La Matematica (2026) 5:17
https://doi.org/10.1007/s44007-025-00183-9

ORIGINAL RESEARCH ARTICLE

Towards Stratified Space Learning: 2-complexes

Yossi Bokor Bleile[1,2]

Received: 29 May 2023 / Revised: 8 October 2025 / Accepted: 27 October 2025

Abstract
In this paper, we consider a simple class of stratified spaces – 2-complexes. We present an algorithm that learns the abstract structure of an embedded 2-complex from a point cloud sampled from it. We use tools and inspiration from computational geometry, algebraic topology, and topological data analysis and prove the correctness of the identified abstract structure under assumptions on the embedding.

Keywords Stratified space learning · Embedded spaces · Applied topology · Computational geometry

Mathematics Subject Classification 55N31 · 68T09 · 51-08

1 Introduction

Recent developments in technology have led to a dramatic increase in the quantity and complexity of data we can collect. These increases require new methods to enable efficient discovery and modelling of the structures underlying them. As the dimension in which we can observe data increases, it becomes more important to be able to reduce the dimensionality of large amounts of data. Often, it is assumed that there is some subset of features that can be used to obtain a low dimensional embedding that is still *true* to the original topological structure. Sometimes, this low-dimensional representation is actually a manifold, or at least locally so, for example McInnes et al. [1]; or each data point is considered to be a vertex and then graph-based methods, such as Laplacians are used to understand the structure, see Kileel et al. [2].

✉ Yossi Bokor Bleile
yossi.bokorbleile@ist.ac.at

1 Institute of Science and Technology Austria, Klosterneuburg 3400, Austria

2 School of Mathematics and Statistics, University of Sydney, Sydney 2006, Australia

Published online: 08 February 2026

Such methods *choose* how low the low-dimensional representation should be, making assumptions that there is some low-dimensional manifold on which the data lies. This does not always reflect reality. There are also a variety of approaches and algorithms to learn manifold structures from (noisy) samples, see Cheng et al. [3], Dey [4], Dey and Wang [5]. These methods often place assumptions on the manifold and the sampling procedure, generally in the form of restrictions on curvatures, as well as on the density of the sample and the type of noise. The assumptions on curvature are not satisfied by data sets arising in many applications, in particular geospatial data sets arising from person and vehicle movement in transportation networks. There are many data sets that arise from spaces that are not manifolds, but instead are *stratified spaces* – that is, a stratified space is a space described by gluing together (manifold) pieces, called strata. There are no restrictions placed upon each stratum's dimension, and the gluing can give rise to a variety of interesting and complex local structures.

In Bokor et al. [6], the authors removed the assumption that the dimension is constant and presented an algorithm for learning the simplest class of stratified spaces – graphs. We extend their work to the identification of the abstract structure underlying a 2-complex. As observed in Bokor et al. [6], manifold learning can be used to detect and model structures underlying data sets. We make a second step in expanding the set of allowable underlying structures to include linear embeddings of stratified spaces of dimension 2. Removing the linearity assumption is left for future work. Bendich et al. [7] focuses on an algorithm to identify when two points have been sampled from the same stratum of a stratified space, but does not present a method for detecting what the dimension of this piece is, or what the global structure is. Stolz et al. [8] presents an algorithm for detecting samples of two intersecting manifolds, which is a first approximation of splitting a space into stratified pieces, and it comes with experimental verification but no theoretical guarantees. In Aanjaneya et al. [9], the focus is on reconstructing a metric on a graph, with the input consisting of intrinsic distance on the metric graph, the associated theoretical guarantees are about the lengths of the edges in the metric, instead of relating to the geometry of the embedding. In particular, they do not need to consider vertices of degree 2, as in their setting these are points on an edge. Chazal et al. [10] presents a method for sampling and reconstructing compact sets in Euclidean space, with a similar focus on samples with bounded Hausdorff noise and a sufficient density. They guarantee a homotopy equivalence under sufficient conditions but do not present a method for learning the stratified structure. Bendich et al. [11] present an algorithm using persistent homology to assess the local homology of a sample at a particular point, using Delauney triangulations, which comes with a great computational cost.

Instead of assuming the space sampled is a manifold, or only determining if two samples have come from the same stratum, or working with intrinsic distances, or only obtaining the homotopy type, or making local decisions, this paper describes a method for learning the abstract 2-complex structure underlying a point cloud sample P, and provides theoretical guarantees in terms of the geometric embedding that has been sampled. In particular, the algorithm can be used to learn the number of cells of each dimension, and how they piece together. The output of this algorithm can then be used as a starting point to learn the particular embedding the sample came from. Previous work has focussed on using persistent homology to approximate the local homology at a point, which comes with significant computational overheads. We avoid this by approximating the local homology at each point using a fixed approximation scale, related to the geometric conditions on the embedding. The decision process easily works in parallel, which significantly reduces the run time on large data sets. While the method only applies to 2-complexes, many data sets arising from applications are 2-dimensional and it provides a foundation for further developments to increase/remove the dimensionality assumption. We acknowledge that from certain perspectives, moving from graphs to 2-complexes is a small step, yet there are many technical and geometric details involved in guaranteeing the accuracy of the structure learnt even for 2-complexes, and this is the limiting factor for removing the dimensionality assumption at this stage.

This article begins with Sect. 2, containing definitions of the main objects and tools we use throughout the article. After this, Sect. 3 consists of geometric lemmas used in Sect. 4, which considers the local geometry and topology we use to partition the sample P. Finally, Sect. 5 presents the decisions processes we use to recover the abstract structure. Section 5 contains a sequence of lemmas (Lemmas 5.9 to 5.24), which cover cases used in Theorem 5.25, also known as the 'Big Theorem' of this article:

Theorem Let P be an ε-sample of an embedded 2-complex $|X| \subset \mathbb{R}^d$ satisfying Assumption 1. Then, we can reconstruct the incidence graph of X, and recover the abstract structure.

With this result, we know what classes of linearly embedded 2-complexes we can distinguish between with certainty. There are more classes where the method presented is this paper *probably* distinguishes between them, but this is not mathematically guaranteed. We could relax assumptions on the embedding of these complexes, however this results in a more complicated procedure for learning the underlying structure.

2 Definitions and Notations

To be accessible to readers from a diverse set of backgrounds, we begin with some definitions and notations that will be used throughout this article. These will be split into two groups: we begin with definitions focused on complexes and their structures, and then provide some more general definitions that establish some notation and more geometric concepts.

2.1 Complexes

Definition 2.1 (Abstract Complex, Definition 2.4 Carlsson [12]**)** An *abstract simplicial complex* X consists of a pair $(V(X), \Sigma(X))$, with $V(X)$ a finite set, and $\Sigma(X)$ a subset of the power set of $V(X)$, such that for all $\sigma \in \Sigma(X)$ and $\emptyset \neq \tau \subseteq \sigma$, we have $\tau \in \Sigma(X)$. We call $V(X)$ the vertices, and $\Sigma(X)$ the *simplices* of X.

For ease of notation and to avoid confusion later in this paper, we will use the following specialised definition/notation for abstract simplicial complexes with top dimension 2.

Definition 2.2 (Abstract 2-Complex) An *abstract* 2-*complex* X consists of

1. a set $V = V(X)$ of vertices,
2. a set $E = \{\, \sigma \in \Sigma(X) \,|\, \sigma \text{ contains 2 distinct elements}\}$ of edges,
3. a set $T = \{\, \sigma \in \Sigma(X) \,|\, \sigma \text{ contains 3 distinct elements}\}$ of triangles,

and an incidence operator $\mathcal{I}$, which acts as follows: for any pair of cells $\sigma, \tau \in X$

$$\mathcal{I}\,(\sigma, \tau) = \begin{cases} 1 & \text{if } \sigma \subsetneq \tau \\ 0 & \text{otherwise} \end{cases}$$

The specification that the elements of E and T contain 2 and 3 *unique* elements respectively are to ensure there are no self-edges and degenerate triangles in the complex. We restrict ourselves to linear embeddings of 2-complexes X in $\mathbb{R}^d$ for some $n \geq 3$. Below, we spell out the definition of a linear embedding for non-experts.

Definition 2.3 (Linear embedding of 2-complex) Fix $d \geq 3$, then a linear embedding of a 2-complex X in $\mathbb{R}^d$, (X, Θ), consists of an abstract 2-complex X and a map

$$\Theta : X \to \mathbb{R}^d$$

such that

1. on vertices $v \in V$, Θ is injective,
2. on edges $\{u, v\} \in E$, Θ is defined by linear interpolation on $\Theta(u)$ and $\Theta(v)$: $\Theta(\{\, u, v \,\}) = \overline{uv}$ is the line segment between $\Theta(u)$ and $\Theta(v)$,
3. on triangles $\{\, u, v, w \,\} \in E$, Θ is defined by linear interpolation on $\Theta(u)$, $\Theta(v)$ and $\Theta(w)$: $\Theta(\{\, u, v, w \,\}) = \triangle uvw$ is the triangle with vertices $\Theta(u)$, $\Theta(v)$ and $\Theta(w)$, and $\Theta(u)$, $\Theta(v)$, $\Theta(w)$ are not co-linear,
4. for any two cells σ, τ of X, we have $\Theta(\sigma) \cap \Theta(\tau) = \Theta(\sigma \cap \tau)$.

We denote the image of Θ in $\mathbb{R}^d$ by $|X|_\Theta$.
We restrict our attention to embedded 2-complexes $|X|_\Theta$ such that

1. if a vertex v is in the boundary of precisely two edges $\{v, u_1\}$ and $\{v, u_2\}$, then $\angle u_1 v u_2 \neq \pi$,
2. if an edge $\{v_0, v_1\}$ is in the boundary of precisely two triangles $\{v_0, v_1, u_1\}$ and $\{v_0, v_1, u_2\}$, then v_0, v_1, u_1, u_2 are not co-planar.

We often talk about the *boundary* of a cell.

Definition 2.4 (Cell boundary) Let X be an abstract 2-complex, and take a cell $\sigma \in X$. Then the *boundary* of τ, $\partial\tau$, consists of the cells $\sigma \in X$ such that $\mathcal{I}(\sigma,\tau) = 1$.

An important property of a cell $\sigma \in X$, is whether it is *locally maximal* or not.

Definition 2.5 (Locally maximal cell) Let σ be a cell in a 2-complex. We say σ is *locally maximal* if there is no cell $\tau \in X, \tau \neq \sigma$ with $\sigma \subset \tau$. That is, there is no cell τ with σ in the boundary of τ.

Remark 1 Consider two cells σ, τ in a complex *X*, we say σ *is a face of* τ if σ is in the boundary of τ.

We can represent the incidence relationships of cells in *X* in a weighted graph *B*.

Definition 2.6 (Incidence graph) Take an abstract 2-complex *X*. The *incidence graph B* of *X* is the weighted graph with

1. a weight 0 node n_v for each vertex *v* of *X*,
2. a weight 1 node n_e for each edge $e = \{u, v\}$ of *X*,
3. a weight 2 node n_t for each triangle $t = \{u, v, w\}$ of *X*,
4. an edge between a weight 2 node n_t and weight 1 node n_e if $e \subset t$,
5. an edge between a weight 2 node n_t and weight 0 node n_v if $v \in t$,
6. an edge between a weight 1 node n_e and weight 0 node n_v if $v \in e$.

All the edges have weight 1.

Remark 2 The incidence graph we have defined can be thought of as a Hasse diagram with weights on the nodes telling us the dimension of each element.

Abusing notation, we usually write |*X*| instead of $|X|_\Theta$ or (X, Θ), use *v* to denote both the abstract vertex and its embedded location $\Theta(v)$, $\overline{uv}$ to denote both the abstract edge and the embedded image $\Theta(\{\, u, v \,\})$, and $\triangle uvw$ to denote both the abstract triangle and the embedded image $\Theta(\{\, u, v, w \,\})$. Whether we are referring to an element of the abstract 2-complex or its image in $\mathbb{R}^d$ should be clear from the context.

A particular method for constructing a complex from a set of points $P \subset \mathbb{R}^d$ is the *Čech* complex.

Definition 2.7 (Čech Complex) Let $P \subset \mathbb{R}^d$ be a finite set of points. The *Čech complex at scale* δ, $\check{C}_\delta(P)$ is the complex with j-cells $\{v_i\}_{i=0}^{j}$ such that the intersection $\bigcap_{i=0}^{j} B_\delta(v_i)$ is non-empty.

2.2 Geometry

We now focus on some more geometric notions. Given two points $x, y \in \mathbb{R}^d$, $\|x - y\|$ is the standard Euclidean distance between *x* and *y*, for a point $x \in \mathbb{R}^d$ and a set $Y \subset \mathbb{R}^d$, we set

$$d(x, Y) := \inf_{y \in Y} \|x - y\|,$$

and for two sets $X, Y \subset \mathbb{R}^d$, we set

$$d(X, Y) := \min \left\{ \inf_{x \in X} d(x, Y), \inf_{y \in Y} d(y, X) \right\},$$
$$d_H(X, Y) := \max \left\{ \sup_{x \in X} d(x, Y), \sup_{y \in Y} d(y, X) \right\}.$$

We also consider thickenings of a subset *X*: we let

$$X^\alpha := \{p \in \mathbb{R}^d \mid d(p, X) \leq \alpha\}.$$

In proofs towards the end of this article, we use the *weak feature size* of *X* to allow us to construct isomorphism, which was introduced in Chazal and Lieutier [13] as the infimum of the positive critical values of the distance function of *X*.

Definition 2.8 (Weak feature size Chazal and Lieutier [13]**)** Consider a bounded subset $X \subset \mathbb{R}^d$, and let dist_X be the distance to *X* on $\mathbb{R} \setminus X$. The *weak feature size* of *X*, $\mathrm{wfs}(X)$ is the minimum distance between the boundary of *X*, ∂X, and the set of singular values of dist_X.

At various moments in the algorithm, we consider the *diameter* of a set of points *X*. The *diameter of X*, $\mathcal{D}(X)$, is the maximum distance between any pair of points $x, y \in X$:

$$\mathcal{D}(X) := \max_{x,y \in X} \|x - y\|.$$

We use $B_r(p)$ to denote the ball of radius *r* centred at a point $p \in \mathbb{R}^d$, by $\partial B_R(p)$ we mean the boundary of such a ball, and let

$$\mathbb{S}^k = \{x \in \mathbb{R}^d \mid \|x\| = 1\}$$

denote the standard *k*-sphere. We also regularly consider points in a *spherical shell*.

Definition 2.9 Fix $a < b$, and let *y* be a point in $\mathbb{R}^d$. The *spherical shell of radii a and b* centered at *p*, $S_a^b(p)$ is the set

$$\left\{q \in \mathbb{R}^d \mid a \leq \|q - p\| \leq b\right\}.$$

We consider dihedral angles between two half-planes.

Definition 2.10 Let H_1, H_2 be two half-planes with a common boundary line L. Then, the *dihedral angle* α between H_1 and H_2 is the angle formed by two vectors $v_1 \in H_1$ and $v_2 \in H_2$ originating from the same point $x \in L$ such that both v_1 and v_2 are perpendicular to L.

We work with ε-samples P of the embedded 2-complex $|X|$.

Definition 2.11 (**ε-sample**) Let $|X| \subset \mathbb{R}^d$ be an embedded 2-complex. An *ε-sample* P of $|X|$ is a finite subset of $\mathbb{R}^d$ such that $d_H(|X|, P) \leq \varepsilon$.

In Bokor et al. [6] the authors use the threshold graph on a set of points, which we will also use.

Definition 2.12 (Threshold graph, Definition 3.1 Bokor et al. [6]**)** Let $P \subset \mathbb{R}^d$ be a finite collection of points, and fix $r > 0$. The *graph at threshold r on P*, $\mathfrak{G}_r(P)$, is the graph with vertices $p \in P$, and edges (p, q) if $\|p - q\| \leq r$.

Now, we can formalise the aim of this article: given an ε-sample P of some linearly embedded 2-complex $|X|$, we want to recover the abstract structure of the 2-complex X. To do this, we need to learn the number of vertices, the number of edges, and the number of triangles, as well as the incidence relations between them. We achieve this by first deciding for each $p \in P$ if it is near a cell that is not locally maximal, or far away from all cells which are not locally maximal. This partitions P into two subsets which intuitively are P_{NLM} containing samples p near non-locally maximal cells, and P_{LM} containing samples p only near locally maximal cells. Rigorous definitions of P_{NLM} and P_{LM} are in Definition 4.6. Part of this process involves approximating the local homology at each $p \in P$ using a radius r. This requires a choice of scale at which to approximate $|X|$ from P. Unlike in Bokor et al. [6], the relationship between clusters in P_{NLM} and P_{LM} to vertices, edges and triangles is not direct. We can, however, still infer the incidence operator.

Remark 3 In this paper, we use *local homology* in very restrictive settings. It is a very general construction: for a space X, the local homology of X at a point $x \in X$ is the relative homology $H\left(X, X \setminus \{x\}\right)$.

3 Geometric Lemmas

We provide some geometric lemmas as motivation for the definitions of local structures and the geometric assumptions we place on the embeddings of a 2-complex. There are two parts to the definition of the local structure of a point cloud P at a sample p: the first is a topological condition relating to the homology of the samples in a spherical shell around p, and the second relates to the geometry of these samples. The geometric lemmas in this section allow us to distinguish between points near cells that are not locally maximal and those that are only near locally maximal cells when the topological structure of P at p does not, see Sect. 4. The proofs of the lemmas in this section can be found in Appendix A.

We begin with a helpful lemma that bounds the distance between a point in a spherical shell within ε of a ray and the point in the ray in the middle of the shell.

