : Q(C;\alpha_i) \subset Q(C;\alpha_j)$ induces a homomorphism $f_* : H_1(Q(C;\alpha_i)) \to H_1(Q(C;\alpha_j))$. These homomorphisms for $\alpha_i < \alpha_j$ are crucial for defining life intervals (from birth to death) of homological classes below.

For $C \subset \mathbb{R}^2$ in Fig. 1 and $\alpha = 1.5$, when 2 horizontal edges of length 3 enter $\mathrm{VR}(C,1.5)$, a first cycle appears, so a homology class γ_1 is born at $\mathrm{birth}(\gamma_1) = 1.5$. For $\alpha = 2$, the horizontal edge of length 4 enters $\mathrm{VR}(C,2)$ and the big cycle splits into 2 smaller cycles. So a homology class γ_2 is born at $\mathrm{birth}(\gamma_2) = 2$. Both γ_1,γ_2 die at $\alpha = 2.5$ when their representative cycles become contractible within $\mathrm{VR}(C;2.5)$.

Definition 1 (births and deaths) For any filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C in a metric space, a homology class $\gamma \in H_1(Q(C;\alpha_i))$ is *born* at $\alpha_i = \mathrm{birth}(\gamma)$ if γ is not in the full image under the induced homomorphism $H_1(Q(C;\alpha)) \to H_1(Q(C;\alpha_i))$ for any $\alpha < \alpha_i$. The class γ *dies* at $\alpha_j = \mathrm{death}(\gamma) \geq \alpha_i$ when the image of γ under $H_1(Q(C;\alpha_i)) \to H_1(Q(C;\alpha_j))$ merges into the full image under $H_1(Q(C;\alpha)) \to H_1(Q(C;\alpha_j))$ for some $\alpha < \alpha_i$.

The births and deaths from Definition 1 are all critical scales $\alpha_1,\ldots,\alpha_m$ when the homology group $H_1(Q(C;\alpha))$ changes, so the induced homomorphisms $H_1(Q(C;\alpha_1)) \to H_1(Q(C;\alpha_2)) \to \ldots \to H_1(Q(C;\alpha_m))$ are not isomorphisms. This sequence of homomorphisms for the complexes $\mathrm{VR}(C;\alpha)$ in Fig. 2 is $0 \to \mathbb{Z}_2 \to \mathbb{Z}_2 \oplus \mathbb{Z}_2 \to 0$ corresponding to the scales $\alpha = 0$, 1.5, 2, 2.5. The persistence diagram below consists of the pairs $(\text{birth}, \text{death})$, which will be called *dots* to distinguish them from points of a cloud C.

Definition 2 (persistence diagram) Fix a filtration $\{Q(C;\alpha)\}$ of complexes on a set C. Let $\alpha_1,\dots,\alpha_m$ be all critical scales when the homology group $H_1(Q(C;\alpha))$ changes. Let μ_{ij} be the number of independent classes in $H_1(Q(C;\alpha))$ that are born at α_i and die at α_j. The *persistence diagram* $\text{PD}\{Q(C;\alpha)\} \subset \mathbb{R}^2$ is the multi-set consisting of all dots $(\alpha_i,\alpha_j) \in \mathbb{R}^2$ with the multiplicities $\mu_{ij} \geq 1$ and all diagonal dots (x,x) with the infinite multiplicity.

For the cloud C in Fig. 1 and the filtration $\{C^\alpha\}$, the homology H_1 has 2 classes that persist over $1.5 \leq \alpha < R$ and $2 \leq \alpha < R$, where $R = \frac{15}{8}\sqrt{17}$ is the circumradius of the largest Delaunay triangle with sides $4, \sqrt{17}, 5$ in Fig. 2. We use the approximate value $R \approx 2.577$ for simplicity. Hence the persistence diagram $\text{PD}\{C^\alpha\}$ has 2 off-diagonal red dots $(1.5, 2.577)$ and $(2, 2.577)$ in the last picture of Fig. 3.

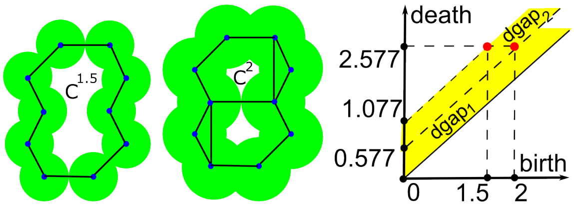

Figure 3: *The* α*-offsets* $C^{1.5}$, C^2 *and the 1-dimensional persistence diagram* $\text{PD}\{C^\alpha\}$ *for the cloud C in Fig. 1.*

Using persistence for noisy data is justified by Stability Theorem [EH10, section VIII.2] roughly saying that perturbing a cloud by ε perturbes the persistence diagram also by at most ε, see [Kur15b, Appendix A]. This stability is crucial for our Reconstruction Theorem 15, while the classical medial axis is unstable for noisy inputs according to [ABE09].

4. A Homologically Persistent Skeleton $\text{HoPeS}(C)$

To motivate the new concept of a Homologically Persistent Skeleton in key Definition 6, we explain below the simpler 0-dimensional case when a Minimum Spanning Tree $\text{MST}(C)$ nicely summarizes the single-edge clusters of a cloud C at all scales α. In sections 4 and 5, we always fix a filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C in a metric space M. So MST and HoPeS depend on any filtration $\{Q(C;\alpha)\}$, but we use the simpler notations $\text{MST}(C)$ and $\text{HoPeS}(C)$.

Definition 3 (MST) Fix a filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C. A *Minimum Spanning Tree* $\text{MST}(C)$ is a connected graph with the vertex set C and the minimum total length of edges relative to the filtration $\{Q(C;\alpha)\}$. For any scale $\alpha \geq 0$, a *forest* $\text{MST}(C;\alpha)$ is obtained from $\text{MST}(C)$ by removing all open edges that are longer than 2α.

A graph G *spans* a possibly disconnected complex Q on a cloud C if G has the vertex set C, every edge of G belongs to Q and the inclusion $G \subset Q$ induces a 1-1 correspondence between connected components. Lemma 4 says that $\text{MST}(C)$ for every $\alpha > 0$ contains the shortest forest that identifies all single-edge clusters of C. The *single-edge clustering* of C at a scale $\alpha > 0$ means that points $p, q \in C$ belong to the same cluster if and only if the distance $D(p,q) \leq \alpha$.

Lemma 4 Fix a filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C in a metric space. For any fixed scale $\alpha \geq 0$, the forest $\text{MST}(C;\alpha)$ has the minimum total length of edges among all graphs that span $Q(C;\alpha)$ at the same scale α.

Proof Let $e_1,\dots,e_m \subset \text{MST}(C)$ be all edges longer than 2α, so $\text{MST}(C) = e_1 \cup \dots \cup e_m \cup \text{MST}(C;\alpha)$. Assume that there is a graph G that spans $Q(C;\alpha)$ and is shorter than $\text{MST}(C;\alpha)$. Then the connected graph $G \cup e_1 \cup \dots \cup e_m$ has the vertex set C and is shorter than a Minimum Spanning Tree $\text{MST}(C)$, which contradicts Definition 3. $\square$

Our first main Theorem 9 extends Lemma 4 from $\text{MST}(C)$ to the skeleton $\text{HoPeS}(C)$ with cycles summarizing all 1-dimensional persistence of C instead of only clusters.

Definition 5 (critical edges) Fix a filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C. Each off-diagonal dot $(\alpha_i,\alpha_j) \in \text{PD}\{Q(C;\alpha)\}$ corresponds to a homology class γ that persists in $H_1(Q(C;\alpha))$ over $\alpha_i \leq \alpha < \alpha_j$. The class γ was formed (born) after adding a last edge e to $Q(C;\alpha_i)$. This edge e is called *critical* and has the associated *label* (α_i,α_j).

For a fixed scale $\alpha > 0$, the above inequalities $\alpha_i \leq \alpha < \alpha_j$ describe the axes-aligned rectangle in the persistence diagram $\text{PD}\{Q(C;\alpha)\}$ with the bottom right corner at the dot (α,α). This rectangle contains $(\text{birth}, \text{death})$ of all classes that are 'alive' at the scale α. That is why at a fixed scale α we will ignore all critical edges e with $\text{death}(e) \leq \alpha$.

Definition 6 (HoPeS) A *Homologically Persistent Skeleton* $\text{HoPeS}(C)$ is the union of a minimum spanning tree $\text{MST}(C)$ and all critical edges with their labels $(\text{birth}, \text{death})$.

For any scale $\alpha \geq 0$, the *reduced* skeleton $\text{HoPeS}(C;\alpha)$ is obtained from $\text{HoPeS}(C)$ by removing all edges that are longer than 2α and all critical edges e with $\text{death}(e) \leq \alpha$.

If some distances between points of C are the same, then $\text{MST}(C)$, hence $\text{HoPeS}(C)$, is not unique. The complex $Q(C;0)$ in any our filtration coincides with a cloud C, so $\text{HoPeS}(C;0) = C$. By Definition 6 a critical edge e belongs to the reduced skeleton $\text{HoPeS}(C;\alpha)$ if and only if $\text{birth}(e) \leq \alpha < \text{death}(e)$. Hence the homology class born due to the addition of e is alive at the scale α. Namely, any critical edge e is added to $\text{HoPeS}(C;\alpha)$ at $\alpha = \text{birth}(e)$ and will be later removed at the larger scale $\alpha = \text{death}(e)$.

When α is increasing, the sequence of $\text{HoPeS}(C;\alpha)$ may not be monotone. But if $\text{HoPeS}(C;\alpha)$ has become connected, it will stay connected for all larger α. Indeed, removing a critical edge destroys only a cycle, not connectivity.

5. Optimality of reduced skeletons $\mathrm{HoPeS}(C;\alpha)$

In this section we prove first properties of the skeleton $\mathrm{HoPeS}(C)$ including Optimality Theorem 9. Proposition 7 says that the persistence diagram $\mathrm{PD}\{Q(C;\alpha)\}$ of isolated dots can be turned into a graph structure on a cloud C. So the skeleton $\mathrm{HoPeS}(C)$ naturally 'structurizes' isolated dots in the 1-dimensional persistence diagram $\mathrm{PD}\{Q(C;\alpha)\}$.

Proposition 7 Fix a filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C in a metric space. The 1-dimensional persistence diagram $\mathrm{PD}\{Q(C;\alpha)\}$ can be canonically reconstructed from a full Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$.

Proof By Definitions 5 and 6 any off-diagonal dot $(\alpha_i,\alpha_j)\in\mathrm{PD}\{Q(C;\alpha)\}$ corresponds to a critical edge of $\mathrm{HoPeS}(C)$ with the label (α_i,α_j). We can easily read the labels on all critical edges of $\mathrm{HoPeS}(C)$ to get all off-diagonal dots in the persistence diagram $\mathrm{PD}\{Q(C;\alpha)\}$. □

Proposition 8 means that $\mathrm{HoPeS}(C)$ can be used for comparing point clouds up to isometries and uniform scaling in a metric space. Indeed, the definitionof the bottleneck distance d_B between persistence diagrams in [Kur15b, Appendix A] is easily re-stated for skeletons with labels $(\mathrm{birth},\mathrm{death})$.

Proposition 8 The topology of a Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$ with all labels is preserved under any isometric transformation of a cloud C. If a cloud C is fixed and a filtration $\{Q(C;\alpha)\}$ is re-parameterized by $\alpha\mapsto\mathrm{constant}\cdot\alpha$, then $\mathrm{HoPeS}(C)$ has the same combinatorial structure, but all labels will be multiplied by the constant.

Proof The complexes $Q(C;\alpha)$ are unaffected by isometries of C. If α is multiplied by a constant, we multiply all births and deaths in $\mathrm{PD}\{Q(C;\alpha)\}$ by the same constant. □

Theorem 9 naturally extends the optimality of $\mathrm{MST}(C)$ of a cloud C in Lemma 4 from dimension 0 (single-edge clusters of C) to dimension 1 (persistent cycles hidden in C).

Theorem 9 (optimality of reduced skeletons at all scales) Fix a filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C in a metric space. For any fixed scale $\alpha>0$, the reduced skeleton $\mathrm{HoPeS}(C;\alpha)$ has the minimum total length of edges over all graphs $G\subset Q(C;\alpha)$ that span $Q(C;\alpha)$ and induce an isomorphism in homology $H_1(G)\to H_1(Q(C;\alpha))$.

Theorem 9 and all further results have proofs in [Kur15b, Appendices B,C]. In section 8 we discuss computations and approximations of $\mathrm{HoPeS}(C)$ in any metric space.

6. The Gap Search strengthens stability of persistence

In sections 6, 7 we consider 1-dimensional persistence diagrams only for α-offsets of a subspace C in a metric space M, where C can be a cloud or a graph $G\subset M$. We strengthen Stability Theorem [EH10, section VIII.2] restricting a bijection between original infinite persistence diagrams to a bijection between finite subdiagrams in Propositions 11, 13.

We will use a Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$ to reconstruct an unknown graph G in a metric space M from a noisy sample C. However, a full skeleton $\mathrm{HoPeS}(C)$ contains too many critical edges that are in a 1-1 correspondence with all off-diagonal dots in the persistence diagram $\mathrm{PD}\{C^\alpha\}$. We will derive smaller skeletons $\mathrm{HoPeS}_{k,l}(C)$ by removing critical edges with (1) a low persistence and (2) a high persistence and large birth times.

Critical edges of type (1) above correspond to homology classes that die shortly after birth. Critical edges of type (2) form cycles that live long, but are born only at large scales α. These large cycles are not present in a graph G given by a sample C, but are formed when distant edges of G start overlapping with each other after a substantial thickening of G. The diagonal gaps quantify our perception of the persistence diagram, because we naturally look for a widest gap to separate persistent features from noisy artefacts.

The persistence diagram $\mathrm{PD}\{C^\alpha\}$ technically includes all diagonal dots $(x,x)\in\mathbb{R}^2$. Below we define smaller subdiagrams of $\mathrm{PD}\{C^\alpha\}$ consisting of finitely many dots.

Definition 10 (gap dgap_k, subdiagram DS_k, scale ds_k) For any subspace C of a metric space, a *diagonal gap* is a strip $\{0\le a<y-x<b\}$ that has dots of $\mathrm{PD}\{C^\alpha\}$ in both boundary lines $\{y-x=a\}$ and $\{y-x=b\}$, but not inside the strip. For any integer $k\ge1$, the k-th *widest diagonal gap* $\mathrm{dgap}_k(C)$ has the k-th largest vertical width $|\mathrm{dgap}_k(C)|=b-a$.

The *diagonal subdiagram* $\mathrm{DS}_k(C)\subset\mathrm{PD}\{C^\alpha\}$ consists of only the dots above the lowest of the first k widest $\mathrm{dgap}_i(C)$, $i=1,\dots,k$. So each $\mathrm{DS}_k(C)$ is bounded below by $y-x=a$ and has the *diagonal scale* $\mathrm{ds}_k(C)=a$.

In Definition 10 if $\mathrm{PD}\{C^\alpha\}$ has different diagonal gaps with the same width, we say that a lower diagonal gap has a larger width. If $\mathrm{PD}\{C^\alpha\}$ has dots only in m different lines $\{y-x=a_i>0\}$, $i=1,\dots,m$, we have exactly m diagonal gaps $\mathrm{dgap}_i(C)$. We ignore the highest gap $\{y-x>\max a_i\}$, so we set $\mathrm{dgap}_i(C)=\emptyset$ and $|\mathrm{dgap}_i(C)|=0$ for $i>m$.

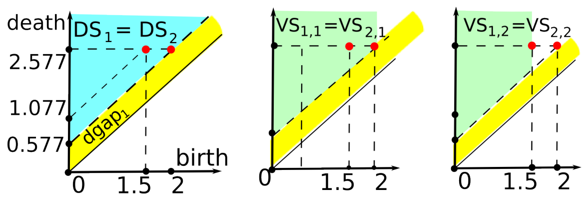

Figure 4: *Subdiagrams* $\mathrm{DS}_k(C),\mathrm{VS}_{k,l}(C)\subset\mathrm{PD}\{C^\alpha\}$ *from Definitions 10 and 12 for the cloud C in Fig. 1.*

The cloud C in Fig. 1 has the persistence diagram $\mathrm{PD}\{C^\alpha\}$ with only 2 off-diagonal dots $(1.5,2.577)$ and $(2,2.577)$. Then $\mathrm{dgap}_1(C)=\{0<y-x<0.577\}$ is below $\mathrm{dgap}_2(C)=\{0.577<y-x<1.577\}$. So $\mathrm{DS}_1(C)=\mathrm{DS}_2(C)$ consists of $(1.5,2.577),(2,2.577)$, $\mathrm{ds}_1(C)=\mathrm{ds}_2(C)=0.577$ in Fig. 4.

Definition 10 makes sense in any persistence diagram, say for α-offsets G^α of a graph $G \subset M$. The graph θ in Fig. 5 has the 1-dimensional persistence diagram $\mathrm{PD}\{\theta^\alpha\}$ containing only one off-diagonal dot $(0, 2.577)$ of multiplicity 2. Then $\mathrm{dgap}_1(\theta) = \{0 < y - x < 2.577\}$, $|\mathrm{dgap}_2(\theta)| = 0$ and $\mathrm{DS}_1(\theta) = \{(0, 2.577), (0, 2.577)\}$ and $\mathrm{ds}_1(\theta) = 2.577$.

Stability Theorem [EH10, section VIII.2] briefly says that, for any ε-sample C of G, there is a bijection between infinite persistence diagrams $\psi : \mathrm{PD}\{G^\alpha\} \to \mathrm{PD}\{C^\alpha\}$ so that the L_∞-distance $||q - \psi(q)||_\infty \leq \varepsilon$ for any $q \in \mathrm{PD}\{G^\alpha\}$.

Proposition 11 below restricts such a bijection ψ to finite subdiagrams of only finitely many dots (with high persistence) above the k-th widest diagonal gap in persistence.

Proposition 11 Let C be any ε-sample of a graph G. If $|\mathrm{dgap}_k(G)| - |\mathrm{dgap}_{k+1}(G)| > 8\varepsilon$, there is a bijection $\psi : \mathrm{DS}_k(G) \to \mathrm{DS}_k(C)$ so that $||q - \psi(q)||_\infty \leq \varepsilon$, $q \in \mathrm{DS}_k(G)$.

If we reconstruct cycles of G from a noisy sample C using a close diagram $\mathrm{PD}\{C^\alpha\}$, we should look for dots $(\mathrm{birth}, \mathrm{death})$ having a high persistence death − birth and small birth. Hence we search for a vertical gap that separates dots $(\mathrm{birth}, \mathrm{death})$ having a small birth and all other dots.

Definition 12 (gap $\mathrm{vgap}_{k,l}$, subdiagram $\mathrm{VS}_{k,l}$, scale $\mathrm{vs}_{k,l}$) In the subdiagram $\mathrm{DS}_k(C)$ from Definition 10 a *vertical gap* is a widest vertical strip $\{a < x < b\}$ whose boundary contains a dot from $\mathrm{DS}_k(C)$ in the line $\{x = a\}$, but not inside the strip. For $l \geq 1$, the l-th *widest vertical gap* $\mathrm{vgap}_{k,l}(C)$ has the l-th largest horizontal width $|\mathrm{vgap}_{k,l}(C)| = b - a$. The *vertical subdiagram* $\mathrm{VS}_{k,l}(C) \subset \mathrm{DS}_k(C)$ consists of only the dots to the left of the first l widest vertical gaps $\mathrm{vgap}_{k,j}(C)$, $j = 1, \ldots, l$. So each $\mathrm{VS}_{k,l}(C)$ is bounded on the right by the line $x = b$ and has the *vertical scale* $\mathrm{vs}_{k,l}(C) = a$.