Lemma 3.1 *Let $L \subset \mathbb{R}^d$ be a ray originating at a point z and fix*

$$R \geq 14\varepsilon > 0.$$

Let $P \subset \mathbb{R}^d$ have $d_H(P, L) \leq \varepsilon$ and take $p \in \mathbb{R}^d$ with

$$\|p - z\| \leq \frac{R}{2}.$$

Let x be the point in L with $\|x - p\| = R$. Then for all $q \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$

$$\|q - x\| \leq \left(1 + \sqrt{2}\right) \varepsilon.$$

Next, Lemma 3.2, which motivates part 3 in Definition 4.4. The lemma considers the distances between triples of points in $S^{R+\varepsilon}_{R-\varepsilon}(p) \cap H^{\varepsilon}$ for some point $p \in H^{\varepsilon}$, where H^{ε} is the thickening of a plane H by ε, with $\varepsilon > 0$.

Lemma 3.2 *Consider an affine 2-hyperplane $H \subset \mathbb{R}^d$ and fix*

$$R \geq 14\varepsilon \geq 0.$$

Let $P \subset \mathbb{R}^d$ be such that $d_H(P, H) \leq \varepsilon$, and take p with $d(p, H) \leq \varepsilon$. Then, for all $q_1 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$, there exists $q_2 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$ with

$$\|q_2 - q_1\| \geq 2\sqrt{R^2 - \varepsilon^2} - (2 + \sqrt{2})\varepsilon.$$

Now that we have a geometric property to test if a point p and the samples in $S^{R+\varepsilon}_{R-\varepsilon}(p)$ are from a subset of a plane. We want to understand what conditions need to be placed on points near an edge in two triangles to guarantee this property does not hold. In particular, Lemma 3.3 motivates part 4 of Definition 4.5.

For ease of reading, we let

$$\Psi(\varepsilon, R) = \arccos\left(\frac{(R + 2\varepsilon)^2 + \left(\frac{3R}{2} - \varepsilon\right)^2 - \left(2\sqrt{R^2 - \varepsilon^2} - \left(4 + 2\sqrt{2}\right)\varepsilon\right)^2}{2(R + 2\varepsilon)\left(\frac{3R}{2} - \varepsilon\right)}\right).$$

The following lemma motivates the conditions we place on the dihedral angle between two triangles with a common boundary edge $\overline{uv}$ (of degree 2). This allows us to guarantee that the geometry of the samples in $S_{R-\varepsilon}^{R+\varepsilon}(p)$ for a sample *p near* $\overline{uv}$ is not the same as the geometry of samples in $S_{R-\varepsilon}^{R+\varepsilon}(p)$ when *p* is near a triangle but *far away* from its boundary.

Lemma 3.3 *Consider two affine 2-half-planes $H_1, H_2 \subset \mathbb{R}^d$ whose boundaries are equal, say L, and fix $R \geq 14\varepsilon > 0$. Let α be the dihedral angle between H_1 and H_2. Let P be a set of points such that $d_H(P, H_1 \cup H_2) \leq \varepsilon$. Further, take p such that $d(p, H_1) \leq \varepsilon$. If*

$$d(L, p) \leq \frac{R}{2} - 2\varepsilon$$

and

$$\alpha \in (0, \Psi(\varepsilon, R))$$

then there exist $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ such that for all $q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon.$$

Next, we investigate the geometry of points near a ray and half-plane, to develop a test for points near not locally maximal cells.

There are several local structures that have the same topological structure: they consist of two connected components with no 1-cycles. In Bokor et al. [6], the authors used the angle between the centroids of the two connected components to distinguish between points near a degree 2 vertex and points near the *interior* of an edge. Unfortunately, this test is not sufficient after introducing triangles. If we first check for the presence of triangles, we can again use the inner-product test. To test for the presence of triangles, we examine the diameters of the two connected components.

So, we first bound the diameter of a set of samples only near a line.

Lemma 3.4 *(Diameter of points near ray) Let $L \subset \mathbb{R}^d$ be a ray originating at a point z, and fix $R > 14\varepsilon > 0$. Let $P \subset \mathbb{R}^d$ have $d_H(P, L) \leq \varepsilon$ and take $p \in \mathbb{R}^d$ with $d(L, p) \leq \varepsilon$ and $\|p - z\| \leq \frac{R-\varepsilon}{2}$. Then $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{5\varepsilon}{2}}$ has 1 connected component c, and the diameter is less than $\left(2 + 2\sqrt{2}\right)\varepsilon$.*

The previous lemma bounds the diameter of a connected component containing points with ε of an edge, that are within $S_{R-\varepsilon}^{R+\varepsilon}(p)$ for a sample *p* near a vertex in the boundary of this edge. We need to guarantee that if *p* is near the interior of an edge, it does not fail the diameter test. To ensure this, we obtain the following as a corollary of Lemma 3.4.

Corollary 1 *Let $L \subset \mathbb{R}^d$ be a line, and fix $R > 3\varepsilon > 0$. Let $P \subset \mathbb{R}^d$ have $d_H(P, L) \leq \varepsilon$ and take $p \in \mathbb{R}^d$ with $d(L, p) \leq \varepsilon$ and*

$$\|p - z\| \leq \frac{R - \varepsilon}{2}.$$

Then $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{5\varepsilon}{2}}$ *has 2 connected components* c_1, c_2*, and their diameters are less than* $\left(2 + 2\sqrt{2}\right)\varepsilon$.

Proof First note that $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap L$ consists of two connected components, C_1, C_2, and the distance between them is $R - \varepsilon$. Hence, we can apply Lemma 3.4, to C_1 and C_2 individually, obtaining a connected component for each, say c_1 and c_2. Further, the diameters of c_1 and c_2 are less than $\left(2 + 2\sqrt{2}\right)\varepsilon$.

The following lemma guarantees that if there are samples in $S_{R-\varepsilon}^{R+\varepsilon}(p)$ that are within ε of a triangle, the diameter test fails.

Lemma 3.5 *Let* $L_1, L_2 \subset \mathbb{R}^d$ *be two rays originating at the same point* z *with the angle* α *between in the interval*

$$\left[\frac{\pi}{6}, \pi\right),$$

and fix $R \geq 14\varepsilon > 0$. *Let* T *be the set between* L_1 *and* L_2. *Take* $p \in \mathbb{R}^d$ *with* $d(T, p) \leq \varepsilon$ *and* $\|p - x\| \leq \frac{R-\varepsilon}{2}$, *and* $P \subset \mathbb{R}^d$ *with* $d_H(P, T) \leq \varepsilon$. *Then, there exist points* q_1, q_2 *in* P *with* $\|q_1 - p\|, \|q_2 - p\| \in [R - \varepsilon, R + \varepsilon]$ *such that* $\|q_1 - q_2\| > \left(2 + 2\sqrt{2}\right)\varepsilon$, *and* q_1, q_2 *are path connected. Furthermore, the connected component containing* q_1 *and* q_2 *has diameter bigger than* $\left(2 + 2\sqrt{2}\right)\varepsilon$.

4 Local Structures

To identify the abstract structure of the 2-complex, the method presented in Sect. 5 first partitions the sample P into sets P_{LM}, containing samples that are only near locally maximal cells, and P_{NLM}, containing samples near cells that are not locally maximal. The decision tree for if a point is in P_{NLM} or P_{LM} is summarised in Fig. 1. After this, we further partition P_{LM} and P_{NLM} to infer the number of cells and their dimensions, as well as the incidence operator.

Take an embedded 2-complex $|X| \subset \mathbb{R}^d$, fix (an appropriate) $0 < \varepsilon \leq R$ and take $p \in \mathbb{R}^d$ with $d(|X|, p) \leq \varepsilon$. Consider the topological and geometric structure of $|X|$ in a neighbourhood of p, beginning with $B_R(p) \cap |X|$. If $B_R(p) \cap |X|$ is disconnected, we restrict to the connected component C_p containing $\mathrm{proj}_{|X|}(p)$. Then, we consider $\partial B_R(p) \cap C_p$. Let $\mathrm{proj}_{|X|}(p)$ be the projection of p to $|X|$, and let σ_p be the cell containing $\mathrm{proj}_{|X|}(p)$. If σ_p is locally maximal and $d(|\partial\sigma_p|, p) > R$, then $\partial B_R(p) \cap C_p$ has one of the following structures:

1. $\partial B_R(p) \cap C_p$ is empty, in which case σ_p is a locally maximal vertex,

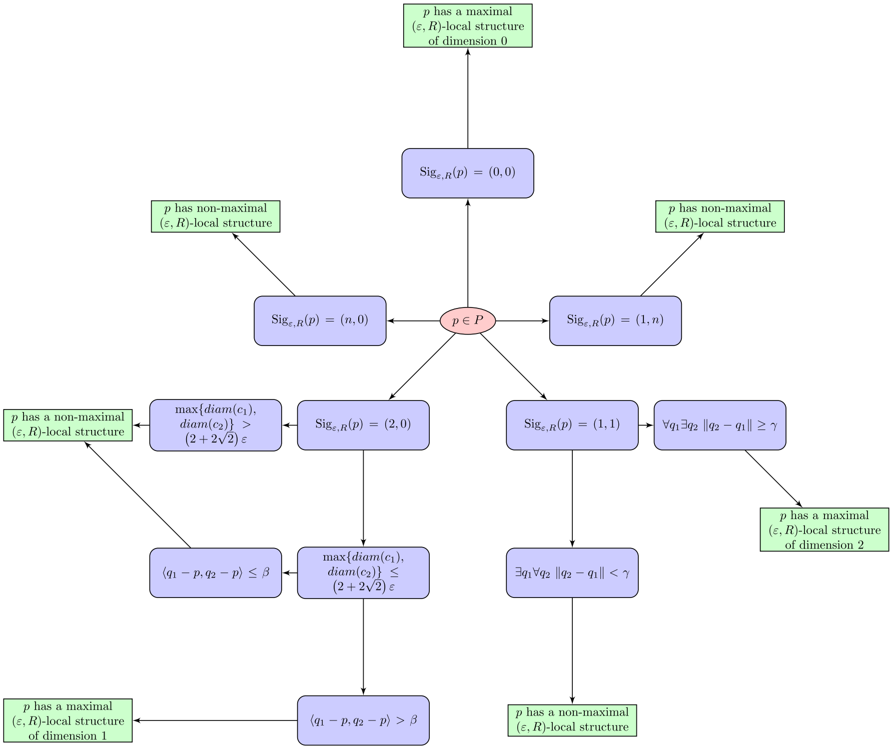


Fig. 1 Flow chart for determining if the (ε, R)-local structure of P at p is maximal or not. If maximal, what the dimension is

2. $\partial B_R(p) \cap C_p$ is a pair of antipodal points, in which case σ_p is a locally maximal 1-cell,
3. $\partial B_R(p) \cap C_p$ is homotopic to $\mathcal{S}^1$ lying in a plane, in which case σ_p is a 2-cell.

The above structures consist of two parts: we examine the topological structure of $\partial B_R(p) \cap C_p$, and then look at its geometry. If p is within R of some cell τ_p (possibly $\tau_p = \sigma_p$) which is not locally maximal, then either the topological structure or the geometric structure is not one of the above cases. As such, we use a two-step process to decide if a given sample p is within R of some not locally maximal cell τ_p: first, we examine the topological structure of $\partial B_R(p) \cap C_p$ by looking at its homology, and then if necessary, we consider its geometric structure. We let

$$\mathcal{H}_{\bullet,R}(p) := H_\bullet\left(\partial B_R(p) \cap C_p\right).$$

As we are restricting ourselves to 2-complexes, we focus on $\mathcal{H}_0(p)$ and $\mathcal{H}_1(p)$.

Definition 4.1 (Local homology signature) Let $|X| \subset \mathbb{R}^d$ be an embedded 2-complex, and fix $R > \varepsilon > 0$. Take a point $p \in \mathbb{R}^d$ with $d(p, |X|) \leq \varepsilon$. The *local homology signature* of $|X|$ at p is

$$\mathrm{Sig}_R(p) := (|\mathcal{H}_{0,R}(p)|, |\mathcal{H}_{1,R}(p)|) .$$

In the above cases, the local homology signature of |*X*| at *p* is as follows.

1. $\mathrm{Sig}_R(p) = (0, 0)$,
2. $\mathrm{Sig}_R(p) = (2, 0)$,
3. $\mathrm{Sig}_R(p) = (1, 1)$.

and so if $\mathrm{Sig}_R(p)$ is not equal to (0, 0), (2, 0) or (1, 1), then *p* is within *R* of a cell τ_p which is not locally maximal. If $\mathrm{Sig}_R(p)$ is (0, 0) then *p* is within ε of a degree 0 vertex. Unfortunately, if $\mathrm{Sig}_R(p)$ is either (2, 0) or (1, 1), we need to examine the geometric structure of $\partial B_R(p) \cap C_p$. When $\mathrm{Sig}_R(p) = (2, 0)$, we can distinguish between the case where σ_p is a locally maximal 1-cell and where σ_p is a vertex of degree 2 as follows: let the two points in $\partial B_R(p) \cap C_p$ be c_1 and c_2. If σ_p is a 1-cell, then $\angle c_1 p c_2 = \pi$, and other $\angle c_1 p c_2 \neq \pi$. When $\mathrm{Sig}_R(p) = (1, 1)$ we need to distinguish between if σ_q is a 2-cell, and if σ_p is in the boundary of 2-cells. We can do so by checking if $\partial B_R(p) \cap C_p$ is contained in a plane: if it is, then σ_p is a 2-cell, if not σ_p is either an edge or a vertex that is not locally maximal.

Recall that we are working with an ε-sample *P* of the embedded 2-complex |*X*| instead of |*X*|. We want to approximate $\mathrm{Sig}_R(p)$ with *P*. As *P* is an ε-sample, we can approximate $\partial B_R(p) \cap C_p$ by first considering the structure of $B_{R+\varepsilon}(p) \cap P$, then the structure of $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$. Before we define the (ε, R)-local structure of *P* at *p* (Definition 4.3), we need the following notation.

Definition 4.2 Let $P \subset \mathbb{R}^d$ be a finite set of points. Then, $\mathrm{rk}_k^{\delta,\gamma}(P)$ is the rank of the map on the k^{th} homology groups induced by the inclusion $P^\delta \hookrightarrow P^\gamma$.

We can now formally define the (ε, R)-local structure of *P* at *p*.

Definition 4.3 (**(ε, R)-local homology signature**) Let $P \subset \mathbb{R}^d$ be an ε-sample of an embedded 2-complex |*X*|, and fix $R \geq 14\varepsilon$. Let $C_p^{\frac{3\varepsilon}{2}}$ be samples in the same connected component of threshold graph $\mathfrak{G}_{3\varepsilon}\left(B_{R+\varepsilon}(p) \cap P\right)$ as *p*. The (ε, R)-*local homology signature* $\mathrm{sig}_{\varepsilon,R}(p)$ of *P* at a sample *p* is

$$\mathrm{Sig}_{\varepsilon,R}(p) := \left(\mathrm{rk}_0^{\frac{5\varepsilon}{2},\frac{9\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{5\varepsilon}{2}}\right), \mathrm{rk}_1^{\frac{5\varepsilon}{2},\frac{9\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{5\varepsilon}{2}}\right)\right).$$

We now define the types of (ε, R)-local structures as we will be working with a ε -sample *P* of some unknown linearly embedded complex |*X*|. We begin with *maximal* (ε, R)-local structures.

Definition 4.4 (**Maximal (ε, R)-local structure**) Let *P* be an ε sample of a linearly embedded 2-complex |*X*| and fix $R \geq 14\varepsilon$. Let $C_p^{\frac{3\varepsilon}{2}}$ be the set of samples in the same connected component of $(B_{R+\varepsilon}(p) \cap P)^{\frac{3\varepsilon}{2}}$ as p. We say the (ε, R)-*local structure of P at p is maximal* if any of the following hold:

1. $\mathrm{Sig}_{\varepsilon,R}(p) = (0,0)$, in which case we say that the (ε, R)-*local structure of P at p is maximal of dimension 0*,
2. $\mathrm{Sig}_{\varepsilon,R}(p) = (2,0)$, and the two connected components c_1, c_2 of $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)^{\frac{3\varepsilon}{2}}$ have diameters less than 5ε and mid-points q_1 and q_2 such that

$$\langle q_1 - p, q_2 - p \rangle \leq -R^2 + 2R\varepsilon + 7\varepsilon^2,$$

   in which case we say that the (ε, R)-*local structure of P at p is maximal of dimension 1*,
3. $\mathrm{Sig}_{\varepsilon,R}(p) = (1,1)$, and for all $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$, $\exists q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ with

$$\|q_2 - q_1\| > 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon.$$

   in which case we say that the (ε, R)-*local structure of P at p is maximal of dimension 2*,

Next, we define *not maximal* (ε, R)-local stuctures.

Definition 4.5 (Not maximal (ε, R)-local structure) Let P be an ε sample of a linearly embedded 2-complex $|X|$ and fix $R \geq 14\varepsilon$. Let $C_p^{\frac{3\varepsilon}{2}}$ be the set of samples in the same connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(S_{R+\varepsilon}(p) \cap P)$ as p. We say that *the* (ε, R)-*local structure of P at* $p \in P$ *is not maximal* if any of the following hold:

1. $\mathrm{Sig}_{\varepsilon,R}(p) = (n,0)$ for some $n \in \mathbb{Z}_{\geq 0}$, $n \neq 0, 2$,
2. $\mathrm{Sig}_{\varepsilon,R}(p) = (1,n)$ for some $n \in \mathbb{Z}_{\geq 0}$, $n \neq 1$,
3. $\mathrm{Sig}_{\varepsilon,R}(p) = (2,0)$ and letting two connected components of

$$\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)^{\frac{3\varepsilon}{2}}$$

   be c_1, c_2, either $\max\{\mathcal{D}(c_1), \mathcal{D}(c_2)\} \leq \left(2 + 2\sqrt{2}\right)\varepsilon$ and letting mid-points of c_1, c_2 be q_1, q_2

$$\langle q_1 - p, q_2 - p \rangle > -R^2 + 2R\varepsilon + 7\varepsilon^2,$$

4. $\mathrm{Sig}_{\varepsilon,R}(p) = (1,1)$ and there exists $q_1 \in P \cap S_{R-\varepsilon}^{R+\varepsilon}$ such that for all $q_2 \in P \cap S_{R-\varepsilon}^{R+\varepsilon}(p)$

$$\|q_2 - q_1\| \leq 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon.$$

Having defined the two classes of (ε, R)-local structures, we can define our initial partition.

Definition 4.6 (P_{LM} **and** P_{NLM}) Let P be an ε-sample of an embedded 2-complex $|X|$. We partition P into two sets P_{LM} and P_{NLM} defined as

$$\begin{aligned} P_{LM} &:= \{p \in P \mid \text{ the } (\varepsilon, R)\text{-local structure at of } P \text{ at } p \text{ is maximal.}\} \\ P_{NLM} &:= \{p \in P \mid \text{ the } (\varepsilon, R)\text{-local structure of } P \text{ at } p \text{ is not maximal.}\} \end{aligned}$$

Remark 4 For all $p \in P$, P either has maximal (ε, R)-local structure at $p \in P$ or it does not. Hence, the partitioning of P into P_{LM} and P_{NLM} defined in Definition 4.6 is disjoint.

Recall that the samples we are working with can contain noise, and we use the homology of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S^{R+\varepsilon}_{R-\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)$ in the definition of (ε, R)-local structure. Hence, we place assumptions on $|X|$ to ensure that we correctly detect when samples are near cells that are not locally maximal. We place assumptions on the distances between any two vertices u and v, the distance between an edge $\overline{uw}$ and a vertex $v \neq u, w$, the angle between any pair of edges with a common boundary vertex. Additionally, we place assumptions on the dihedral angle between any two 2-cells which have common boundary components. So that we can infer the incidence operator, we will require an upper bound on the relationship between R and ε, and so we also restrict out choice of R in terms of ε. We use the following notation in the decicision flow chart (Fig. 1):

$$\begin{aligned} \beta &= -R^2 + 2R\varepsilon + 7\varepsilon^2, \\ \gamma &= 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon. \end{aligned}$$

To increase the readability of this article, we define the following functions.

Definition 4.7 Fix $R > 14\varepsilon > 0$. We define the following functions:

1.

$$\begin{aligned} \Psi_1(\varepsilon, R) &= \arccos\left(\frac{\left(\frac{R}{2} - \varepsilon\right)^2 - 18\varepsilon^2}{\left(\frac{R}{2} - \varepsilon\right)^2}\right) \\ &\geq \arccos\left(\frac{(R-\varepsilon)^2 - 18\varepsilon^2}{(R-\varepsilon)^2}\right) + 2\arcsin\left(\frac{2\varepsilon}{(R-\varepsilon)}\right) \end{aligned}$$

2.
$$\Psi_2(\varepsilon, R) = \pi - \arctan\left(\frac{R+3\varepsilon}{6\varepsilon}\right) + \arcsin\left(\frac{R^2 - 4R\varepsilon - 9\varepsilon^2}{(R+\varepsilon)\sqrt{R^2 + 6R\varepsilon + 34\varepsilon^2}}\right)$$

3.
$$\Psi_3(\varepsilon, R) = \arccos\left(\frac{(R+2\varepsilon)^2 + \left(\frac{3R}{2} - \varepsilon\right)^2 - \left(2\sqrt{R^2 - \varepsilon^2} - \left(4 + 2\sqrt{2}\right)\varepsilon\right)^2}{2(R+2\varepsilon)\left(\frac{3R}{2} - \varepsilon\right)}\right)$$

For those wishing to improve intuition of these functions, they can find graphs (Figs. 12 to14) of these functions in the appendixes. The quantity Ψ_1 ensures that we can distinguish between a degree 2 vertex and the middle of an edge, Ψ_2 ensures that we can distinguish between degree 2 and degree 1 vertices, and Ψ_3 ensures that we can distinguish vertices which are have a local region heomeomorphic to $\mathbb{R}^2$ from the middle of triangles. Note they are effectively a function of $\frac{R}{\varepsilon}$ as they are invariant to scaling both R and ε by the same amount.

We now state the assumptions we place on $|X|$.

Assumption 1 Fix $R \geq 14\varepsilon > 0$. We restrict to embedded 2-complexes $|X| = (X, \pi)$ which satisfy the following.

i For all vertices u, v,
$$\|u - v\| > 6(R + \varepsilon).$$

ii For a vertex v and edge $\overline{uw}$ with $v \neq u, w$,
$$d(\overline{uw}, v) > 6(R + \varepsilon).$$

iii For a vertex v and a triangle $\triangle uwx$ with $v \neq u, w, x$,
$$d(\triangle uwx, v) > 6(R + \varepsilon).$$

iv For an edge $\overline{uv}$ and a triangle $\triangle wxy$ with $v, u \neq w, x, y$,
$$d(\triangle wxy, \overline{uv}) > 6(R + \varepsilon).$$

v For any triangle $\triangle uvw$,
$$\angle uvw, \angle vwu, \angle wuv \geq \frac{\pi}{6}.$$

vi For any pair of edges $\overline{uv}, \overline{xy}$ with no common vertex,
$$d(\overline{uv}, \overline{xy}) > 6(R + \varepsilon).$$

vii For any triangles $\triangle uwv, \triangle xyz$,

$$d(\triangle uwv, \triangle xyz) > 6(R+\varepsilon).$$

viii For any pair of edges $\overline{uv}, \overline{wv}$,

$$\angle uvw \geq \Psi_1(\varepsilon, R).$$

ix For all degree 2 vertices *v* with edges $\overline{uv}, \overline{wv}$ and no triangle $\triangle uvw$,

$$\angle uvw \leq \Psi_2(\varepsilon, R).$$

x For any pair of triangles $\triangle uvw_1, \triangle uvw_2$, the dihedral angle between them is bounded below by $\Psi_1(\varepsilon, R)$.

xi For any pair of triangles $\triangle uvw_1, \triangle uvw_2$, with $\overline{uv}$ of degree 2, the dihedral angle between them is bounded above by $\Psi_2(\varepsilon, R)$.

xii For any triangle $\triangle wvw_2$ and edge $\overline{uv}$ the angle between $\overline{uv}$ and and ray L in $\triangle w_1vw_2$ at *v* is bounded below by $\Psi_1(\varepsilon, R)$ and the radius of the largest circle inscribed by $\triangle uvw$ is at least $2R + 3\varepsilon$.

xiii For any vertex *v* such that

$$|H_0\left(B_R(v) \cap |X|\right)| = 1, \text{ and } |H_1\left(B_R(v) \cap |X|\right)| = 1,$$

the angle between any two rays $L_1, L_2 \in |X|$ at *v* is bounded above $\Psi_3(\varepsilon, R)$.

Remark 5 The reasons behind some of the conditions in Assumption 1 are relatively clear, while others are a bit more obscure. In particular, the roles of Assumption xi and Assumption xii are not immediately clear. Condition 12 allows us to detect the vertex *v* in our algorithms. In particular, it is used in Proposition 4.11 show that we obtain $\mathrm{Sig}_{\varepsilon,R} = (n, \bullet), \;\; n \geq 2$. Assumption xiii allows us to detect which topologically looks similar to an edge of degree 2 or a triangle, and so we place restrictions on the formation of the *cone*, potentially with *fins*, so that we can detect the vertex (Proposition 4.11). This condition is equivalent to bounding the angle at *v* of the convex hull which contains the triangles with vertex *v*. Assumption v is used only to control certain geometric constants in the explicit bounds (see Proof 5). The arguments would remain valid for any fixed lower bound $\alpha > 0$, provided the sampling radius ε is scaled by a factor proportional to $\frac{1}{1-\cos\alpha}$. Hence, while the choice of $\frac{\pi}{6}$ simplifies the analysis, it does not represent a fundamental limitation of the reconstruction result.

The following Propositions provide us with ‘regions’ near locally maximal *i*-cells σ (for $i = 0,\ 1,\ 2$), where we can guarantee that at any sample in this region, the (ε, R)-local structure of *P* at *p* is maximal of dimension *i*.

We begin with the region around a locally maximal vertex.

Proposition 4.8 *Let v be a vertex of* $|X| \subset \mathbb{R}^d$*, which is locally maximal, and let P be an* ε*-sample of* $|X|$*. Then, for all* $p \in P$ *with* $\|p - v\| \leq 4\varepsilon$*, the* (ε, R)*-local structure of P at p is maximal of dimension 0.*

Proof As v is locally maximal, it is not in the boundary of any other cell, and from Assumption i for all vertices $u \neq v$, $\|u - v\| > 6(R + \varepsilon)$, by Assumption ii for all edges $\overline{uw}$ with $v \neq u, w$,

$$d(\overline{uv}, v) > 6(R + \varepsilon),$$

and by Assumption iii for all triangles $\triangle uwx$ with $v \neq u, w, x$,

$$d(\triangle uwx, v) > 6(R + \varepsilon).$$

Hence, any sample $p \in P$ within 4ε of v is within ε of v. Thus, $(B_{R+\varepsilon}(p) \cap P)^{\frac{3\varepsilon}{2}}$ consists of a single connected component, and $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P = \emptyset$.

Thus, $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$, $\mathrm{Sig}_{\varepsilon,R}(p) = (0, 0)$, and the (ε, R)-local structure of P at p is maximal of dimension 0. □

Next, we bound the region near a locally maximal edge.

Proposition 4.9 *Let* $\overline{uv}$ *be an edge of* $|X| \subset \mathbb{R}^d$*, which is locally maximal, and let P be an* ε*-sample of* $|X|$*. Then, for all* $p \in P$ *with* $d(\overline{uv}, p) \leq \varepsilon$*, and* $\|p - u\|, \|p - v\| \geq \frac{3R}{2} + \varepsilon$*, the* (ε, R)*-local structure of P at p is maximal of dimension 1.*

Proof By Assumption ii, for any vertex $w \neq u, v$

$$d(\overline{uv}, w) > 6(R + \varepsilon),$$

Assumption vi for any edge $\overline{wx}$, with $w, x \neq u, v$,

$$d(\overline{uv}, \overline{wx}) > 6(R + \varepsilon),$$

Assumption iv for any triangle $\triangle wxy$, with $w, x, y \neq u, v$,

$$d(\triangle wxy, \overline{uv}) > 6(R + \varepsilon),$$

and so the connected component $C_p^{\frac{3\varepsilon}{2}}$ of $(B_{R+\varepsilon}(p) \cap P)^{\frac{3\varepsilon}{2}}$ which contains p, contains only points $q \in P$ with $d(q, \overline{uv}) \leq \varepsilon$.

Hence, $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)$ consists of two connected components, c_1 and c_2. By Lemma 3.4, the diameters of c_1 and c_2 are less than $(2 + 2\sqrt{2})\varepsilon$. Let x_1 and x_2 be the centroids of c_1 and c_2. Then, applying Lemma 2.1 in Bokor et al. [6],

$$\langle x_1 - p, x_2 - p\rangle \leq -R^2 + 2R\varepsilon + 7\varepsilon^2,$$

so the (ε, R)-local structure of P at p is maximal of dimension 1.

Finally, we bound the region near (locally maximal) triangles.

Proposition 4.10 *Let $\triangle uvw$ be an triangle of $|X| \subset \mathbb{R}^d$, and let P be an ε-sample of $|X|$. Then, for all $p \in P$ with $d(\triangle uvw, p) \leq \varepsilon$, and $d(\partial\triangle uvw, p) \geq \frac{3R}{2} + \varepsilon$, the (ε, R)-local structure of P at p is maximal of dimension 2.*

Proof From Assumption vii, for all triangles $\triangle xyz$, with $x, y, z \neq u, v, w$,

$$d(\triangle uwv, \triangle xyz) > 6(R + \varepsilon),$$

and hence the connected component $C_p^{\frac{3\varepsilon}{2}}$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(B_{R+\varepsilon}(p) \cap P\right)$ containing p, consists only of samples $q \in P$ with $d(q, \triangle uvw) \leq \varepsilon$, as the angle between triangles is bounded below by Assumption x.

First, we need to show that $\mathrm{Sig}_{\varepsilon,R}(p) = (1, 1)$, after which Lemma 3.2 implies that for all $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$, there exists $q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ such that

$$\|q_2 - q_1\| \geq 2\sqrt{R^2 - \varepsilon^2} - (1 + \sqrt{2})\varepsilon.$$

As $d(\partial\triangle uvw, p) > \frac{3R}{2} + \varepsilon$, we have the following inclusions

$$\begin{aligned} S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{5\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{7\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{9\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{11\varepsilon}{2}}. \end{aligned}$$

By the bounds in Assumptions iii and iv on the distances between a triangle and cells not in its boundary, the weak feature size of $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw$ is greater than 5ε, and so the inclusion maps induce isomorphisms

$$H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right) \cong H_\bullet\left(\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{7\varepsilon}{2}}\right) \cong H_\bullet\left(\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{9\varepsilon}{2}}\right).$$

The above homology factors through $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{5\varepsilon}{2}}$ and $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{9\varepsilon}{2}}$ so we have

$$\mathrm{rk}_\bullet^{\frac{5\varepsilon}{2}, \frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right) = \left|H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)\right|,$$

and as

$$|H_0\left(S^{R+\varepsilon}_{R-\varepsilon}(p)\cap\triangle uvw\right)|=1, |H_1\left(S^{R+\varepsilon}_{R-\varepsilon}(p)\cap\triangle uvw\right)|=1,$$

it follows that $\mathrm{Sig}_{\varepsilon,R}(p)=(1,1)$. Now we apply Lemma 3.2 and conclude that the (ε,R)-local structure of P at p is maximal of dimension 2.

Now, we obtain the regions around not locally maximal i-cells σ ($i=0$, 1) in which we can guarantee that the (ε,R)-local structure of P at a sample p in this region is not locally maximal. Again, we begin with non-locally maximal vertices.

Remark 6 As we have restricted our considerations to 2-complexes, every triangle σ is locally maximal; hence, we need only to consider vertices and edges that are not locally maximal.

Proposition 4.11 *Let v be a vertex of $|X|\subset\mathbb{R}^d$, which is not locally maximal, and let P be an ε-sample of $|X|$. Then, for all $p\in P$ with*

$$\|p-v\|\leq\frac{R}{2}-2\varepsilon,$$

the (ε,R)-local structure of P at p is not maximal.

Proof There are several cases we need to consider, which we can classify by the homology of $\partial B_R(v)\cap|X|$:

1. $|H_0\left(\partial B_R(v)\cap|X|\right)|=n$, $|H_1\left(\partial B_R(v)\cap|X|\right)|=0$, $n\neq 2$,
2. $|H_0\left(\partial B_R(v)\cap|X|\right)|=2$, $|H_1\left(\partial B_R(v)\cap|X|\right)|=0$,
3. $|H_0\left(\partial B_R(v)\cap|X|\right)|=1$, $|H_1\left(\partial B_R(v)\cap|X|\right)|=1$,
4. $|H_0\left(\partial B_R(v)\cap|X|\right)|=1$, $|H_1\left(\partial B_R(v)\cap|X|\right)|=n$, $n\geq 2$.

In each of these cases, the following argument holds. Let C_p be the connected component of $B_{R+\varepsilon}(p)\cap|X|$ which contains the projection of p to $|X|$, and let $C_p^{\frac{5\varepsilon}{2}}$ be the connected component of $\left(S^{R+\varepsilon}_{R-\varepsilon}(p)\cap P\right)^{\frac{5\varepsilon}{2}}$. As P is a ε-sample of $|X|$, we have the following inclusions

$$\begin{aligned}S^{R+\varepsilon}_{R-\varepsilon}(p)\cap\triangle uvw &\hookrightarrow \left(S^{R+\varepsilon}_{R-\varepsilon}(p)\cap P\right)^{\frac{5\varepsilon}{2}}\\ &\hookrightarrow \left(S^{R+\varepsilon}_{R-\varepsilon}(p)\cap \triangle uvw\right)^{\frac{7\varepsilon}{2}}\\ &\hookrightarrow \left(S^{R+\varepsilon}_{R-\varepsilon}(p)\cap P\right)^{\frac{9\varepsilon}{2}}\\ &\hookrightarrow \left(S^{R+\varepsilon}_{R-\varepsilon}(p)\cap \triangle uvw\right)^{\frac{9\varepsilon}{2}}.\end{aligned}$$

By the bounds on

- the distance between vertices and cells they do not intersect with (Assumptions ii and iii),
- the distance between vertices Assumption i),

- the angle betwen edges at a common vertex (Assumptions viii and ix),
- the angles between triangles with a common vertex or edge (Assumptions x to xiii),

the weak feature size of $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{5\varepsilon}{2}}$ is greater than 5ε, and we have the following isomorphism on homology induced by the inclusions above

$$H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right) \cong H_\bullet\left(\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right)^{\frac{7\varepsilon}{2}}\right) \cong H_\bullet\left(\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right)^{\frac{9\varepsilon}{2}}\right).$$

The above homology factors through $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{5\varepsilon}{2}}$ and $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{9\varepsilon}{2}}$ so we have

$$\mathrm{rk}_\bullet^{\frac{5\varepsilon}{2},\frac{9\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right) = \left|H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right)\right|.$$

As $\|p-v\| \leq \frac{R}{2} - 2\varepsilon$, we have

$$\left|H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right)\right| = \left|H_\bullet\left(\partial B_R(v) \cap |X|\right)\right|,$$

giving

$$\mathrm{rk}_\bullet^{\frac{5\varepsilon}{2},\frac{9\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right) = \left|H_\bullet\left(\partial B_R(v) \cap |X|\right)\right|.$$

Case 1: $|H_0\left(\partial B_R(v) \cap |X|\right)| = n$, $|H_1\left(\partial B_R(v) \cap |X|\right)| = 0$, $n \neq 2$

By the above, we have $\mathrm{Sig}_R(p) = (n, 0)$, $n \neq 2$, and so the (ε, R)-local structure of *P* at *p* is not maximal.