In Definition 12 if there are different vertical gaps with the same horizontal width, we say that the leftmost vertical gap has a larger width. So we prefer the leftmost vertical gap, while in Definition 10 we prefer the lowest diagonal gap.

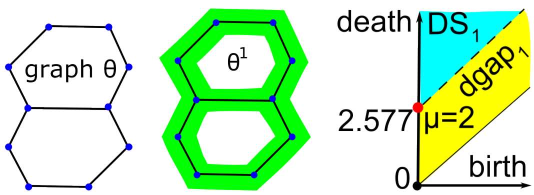

Figure 5: *The graph $\theta \subset \mathbb{R}^2$ with the α-offset θ^1 and the 1-dimensional persistence diagram* $\mathrm{PD}\{\theta^\alpha\}$.

We allow the case $b = +\infty$, so the widest vertical gap $\mathrm{vgap}_{k,1}(C)$ always has the form $\{x > a\}$, $VS_{1,1}(C) = \mathrm{DS}_1(C)$ and we set $|\mathrm{vgap}_{k,1}(C)| = +\infty$. If $\mathrm{DS}_k(C)$ has dots in $m \geq 1$ different lines $\{x = b_l \geq 0\}$, $l = 1, \ldots, m$, then $\mathrm{DS}_k(C)$ has exactly m vertical gaps $\mathrm{vgap}_{k,l}(C)$.

For the cloud C in Fig. 1, the diagonal subdiagram $\mathrm{DS}_1(C)$ has $\mathrm{vgap}_{1,1}(C) = \{x > 2\}$, $\mathrm{vgap}_{1,2}(C) = \{1.5 < x < 2\}$, $\mathrm{VS}_{1,1}(C) = \{(1.5, 2.577), (2, 2.577)\}$ and $\mathrm{VS}_{1,2}(C) = \{(1.5, 2.577)\}$, so $\mathrm{vs}_{1,1}(C) = 2$, $\mathrm{vs}_{1,2}(C) = 1.5$, see Fig. 4. Any finite cloud C has no cycles at $\alpha = 0$, hence $\mathrm{PD}\{C^\alpha\}$ has no dots in the vertical death axis, so any $\mathrm{vgap}_{k,l}(C)$ has a left boundary line $\{x = a > 0\}$ and all $\mathrm{vs}_{k,l}(C) > 0$.

Definition 12 makes sense for any persistence diagram, say for α-offsets G^α of a graph G in a metric space. The diagonal subdiagram $\mathrm{DS}_1(\theta)$ for the graph θ in Fig. 5 consists of the doubled dot $(0, 2.577)$. The only vertical subdiagram $\mathrm{VS}_{1,1}(\theta)$ is within the vertical death axis and consists of the same doubled dot $(0, 2.577)$, so $\mathrm{vs}_{1,1}(\theta) = 0$.

Proposition 13 Let $\psi : \mathrm{DS}_k(G) \to \mathrm{DS}_k(C)$ be a bijection such that $||q - \psi(q)||_\infty \leq \varepsilon$ for all dots $q \in \mathrm{DS}_k(G)$ as in Proposition 11. If $|\mathrm{vgap}_{k,l}(G)| - |\mathrm{vgap}_{k,l+1}(G)| > 4\varepsilon$ for some $l \geq 1$, then ψ restricts to a bijection $\mathrm{VS}_{k,l}(G) \to \mathrm{VS}_{k,l}(C)$ between smaller vertical subdiagrams.

7. Reconstructions by derived skeletons $\mathrm{HoPeS}_{k,l}(C)$

Definition 14 (derived skeletons $\mathrm{HoPeS}_{k,l}$) Let C be a finite cloud in a metric space. For integers $k, l \geq 1$, the *derived skeleton* $\mathrm{HoPeS}_{k,l}(C)$ is obtained from a full Homologically Persistent Skeleton $\mathrm{HoPeS}(C)$ by removing all edges that are longer than $2\mathrm{vs}_{k,l}(C)$ and keeping only critical edges with $(\mathrm{birth}, \mathrm{death}) \in \mathrm{VS}_{k,l}(C)$ and with death $> \mathrm{vs}_{k,l}(C)$.

For the cloud C in Fig. 1, we have $\mathrm{vs}_{1,1}(C) = 2$ and $\mathrm{vs}_{1,2}(C) = 1.5$. Hence the derived skeleton $\mathrm{HoPeS}_{1,1}(C)$ coincides with full $\mathrm{HoPeS}(C)$, while $\mathrm{HoPeS}_{1,2}(C)$ misses only the middle edge of length 4. The critical scale $\mathrm{vs}_{k,l}(C)$ is defined in such a way that the derived skeleton $\mathrm{HoPeS}_{k,l}(C)$ is within the reduced skeleton $\mathrm{HoPeS}(C; \mathrm{vs}_{k,l}(C))$.

Our second main Theorem 15 guarantees a reconstruction of a good enough graph G from any noisy ε-sample C up to homotopy equivalence. The homotopy type of a connected graph G is completely determined by its homology $H_1(G)$. Namely, G is homotopy equivalent to a wedge of $\beta(G)$ loops, where the first Betti number is $\beta(G) = \dim H_1(G)$.

Theorem 15 (reconstruction by derived skeletons) Let C be any ε-sample of an unknown graph G in a metric space, so $C \subset G^\varepsilon$ and $G \subset C^\varepsilon$. Let G satisfy the following conditions.

(1) All cycles $L \subset G$ are 'persistent', namely $\mathrm{death}(L) \geq \mathrm{ds}_k(G)$ for some index $k \geq 1$.

(2) The width $|\mathrm{dgap}_k(G)|$ 'jumps', namely $|\mathrm{dgap}_k(G)| - |\mathrm{dgap}_{k+1}(G)| > 8\varepsilon$ for the same k as in (1).

(3) No cycles are born in α-offsets G^α for 'small' $\alpha > 0$, namely $\mathrm{vs}_{k,l}(G) = 0$ for some $l \geq 1$.

(4) The width $|\mathrm{vgap}_{k,l}(G)|$ 'jumps', so $|\mathrm{vgap}_{k,l}(G)| - |\mathrm{vgap}_{k,l+1}(G)| > 4\varepsilon$ for the same k, l as above.

Then we get the lower bound for noise $\mathrm{vs}_{k,l}(C) \leq \varepsilon$ and the derived skeleton $\mathrm{HoPeS}_{k,l}(C) \subset G^{2\varepsilon}$ has the same H_1 as G.

Conditions (1)–(4) of Theorem 15 are stated in terms of the persistence diagram $\mathrm{PD}\{G^\alpha\}$ for simplicity. However, all characteristics like $|\mathrm{dgap}_k(G)|$ and $|\mathrm{vgap}_{k,l}(G)|$ can be defined purely in terms of α-offsets $G^\alpha \subset M$. Indeed, any dot $(\mathrm{birth}, \mathrm{death}) \in \mathrm{PD}\{G^\alpha\}$ corresponds to the life span of a homology class in $H_1(G^\alpha)$ over $\mathrm{birth} \le \alpha < \mathrm{death}$.

For a fixed graph G, Theorem 15 can be satisfied by many different indices $k, l \ge 1$. Hence, starting from only a cloud C, we get many approximations by $\mathrm{HoPeS}_{k,l}(C)$ to an unknown graph G within the 2ε-offset $G^{2\varepsilon} \subset M$. Moreover, we can estimate the closeness of our approximation due to the lower bound $\mathrm{vs}_{k,l}(C) \le \varepsilon$ of an unknown noise level ε.

In the proof of Theorem 15 condition (1) implies that the homology class of each cycle $L \subset G$ is captured in the subdiagrams $\mathrm{DS}_k(G), \mathrm{VS}_{k,l}(G)$. The diagonal scales $\mathrm{ds}_k(G)$ are expected to be low, because the diagonal subdiagram $\mathrm{DS}_k(G)$ is above the *lowest* of the first k widest $\mathrm{dgap}_i(G)$, $i = 1, \dots, k$. If we consider C in Fig. 1 as an ε-sample of the graph θ in Fig. 5, then condition (1) holds, because all cycles of θ have death $= 2.577$ and both $\mathrm{ds}_1(\theta) = \mathrm{ds}_2(\theta) = 2.577$.

Let $G \subset \mathbb{R}^2$ be a distant union of 2 circles with radii $R_1 < R_2$. Then $\mathrm{PD}\{G^\alpha\}$ has only 2 off-diagonal dots $(0, R_1), (0, R_2)$. If $R_2 > 2R_1$, then $\mathrm{dgap}_1(G) = \{R_1 < y - x < R_2\}$, so $\mathrm{ds}_1(G) = R_2$ is larger than the death radius of the 1st circle. Here condition (1) fails for $k = 1$, because the 1st circle is 'too small' in comparison with the 2nd circle.

However, the 2nd widest diagonal gap is $\mathrm{dgap}_2(G) = \{0 < y - x < R_1\}$, so $\mathrm{ds}_2(G) = R_1$ and condition (1) holds. Namely both dots $(0, R_1)$ and $(0, R_2)$ are captured in the diagonal subdiagram $\mathrm{DS}_2(G)$. For radii $R_1 < R_2 \le 2R_1$, condition (1) holds for both possible indices $k = 1$, $k = 2$.

Conditions (2) and (4) are the same as in Propositions 11, 13, which will imply that the vertical subdiagrams $\mathrm{VS}_{k,l}(G), \mathrm{VS}_{k,l}(C)$ are ε-close in the bottleneck distance d_B. All diagonal gaps are ordered by their vertical widths: $|\mathrm{dgap}_1(G)| \ge |\mathrm{dgap}_2(G)| \ge \dots$, where we can set the last width as 0, say for the empty diagonal gap $\{0 < y - x < 0\}$.

So if $\mathrm{PD}\{G^\alpha\}$ isn't empty, we can always find different successive widths $|\mathrm{dgap}_k(G)| > |\mathrm{dgap}_{k+1}(G)|$. Hence Condition (2) holds for any $\varepsilon < \frac{1}{8}(|\mathrm{dgap}_k(G)| - |\mathrm{dgap}_{k+1}(G)|)$. The same arguments apply to condition (4), which certainly holds for $l = 1$. Indeed, the 1st widest vertical gap has the infinite width $|\mathrm{vgap}_{k,1}(G)| = +\infty$ by Definition 12.