Case 2: $|H_0\left(\partial B_R(v) \cap |X|\right)| = 2$, $|H_1\left(\partial B_R(v) \cap |X|\right)| = 0$

By the above, we have $\mathrm{Sig}_R(p) = (2, 0)$. Let $C_p^{2\varepsilon}$ be the connected component of $\check{C}_{\frac{5\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)$ containing *p*.

Assume that *v* is a face of some triangle $\triangle uvw$. Then by the bounds placed on angles between edges, and distances between edges without a common face, edges and vertices which are not faces, and vertices and triangles they are not a face of (see Assumption i), and Lemma 3.5 at least one connected component in $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{5\varepsilon}{2}}\right)^{\frac{5\varepsilon}{2}}$ has a diameter greater than $\left(2+2\sqrt{2}\right)\varepsilon$. Thus, the (ε, R)-local structure of *P* at *p* is not maximal.

If *v* is only the face of edges, then by the bounds placed on angles between edges (Assumptions v, viii and ix), and distances between edges without a common face (Assumption vi), edges and vertices which are not faces (Assumption ii), and vertices and triangles they are not a face of (see Assumption iii), both connected components come from two edges *uv* and *wv*, Lemma 2.1, in Bokor et al. [6] and Lemma 3.4 give that the (ε, R)-local structure of *P* at *p* is not maximal.

Case 3: $|H_0\left(\partial B_R(v) \cap |X|\right)| = 1$ $|H_1\left(\partial B_R(v) \cap |X|\right)| = 1$

Again, we have $\mathrm{Sig}_R(p) = (1, 1)$ so there are at least three triangles having *v* as a common vertex. Let p_X be the closest point in |*X*| to *p*, and let $x_1 \in \partial B_R(p) \cap |X|$ be colinear with *v* and p_X, then there is $q_1 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$ with $\|q_1 - x_1\| \leq \varepsilon$.

Now take any $q_2 \in S^{R+\varepsilon}_{R-\varepsilon} \cap P$, and let x_2 be the point in $|X| \cap \partial B_R(p)$ closest to q_2. Using calculations similar to the proof of Lemma 3.1 we obtain

$$\|q_2 - x_2\| \leq \sqrt{2}\varepsilon.$$

Consider the rays L_1, L_2 from *v* through x_1, x_2 respectively, and assume $d(p, L_1) \leq \varepsilon$, see Fig. 2.

We have

$$\begin{aligned}\|x_1 - v\| &= \|x_1 - p_X\| + \|p_X - v\| \leq \frac{3R}{2} - 2\varepsilon,\\ \|x_2 - v\| &\leq R + \varepsilon,\end{aligned}$$

and so

$$\begin{aligned}\|x_2 - x_1\| &= \|x_2 - v\|^2 + \|x_1 - v\|^2 - 2\|x_2 - v\|\|x_1 - v\|^2 \cos \angle x_1 v x_2\\ &\leq \left(\frac{3R}{2} - 2\varepsilon\right)^2 + (R + \varepsilon)^2 - \left(\frac{3R}{2} - 2\varepsilon\right)(R + \varepsilon) \cos x_1 v x_2.\end{aligned}$$

By Assumption xiii the angle between them is bounded above by $\Psi_3(\varepsilon, R)$, so

$$\|x_2 - x_1\| \leq 2\sqrt{R^2 - \varepsilon^2} - (2 + 2\sqrt{2})\varepsilon,$$

and so

$$\|q_2 - q_1\| \leq 2\sqrt{R^2 - \varepsilon^2} - (1 + \sqrt{2})\varepsilon.$$

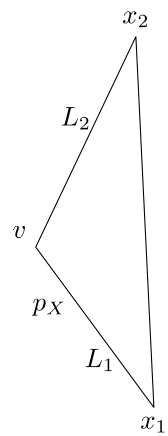


Fig. 2 An illustration of the case where there are two rays L_1, L_2 originating at a point *v* through x_1, x_2 respectively. We have assumed that $d(p, L_1) \leq \varepsilon$, and so the projection p_X to |*X*| is in L_1

Thus, the (ε, R)-local structure of P at p is not maximal.

Case 4: $|H_0\left(\partial B_R(v) \cap |X|\right)| = 1,\ |H_1\left(\partial B_R(v) \cap |X|\right)| = n,\ n \geq 2$

By the argument at the start of this proof, $\mathrm{Sig}_R(p) = (1, n),\ n \geq 2$ and so the (ε, R)-local structure of P at p is not maximal.

Next, we bound the region near edges that are not locally maximal.

Proposition 4.12 *Let $\overline{uv}$ be an edge of $|X| \subset \mathbb{R}^d$, which is not locally maximal, and let P be an ε-sample of |X|. Then, for all $p \in P$ with $d(\overline{uv}, p) \leq \frac{R}{2} - 2\varepsilon$, the (ε, R)-local structure of P at p is not maximal.*

Proof If an edge $\overline{uv}$ is not locally maximal, then there is at least one triangle $\triangle uvw$.

We consider 3 cases:

1. there is a unique triangle $\triangle uvw$ with $\overline{uv}$ in the boundary,
2. there are exactly two triangles $\triangle uvw_1$ and $\triangle uvw_2$ with $\overline{uv}$ in their boundaries,
3. there are three or more triangles $\triangle uvw_1, \triangle uvw_2$ and $\triangle uvw_3$ with $\overline{uv}$ in their boundaries.

Recall that we restrict our attention to the connected components $C_p, C_p^{\frac{5\varepsilon}{2}}$ of $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|$ and $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{5\varepsilon}{2}}$ which contains p.

By the bounds on

- the angle betwen edges at a common vertex (Assumptions viii and ix),
- the distance between edges that do not have a common face (Assumption vi),
- the angles between triangles with a common edge (Assumptions xi and xii),
- the distance between edges and cells they do not intersect with (Assumptions ii, iv and vi),

the weak feature size of C_p is greater than 5ε. Hence by the same argument as at the start of the poof of Proposition 4.11,

$$\mathrm{Sig}_{\varepsilon,R}(p) = \left(\left|H_0\left(\partial B_R(m) \cap |X|\right)\right|,\ \left|H_1\left(\partial B_R(m) \cap |X|\right)\right|\right).$$

Thus, in cases 1 and 3, we get $\mathrm{Sig}_R(p) = (1, 0)$ and $\mathrm{Sig}_R(p) = (1, n)$ for $n \geq 3$ respectively.

In case 2, we get $\mathrm{Sig}_R(p) = (1, 1)$, and so need to check the geometric condition. By Lemma 3.3, there is a $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ such that for all $q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - (1 + \sqrt{2})\varepsilon,$$

and so the (ε, R)-local structure of P at p is not maximal.

Hence, in all 3 cases, the (ε, R) local structure of P at p is not maximal. □

5 2-Complex Decision Process and Correctness

In this section, we present a set of decision processes, which together, recover the structure of X from an ε-sample P of a linear embedding $(X, \Theta) \subset \mathbb{R}^d$. Theorem 5.25 states that given an ε-sample P of a linearly embedded 2 complex $|X| = (X, \Theta_X) \subset \mathbb{R}^d$ satisfying Assumption 1, we can recover the structure of X using this process. There is a sequence of lemmas (Lemmas 5.9 to 5.24), which culminates in the 'big theorem' (Theorem 5.25). The proofs of the lemmas are in Appendix B.

We begin by partitioning P into $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$ and P_{NLM}, such that for each $p \in P_{LM,d}$ the (ε, R)-local structure of P at p is maximal of dimension d, and for each $p \in P_{NLM}$ the (ε, R)-local structure of P at p is not maximal. We then *learn* the number of vertices, the number of edges, the number of triangles and the incidence operator. The decision process we use to detemine which partition of p each point $p \in P$ should be placed in, is visualised in Fig. 3.

Now, given some $p \in P$, let $\mathcal{C}_p$ be the samples $q \in P$ in the connected component containing p in the threshold graph

$$\mathcal{G}_p = \mathfrak{G}_{3\varepsilon}\left(B_{R+\varepsilon}(p) \cap P\right)$$

with $\|q - p\| \in [R - \varepsilon, R + \varepsilon]$. In the definitions of (ε, R)-local structure (Definitions 4.4 and 4.5), we used

$$\mathrm{rk}_{\bullet}^{\frac{5\varepsilon}{2}, \frac{9\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right),$$

which by the Nerve Lemma (Corollary 4 G.3 Hatcher [14]) is equal to the rank, $\mathcal{RK}_{\bullet}$, of the map

$$H_{\bullet}\left(\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \mathcal{C}_p\right)\right) \to H_{\bullet}\left(\check{\mathcal{C}}_{\frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \mathcal{C}_p\right)\right)$$

induced by the inclusion

$$\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right) \hookrightarrow \check{\mathcal{C}}_{\frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right).$$

Hence, when can replace $\mathrm{Sig}_{\varepsilon,R}(p)$ with $(\mathcal{RK}_0, \mathcal{RK}_1)$ in Fig. 1 when considering each point.

Remark 7 We can appeal to the Nerve Lemma, as the balls used in the construction of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \mathcal{C}_p\right)$ and $\check{\mathcal{C}}_{\frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \mathcal{C}_p\right)$ lead us to *good covers* of $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{5\varepsilon}{2}}$ and $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{9\varepsilon}{2}}$ respectively. To see that these covers satisfy the 'every non-empty intersection is contractible' condition required to be a good cover, note that we are using the Čech complex, rather than the Viertoris-Rips complex. Combining this with the linearity of the embedding and the assumptions placed on both ε and R, we have covers that satisfy the Nerve Lemma.

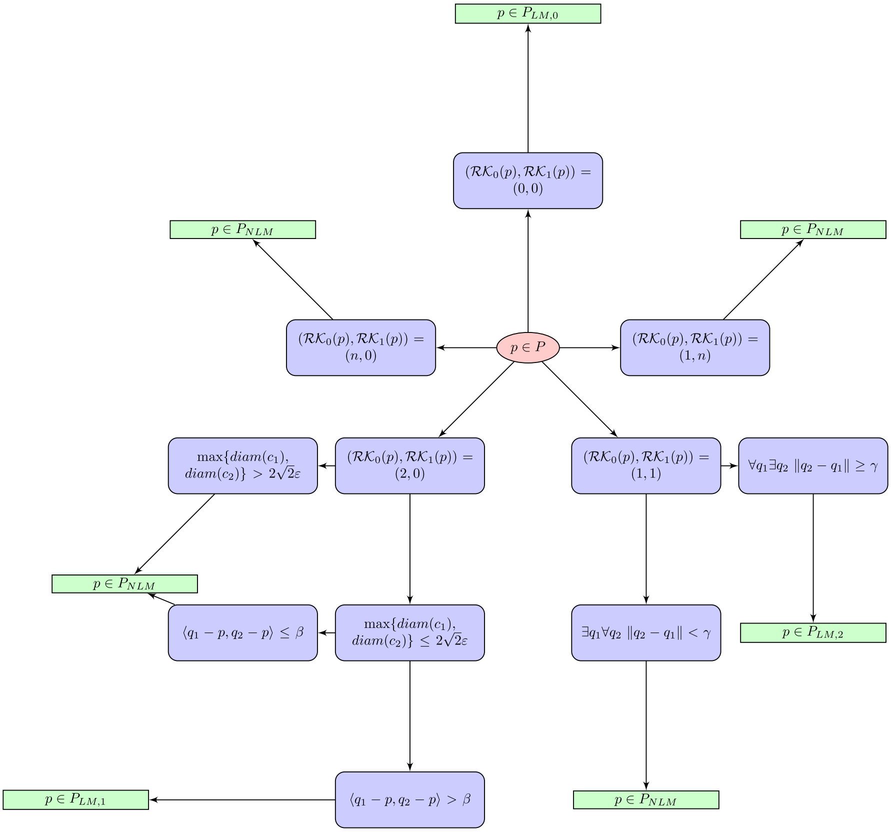


Fig. 3 Flow chart for determining which partition of P a point p is assigned to

Recall that our end goal is to learn the combinatorial structure of X. We begin by learning the number of triangles, locally maximal edges, and locally maximal vertices.

When partitioning P, there is a *grey* region around each locally maximal d-simplex, for $d = 1, 2$, in which a sample p could be in either of $P_{LM,d}$ or P_{NLM}. This presents a problem for learning the combinatorics of X from this partitioning by just looking at connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ and $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$. We overcome this by introducing the notion of a connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ *spanning* an edge, and the notion of a connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ *spanning* a triangle.

Definition 5.1 (Spanning an edge) We say a connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ *spans* a locally maximal edge $\overline{uv}$ if it contains a sample p within ε of the midpoint of $\overline{uv}$.

Definition 5.2 (Spanning a triangle) We say a connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ *spans* a triangle $\triangle uvw$ if it contains a sample *p* within ε of the midpoint of $\triangle uvw$.

We require some geometric conditions on when a connected component spans an edge or a triangle. For an edge, we will use the diameter of the connected component as a condition.

Proposition 5.3 *A connected component C of* $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ *spans a locally maximal edge* $\overline{uv}$ *if and only if* $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$.

Unfortunately, it is not immediately clear that such a test is suitable for detecting components that span triangles. For instance, consider a complex which consists of a single triangle, its three edges, and the three required vertices. While heuristically, it is unlikely to occur, the sampling could lead to 2 connected components $C_1, C_2 \in \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$: one which is far away from the boundary of the triangle, and one that is *surrounded* by points in P_{NLM}, both with large diameters. In fact, the one we wish to say is spanning, say C_1, will have a smaller diameter than the other one, C_2. Note, however, that as C_2 does not contain a sample *p* near the midpoint of $\triangle uvw$, if $\mathcal{D}(C_1) \leq \mathcal{D}(C_2)$, then C_2 contains a non-contractible loop. However, a sample $p \in P$ near the midpoint $m_{\triangle uvw}$ of a triangle $\triangle uvw$ is not near any samples $q \notin P_{LM,2}$, and so we can exploit this fact to obtain a geometric test.

Proposition 5.4 *A connected component C of* $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ *spans a triangle* $\triangle uvw$ *if and only if there is a point* $p \in C$ *such that*

$$B_{\frac{R}{2}+\varepsilon}(p) \cap P \subset P_{LM,2}.$$

We now have geometric conditions for determining if a connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})/\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ spans a triangle/edge respectively. Next, show that the locally maximal vertices of *X* are in bijection with connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,0})$, the locally maximal edges of *X* are in bijection with the spanning connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$, and that the triangles of *X* are in bijection with the spanning connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$.

We begin with the locally maximal vertices.

Proposition 5.5 *The connected components of* $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,0})$ *are in bijection with the set* V_{LM} *of locally maximal vertices of X.*

Next, we show that the edge spanning components are in bijection with the locally maximal edges.

Proposition 5.6 *The spanning components of* $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ *are in bijection with the set* E_{LM} *of locally maximal edges of X.*

Finally, we show that the spanning components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ are in bijection with the triangles of *X*.

Proposition 5.7 *The spanning components of* $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ *are in bijection with the set T of triangles in X.*

These propositions allow us to learn the number of locally maximal vertices, locally maximal edges and triangles from the cardinality of the sets

$$\mathcal{V}_{LM} = \left\{\mathcal{V} \mid \mathcal{V} \text{ is a connected component of } \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,0})\right\},$$
$$\mathcal{E}_{LM} = \left\{\mathcal{E} \mid \mathcal{E} \text{ is a connected component of } \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})\right\},$$
$$\mathcal{T}_{LM} = \left\{\mathcal{T} \mid \mathcal{T} \text{ is a connected component of } \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})\right\}.$$

Having identified the locally maximal cells X_{LM} of *X*, we could learn the combinatorial structure of *X* by identifying the structure of X_{NLM} from P_{NLM}, and combining this with what we know about X_{LM} from P_{LM}. The process in Bokor et al. [6] could be applied, but this requires the existence of some $\widehat{\varepsilon}$ such that P_{NLM} is a $\widehat{\varepsilon}$-sample of X_{NLM} satisfying Assumptions 1 in Bokor et al. [6] This would impose stricter assumptions than Assumption 1, but after ensuring these new assumptions are satisfied, works out of the box.

To avoid placing stricter assumptions on |*X*|, we use the idea of *witness points* to discover the combinatorics. For each sample $p \in P_{NLM}$, we can examine the spanning connected components C_{LM} of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ and $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ such that $C_{LM} \cap B_{R+3\varepsilon}(p) \neq \emptyset$. In particular, we can use $\mathfrak{D}_{\varepsilon,R}(q)$ for some $q \in C_{LM}$, to determine of what dimension the local structure is maximal. If there is a *q* in $C_{LM} \cap S^{R+\varepsilon}_{R-\varepsilon}(p)$ such that $\mathfrak{D}_{\varepsilon,R}(q) = 1$, then *p* is near a vertex.

If there are no connected components C_{LM} which are (ε, R)-locally maximal of dimension 1, then *p* only *witnesses* samples $q \in P_{LM}$ such that the (ε, R)-local structure of *P* at *q* is maximal of dimension 2. Hence, we need to understand the combinatorics of $|X| \setminus (\mathcal{E}_{LM} \cup \mathcal{V}_{LM})$.

In Assumption v, we assumed that for any triangle $\triangle uvw$,

$$\angle uvw, \angle vwu, \angle wuv \geq \frac{\pi}{6}.$$

This means that for any sample $p \in P_{NLM}$ with $d(\partial\triangle uvw, p) < R + \varepsilon$ for some $\triangle uvw$, there is some sample $q \in P_{LM,2}$ with $d(\triangle uvw, q) \leq \varepsilon$ and $d(\partial\triangle uvw, p) \geq R + \varepsilon$, such that $\|q - p\| \leq \frac{2\sqrt{2}(R+2\varepsilon)}{\sqrt{3}-1}$. Further, *q* is in a triangle spanning component $\mathcal{T}$.

Similarly, for any sample $p \in P_{NLM}$ with $d(\partial\overline{uv}, p) < \frac{3R}{2} + \varepsilon$ for some edge $\overline{uv}$, there is a sample $q \in P_{LM,1}$ with $d(\partial\overline{uv}, p) \geq \frac{3R}{2} + \varepsilon$ such that $\|q - p\| \leq \frac{2\sqrt{2}(R+2\varepsilon)}{\sqrt{3}-1}$. Further, q is in an edge spanning component $\mathcal{E}$.

We say a sample $p \in P_{NLM}$ *witnesses* a spanning connected component $\mathcal{C}$ if

$$B_{\frac{2\sqrt{2}(R+2\varepsilon)}{\sqrt{3}-1}}(p) \cap \mathcal{C} \neq \emptyset.$$

For ease of reading, we set $\kappa = \frac{2\sqrt{2}}{\sqrt{3}-1}$. Further, every $p \in P_{NLM}$ witnesses at least one spanning connection component $\mathcal{C}$.

Definition 5.8 (Witnessing a spanning component) Let P be an ε-sample P of an embedded 2-complex $|X|$ satisfying Assumption 1. Then a sample $p \in P_{NLM}$ *witnesses* an edge/triangle spanning component $\mathcal{C}$ if

$$B_{\kappa(R+\varepsilon)}(p) \cap \mathcal{C} \neq \emptyset.$$

To determine the final combinatorial structure of X, we examine a local neighbourhood of each $p \in P_{NLM}$ and look at both

$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1}) \text{ and } B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2}).$$

If

$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1}) \neq \emptyset$$

then we know that p is *near* a vertex, and the spanning components $\mathcal{E}$ of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ that p witnesses, share a boundary vertex. Further, if

$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2}) \neq \emptyset$$

as well, then there are spanning components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ that p witnesses, which have a vertex in common with the edges.

If only

$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2}) \neq \emptyset$$

we examine how many spanning components $\mathcal{T}$ are seen by p, as well as if samples $p \in P_{NLM}$ that witness $\mathcal{T}$, also witness any other spanning components $\mathcal{T}'$. We use this information to partition P_{NLM} into $\{P_i\}$, with a final clean of the partitions to account for some special cases, as follows. For each $p \in P_{NLM}$, consider the following sets:

$$S_E(p) = \left\{\mathcal{E} \text{ spanning edge component} \mid \mathcal{E} \cap B_{(R+2\varepsilon)\kappa}(p) \neq \emptyset\right\},$$
$$S_T(p) = \left\{\mathcal{T} \text{ spanning triangle component} \mid \mathcal{T} \cap B_{(R+2\varepsilon)\kappa}(p) \neq \emptyset\right\}.$$

First, we partition P_{NLM} in $\{P_i\}$ such that for every $p, p' \in P_i$

$$S_E(p) = S_E(p'),$$
$$S_T(p) = S_T(p'),$$

and label each P_i with

$$S_E(P_i) = S_E(p) \text{ for some } p \in P_i,$$
$$S_T(P_i) = S_T(p) \text{ for some } p \in P_i.$$

Further, to assure no double counting, we find any P_i, P_j with $S_E(P_j) \subset S_E(P_i)$ and $S_T(P_j), S_T(P_i)$, and do the following (see Fig. 4): if $S_E(P_j), S_T(P_j) \neq \emptyset$ merge P_j into P_i with labels $S_E(P_i), S_T(P_i)$; if $S_E(P_j) = \emptyset$ and $|S_T(P_j)| \geq 2$ merge P_j into P_i with labels $S_E(P_i), S_T(P_i)$; otherwise leave P_j and P_i as distinct partitions.

Having partitioned P_{NLM} into $\{P_i\}$, we need a method to learn the number of non-locally maximal edges and vertices. Given a partition P_i of P_{NLM}, if there is some edge spanning connected component $\mathcal{E}$ that P_i witnesses, then by a dimensionality argument, P_i corresponds to a vertex or pair of vertices. If P_i witnesses only triangle spanning components, then there are several ppossibilities in terms of the non-locally maximal simplices P_i corresponds to. Depending on the geometry of the simplicial complex, P_i could correspond to a vertex, a vertex and an edge, a vertex and two edges, a vertex and three edges, two vertices and an edge, two vertices and three edges, three vertices and three edges, an edge, two edges, or three edges. To determine which how many vertices and edges P_i corresponds to, we examine its local neighbourhood, labelling it as follows:

V 0 if P_i corresponds to a vertex,
VV −1 if P_i corresponds to 2 vertices,
VE 1 if P_i corresponds to a vertex and an edge,
VEE 4 if P_i corresponds to a vertex and two edges,
VEEE 6 if P_i corresponds to a vertex three edges,
VVE 2 if P_i corresponds to two vertices and an edge,
VVEEE 5 if P_i corresponds to two vertices and three edges,
VVVEEE 7 if P_i corresponds to three vertices and three edges,

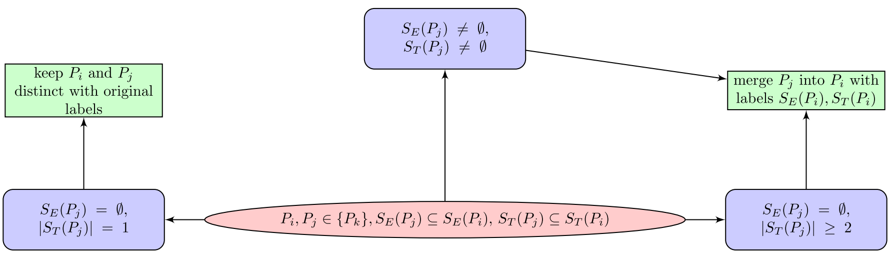


Fig. 4 Flow chart for determining when to partitions P_i, P_j should be merged and which labels to use

E 3 if P_i corresponds to an edge,
EE 9 if P_i corresponds to two edges,
EEE 8 if P_i corresponds to three edges,

Unfortunately, there are several geometric compositions of local neighbourhood where we can't determine which label to give P_i without possessing information about other partitions it is near. Hence, we split $\{P_i\}$ into two sets, P^1, P^2, and first label $P_i \in P^1$ before we label $P_i \in P^2$. We use the decision process in Fig. 5 to obtain P^1 and P^2.

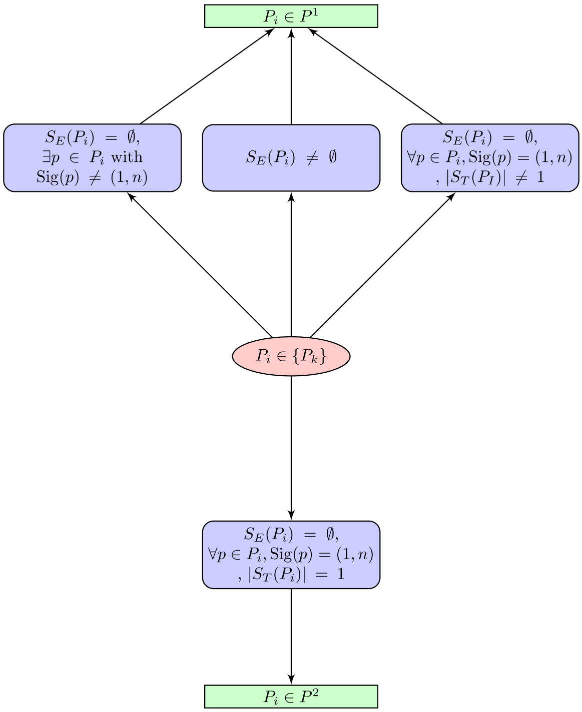
Now we can begin labelling those partitions in P^1, using the flow chart in Fig. 6, if there are any $P_i \in P^1$ which are not labelled after going through P^1, we give them the label *VVE*. Then, we use the flow chart in Fig. 7 to label $P_i \in P^2$.

If, after going through P^1 once, there are $P_i \in P^1$ without a label, we give them the label *VVE*.

The following lemmas together show that the decision processes in Figs. 3 to 7 correctly partition P_{NLM} and label the partitions P_i appropriately. Each Lemma concerns a particular case. We begin with considering simplices around which the structure of X is uncomplicated, and then progress to more intricate configurations. To be precise, we begin by considering a locally maximal edge that is disconnected from the rest of the complex in Lemma 5.9.

Lemma 5.9 *Let $\overline{uv}$ be a locally maximal edge of X, such that u, v are only faces of $\overline{uv}$. Then, there is a unique partition P_1 of P_{NLM} which witnesses $\mathcal{E}$, where $\mathcal{E}$ is the edge spanning component corresponding to $\overline{uv}$. Further, P_1 is assigned label VV.*

Fig. 5 Flow chart for determining when if $P_i \in P^1$ or $P_i \in P^2$



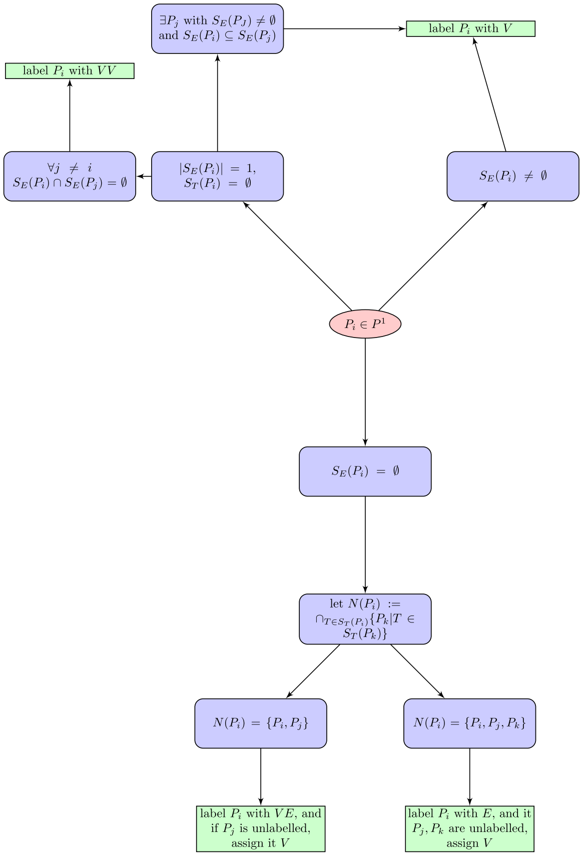


Fig. 6 Flow chart for labelling $P_i \in P^1$

Next, we consider a locally maximal edge that shares at least 1 vertex with some other edge.

Lemma 5.10 *Let $\overline{uv}$ be a locally maximal edge of X, such that u and/or v is the face of some locally maximal cell $\sigma \in X$, $\sigma \neq \overline{uv}$. Then, there are partitions P_1, P_2 of P_{NLM}, which witness $\mathcal{E}$, where $\mathcal{E}$ is the edge spanning component corresponding to $\overline{uv}$. Further, P_1 and P_2 are assigned label V.*

Next, we consider a triangle;$\triangle uvw$ which forms its own connected component of X.

Lemma 5.11 *Let $\triangle uvw$ be a triangle of X, such that for all locally maximal cells $\sigma \in X$ with $\sigma \neq \triangle uvw$, we have*

$$u, v, w \notin \sigma.$$

Then, there is a unique partition P_1 of P_{NLM} which witnesses $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1 is given label

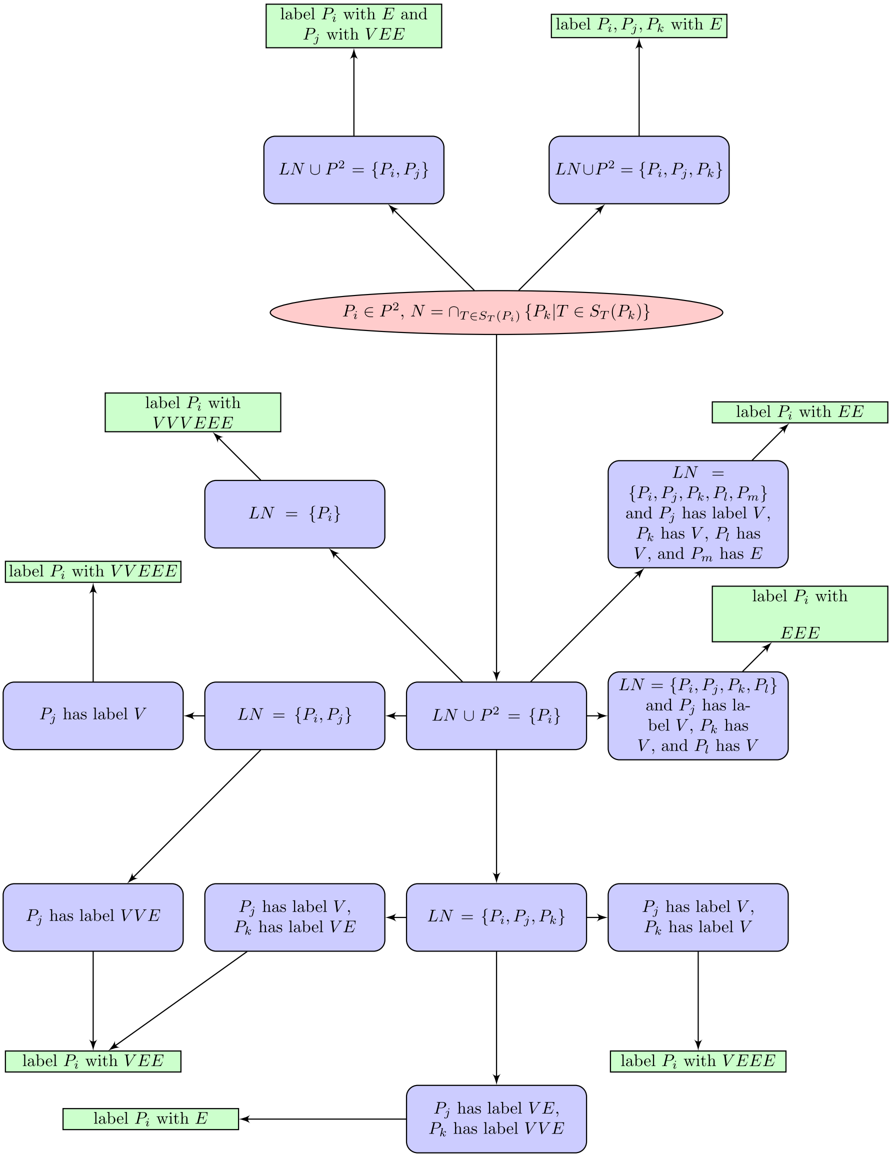


Fig. 7 Flow chart for labelling $P_i \in P^2$

VVVEEE.

Now, we consider a triangle $\triangle uvw$ which shares one vertex with some other locally maximal simplex σ.

Lemma 5.12 *Let $\triangle uvw$ be a triangle of X, such that there is some locally maximal cell $\sigma \in X$ with $\sigma \neq \triangle uvw$, such that $v \in \sigma$, without loss of generality, and for all locally maximal $\tau \in X$, $\tau \neq \sigma, \triangle uvw$, either $\triangle uvw \cap \tau = v$ or $\triangle uvw \cap \tau = \emptyset$.*

Then, there are exactly two partitions P_1, P_2 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1 is given label V and P_2 label VVEEE.

Next, consider the case where $\triangle uvw$ shares an edge with some other triangle $\triangle uvw'$.

Lemma 5.13 *Let $\triangle uvw$ be a triangle of X, such that there is some locally maximal cell $\sigma \in X$ with $\sigma \neq \triangle uvw$, such that $v \in \sigma$, without loss of generality, and for all locally maximal $\tau \in X$, $\tau \neq \sigma, \triangle uvw$, either $\triangle uvw \cap \tau = \overline{uv}$ or $\triangle uvw \cap \tau = \emptyset$.*

Then, there are exactly two partitions P_1, P_2 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1 is given label V and P_2 label VEE.

Already, the geometry is becoming more complicated. We next consider what happens when there are two locally maximal cells σ_1, σ_2 that each share a distinct vertex with $\triangle uvw$.

Lemma 5.14 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1 \neq \sigma_2 \in X$ with $\sigma_1, \sigma_2 \neq \triangle uvw$, such that*

$$\sigma_1 \cap \triangle uvw = v$$
$$\sigma_2 \cap \triangle uvw = u$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = v$,
2. $\tau \cap \triangle uvw = u$,
3. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly three partitions P_1, P_2, P_2 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1, P_2 are given label *V* and P_3 label *VEEE*.

Next, we consider what happens when there are two locally maximal cells σ_1, σ_2, one that shares a vertex and one an edge with $\triangle uvw$, and σ_1 and σ_2 also share a vertex.

Lemma 5.15 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1 \neq \sigma_2 \in X$ with $\sigma_1, \sigma_2 \neq \triangle uvw$, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = v$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = \overline{uv}$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly three partitions P_1, P_2, P_2 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1 has label *V*, P_2 label *VE* and P_3 label *VEE*.

Next, we consider what happens when there are two locally maximal cells σ_1, σ_2, one that shares a vertex and one an edge with $\triangle uvw$, and σ_1 and σ_2 do not share a face.

Lemma 5.16 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1 \neq \sigma_2 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = w$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = \overline{uv}$,
2. $\tau \cap \triangle uvw = w$,
3. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly three partitions P_1, P_2, P_2 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1 has label *V*, P_2 label *VVE* and P_3 label *EE*.