If a graph G is a tree without cycles, so $H_1(G) = 0$ and no 1-dimensional homology classes are born at $\alpha = 0$, then $\mathrm{PD}\{G^\alpha\}$ has no dots in the vertical axis and any $\mathrm{vs}_{k,l}(G) > 0$, hence condition (3) fails. Actually, our attempt to approximate a tree G from only a cloud C fails naturally, because the dot closest to the vertical death axis in $\mathrm{PD}\{G^\alpha\}$ is considered as a 'noisy' perturbation of the dot $(0, \mathrm{death}(L))$ representing a non-existing cycle L of the tree G.

For a graph G with cycles, we explain why condition (3) isn't restrictive. Let $\{0 < x < w\}$ be the leftmost vertical gap in the diagonal subdiagram $\mathrm{DS}_k(G)$. Denote by $l_{\min}$ the minimum index with $|\mathrm{vgap}_{k,l}(G)| \le w$. Then all further vertical gaps can be only thinner, hence the leftmost of any first $l \ge l_{\min}$ vertical gaps is $\{0 < x < w\}$, so $\mathrm{vs}_{k,l}(G) = 0$.

Corollary 16 In the conditions of Theorem 15 if another cloud $\tilde{C}$ is δ-close to C, then the perturbed derived skeleton $\mathrm{HoPeS}_{k,l}(\tilde{C})$ is $(2\delta + 4\varepsilon)$-close to $\mathrm{HoPeS}_{k,l}(C)$.

The skeletons $\mathrm{HoPeS}_{k,l}(C)$ have vertices at all points of C and are locally sensitive to perturbations of C. However, Corollary 16 justifies that $\mathrm{HoPeS}_{k,l}(C)$ are globally stable in the most practical case for noisy samples C of graphs G.

8. Algorithms, experiments, discussion and problems

Fig. 6–8 show the derived skeletons $\mathrm{HoPeS}_{1,1}(C)$, which were computed for clouds $C \subset \mathbb{R}^2$ of n points in time $O(n \log n)$, see details in [Kur15b, Appendix D]. This algorithm uses a duality between 0-dimensional and 1-dimensional persistence for α-complexes obtained from a Delaunay triangulation $\mathrm{Del}(C) \subset \mathbb{R}^2$ [EH10]. Starting from $\mathrm{Del}(C)$, we consider all edges in the decreasing order of length and find all critical edges when two components of $\mathbb{R}^2 - C^\alpha$ merge. At the end $\mathbb{R}^2 - C(\alpha)$ becomes a single component, we get $\mathrm{MST}(C)$ and add all critical edges. The simplified graphs are obtained by collapsing edges of half-lengths shorter than $\mathrm{vs}_{k,l}(C) \le \varepsilon$ justified by Theorem 15.

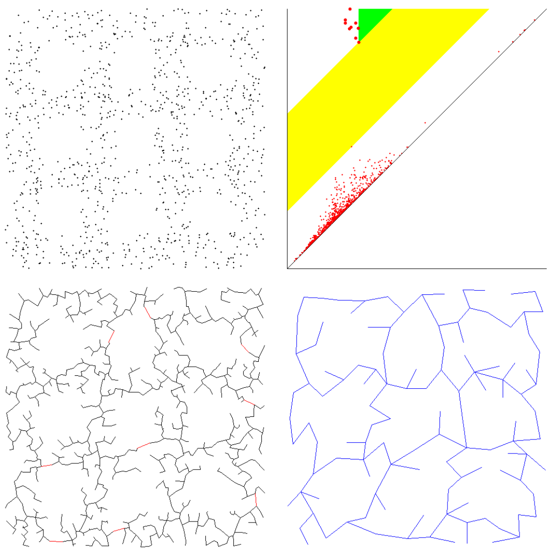

Figure 6: *Top: cloud C of 999 points and $\mathrm{PD}\{C^\alpha\}$. Bottom: derived skeleton $\mathrm{HoPeS}_{1,1}(C)$ and its rough simplification.*

For a cloud C of n points in any metric space, the 1-dimensional persistence diagram is found [EH10] in time

$O(m^3)$, where $m = O(n^3)$ is the size of the largest 2-dimensional complex in a given filtration on C. This algorithm allows us in parallel to record in all critical edges, usually called *positive*, because they create 1D cycles, while edges of $\mathrm{MST}(C)$ are *negative*, because they lead to a merger of components. Hence the skeletons $\mathrm{HoPeS}(C)$, $\mathrm{HoPeS}(C;\alpha)$, $\mathrm{HoPeS}_{k,l}(C)$ are found in the same time.

In Fig. 9–12 clouds C are extracted from images using the Canny edge detector with the low threshold 150 and ratio 3.

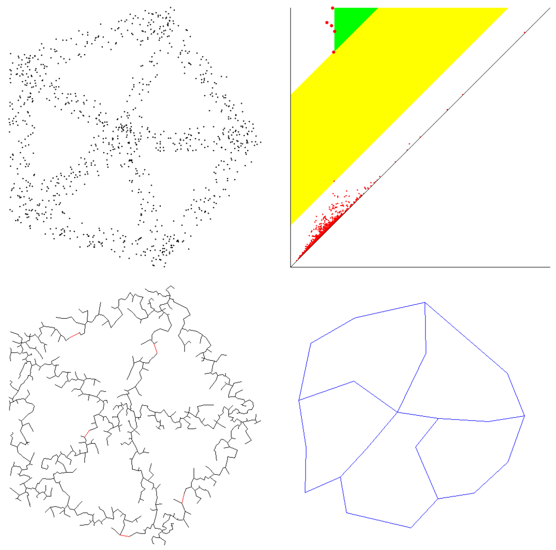

Figure 7: *Top: cloud C of 999 points and* $\mathrm{PD}\{C^\alpha\}$. *Bottom: derived skeleton* $\mathrm{HoPeS}_{1,1}(C)$ *and its rough simplification.*

This work was supported by the EPSRC-funded secondment at Microsoft Research Cambridge, UK. A C++ code is at http://kurlin.org/persistent-skeletons.php. The author thanks all reviewers for their helpful suggestions and is open for collaboration on problems below and any related projects.

• Simplify $\mathrm{HoPeS}(C)$ to get a smoother and locally stable skeleton that has fewer vertices than C, but still guarantees a close geometric approximation as in Theorem 15.

• Generalize Theorem 15 for unbounded noise similarly to the recent advance in dealing with outliers [BCD*15].

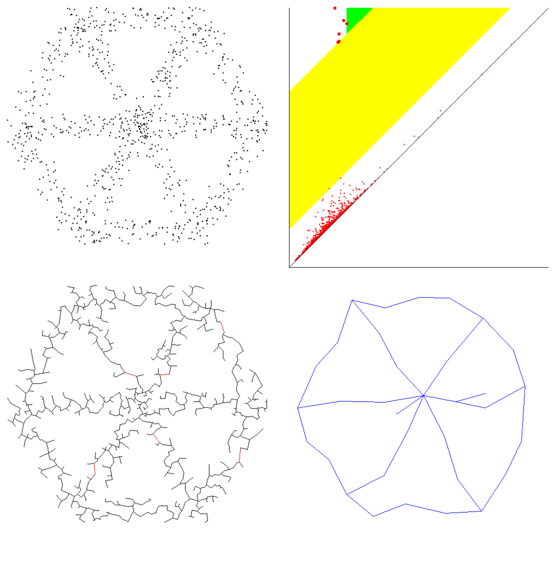

Figure 8: *Top: cloud C of 999 points and* $\mathrm{PD}\{C^\alpha\}$. *Bottom: derived skeleton* $\mathrm{HoPeS}_{1,1}(C)$ *and its rough simplification.*

References

[ABE09] Attali D., Boissonnat J.-D., Edelsbrunner H.: Stability and computation of medial axes - a state-of-the-art. In *Math. Foundations of Scientific Visualization, Comp. Graphics, Massive Data Exploration* (2009), Springer, pp. 109–125. 4

[ACC*12] Aanjaneya M., Chazal F., Chen D., Glisse M., Guibas L., Morozov D.: Metric graph reconstruction from noisy data. *Int. J. Comp. Geometry Appl. 22* (2012), 305–325. 2

[BCD*15] Buchet M., Chazal F., Dey T. K., Fan F., Oudot S., Wang Y.: Topological analysis of scalar fields with outliers. In *Proceedings of SoCG* (2015). 8

[CdSO14] Chazal F., de Silva V., Oudot S.: Persistence stability for geometric complexes. *Geometriae Dedicata 173* (2014), 193–214. 9

[CHS15] Chazal F., Huang R., Sun J.: Gromov-hausdorff approximation of filament structure using reeb-type graph. *Discrete Computational Geometry 53* (2015), 621–649. 2

[CK13] Chernov A., Kurlin V.: Reconstructing persistent graph structures from noisy images. *Image-A 3* (2013), 19–22. URL: http://munkres.us.es/image-a. 2

[DFW13] Dey T. K., Fan F., Wang Y.: Graph induced complex on data points. In *Proceedings of SoCG* (2013), pp. 107–116. 2

[EH10] Edelsbrunner H., Harer J.: *Computational topology. An introduction*. AMS, Providence, 2010. 4, 5, 6, 7

[EHZ03] Edelsbrunner H., Harer J., Zomorodian A.: Hierarchical morse-smale complexes for piecewise linear 2-manifolds. *Discrete Comput. Geometry* (2003), 87–107. 2

[GSBW11] Ge X., Safa I., Belkin M., Wang Y.: Data skeletonization via reeb graphs. In *Proceedings of NIPS* (2011). 2

[Hat01] Hatcher A.: *Algebraic Topology*. Cambridge University Press, 2001. 9

[KK02] Kégl B., Krzyzak A.: Piecewise linear skeletonization using principal curves. *Tran. PAMI 24* (2002), 59–74. 2