Again, the geometries we are considering become more intricate. We now consider situations where there are (at least) three locally maximal cells $\sigma_1, \sigma_2, \sigma_3$ which share (at least) a face with $\triangle uvw$. We begin by restricting to the situation where each intersection is a distinct vertex.

Lemma 5.17 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = u$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = w$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly four partitions P_1, P_2, P_3, P_4 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1, P_2 and P_3 are labelled with *V* and P_4 with *EEE*.

Next we consider what happens when σ_1 shares an edge with $\triangle uvw$, and σ_2, σ_3 a vertex each.

Lemma 5.18 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = w$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = \overline{uv}$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly four partitions P_1, P_2, P_3, P_4 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1 is labelled with *VE*, P_2, P_3 with *V* and P_4 with *EE*.

And now we consider what happens when σ_1 shares an edge with $\triangle uvw$, and σ_2, σ_3 each share a vertex with $\triangle uvw$ and σ_1.

Lemma 5.19 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = u$$
$$\sigma_3 \cap \triangle uvw = v$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = \overline{uv}$,
2. $\tau \cap \triangle uvw = u$,
3. $\tau \cap \triangle uvw = v$,
4. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly four partitions P_1, P_2, P_3, P_4 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1 is labelled with *E*, P_2, P_3 with *V* and P_4 with *VEE*.

We are slowly exhausting the, thankfully finite, list of geometric scenarios we need to consider. We now consider what happens when σ_1, σ_2 each share an edge with $\triangle uvw$, and σ_3 a vertex.

Lemma 5.20 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = \overline{vw}$$
$$\sigma_3 \cap \triangle uvw = v$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = \overline{uv}$,
2. $\tau \cap \triangle uvw = \overline{vw}$,
3. $\tau \cap \triangle uvw = v$,
4. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly four partitions P_1, P_2, P_3, P_4 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1 is labelled with *V*, P_2, P_3 with *VE* and P_4 with *E*.

In the following lemma, we consider what happens where there are three locally maximal cells $\sigma_1, \sigma_2, \sigma_3$ that share a distinct vertex with $\triangle uvw$ and a fourth locally maximal cell σ_4 which shares an edge.

Lemma 5.21 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = u$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = w$$
$$\sigma_4 \cap \triangle uvw = \overline{uv}$$

and for all other locally maximal cells $\tau \in X$, either

1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \overline{uv}$
5. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly five partitions P_1, P_2, P_3, P_4, P_5 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1, P_2, P_3 are labelled with *V*, P_4 with *E*, and P_5 with *EE*.

The next combination to consider is again four locally maximal cells, where σ_1, σ_2 share a vertex each, and σ_3, σ_4 each share an edge, and one of these edges is between the vertices of σ_1 and σ_2.

Lemma 5.22 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = u$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = \overline{vw}$$
$$\sigma_4 \cap \triangle uvw = \overline{uv}$$

and for all other locally maximal cells $\tau \in X$, either

1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \overline{uv}$
5. $\tau \cap \triangle uvw = \overline{vw}$
6. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly five partitions P_1, P_2, P_3, P_4, P_5 of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1, P_2 are labelled with *V*, P_3 with *VE*, and P_4, P_5 with *E*.

Now, we add one more locally maximal cell to the mix, and consider what happens when $\sigma_1, \sigma_2, \sigma_3$ each share a vertex, and σ_4, σ_4 an edge each.

Lemma 5.23 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\begin{aligned}\sigma_1 \cap \triangle uvw &= u\\ \sigma_2 \cap \triangle uvw &= v\\ \sigma_3 \cap \triangle uvw &= w\\ \sigma_4 \cap \triangle uvw &= \overline{uv}\\ \sigma_5 \cap \triangle uvw &= \overline{vw}\end{aligned}$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \overline{uv}$
5. $\tau \cap \triangle uvw = \overline{vw}$
6. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly five partitions $P_1, P_2, P_3, P_4, P_5, P_6$ of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1, P_2, P_3 are labelled with V, P_4, P_5, P_6 with E.

The final case to consider before we can prove Theorem 5.25, is when there are six locally maximal cells that intersect $\triangle uvw$, with $\sigma_1, \sigma_2, \sigma_3$ each sharing a vertex, and $\sigma_4, \sigma_5, \sigma_6$ an edge each.

Lemma 5.24 *Let $\triangle uvw$ be a triangle of X, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\begin{aligned}\sigma_1 \cap \triangle uvw &= u\\ \sigma_2 \cap \triangle uvw &= v\\ \sigma_3 \cap \triangle uvw &= w\\ \sigma_4 \cap \triangle uvw &= \overline{uv}\\ \sigma_5 \cap \triangle uvw &= \overline{vw}\\ \sigma_6 \cap \triangle uvw &= \overline{uw}\end{aligned}$$

and for all other locally maximal cells $\tau \in X$, either
1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \overline{uv}$,
5. $\tau \cap \triangle uvw = \overline{vw}$,
6. $\tau \cap \triangle uvw = \overline{uw}$,
7. $\tau \cap \triangle uvw = \emptyset$.

Then, there are exactly six partitions $P_1, P_2, P_3, P_4, P_5, P_6$ of P_{NLM} which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, P_1, P_2, P_3 are labelled with *V*, P_4, P_5, P_6 with *E*.

Now that we have covered our bases, we can state and prove the 'Big Theorem' of this paper. It is important to note that this theorem is not about how to sample a given linearly embedded 2-complex |*X*| so that we can recover the abstract structure. It is about demonstrating what classes of samples of linearly embedded 2-complexes we can distinguish between.

Theorem 5.25 Let *P* be an ε-sample of an embedded 2-complex $|X| \subset \mathbb{R}^d$ satisfying Assumption 1. Then, we can reconstruct the incidence graph of *X*, and recover the abstract structure.

Proof We first construct a labelled graph B which we will then complete to be the incidence graph as follows. We start by adding nodes which correspond to the locally maximal simplices.

- for each spanning connected component *T* of $P_{LM,2}$, add a weight 2 node to B labelled with *T*,
- for each spanning connected component *E* of $P_{LM,1}$, add a weight 1 node to B labelled with *E*,
- for each connected component *V* of $P_{LM,0}$, add a weight 2 node to B labelled with *V*,

Next, we need to add some nodes for each partition P_i of P_{NLM}. In fact, we can easily read off the number of nodes and their weights from the labels. If P_i is labelled with
V: add one weight 0 node labelled with P_i,
VE: add one weight 0 node, and one weight 1 node, labelled with P_i,
VEE: add one weight 0 node, and two weight 1 nodes, labelled with P_i,
VEEE: add one weight 0 node, and three weight 1 nodes, labelled with P_i,
VV: add two weight 0 nodes labelled with P_i,
VVE: add two weight 0 nodes and one weight 1 node, labelled with P_i,
VVEEE: add two weight 0 nodes and three weight 1 nodes, labelled with P_i,
VVVEEE: add three weight 0 nodes and three weight 1 nodes, labelled with P_i,
E: one weight 1 node, labelled with P_i,
EE: two weight 1 nodes, labelled with P_i,
EEE: three weight 1 nodes, labelled with P_i.

From Propositions 5.5 to 5.7, we correctly identify the locally maximal components of *X*. So we need to show that we correctly learn the number of not locally maximal cells, and the incidence relationship.

For a locally maximal edge, we need to identify two vertices as its faces. To do so, we must identify which partition(s) of P_{NLM} correspond to these vertices.

Take a spanning edge component $\mathcal{E}$. Then there is some locally maximal edge $\overline{uv}$ corresponding to $\mathcal{E}$. There are two cases to consider:

A: $\overline{uv}$ is disconnected from every other part of X,
B: $\overline{uv}$ is not disconnected every other part of X.Case A: From Assumptions i to iv and vi to xiii and Propositions 4.8 to 4.12, there is a single partition $P_i \subset P_{NLM}$ which contains points p such that $\mathcal{E} \cap B_{(R+\varepsilon)/\kappa+3\varepsilon}(p) \neq \emptyset$. Hence, P_i contains samples p such that either $\|v-p\| \leq \frac{3R}{2}+\varepsilon$ or $\|u-p\| \leq \frac{3R}{2}+\varepsilon$, and P_i corresponds to u and v. In this case, P_i is labelled with *VV*. This occurs only when $\overline{uv}$ is disconnected from the rest of $|X|$; hence, we infer the two boundary vertices.

Case B: As $\overline{uv}$ is not disconnected, there is some locally maximal cell $\sigma \in X$, $\sigma \neq \overline{uv}$ such that either u or v is a vertex of σ. Without loss of generality, let $v \in \sigma$. For the vertices u and v let the set of locally maximal faces they see be $S(u)$ and $S(v)$, respectively. As X is a 2-complex, and $\overline{uv}$ a locally maximal edge, $\sigma \notin S(u)$. Hence, there are two partitions, P_u, P_v, which correspond to the vertices u and v, respectively. In this case, P_u and P_v are labelled with *V*.

We now need to examine how we identify the faces of triangles.

For a triangle spanning component $\mathcal{T}$, let $\mathcal{P}_{\mathcal{T}}$ be the set of partitions P_i of P_{NLM} such that $d(\mathcal{T}, P_i) \leq 3\varepsilon$. There are a few cases we need to consider to ensure we correctly recover the structure of X:

1. $|\mathcal{P}_{\mathcal{T}}| = 1$,
2. $|\mathcal{P}_{\mathcal{T}}| = 2$,
3. $|\mathcal{P}_{\mathcal{T}}| = 3$,
4. $|\mathcal{P}_{\mathcal{T}}| = 4$,
5. $|\mathcal{P}_{\mathcal{T}}| = 5$,
6. $|\mathcal{P}_{\mathcal{T}}| = 6$.

Let the weight 2 node labelled with $\mathcal{E}$ be t.

Case 1 $|\mathcal{P}_{\mathcal{T}}| = 1$: Let P_1 be the single partition in $\mathcal{P}_{\mathcal{T}}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ does not share any faces with another cell. Then, P_1 corresponds to three edges and three vertices and is correctly labelled with *VVVEEE*. Let the corresponding weight 1 nodes of B be e_1, e_2, e_3 and the weight 0 nodes be v_1, v_2, v_3. We add an edge between t and $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

Case 2 $|\mathcal{P}_{\mathcal{T}}| = 2$: Let $\mathcal{P}_{\mathcal{T}} = \{P_1, P_2\}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares either a vertex, or an edge and two vertices with other triangles or locally maximal edges. Thus, either P_1 is labelled with *V* and P_2 with *VVEEE*, or P_1 is labelled with *VVE* and P_2 with *VEE*.

If P_1 has label *V* and P_2 has label *VVEEE*, we find the weight 0 node v_1 with label P_1 and the three weight 1 nodes e_1, e_2, e_3 and two weight 0 nodes v_2, v_3 with label P_2. Then, we add an edge between t and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If P_1 has label *VVE* and P_2 has label *VEE*, we find the weight 1 note e_1 and two weight 0 node v_1, v_2 with label P_1, the two weight 1 nodes e_2, e_3 and one weight 0 nodes v_3 with label P_2. We add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

Case 3 $|\mathcal{P}_\mathcal{T}| = 3$: Let $\mathcal{P}_\mathcal{T} = \{P_1, P_2, P_3\}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares either two vertices, or two vertices and an edge with other triangles or locally maximal edges. Thus, either P_1 and P_2 are labelled with *V* and P_2 with *VEEE*; or P_1 is labelled with *V*, P_2 with *VE* and P_3 with *VEE*; or P_1 is labelled 0, P_2 with *VVE* and P_3 with *EE*.

If P_1, P_2 have label *V* and P_3 has label *VEEE*, we find the weight 0 node v_1 with label P_1, the weight 0 node v_2 with label P_2, the three weight 1 nodes e_1, e_2, e_3 and the weight 0 node v_3 with label P_3. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If P_1 has label *V*, P_2 label *VE* and P_3 label *VEE*, we find the weight 0 node v_1 with label P_1, the weight 0 node v_2 with label P_2, weight 1 node e_1 with label P_2, the weight 0 node v_3 with label P_3, and the two weight 1 nodes e_2, e_3 with label P_3. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If P_1 has label *V*, P_2 label *VVE* and P_3 label 9, we find the weight 0 node v_1 with label P_1, the weight 0 node v_2 and weight 1 node e_1 with label P_2, and the weight 1 nodes e_2, e_3 and weight 0 node v_3 with label P_3. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

Case 4 $|\mathcal{P}_\mathcal{T}| = 4$: Let $\mathcal{P}_\mathcal{T} = \{P_1, P_2, P_3, P_4\}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares three vertices, or three vertices and an edge, or three vertices and two edges with other triangles or locally maximal edges. Thus, either P_1, P_2 and P_3 are labelled with *V* and P_4 with *EEE*; or P_1 is labelled with *VE*, P_2, P_3 with *V* and P_3 with *EE*; or P_1 with *E*, P_2, P_3 with *V* and P_4 with *VEE*; or P_1 is labelled with *V*, P_2, P_3 with *VE*, and P_3 with *E*.

If P_1, P_2, P_3 have label *V* and P_4 has label 8, find the weight 0 node v_1 with label P_1, weight 0 node v_2 with label P_2, weight 0 node v_3 with label P_3, and the three weight 1 nodes e_1, e_2, e_3 with label P_4. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If P_1 has label *VE*, P_2, P_3 have label *V*, and P_4 has label 9, find the weight 0 node v_1 and weight 1 node e_1 with label P_1, weight 0 node v_2 with label P_2, weight 0 node v_3 with label P_3, and the two weight 1 nodes e_2, e_3 with label P_4. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If P_1 has 3, P_2, P_3 label *V* and P_4 label *VEE*;, find the weight 1 node e_1 with label P_1, weight 0 node v_1 with label P_2, weight 0 node v_2 with label P_3, and the two weight 1 nodes e_2 and weight 0 node e_3 with label P_4. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If P_1 has label *V*, P_2, P_3 have label *VE*, and P_4 has label *E*, find the weight 0 node v_1 with label P_1, weight 0 node v_2 and weight 1 node e_1 with label P_2, weight 0 node v_3 and weight 1 node e_3 with label P_3, and the two weight 1 nodes e_2 with label P_4. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

Case 5 $|\mathcal{P}_\mathcal{T}| = 5$: Let $\mathcal{P}_\mathcal{T} = \{P_1, P_2, P_3, P_4, P_5\}$.

This can occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares three vertices and two edges; or three vertcies and one edge with other triangles or locally maximal edges. Thus, P_1, P_2 are labelled with *V*, P_3 with *VE* and P_4, P_5 with *E*; or P_1, P_2, P_3 are labelled with *V*, P_4 with *E* and P_5 with *EE*.

If P_1, P_2 are labelled with *V*, P_3 with *VE* and P_4, P_5 with *E* we find the weight 0 node v_1 with label P_1, find the weight 0 node v_2 with label P_2, find the weight 1 node e_1 and weight 0 node v_3 with label P_3, find the two weight 1 nodes e_2, e_3 with label P_4, and the two weight 1 nodes e_2, e_3 with label P_5. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between e_i with label P_i and v_j with label P_j if $d(P_i, P_j) \leq 3\varepsilon$.

If P_1, P_2, P_3 are labelled with *V*, P_4 with *E* and P_5 with *EE* we find the weight 0 node v_1 with label P_1, find the weight 0 node v_2 with label P_2, find the weight 0 node v_3 with label P_3, find the weight 1 node e_1 with label P_4, and the two weight 1 nodes e_2, e_3 with label P_5. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between e_i with label P_i and v_j with label P_j if $d(P_i, P_j) \leq 3\varepsilon$.

Case 6 $|\mathcal{P}_\mathcal{T}| = 6$: Let $\mathcal{P}_\mathcal{T} = \{P_1, P_2, P_3, P_4, P_5, P_6\}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares three vertices and two edges, or three vertices and three edges with other triangles or locally maximal edges. In either case, P_1, P_2, P_3 are labelled with *V*, P_4, P_5, P_6 with *E*.

So we find the weight 0 node v_1 with label P_1, find the weight 0 node v_2 with label P_2, find the weight 0 node v_3 with label P_3, find the weight 1 node e_1 with label P_4, the weight 1 node e_2 with label P_5, and the weight 1 node e_3 with label P_6. Then add an edge between *t* and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between e_i with label P_i and v_j with label P_j if $d(P_i, P_j) \leq 3\varepsilon$.

In each of these 6 cases, we have connected the weight 2 node *t* corresponding to the cell τ to each weight 1 node *e* corresponding to an edge σ_e of τ, as well as to each weight 0 node *v* corresponding to a vertex σ_v of τ. Further, in the process, we also connect the weight 1 node *e* and weight 0 node *v* if σ_v is a vertex of σ_e.

We have shown that the weight 2 nodes of *B* correspond bijectively to the triangles of *X*, the weight 1 nodes of *B* correspond bijectively to the edges of *X*, and the weight 0 nodes of *B* correspond bijectively to the vertices of *X*. We have also shown that for any pair of nodes n_1, n_2 with corresponding cells σ_1, σ_2, there an edge between them if and only if $\sigma_1 \subset \sigma_2$ or $\sigma_2 \subset \sigma_1$.

Hence, *B* is the incidence graph of *X*.

This thereom tells us which classes of linearly embedded 2-complexes we are guaranteed to distinguish between given an ε-sample of them. The class of embeddgins we *probably* can distinguish between is larger, yet we can not *prove* that we will always distinguish between them.

We have presented the process mostly as decision flowcharts, but they can be converted to pseudocode (see Section D). At this stage, there is no implementation. The *algorithm* we directly obtain from the flowcharts is (highly) non-optimal. The first part is polynomial in the number of points, and the second is polynomial in the number of simplices. As the number of sample points is much larger than the number of simplices, the entire algorithm is polynomial in the size of the ε-sample *P*.