[Kur14a] Kurlin V.: Auto-completion of contours in sketches, maps and sparse 2d images based on topological persistence. In *Proceedings of SYNASC workshop CTIC: Computational Topology in Image Context* (2014), pp. 594–601. 2

[Kur14b] Kurlin V.: A fast and robust algorithm to count topologically persistent holes in noisy clouds. In *Proceedings of CVPR* (2014), pp. 1458–1463. URL: http://kurlin.org/projects/holes-in-noisy-clouds.php. 2

[Kur15a] KURLIN V.: A homologically persistent skeleton is a fast and robust descriptor for a sparse cloud of interest points in noisy 2d images. In *Proceedings of CAIP* (2015). URL: http://kurlin.org/projects/hopes.pdf. 2

[Kur15b] KURLIN V.: A one-dimensional homologically persistent skeleton of an unstructured point cloud in any metric space (full version of this paper). URL: http://kurlin.org/projects/persistent-skeleton.pdf. 3, 4, 5, 7

[LLT04] LEWINER T., LOPES H., TAVARES G.: Applications of forman's discrete morse theory to topology visualization and mesh compression. *IEEE Transactions on Visuzalization and Computer Graphics 10* (2004), 499–508. 2

[SCP00] SINGH R., CHERKASSKY V., PAPANIKOLOPOULOS N.: Self-organizing maps for the skeletonization of sparse shapes. *IEEE Tran. Neural Networks 11* (2000), 241–248. 2

Appendix A: background on complexes and homology

To avoid any confusion, we continue numbering definitions, claims and figures in the appendices below as in the paper.

Definition 17 (simplicial complex) A simplicial complex Q with a vertex set V is a collection of finite subsets $\{v_0,\dots,v_k\} \subset V$ (called *simplices*) such that

- any subset (called a *face*) of a simplex is also a simplex and is included in Q;
- any non-empty intersection of simplices is their common face included in Q.

Any simplex on $k+1$ vertices has this geometric realization:

$$\Delta^k = \{(t_0,t_1,\dots,t_k) \in \mathbb{R}^k \mid t_0+t_1+\dots+t_k \leq 1,\ t_i \geq 0\} \subset \mathbb{R}^k.$$

We define the *geometric realization* of any simplicial complex Q by gluing realizations of all its simplices along their common faces. Hence Q inherits the Euclidean topology.

The dimension of a simplex spanned by $k+1$ vertices is k. The dimension of a complex Q is the maximum dimension of its simplices. A 1-dimensional cycle in a complex Q is a sequence of edges $e_1,\dots,e_k$ such that e_i and e_{i+1} have the same endpoint, where $e_{k+1} = e_1$. Below we define the homology group $H_1(Q)$ of a simplicial complex Q only with coefficients in the group $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z} = \{0,1\}$.

Definition 18 (homology $H_1(Q)$ of a complex) Cycles of a complex Q can be algebraically written as finite linear combinations of edges (with coefficients 0 or 1) and generate the vector space C_1 of cycles. The boundaries of all triangles in Q are cycles of 3 edges and generate the space $B_1 \subset C_1$. The quotient space C_1/B_1 is the *homology* group $H_1(Q)$. The operation is the addition of cycles, the empty cycle is the zero.

The sphere S^2 has trivial $H_1(S^2) = 0$, the torus T has $H_1(T) = \mathbb{Z}_2 \oplus \mathbb{Z}_2$. Even if a simplicial complex Q is infinite, say the Vietoris-Rips complex of an infinite set, $H_1(Q)$ is well-defined in terms of finite linear combinations of edges.

Definition 19 (homotopy equivalence) Topological spaces are called *homotopy equivalent* $X \sim Y$ if there is a pair of continuous maps $f: X \to Y$ and $g: Y \to X$ such that both compositions $f \circ g: Y \to X \to Y$, $g \circ f: X \to Y \to X$ can be deformed to the identities $\mathrm{id}_Y, \mathrm{id}_X$ via continuous families of intermediate continuous maps $Y \to Y$, $X \to X$, respectively.

Definition 20 (Čech complex) Let C be any set in a metric space M. The *ambient Čech complex* $\check{\mathrm{C}}\mathrm{h}(C,M;\alpha)$ has a simplex on $v_0,\dots,v_k \in C$ if the full intersection of $k+1$ closed balls with the radius α and centers $v_0,\dots,v_k$ has a point from the ambient space M, see [CdSO14, section 4.2.3].

Definition 20 makes sense for any infinite subset C of a metric space M, say for the graph $\theta \subset \mathbb{R}^2$ in Fig. 5. Then $\check{\mathrm{C}}\mathrm{h}(\theta,M;\alpha)$ has the homotopy type of the α-offset $\theta^\alpha \subset M$ by Nerve Lemma 21. Both cycles of θ are born at $\alpha = 0$ and die at $\alpha = 2.577$. Hence $\mathrm{PD}\{\check{\mathrm{C}}\mathrm{h}(\theta,\mathbb{R}^2;\alpha)\}$ has only one off-diagonal dot $(0, 2.577)$ with multiplicity $\mu = 2$, see Fig. 5.

Lemma 21 [Hat01, Corollary 4G.3] Let C be a subspace of a metric space M. Then any α-offset $C^\alpha \subset M$ is homotopy equivalent to the ambient Čech complex $\check{\mathrm{C}}\mathrm{h}(C,M;\alpha)$.

We define the bottleneck distance d_B between persistence diagrams needed for Stability Theorem 23 below.

Definition 22 (bottleneck distance d_B) For points $p = (x_1,y_1)$, $q = (x_2,y_2)$ in $\mathbb{R}^2$, we recall the L_∞-distance $||p-q||_\infty = \max\{|x_1-x_2|,|y_1-y_2|\}$. The *bottleneck distance* between persistence diagrams PD,PD' is defined by $d_B = \inf_{\psi} \sup_{q\in \mathrm{PD}} ||q-\psi(q)||_\infty$ for all bijections $\psi: \mathrm{PD} \to \mathrm{PD}'$.

Stability Theorem 23 below informally says that any small perturbation of original data leads to a small perturbation of the persistence diagram. A metric space M is *totally bounded* if M has a finite ε-sample $C \subset M$ for any $\varepsilon > 0$.

Theorem 23 [CdSO14, simplified Theorem 5.6] Let C be an ε-sample of a graph G in a totally bounded metric space M. Then the 1-dimensional persistence diagrams of Čech filtrations on G and C are ε-close, namely $d_B(\mathrm{PD}\{\check{\mathrm{C}}\mathrm{h}(G,M;\alpha)\},\mathrm{PD}\{\check{\mathrm{C}}\mathrm{h}(C,M;\alpha)\}) \leq \varepsilon$.

The same inequality holds for the filtrations of α-offsets by Nerve Lemma 21, namely $d_B(\mathrm{PD}\{G^\alpha\},\mathrm{PD}\{C^\alpha\}) \leq \varepsilon$.

Appendix B: detailed proof of Optimality Theorem 9

Theorem 9 will follow from Lemma 24 and Propositions 28 and 30, which require Lemmas 25, 26, 27 below. Recall that we fix a filtration $\{Q(C;\alpha)\}$ of complexes on a cloud C, but use the simpler notation $\mathrm{HoPeS}(C)$ for the skeleton.

Lemma 24 Let a class $\gamma \in H_1(Q(C;\alpha))$ be born due to a critical edge e added to $Q(\alpha)$. Then $2\mathrm{birth}(\gamma)$ equals the length $|e|$ relative to the filtration $\{Q(C;\alpha)\}$, see section 3.

Proof By Definition 5 the critical edge $e(\gamma)$ is the last edge added to a cycle $L \subset Q(C;\alpha)$ giving birth to the homology class γ at $\alpha = \mathrm{birth}(\gamma)$. The length $|e|$ equals the doubled scale 2α when e enters $Q(C;\alpha)$, so $|e(\gamma)| = 2\mathrm{birth}(\gamma)$. □

Lemma 25 For any scale $\alpha > 0$, the reduced skeleton $\mathrm{HoPeS}(C;\alpha)$ is a subgraph of the complex $Q(C;\alpha)$.

Proof By Definition 6, for any $\alpha > 0$, the skeleton $\mathrm{HoPeS}(C;\alpha)$ consists of the forest $\mathrm{MST}(C;\alpha)$ and all critical edges with lengths $|e| \leq 2\alpha$. Lemma 4 implies that $\mathrm{MST}(C;\alpha) \subset Q(C;\alpha)$. Any critical edge $e \subset \mathrm{HoPeS}(C)$ belongs to the complex $Q(C;\alpha)$ for $2\alpha = |e|$. Hence $\mathrm{HoPeS}(C;\alpha) \subset Q(C;\alpha)$. □

The inclusion $\mathrm{HoPeS}(C;\alpha) \subset Q(C;\alpha)$ in Lemma 25 induces a homomorphism f_* in H_1. Lemmas 26 and 27 analyze what happens with f_* when a critical edge e is added to (or deleted from) $Q = Q(C;\alpha)$ and $G = \mathrm{HoPeS}(C;\alpha)$.

Lemma 26 (addition of a critical edge) Let an inclusion $f : G \to Q$ of a graph G into a simplicial complex Q induce an isomorphism $f_* : H_1(G) \to H_1(Q)$. Between some vertices $u, v \in G$ let us add an edge e to both G and S that creates a new homology class $\gamma \in H_1(Q \cup e)$. Then f_* extends to an isomorphism $H_1(G \cup e) \to H_1(Q \cup e)$.

Proof Let $L \subset G \cup e$ be a cycle containing the added edge e. Then f extends to the inclusion $G \cup e \to Q \cup e$ and induces the isomorphism $H_1(G \cup e) \cong H_1(G) \oplus \langle [L] \rangle$. Now $f(L)$ is a cycle in $Q \cup e$ and $H_1(Q \cup e) \cong H_1(S) \oplus \langle [f(L)] \rangle$. Mapping $[L]$ to $[f(L)] \in H_1(Q \cup e)$, we extend f_* to a required isomorphism $H_1(G) \oplus \langle [L] \rangle \to H_1(Q) \oplus \langle [f(L)] \rangle$. □

Lemma 27 (deletion of a critical edge) Let an inclusion $f : G \to Q$ of a graph G into a simplicial complex Q induce an isomorphism $f_* : H_1(G) \to H_1(Q)$. Let a homology class $\gamma \in H_1(Q)$ die after adding a triangle T to the complex Q. Let a e be a longest open edge of the triangle T. Then f_* descends to an isomorphism $H_1(G - e) \to H_1(Q \cup T)$.

Proof Adding the triangle T to the complex Q kills the homology class ∂T of the boundary $\partial T \subset G$. Then $H_1(Q \cup T) \cong H_1(Q)/\langle [\partial T] \rangle$. Deleting the open edge e from the boundary $\partial T \subset G$ makes the homology group smaller: $H_1(G - e) \cong H_1(G)/\langle [\partial T] \rangle$. The isomorphism f_* descends to the isomorphism $H_1(G)/\langle [\partial T] \rangle \to H_1(Q)/\langle [\partial T] \rangle$. □

Proposition 28 The inclusion $\mathrm{HoPeS}(C;\alpha) \to Q(C;\alpha)$ from Lemma 25 induces an isomorphism of 1-dimensional homology groups: $H_1(\mathrm{HoPeS}(C;\alpha)) \to H_1(Q(C;\alpha))$.

Proof At the scale $\alpha = 0$, the reduced skeleton $\mathrm{HoPeS}(C;0)$ and the complex $Q(C;0)$ coincide with the cloud C, so their 1-dimensional homology groups are trivial. Each time when a homology class is born or dies in $H_1(Q(C;\alpha))$, the isomorphism in H_1 induced by the inclusion $\mathrm{HoPeS}(C;\alpha) \to Q(C;\alpha)$ is preserved by Lemmas 26 and 27. □

Lemma 29 Let a cycle $L \subset Q(C;\alpha)$ represent a homology class $\gamma \in H_1(Q(C;\alpha))$ in the diagram $\mathrm{PD}\{Q(C;\alpha)\}$. Then any longest edge $e \subset L$ has the length $|e| \geq 2\mathrm{birth}(\gamma)$.

Proof Let a longest edge e of a cycle $L \subset Q(C;\alpha)$ representing the class γ have a half-length $\alpha < \mathrm{birth}(\gamma)$. Then L enters the complex $Q(C;\alpha)$ earlier than $\mathrm{birth}(\gamma)$ and can not represent the class γ that starts living from $\alpha = \mathrm{birth}(\gamma)$. □

Recall that the forest $\mathrm{MST}(C;\alpha)$ on cloud C at a scale α is obtained from a minimum spanning tree $\mathrm{MST}(C)$ by removing all open edges that are longer than 2α.

An edge e is *splitting* a graph G if removing the open edge e makes G disconnected. Otherwise the edge e will be called *non-splitting* and should be in a cycle of the graph G.

Proposition 30 Let a graph $G \subset Q(C;\alpha)$ span $Q(C;\alpha)$ and $H_1(G) \to H_1(Q(C;\alpha))$ be an isomorphism induced by the inclusion. Let (b_i, d_i), $i = 1, \dots, m$, be all dots in the persistence diagram $\mathrm{PD}\{Q(C;\alpha)\}$ such that $\mathrm{birth} \leq \alpha < \mathrm{death}$. Then the total length of G is bounded below by the total length of edges of the forest $\mathrm{MST}(C;\alpha)$ plus $2\sum_{i=1}^{m} b_i$.

Proof Let the subgraph $G_1 \subset G$ consist of all non-splitting edges of G and $e_1 \subset G_1$ be a longest open edge. Removing e_1 from G makes $H_1(G)$ smaller. Hence there is a cycle $L_1 \subset G_1$ containing e_1 and representing a class $\gamma_1 \in H_1(Q(C;\alpha))$ that corresponds to some off-diagonal dot in $\mathrm{PD}\{Q(C;\alpha)\}$, say (b_1, d_1). Then γ_1 lives over $b_1 = \mathrm{birth}(\gamma_1) \leq \alpha < d_1 = \mathrm{death}(\gamma_1)$. Lemma 29 implies that $|e_1| \geq 2b_1$.

Let the graph $G_2 \subset G - e_1$ consist of all non-splitting edges and $e_2 \subset G_2$ be a longest open edge. Similarly, find the corresponding point (b_2, d_2), conclude that $|e_2| \geq 2b_2$ and so on until we get $\sum_{i=1}^{m} |e_i| \geq 2\sum_{i=1}^{m} b_i$. After removing all open edges $e_1, \dots, e_m$, the remaining graph $G - (e_1 \cup \dots \cup e_m)$ still spans the (possibly disconnected) complex $Q(\alpha)$. Indeed, each time we removed a non-splitting edge. So the total length of $G - (e_1 \cup \dots \cup e_m)$ is not smaller than the total length of $\mathrm{MST}(C;\alpha)$ by Lemma 4. □

Proof of Theorem 9. For any $\alpha > 0$, the inclusion $\mathrm{HoPeS}(C;\alpha) \to Q(C;\alpha)$ induces an isomorphism in H_1 by Proposition 28. Let classes $\gamma_1, \dots, \gamma_m$ correspond to all m dots counted with multiplicities in the 'rectangle' $\{\mathrm{birth} \leq \alpha < \mathrm{death}\} \subset \mathrm{PD}\{Q(C;\alpha)\}$. Then $\gamma_1, \dots, \gamma_m$ form a basis of $H_1(Q(C;\alpha)) \cong H_1(\mathrm{HoPeS}(C;\alpha))$ by Definition 2.

The total length of $\mathrm{HoPeS}(C;\alpha)$ equals the total length of $\mathrm{MST}(C;\alpha)$ plus $2\sum_{i=1}^{m} \mathrm{birth}(\gamma_i)$ by Lemma 24. By Proposition 30 this length is minimal over all graphs $G \subset Q(C;\alpha)$ that span $Q(C;\alpha)$ and have the same H_1 as $Q(C;\alpha)$. □

Appendix C: proofs of Theorem 15 and Corollary 16

Reconstruction Theorem 15 will follow from Lemma 31, Propositions 11, 13, 32 and will imply Corollary 16.

Proof of Proposition 11. By Stability Theorem 23 there is a bijection $\psi : \mathrm{PD}\{G^\alpha\} \to \mathrm{PD}\{C^\alpha\}$ such that $q, \psi(q)$ are ε-close in the L_∞ distance on $\mathbb{R}^2$ for all $q \in \mathrm{PD}\{G^\alpha\}$. The ε-neighborhood of a dot $q = (x, y)$ in the L_∞ distance is the square $[x - \varepsilon, x + \varepsilon] \times [y - \varepsilon, y + \varepsilon]$. Under the diagonal projection pr to the vertical death axis, this square maps to the interval $[y-x-2\varepsilon, y-x+2\varepsilon]$. Hence any diagonal gap $\{a<y-x<b\}$ in $\mathrm{PD}\{G^\alpha\}$ can become thinner or wider in $\mathrm{PD}\{C^\alpha\}$ by at most 4ε due to dots q 'jumping' to $\psi(q)$ by at most 2ε relative to the projection $\mathrm{pr}(x,y)=y-x$.

By the given inequality the first k widest gaps $\mathrm{dgap}_i(G)$ in $\mathrm{PD}\{G^\alpha\}$ for $i=1,\dots,k$ are wider by at least 8ε than all other $\mathrm{dgap}_i(G)$ for $i>k$. Hence all dots between any two successive gaps from the first k widest can not 'jump' over these wide gaps and remain 'trapped' between corresponding diagonal gaps in the perturbed diagram $\mathrm{PD}\{C^\alpha\}$.

Despite the order of the first k widest gaps $\mathrm{dgap}_i(G)$, $i=1,\dots,k$, may not be preserved under ψ, the lowest $\{a<y-x<b\}$ of these k diagonal gaps is respected by ψ as follows. The thinner strip $S=\{a+2\varepsilon<y-x<b-2\varepsilon\}$ has no dots from $\mathrm{PD}\{C^\alpha\}$ and has the vertical width $|S|\geq|\mathrm{dgap}_k(G)|-4\varepsilon>|\mathrm{dgap}_{k+1}(G)|+4\varepsilon\geq|\mathrm{dgap}_{k+1}(C)|$.

Then the diagonal strip $S=\{a+2\varepsilon<y-x<b-2\varepsilon\}$ is within the lowest gap of the first k widest gaps $\mathrm{dgap}_i(C)$, $i=1,\dots,k$. Hence all dots above S remain above S under the bijection ψ. By Definition 10 all these dots above S in $\mathrm{PD}\{G^\alpha\}$ and in $\mathrm{PD}\{C^\alpha\}$ form the diagonal subdiagrams $\mathrm{DS}_k(G)$ and $\mathrm{DS}_k(C)$, respectively. So ψ descends to a bijection $\mathrm{DS}_k(G)\to\mathrm{DS}_k(C)$ between finite subdiagrams. □

Proof of Proposition 13. The x-coordinate of any dot $q\in\mathrm{DS}_k(G)$ changes under the given bijection ψ by at most ε. Similarly to the proof of Proposition 11 each vertical gap $\mathrm{vgap}_{k,l}(G)$ becomes thinner or wider by at most 2ε.

By the given inequality the first l gaps $\mathrm{vgap}_{k,j}(G)$ in $\mathrm{PD}_k\{G^\alpha\}$ for $j=1,\dots,l$ are wider by at least 4ε than all other $\mathrm{vgap}_{k,j}(G)$ for $j>l$. Hence all dots between any two successive gaps from the first k widest can not 'jump' over these wide gaps and remain 'trapped' between corresponding vertical gaps in the perturbed diagram $\mathrm{PD}\{C^\alpha\}$.