6 Conclusion & Future directions

In this article, we presented a method for learning the abstract structure X underlying an embedded 2-simplicial complex $|X| = (X, \Theta)$ (satisfying Assumption 1) from an ε-sample *P*. Impleteing this process, as well as modelling the embedding, is future work. In particular, to model embeddings that are not linear or where we allow for cells of dimension 2, which are not triangles (along the lines of CW-complexes), we need to develop the process for learning the faces of locally maximal cells further.

There are several natural paths for the work in this article to be extended beyond pure computational implementations. In particular, removing the assumption that the maximal dimension of a cell in the complex is 2 is a direct next step. It is also natural to consider how to modify the algorithm to allow for non-linear embeddings, in particular using semi-algebraic sets, as well as what happens when the noise is not assumed to be Hausdorff. These directions form a sort of 'orthogonal' basis for future research, as they can be thought of as independent problems, but when combined present a rather significant development towards learning stratified spaces.

(A) When p and q are on the same side of x.

(B) When p and q are on different sides of x.

Fig. 8 The two possible geometric configurations of the points

Appendix A Proofs of Geometric Lemmas

Proof 1 (Proof of Lemma 3.1) Consider a point $q \in S_{R-\varepsilon}^{R+\varepsilon}(p)$ with $d(L, q) \leq \varepsilon$. Let q_L be the projection of q to L, and p_L the projection of p to $\overline{L}$ fig. 8.

There are two cases we need to consider,

1. $\|x - p_L\| = \|p_L - q_L\| + \|q_L - x\|$,
2. $\|q_L - p_L\| = \|x - p_L\| + \|q_l - x\|$.

We begin with case 1. We want to bound $\|x - q\|$. Let x' be the point in L with $\|p - x'\| = \|p - q\|$. Note that

$$\begin{aligned}
\|q - x\| &\leq \|q - x'\| + \|x' - x\| \\
\|q - x'\|^2 &= \|q - q_L\|^2 + \|q_L - x'\|^2, \\
\|q_L - x'\| &= \|p_L - x'\| - \|p_L - q_L\|, \\
\|p_L - q_L\|^2 &= \|q - p\|^2 - \left(\|p - p_L\| + \|q - q_L\|\right)^2, \\
\|p_L - x'\|^2 &= \|x' - p\|^2 - \|p_L - p\|^2,
\end{aligned}$$

and we have

$$\begin{aligned}
\|q - x'\|^2 &= \|q - q_L\|^2 + (\|p_L - x'\| - \|p_L - q_L\|)^2 \\
&= \|q - q_L\|^2 + \left(\sqrt{\|x' - p\|^2 - \|p_L - p\|^2} - \sqrt{\|q - p\|^2 - (\|p - p_L\| + \|q - q_L\|)^2}\right)^2 \\
&= \|q - q_L\|^2 + \left(\sqrt{\|x' - p\|^2 - \|p_L - p\|^2} - \sqrt{\|q - p\|^2 - \|p - p_L\|^2 - (\|q - q_L\|^2 + 2\|p - p_L\|\|q - q_L\|)}\right)^2.
\end{aligned}$$

Recall that $\|p - x'\| = \|p - q\|$, and so

$$\begin{aligned}
\|q - x'\|^2 &= \|q - q_L\|^2 \\
&\quad + \left(\sqrt{\|q - p\|^2 - \|p_L - p\|^2} - \sqrt{\|q - p\|^2 - \|p - p_L\|^2 - (\|q - q_L\|^2 + 2\|p - p_L\|\|q - q_L\|)}\right)^2.
\end{aligned}$$

Further, by assumption

$$
\begin{aligned}
\|q - q_L\| &\leq \varepsilon, \\
\|p - p_L\| &\leq \frac{R}{2}, \\
\|q - p\| &\geq R - \varepsilon, \\
14\varepsilon &\leq R
\end{aligned}
$$

so

$$\left(\sqrt{\|q-p\|^2 - \|p_L - p\|^2} - \sqrt{\|q-p\|^2 - \|p-p_L\|^2 - (\|q-q_L\|^2 + 2\|p-p_L\|\|q-q_L\|)}\right)^2 \leq \varepsilon$$

and

$$\|q - x'\| \leq \varepsilon^2 + \varepsilon^2.$$

Finally, as x, x' are both in L, and

$$
\begin{aligned}
\|p - x\| &= R, \\
\|p - x\| &\geq \|p - x'\|, \\
\|p - x'\| &\geq R - \varepsilon,
\end{aligned}
$$

we have $\|x - x'\| \leq \varepsilon$, which when combined with the above, gives

$$
\begin{aligned}
\|q - x\| &\leq \|q - x'\| + \|x' - x\| \\
&\leq \sqrt{2}\varepsilon + \varepsilon = (1 + \sqrt{2})\varepsilon.
\end{aligned}
$$

Similar calculations in the other cases give bounds that are equal or smaller, so

$$\|q - x\| \leq (1 + \sqrt{2})\varepsilon.$$

□

Proof 2 (Proof of Lemma 3.2) First, let p_H be the projection of p to H, and note that $\|p_H - p\| \leq \varepsilon$. Take $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$. Let x_1 be the point in $\partial B_R(p) \cap H$ closest to q_1, and q_H the projection of q_1 to H. Note that p_H, q_H, x_1 are co-linear, lying on the ray L from p_H, and $\|q_1 - q_H\| \leq \varepsilon$. By Lemma 3.1, $\|q_1 - x_1\| \leq (1 + \sqrt{2})\varepsilon$.

As $H \cap \partial B_R(p)$ is a circle with radius $\sqrt{R^2 - \|p_H - p\|^2}$, there is a point $x_2 \in H \cap \partial B_R(p)$ such that $\|x_2 - x_1\| = 2\sqrt{R^2 - \|p_H - p\|^2}$. As $d_H(p, H) \leq \varepsilon$, we have

$$\|x_2 - x_1\| \geq 2\sqrt{R^2 - \varepsilon^2},$$

and as $d_H(P, H) \leq \varepsilon$, there is $q_1 \in P$ with $\|q_1 - x_1\| \leq \varepsilon$. Hence

$$\|q_2 - q_1\| \geq 2\sqrt{R^2 - \varepsilon^2} - (2 + \sqrt{2})\varepsilon.$$

□

Proof 3 (Proof of Lemma 3.3) First, let H_1' be the half plane containing H_1 with bounding line L' such that $D(L, L') = \varepsilon$, p_H be the projection of p onto H_1' and p_L the projection of p to L. Then take $x_1 \in H_1$ such that $\|p - x_1\| = R$ and p_H, p_L and x_1 are co-linear. Take $q_1 \in P$ with $\|q_1 - x_1\| \leq \varepsilon$, so $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$.

Let q_2 be a point in $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$. There are two cases to consider: $d(q_2, H_1') \leq \varepsilon$ and $d(q_2, H_2) \leq \varepsilon$.

If $d(q_2, H_1') \leq \varepsilon$, take $x_2 \in \partial B_R(p) \cap H_1'$ such that x_2, p_H and the projection of q_2 to H_1' are co-linear. Then by Lemma 3.1 $\|q_2 - x_2\| \leq \sqrt{2}\varepsilon$ Fig. 9.

Consider the triangle formed by x_1, p_h, x_2, and let $\widehat{R} = \sqrt{R^2 - \|p_H - p\|^2}$. By assumption,

$$\begin{aligned} \|p_L - p_H\| &< \frac{R}{2} < R - 7\varepsilon, \\ \|x_2 - p_H\| = \|x_1 - p_H\| &= \widehat{R} \\ \|\widetilde{x} - p_L\| &< \varepsilon. \end{aligned}$$

Further,

$$\begin{aligned} \|x_2 - x_1\|^2 &= \|x_1 - p_H\|^2 + \|x_2 - p_H\|^2 - 2\|x_1 - p_H\|\|x_2 - p_h\| \cos \angle x_2 p_h x_1 \\ &= \|x_1 - p_H\|^2 + \|x_2 - p_H\|^2 + 2\|x_1 - p_H\|\|x_2 - p_h\| \cos \angle x_2 p_h \widetilde{x} \\ &= 2\widehat{R}^2 + 2\widehat{R}^2 \left(\frac{\|\widetilde{x} - p_H\|}{\widehat{R}} \right) \end{aligned}$$

Now, as $R \geq 14\varepsilon$, we have

$$\begin{aligned} 2\widehat{R}^2 + 2\widehat{R}\|p_H - \widetilde{x}\| &< 4\widehat{R}^2 - 4\widehat{R}\left(4 + 2\sqrt{2}\right)\varepsilon + \left(4 + 2\sqrt{2}\right)^2 \varepsilon^2 \\ &= \left(2\widehat{R} - \left(4 + 2\sqrt{2}\right)\varepsilon\right)^2 \end{aligned}$$

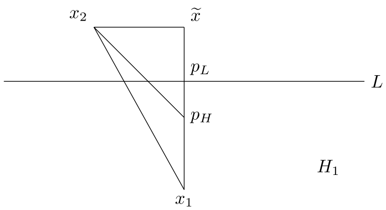


Fig. 9 Understanding the behaviour of points near the common boundary of two half-planes

so

$$\|x_2 - x_1\| < 2\widehat{R} - (4 + 2\sqrt{2})\varepsilon.$$

which implies

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - (2 + \sqrt{2})\varepsilon.$$

Fig. 10.

Now assume $d(q_2, H_2) \leq \varepsilon$. Let H_2' be the half-plane which contains H_2 and has boundary L' with $d(L, L') = \varepsilon$. As $d(q_2, H_2) \leq \varepsilon$, then there is $x_2 \in \partial B_R(p) \cap H_2'$ with $\|q_2 - x_2\| \leq \sqrt{2}\varepsilon$. Hence,

$$\|x_1 - x_2\| \geq 2\sqrt{R^2 - \varepsilon^2} - (2 + 2\sqrt{2})\varepsilon.$$

If $x_2 \in H_2' \backslash H_2$, then by a similar argument to above,

$$\|x_1 - x_2\| \geq 2\sqrt{R^2 - \varepsilon^2} - (2 + 2\sqrt{2})\varepsilon.$$

If $x_2 \in H_2 \subsetneq H_2'$, by the cosine rule we have

$$\|x_2 - x_1\|^2 = \|x_2 - p_L\|^2 + \|x_1 - p_L\|^2 - 2\|x_2 - p_L\|\|x_1 - p_L\| \cos \angle x_1 p_L x_2.$$

Note $\|x_1 - p_L\| = \|x_1 - p_H\| + \|p_H - p_L\|$, and $\|x_2 - x_1\|$ is bounded above by the case when

$$\begin{aligned} \angle x_1 p_L x_2 &= \alpha, \\ \|x_2 - p_L\| &= R + 2\varepsilon, \\ \|x_1 - p_L\| &= \|x_1 - p_H\| + \|p_H - p_L\| = \frac{3R}{2} - \varepsilon. \end{aligned}$$

Hence, we have

$$\|x_2 - x_1\| < (R + 2\varepsilon)^2 + \left(\frac{3R}{2} - \varepsilon\right)^2 - (R + 2\varepsilon)\left(\frac{3R}{2} - \varepsilon\right) \cos \alpha.$$

By assumption, $\alpha \in (0, \Psi(\varepsilon, R))$, and so

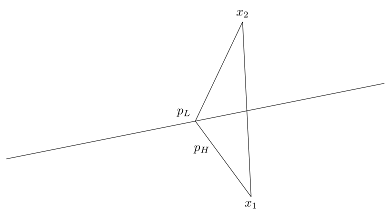


Fig. 10 $d(q_2, H_2) \leq \varepsilon$

$$\|x_2 - x_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(4 + 2\sqrt{2}\right)\varepsilon,$$

which implies that

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(2 + \sqrt{2}\right)\varepsilon.$$

Hence, there is a $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ such that for all $q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(2 + 2\sqrt{2}\right)\varepsilon.$$

□

Proof 4 (Proof of Lemma 3.4) By Lemma 3.1, every $q \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ is with in $1 + \sqrt{2}\varepsilon$ of the point x in L with $\|x - p\| = R$. Hence, $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{5\varepsilon}{2}}$ consists of a single connected component and it has diameter less than $(2 + 2\sqrt{2})\varepsilon$. □

Proof 5 (Proof of Lemma 3.5) As $\|p - z\| \leq \frac{R-\varepsilon}{2}$, the intersection $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap T$ is not empty, connected, and $\mathcal{H}_1\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap T\right) = 0$. Further, the intersections $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap L_1$ and $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap L_2$ are also connected.

Now, let x_1 be the point on L_1 with $\|q_1 - p\| = R$ and let x_2 be the point on L_2 with $\|x_2 - p\| = R$. As $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap T$ is path connected, x_1 and x_2 are path connected in T Fig. 11.

Consider the triangle $\triangle x_1 p x_2$, we have

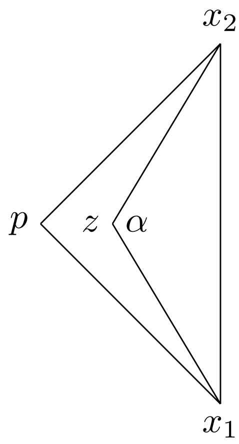
$$\begin{aligned}\|x_1 - x_2\|^2 &= \|x_1 - z\|^2 + \|x_2 - z\|^2 - 2\|x_1 - z\|\|q_2 - z\| \cos\alpha \\ &\geq \left(R - \frac{R-\varepsilon}{2}\right)^2 + \left(R - \frac{R-\varepsilon}{2}\right)^2 - 2\left(R - \frac{R-\varepsilon}{2}\right)^2 \cos\alpha \\ &= 2\left(\frac{R+\varepsilon}{2}\right)^2 (1 - \cos\alpha).\end{aligned}$$

Now, as $d_H(P, T) \leq \varepsilon$, there are points $q_1, q_2 \in P$ with

$$\|q_1 - x_1\|, \|q_2 - x_2\| \leq \varepsilon.$$

Then by the triangle inequality

Fig. 11 Bounding the diameter of a set of points



$$\|q_1 - q_2\|^2 = 2\left(\frac{R+\varepsilon}{2}\right)^2 (1-\cos\alpha) - 2\varepsilon$$
$$> \left(2 + 2\sqrt{2}\right)\varepsilon, \text{ as } \alpha \in \left[\frac{\pi}{6}, \pi\right)$$

$$\arccos\left(\frac{\left(\frac{R}{2}-\varepsilon\right)^2 - 18\varepsilon^2}{\left(\frac{R}{2}-\varepsilon\right)^2}\right)$$

□

Appendix B Proof of Correctness Lemmas

Proof 6 (Proof of Proposition 5.3) Let C be a connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ which spans a locally maximal edge $\overline{uv}$, with midpoint m_{uv}. Then, there is a sample $p_m \in C$ such that $\|p_m - m_{uv}\| \leq \varepsilon$.

To show that $\mathcal{D}(C) \geq \frac{9R}{2}$, we show that there are two points $x_u, x_y \in \overline{uv}$ such that

1. $\|u - x_u\| > \frac{3R}{2} + 2\varepsilon$,
2. $\|v - x_v\| > \frac{3R}{2} + 2\varepsilon$,
3. $\|x_u - x_v\| \geq \frac{3R}{2}$.Without loss of generality, we show that x_u exists, and

$$\|x_u - m_{uv}\| \geq \frac{3R}{4} + \varepsilon.$$

By Assumption i, $\|u - v\| \geq 6(R+\varepsilon)$. As $\overline{uv}$ is a line segment, for all $\eta \in [0, \frac{9R}{4} + 3\varepsilon]$ there is a point $x_\eta \in \overline{uv}$ such that $\|x_\eta - u\| = \eta$. Letting $\eta = \frac{3R}{2} + 2\varepsilon$, there is a point, namely x_u such that $\|x_u - u\| = \frac{3R}{2} + 2\varepsilon$. As P is an ε-sample, there is a sample p_u such that $\|x_u - p_u\| \leq \varepsilon$, and hence $\|p_u - u\| > \frac{3R}{2} + \varepsilon$. Thus, the (ε, R) -local structure of P at p_u is maximal of dimension 1.

We can repeat this argument for all $\eta \in [\frac{3R}{2} + 2\varepsilon, \frac{9R}{4} + 3\varepsilon]$, and obtain a path of points $x_\eta \in \overline{uv}$ and samples $p_\eta \in P$ connecting p_u to p_m.

This also holds when we replace *u* with *v*, and hence we have p_u and p_v. Finally, we have

$$\begin{aligned}\|p_u - p_v\| &\geq \|x_u - x_v\| - \|p_u - x_u\| - \|p_v - x_v\| \\ &\geq \frac{3R}{2} - 2\varepsilon,\end{aligned}$$

and hence $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$.

Now, we show that if $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$, then *C* spans some locally maximal edge. If $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$, then there are points $p, q \in C$ with

$$\|p - q\| \geq \frac{3R}{2} - 2\varepsilon.$$

As *P* is an ε-sample of |*X*|, there are points $x_p, x_q \in |X|$, with

$$\|x_p - p\|, \; \|x_q - q\| \leq \varepsilon.$$

Let m_{pq} be the midpoint of x_p and x_q. As *p* and *q* are in the same connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$, we know there is a sequence of points $\{q_i\}_{i=0}^m$ with $q_0 = p$, $q_m = q$ and for all $0 < i \leq m$, $\|q_i - q_{i-1}\| \leq 3\varepsilon$. Again, *P* is an ε-sample of |*X*|, and as $q_i \in P_{LM,1}$, $\forall 0 \leq i \leq m$, for each q_i there is some $x_i \in |X|$ which is on a locally maximal edge, and $\|q_i - x_i\| \leq \varepsilon$. From Assumptions ii, iv and vi and Proposition 4.9, there is a locally maximal edge, say $\overline{uv}$ such $x_i \in \overline{uv}, \forall 0 \leq i \leq m$. Let the midpoint of $\overline{uv}$ be x_{uv}.