Despite the order of the first l widest $\mathrm{vgap}_{k,j}(G)$, $j=1,\dots,l$, may not be preserved under ψ, the leftmost $\{a<x<b\}$ of these l vertical gaps is respected by ψ in the following sense. The thinner strip $S=\{a+\varepsilon<x<b-\varepsilon\}$ contains no dots from $\mathrm{DS}_k(C)$ and has the horizontal width $|S|\geq|\mathrm{vgap}_{k,l}(G)|-2\varepsilon>|\mathrm{dgap}_{k,l+1}(G)|+2\varepsilon\geq|\mathrm{dgap}_{k,l+1}(C)|$.

Then the vertical strip $S=\{a+\varepsilon<x<b-\varepsilon\}$ is within the leftmost of the first l widest $\mathrm{vgap}_{k,j}(C)$, $j=1,\dots,l$. Hence all dots to the left of S remain to the left of S under the bijection ψ. By Definition 12 all these dots to the left of S in $\mathrm{PD}\{G^\alpha\}$ and in $\mathrm{PD}\{C^\alpha\}$ form the vertical subdiagrams $\mathrm{VS}_{k,l}(G)$ and $\mathrm{VS}_{k,l}(C)$, respectively. So ψ descends to a bijection $\mathrm{VS}_{k,l}(G)\to\mathrm{VS}_{k,l}(C)$ between finite sets. □

Lemma 31 The derived skeleton $\mathrm{HoPeS}_{k,l}(C)$ is a subgraph of the reduced skeleton $\mathrm{HoPeS}(C;\mathrm{vs}_{k,l}(C))$.

Proof By Definition 6 all edges of the reduced skeleton $\mathrm{HoPeS}(C;\mathrm{vs}_{k,l}(C))$ have a half-length at most $\mathrm{vs}_{k,l}(C)$ and death $>\mathrm{vs}_{k,l}(C)$ for all critical edges. Definition 14 also imposes the extra restriction on critical edges of $\mathrm{HoPeS}_{k,l}(C)$, namely each dot (birth, death) is in the vertical subdiagram $\mathrm{VS}_{k,l}(C)$. So $\mathrm{HoPeS}_{k,l}(C)\subset\mathrm{HoPeS}(C;\mathrm{vs}_{k,l}(C))$. □

Proposition 32 (approximation by reduced HoPeS) Let C be any finite ε-sample of a subspace G in a metric space M. Then the reduced skeleton $\mathrm{HoPeS}(C;\alpha)$ for any scale $\alpha>0$ is contained within the $(\varepsilon+\alpha)$-offset $G^{\varepsilon+\alpha}\subset M$.

Proof Since the cloud C is an ε-sample of G, we get $C\subset G^\varepsilon$. Every edge of $\mathrm{HoPeS}(C;\alpha)$ has a half-length at most α by Definition 6. The edge between any points $p,q\in C$ is covered by the balls with the radius α and the centers p,q. Hence $\mathrm{HoPeS}(C;\alpha)\subset C^\alpha\subset G^{\varepsilon+\alpha}$. □

Proof of Theorem 15. Proposition 11 due to condition (2) implies that there is a bijection $\psi:\mathrm{DS}_k(G)\to\mathrm{DS}_k(C)$ so that $||q-\psi(q)||_\infty\leq\varepsilon$ for all $q\in\mathrm{DS}_k(G)$. Proposition 13 due to condition (4) implies that ψ descends to a bijection $\mathrm{VS}_{k,l}(G)\to\mathrm{VS}_{k,l}(C)$ between vertical subdiagrams.

In general, all cycles in a graph G give birth to corresponding homology classes in $H_1(G^\alpha)$ at the scale $\alpha=0$. These classes may split later at $\alpha>0$, but will eventually die and always give dots $(0,\mathrm{death})\in\mathrm{PD}\{G^\alpha\}$ in the vertical death axis. For any cycle $L\subset G$, let $\mathrm{death}(L)$ be the maximum α such that $H_1(G^\alpha)$ contains the class $[L]$. The graph θ in Fig. 5 has 3 cycles with the same $\mathrm{death}(L)=2.577$.

Let $L_1,\dots,L_m\subset G$ be all m cycles generating $H_1(G)$. Then the 1-dimensional persistence diagram $\mathrm{PD}\{G^\alpha\}$ contains m dots $(0,\mathrm{death}(L_i))$, $i=1,\dots,m$, because each class $[L_i]$ persists over $0\leq\alpha<\mathrm{death}(L_i)$ by Definition 2. Condition (1) implies that all dots $(0,\mathrm{death}(L_i))$ belong to the diagonal subdiagram $\mathrm{DS}_k(G)$, hence to $\mathrm{VS}_{k,l}(C)$.

Condition (3) $\mathrm{vs}_{k,l}(G)=0$ means that the leftmost of the first l widest $\mathrm{vgap}_{k,j}(G)$, $j=1,\dots,l$, is attached to the vertical death axis in $\mathrm{PD}\{G^\alpha\}$, which should contain the vertical subdiagram $\mathrm{VS}_{k,l}(G)$. So $\mathrm{VS}_{k,l}(G)$ consists of the m dots $(0,\mathrm{death}(L_i))$, $i=1,\dots,m$. Then the vertical subdiagram $\mathrm{VS}_{k,l}(C)$ for the cloud C also has exactly m dots, which are 'noisy' images $\psi(0,\mathrm{death}(L_i))$, $i=1,\dots,m$. Moreover, all these dots of $\mathrm{VS}_{k,l}(C)$ are at most ε away from the vertical axis, so the vertical scale $\mathrm{vs}_{k,l}(C)$ is at most ε.

The lowest of the points $\psi(0,\mathrm{death}(L_i))$ has a death $\geq\min\limits_{i=1,\dots,m}\mathrm{death}(L_i)-\varepsilon\geq\mathrm{DS}_k(G)-\varepsilon>|\mathrm{dgap}_k(G)|-\varepsilon>7\varepsilon>\mathrm{vs}_{k,l}(C)$. So the condition death $>\mathrm{vs}_{k,l}(C)$ from Definition 14 is satisfied. Hence the reduced skeleton $\mathrm{HoPeS}_{k,l}(C)$ contains exactly m critical edges corresponding to all m dots of $\mathrm{VS}_{k,l}(C)$ All these m critical edges of $\mathrm{HoPeS}_{k,l}(C)$ generate H_1 of the required rank m. The geometric approximation $\mathrm{HoPeS}_{k,l}(C)\subset G^{2\varepsilon}$ follows from Proposition 32 and Lemma 31 for the vertical scale $\alpha=\mathrm{vs}_{k,l}(C)\leq\varepsilon$. □

Proof of Corollary 16. The condition that the perturbed cloud $\tilde{C}$ is δ-close to the original cloud C, which is ε-closed to the graph G, implies that $\tilde{C}$ is $(\delta+\varepsilon)$-close to G.

Reconstruction Theorem 15 for the ε-sample C and $(\delta+\varepsilon)$-sample $\tilde{C}$ of G says that $\mathrm{HoPeS}_{k,l}(C)$ is 2ε-close to G and $\mathrm{HoPeS}_{k,l}(\tilde{C})$ is $(2\delta+2\varepsilon)$-close to G. Hence $\mathrm{HoPeS}_{k,l}(C)$ and $\mathrm{HoPeS}_{k,l}(\tilde{C})$ are $(2\delta+4\varepsilon)$-close as required. □

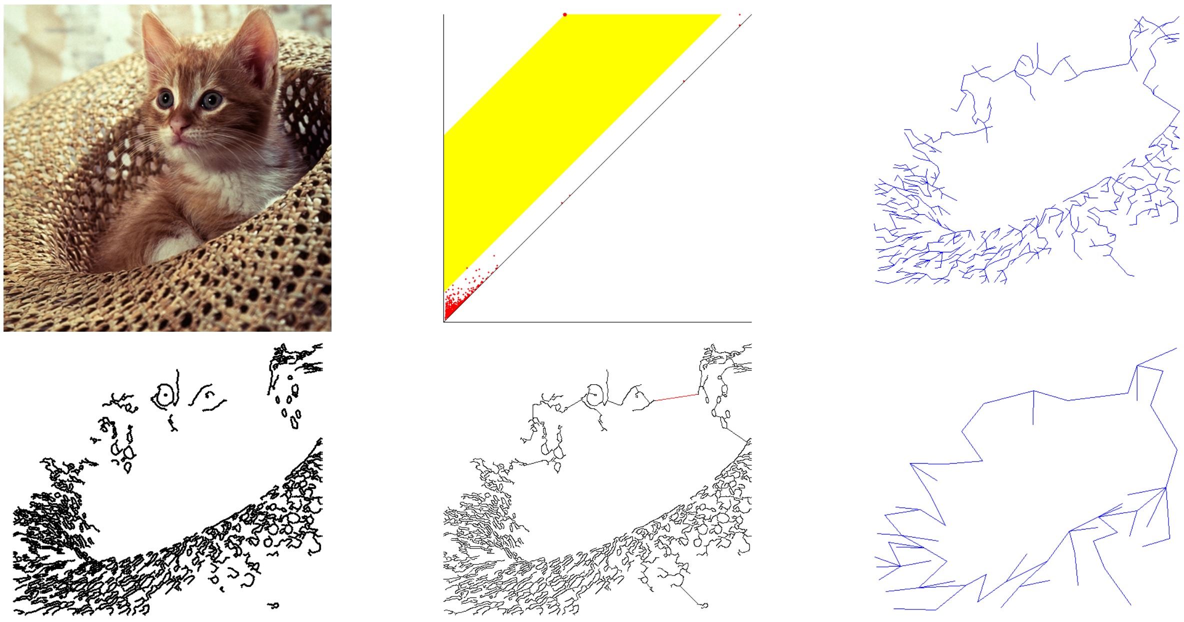

Figure 9: ***Left column***. Top: *cat.png*. Bottom: *cloud C of $n = 14272$ Canny edge points.* ***Middle column***. Top: *persistence diagram* $\mathrm{PD}\{C^{\alpha}\}$ *with the diagonal gap for $k = 1$ and vertical gap for $l = 1$.* Bottom: $\mathrm{HoPeS}_{1,1}(C)$ *with 14272 edges including 1 red critical edge. Despite many noisy cycles the round hat in the image is successfully detected as the most persistent cycle.* ***Right column***. Top: *simplified skeleton* $\mathrm{HoPeS}_{1,1}(C)$ *with 645 edges.* Bottom: *simplified* $\mathrm{HoPeS}_{1,1}(C)$ *with 65 edges.*

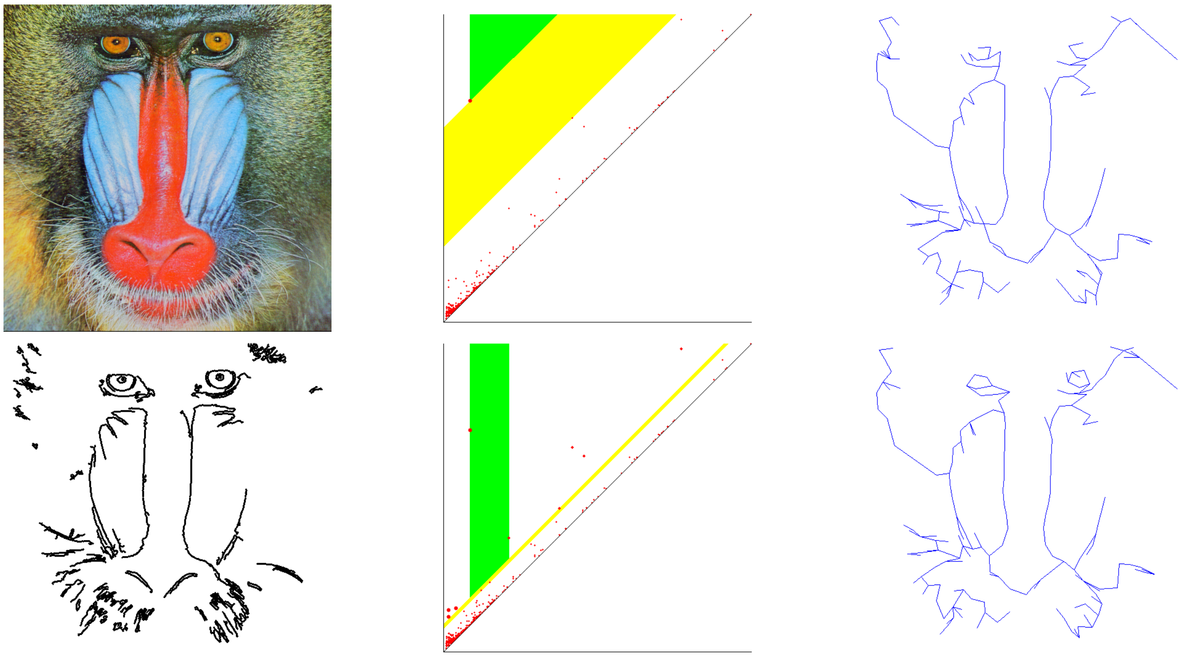

Figure 10: ***Left column***. Top: *mandrill.png*. Bottom: *cloud C of $n = 7490$ Canny edge points.* ***Middle column***. Top: *persistence diagram* $\mathrm{PD}\{C^{\alpha}\}$ *with colored gaps for $k = 1$ and $l = 1$.* Bottom: *persistence diagram with colored gaps for $k = 7$ and $l = 4$.* ***Right column***. Top: *simplified skeleton* $\mathrm{HoPeS}_{1,1}(C)$ *with 177 edges.* Bottom: *simplified* $\mathrm{HoPeS}_{7,4}(C)$ *with 186 edges.*

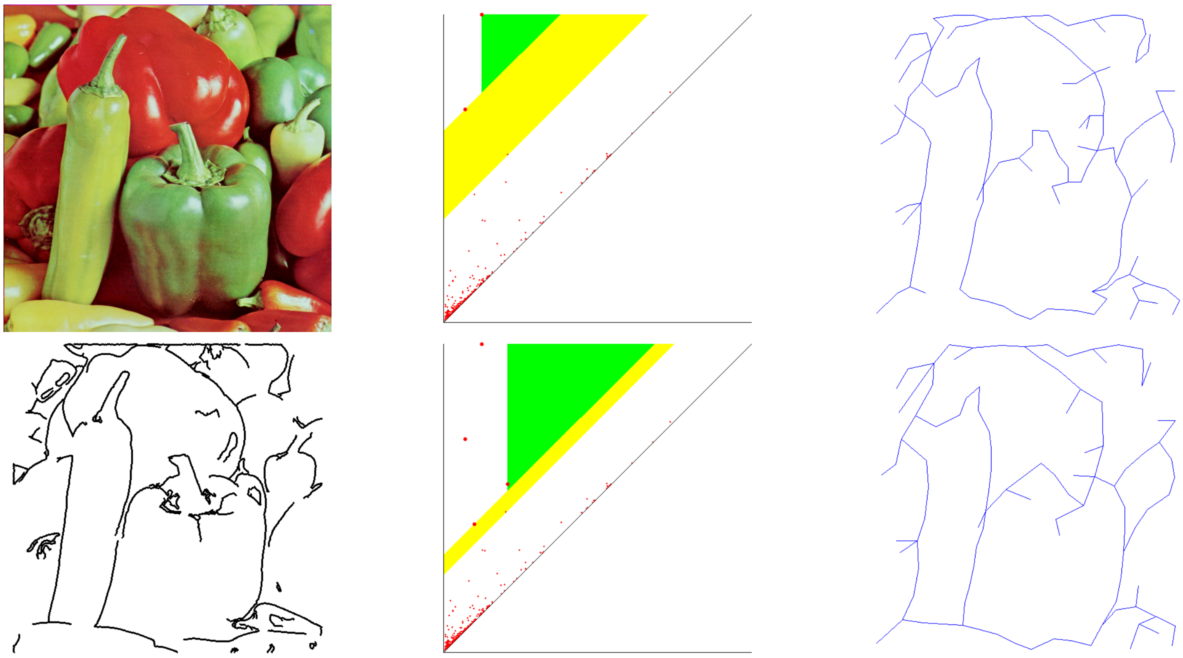

Figure 11: ***Left column***. Top: *peppers.png*. Bottom: *cloud C of $n = 7128$ Canny edge points.* ***Middle column***. Top: *persistence diagram* $\mathrm{PD}\{C^{\alpha}\}$ *with colored gaps for $k = 2$ and $l = 1$ (two largest peppers are detected as two most persistent cycles).* Bottom: *persistence diagram with colored gaps for $k = 3$ and $l = 1$ (four largest peppers are detected as four most persistent cycles).* ***Right column***. Top: *simplified skeleton* $\mathrm{HoPeS}_{2,1}(C)$ *with 141 edges.* Bottom: *simplified* $\mathrm{HoPeS}_{3,1}(C)$ *with 106 edges.*

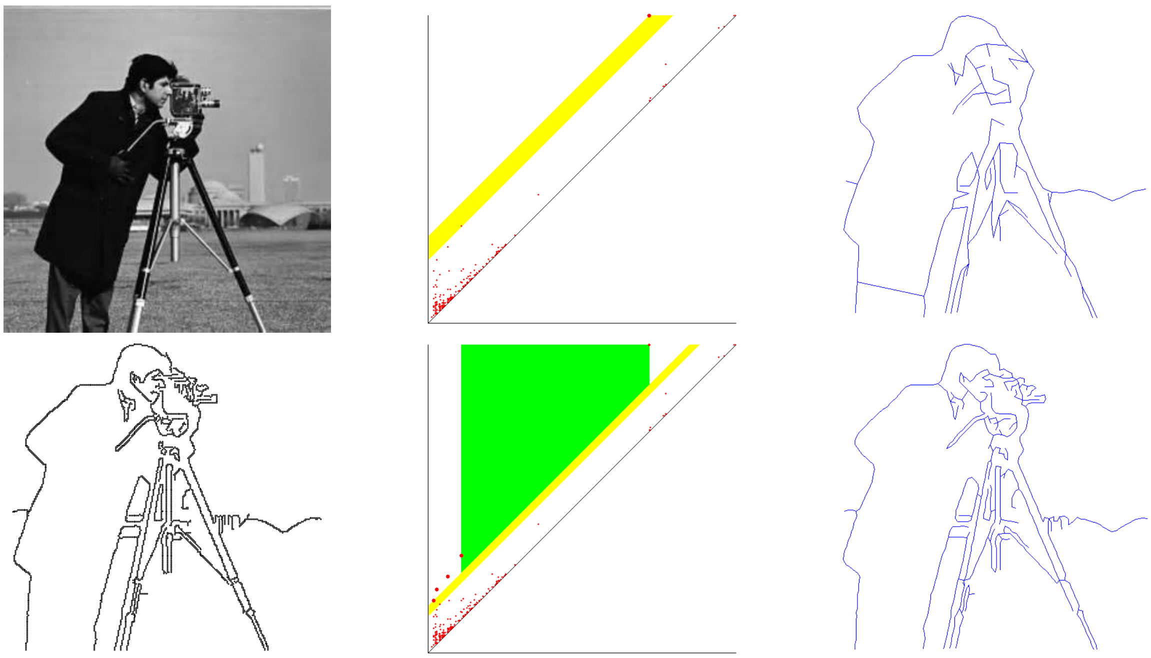

Figure 12: ***Left column***. Top: *cameraman.jpg*. Bottom: *cloud C of $n = 3151$ Canny edge points.* ***Middle column***. Top: *persistence diagram* $\mathrm{PD}\{C^{\alpha}\}$ *with colored gaps for $k = 1$ and $l = 1$.* Bottom: *persistence diagram with colored gaps for $k = 4$ and $l = 2$.* ***Right column***. Top: *simplified skeleton* $\mathrm{HoPeS}_{1,1}(C)$ *with 201 edges.* Bottom: *simplified* $\mathrm{HoPeS}_{4,2}(C)$ *with 678 edges.*