We now split into two cases:

I there is some *i* such that $x_i = x_{uv}$,

II for all *i* we have $x_i \neq x_{uv}$.<u>Case I:</u> The connected component *C* is a spanning connected component, as it contains a sample which is within ε of the midpoint x_{uv} of the locally maximal edge $\overline{uv}$.

<u>Case II:</u> As no q_i is within ε of m_{uv}, we know that $q_i \; \forall 0 \leq i \leq m$ are on the same side of $\overline{uv}$. That is, for all q_i, without loss of generality,

$$\begin{aligned}\|q_i - x_{uv}\| \leq \|q_i - u\| &\geq \frac{3\sqrt{3}}{2}R + 3\varepsilon \\ \|q_m - v\| &\geq \frac{3R}{2} + \varepsilon.\end{aligned}$$

Further, assume that

$$\|q_0 - m_{uv}\| \leq \|q_m - x_{uv}\|.$$

There is another sequence of points $\{x'_j\}_{j=0}^{m'}$ in $\overline{uv}$ with $x'_0 = x_m$ and $x'_{m'} = x_{uv}$, and for $0 < j \leq m'$

$$\|x'_j - x'_{j-1}\| \leq \varepsilon.$$

Then, there exists $q'_j \in P$ with

$$\|q'_j - x'_j\| \leq \varepsilon$$
$$\|q'j - q'_{j-1}\| \leq \varepsilon\, \forall 0 < j \leq m'$$
$$\|q'_j - v\| \geq \frac{3R}{2} + \varepsilon.$$

By Assumptions ii, iv and vi and Proposition 4.9, $q'_j \in P_{LM,1}$ for all $0 \leq j \leq m'$. Hence, each q'_j is in the same connected component C as q_m.

Thus, C contains a sample $q'_{m'}$ which is within ε of the midpoint of the locally maximal edge $\overline{uv}$. Hence, C is a spanning connected component.

Thus a component C of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ spans a locally maximal edge $\overline{uv}$ if and only if $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$. □

Proof 7 (Proof of Proposition 5.4) First, let C be a connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}$ which spans some triangle $\triangle uvw$ with midpoint m. As P is an ε-sample of X, there is a sample $p_m \in P$ with $\|p_m - m\| \leq \varepsilon$. As the radius of the inscribed circle of $\triangle uvw$ is at least $2R + 3\varepsilon$, m is at least $2R + 3\varepsilon$ from $\partial\triangle uvw$. Thus, $d(p_m, \partial\triangle uvw) \geq 2R + 2\varepsilon$.

Hence, for all $q \in B_{\frac{R}{2}+2\varepsilon}(p) \cap P$, $d(q, \partial\triangle uvw) \geq \frac{3R}{2} + \varepsilon$, and so $q \in P_{LM,2}$.

Now, take $p \in P_{LM,2}$ such that $B_{\frac{R}{2}+\varepsilon}(p) \cap P \subset P_{LM,2}$. Then, there is some triangle $\triangle uvw$ with $d(\triangle uvw, p) \leq \varepsilon$. As $p \in P_{LM,2}$, we know that $d(\partial\triangle uvw, p) > \frac{R}{2} - \varepsilon$. By assumption, for all $q \in B_{\frac{R}{2}+\varepsilon}(p) \cap P$, we have $d(\partial\triangle uvw, q) > \frac{R}{2} - \varepsilon$. Recall that P is an ε-sample of $|X|$, so there is a point $x \in X$ such that $\|p - x\| \leq \varepsilon$. As $\triangle uvw$ is convex, and every $B_{\frac{R}{2}+\varepsilon}(p) \cap P \subset P_{LM,2}$, we have

$$d(\partial\triangle uvw, x) \geq \frac{R}{2} + 2\varepsilon + \frac{R}{2} - 2\varepsilon = R.$$

Hence, is a point $y \in B_{\frac{R}{2}+2\varepsilon}(p) \cap \triangle uvw$ with

$$d(\partial\triangle uvw, y) \geq \frac{R}{2} + 2\varepsilon.$$

and a sample $q \in B_{\frac{R}{2}+2\varepsilon}(p) \cap P_{LM,2}$ with $\|q - y\| \leq \varepsilon$.

Now, we can construct a sequence of points $\{y_i\}_{i=0}^{m} \subset \triangle uvw$ such that $\|y_i - y_{i-1}\| \leq \varepsilon$ for $1 \leq i \leq m$, and $y_0 = x$, $y_m = y$. Further, for each y_i there is a

$q_i \in P$ with $\|q_i - y_i\| \le \varepsilon$, and $q_i \in P_{LM,2}$. Note, that this means p and q_m are in the same connected component C of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$.

Finally, we construct a similar sequence of points $\{\widetilde{y}_j\}_{j=0}^{\widetilde{m}}$ in $|X|$ from y to $m_{\triangle uvw}$ with $\widetilde{y}_0 = y$, $\widetilde{y}_{\widetilde{m}} = m_{\triangle uvw}$. Again, for each $\widetilde{y}_j$, there is a $\widetilde{q}_j \in P$ with $\|\widetilde{y}_j - \widetilde{q}_j\| \le \varepsilon$ and $\widetilde{q}_j \in P_{LM,2}$. Hence, the $\widetilde{q}_j$ are in the same connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$, and further, this connected component is C. □

Proof 8 (Proof of Proposition 5.5) Let V_{LM} be the set of locally maximal vertices of X. Let v be a locally maximal vertex, then by Assumptions i to iii and Proposition 4.8, $\forall p \in P$ with $\|p - v\| \le 4\varepsilon$, $p \in P_{LM,0}$. In fact, by Assumptions i to iii, any $p \in P$ with $\|p - v\| \le 4\varepsilon$ is actually within ε of v. Hence, every $p \in P_{LM,0}$ within ε of v are in the same connected component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,0})$.

Now, take a connected component C of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,0})$. Each $p \in C$ is within ε of a locally maximal vertex v_p of X. By Assumptions i to iii, every locally maximal vertex v is at least 5ε away from any other cell of X, and hence $\forall p \in C$, v_p is the same.

Hence, the connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,0})$ correspond bijectively to the locally maximal vertices of X. □

Proof 9 (Proof of Proposition 5.6) Let $E_{LM} \subset E$ be the set of locally maximal edges in X. By Proposition 5.3, a connected component C of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$ spans an edge $\overline{uv}$ if and only if it contains a sample p within ε of the midpoint m of $\overline{uv}$.

If a connected component C is a spanning component, then there is some locally maximal edge $\overline{uv}$ with midpoint m such that there is a sample $p \in C$ with $\|m - p\| \le \varepsilon$.

For any locally maximal $\overline{uv} \in E_{LM}$ with midpoint m, there is some sample $p \in P$ such that $\|m - p\| \le \varepsilon$. Then, by Assumptions ii, iv and vi and Proposition 4.9, $p \in P_{LM,1}$, and so there is some spanning connected component $C_{\overline{uv}}$ in $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$.

Now, consider a locally maximal edge $\overline{uv'}$, $v' \ne v$, and take samples $p, q \in P_{LM,2}$ such that $d(\overline{uv}, p)$, $d(\overline{uv'}, q) \le \varepsilon$. By Assumption i, $\|p - q\| > 6\varepsilon$, and so p and q are in different connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$.

Finally, consider a locally maximal edge $\overline{u'v'}$ such that $\overline{uv}$ and $\overline{u'v'}$ do not have a common vertex. Take samples $p, q \in P_{LM,2}$ such that

$$d(\overline{uv}, p),\ d(\overline{u'v'}, q) \le \varepsilon.$$

Again, by Assumption vi, $\|p - q\| > 6\varepsilon$, and so p and q are in different connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,1})$.

Hence, each connected component C only consists of samples p with $d(\overline{uv}, p) \le \varepsilon$ for a single locally maximal edge $\overline{uv}$.

Thus, the spanning connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ are in bijection with the locally maximal edges of $|X|$.

Proof 10 (Proof of Proposition 5.7) From Proposition 5.4, a connected component C of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ spans a triangle $\triangle uvw$ if and only if it contains a sample p within ε of the midpoint m of $\triangle uvw$.

As P is a ε-sample of $|X|$, for every $\triangle uvw$ with midpoint m, there is a sample $p \in P$ such that $\|p - m\| \leq \varepsilon$. Hence, there is a spanning connected component C in $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$.

Now, consider C a spanning component of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$. Then, as P is a ε-sample, there is some $\triangle uvw$ with midpoint m such that there is a sample $p \in C$ with $\|p - m\| \leq \varepsilon$.

Consider two triangles $\triangle uvw,\ \triangle u'v'w'$, and take two samples $p, p' \in P_{LM,2}$ with

$$d(\triangle uvw, p),\ d(\triangle u'v'w', p') \leq \varepsilon.$$

As $p, p' \in P_{LM,2}$, we know that

$$d(\partial\triangle uvw, p),\ d(\partial\triangle u'v'w', p') > R + \varepsilon,$$

and so by Assumption vii, $\|p - p'\| > 6\varepsilon$.

Hence, the spanning components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$ are in bijection with the triangles of X. □

Proof 11 (Proof of Lemma 5.9) As $\overline{uv}$ is a locally maximal edge, there is a corresponding edge spanning component $\mathcal{E}$. As u, v are not faces of any other cell $\sigma \in X$, by Assumptions i to iv, vi and viii to xiii and Propositions 4.9 and 4.11, the points $p \in P_{NLM}$ which witness $\mathcal{E}$ do not witness any other edge spanning component $\mathcal{E}'$ or any triangle spanning component $\mathcal{T}$.

Thus, there is a single partition P_1 of P_{NLM} which contains all the samples p that witness $\mathcal{E}$. Furthermore, there is no other partition P_2 of P_{NLM} that witnesses $\mathcal{E}$. Hence, P_1 is assigned label VV. □

Proof 12 (Proof of Lemma 5.10) As $\overline{uv}$ is a locally maximal edge, there is a corresponding edge spanning component $\mathcal{E}$. Without loss of generality, assume v is the face of some locally maximal cell $\sigma \neq \overline{uv}$.

By Assumptions i to iv and vi to xiii and Propositions 4.9 to 4.11, there are samples $p_u, pv_v \in P_{NLM}$ such that

$$\|p_u - u\|,\ \|p_v - v\| \leq \varepsilon.$$

Further, there is a spanning connected component $\mathcal{C}$ which p_v also witnesses but p_u does not witness. Hence, there are two partitions P_v, P_u which witness $\mathcal{E}$. By Algorithm 4, there are no other partitions which witness $\mathcal{E}$.

Hence, both P_v and P_u are labelled with V. □

Proof 13 (Proof of Lemma 5.11) Let $\mathcal{T}$ be the triangle spanning component that corresponds to $\triangle uvw$. By Assumption 1 and propositions 4.8 to 4.10 and 4.12, the samples $p \in P_{NLM}$ that witness $\mathcal{T}$ do not witness any spanning connected component $\mathcal{C} \neq \mathcal{T}$. Furthermore, from Algorithm 4 there is a unique connected component P_1 that witnesses $\mathcal{T}$.

As P is an ε-sample of $|X|$, and from Propositions 4.8 and 4.12, there are samples $p_u, p_v, p_w, p_{uv}, p_{vw}, p_{uw} \in P_1$ such that

$$\|p_u - u\|,\ \|p_v - v\|,\ \|p_w - w\| \leq \varepsilon,$$
$$d(\overline{uv}, p_{uv}),\ d(\overline{vw}, p_{vw}),\ d(\overline{uw}, p_{uw}) \leq \varepsilon.$$

Hence, P_1 is assigned label *VVVEEE*. □

Proof 14 (Proof of Lemma 5.12) Let $\mathcal{T}$ be the triangle spanning component that corresponds to $\triangle uvw$. By Assumptions i to iv and vi to xi and propositions 4.8 to 4.10 and 4.12, any spanning connected component $\mathcal{C}$ witnessed by samples $p \in P_{NLM}$ that witness $\mathcal{T}$ corresponds to a locally maximal cell τ such that $\triangle uvw \cap \tau \neq \emptyset$.

We need to split into two cases:

1. there is a unique locally maximal cell $\tau \in X$ with $\triangle uvw \cap \tau = v$
2. there are at least two locally maximal cells $\tau, \sigma \in X$, $\tau \neq \sigma$ with $\triangle uvw \cap \tau = \triangle uvw \cap \sigma = v$.Case 1: We assumed there was a unique locally maximal τ with $\triangle uvw \cap \tau = v$, and hence, by Propositions 5.6 and 5.7 there is some spanning component $\mathcal{C}_\tau$ which corresponds to τ. with By Assumptions i to iv and vi to xi and propositions 4.8 to 4.10 and 4.12, in Algorithm 4 there is a single partition P_1 of P_{NLM} which witnesses $\mathcal{T}$ and $\mathcal{C}_\tau$, and there is a unique partition P_2 which witnesses just $\mathcal{T}$. Further, P_1 is assigned label V and P_2 label *VVEEE*.

Case 2: From our assumptions, there are two locally maximal cells $\tau, \sigma \in X$, $\tau \neq \sigma$ such that

$$\tau \cap \triangle uvw = v = \sigma \cap \triangle uvw.$$

By Propositions 5.6 and 5.7 there is some spanning component $\mathcal{C}_\tau$ which corresponds to τ, and some spanning component $\mathcal{C}_\sigma$ which corresponds to σ.

Further, from Algorithm 4, there is a single partition P_1 of P_{NLM} which witnesses $\mathcal{T}, \mathcal{C}_\tau, \mathcal{C}_\sigma$, and no partitions which witness a subset of these spanning components. This holds, by induction, for any locally maximal cell $\tau' \in X, \tau' \neq \tau, \sigma$ with

$\tau' \cap \triangle uvw = v$. Similarly, there is a single partition P_2 of P_{NLM} which witnesses only $\mathcal{T}$. Further, P_1 is assigned label *V* and P_2 label *VVEEE*. □

Proof 15 (Proof of Lemma 5.13) Let $\mathcal{T}$ be the triangle spanning component that corresponds to $\triangle uvw$. By Assumptions i to iv and vi to xi and propositions 4.8 to 4.10 and 4.12, any spanning connected component $\mathcal{C}$ witnessed by samples $p \in P_{NLM}$ that witness $\mathcal{T}$ corresponds to a locally maximal cell τ such that $\triangle uvw \cap \tau \neq \emptyset$.

We need to split into two cases:

1. there is a unique locally maximal cell $\tau \in X$ with $\triangle uvw \cap \tau = \overline{uv}$
2. there are at least two locally maximal cells $\tau, \sigma \in X$, $\tau \neq \sigma$ with $\triangle uvw \cap \tau = \triangle uvw \cap \sigma = \overline{uv}$.Case 1: We assumed there was a unique locally maximal τ with $\triangle uvw \cap \tau = \overline{uv}$, and hence, by Propositions 5.6 and 5.7 there is some spanning component $\mathcal{C}_\tau$ which corresponds to τ. with By Assumptions i to iv and vi to xi and propositions 4.8 to 4.10 and 4.12, in Algorithm 4 there is a single partition P_1 of P_{NLM} which witnesses $\mathcal{T}$ and $\mathcal{C}_\tau$, and there is a unique partition P_2 which witnesses just $\mathcal{T}$. Further, P_1 is assigned label *VE* and P_2 label *VEE*.

Case 2: From our assumptions, there are two locally maximal cells $\tau, \sigma \in X$, $\tau \neq \sigma$ such that

$$\tau \cap \triangle uvw = \overline{uv} = \sigma \cap \triangle uvw.$$

By Propositions 5.6 and 5.7 there is some spanning component $\mathcal{C}_\tau$ which corresponds to τ, and some spanning component $\mathcal{C}_\sigma$ which corresponds to σ.

And from Algorithm 4, there is a single partition P_1 of P_{NLM} which witnesses $\mathcal{T}, \mathcal{C}_\tau, \mathcal{C}_\sigma$, and no partitions which witness a subset of these spanning components. This holds, by induction, for any locally maximal cell $\tau' \in X, \tau' \neq \tau, \sigma$ with $\tau' \cap \triangle uvw = v$. Similarly, there is a single partition P_2 of P_{NLM} which witnesses only $\mathcal{T}$. Further, P_1 is assigned label *VE* and P_2 label *VEE*. □

Proof 16 (Proof of Lemma 5.14) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemma 5.12. By combining the arguments at the two shared vertices, there are three partitions P_1, P_2, P_3 from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2$ such that P_1 witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2$, and P_2 witnesses $\mathcal{C}_2$ but not $\mathcal{C}_1$. Further, P_3 only witnesses $\mathcal{T}$. Hence, P_1, P_2 are labelled with *V* and P_3 with *VEEE*. □

Proof 17 (Proof of Lemma 5.15) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions P_1, P_2, P_3 from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2$

such that P_1 witnesses $\mathcal{C}_1$ and $\mathcal{C}_2$, and P_2 witnesses $\mathcal{C}_2$ but not $\mathcal{C}_1$. Further, P_3 only witnesses $\mathcal{T}$. Hence, P_1 is labelled with *V*, P_2 with *VE* and P_3 with *E*. □

Proof 18 (Proof of Lemma 5.16) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions P_1, P_2, P_3 from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2$ such that P_1 witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2$, and P_2 witnesses $\mathcal{C}_2$ but not $\mathcal{C}_1$. Further, P_3 only witnesses $\mathcal{T}$. Hence, P_1 is labelled with *V*, P_2 with *VVE* and P_3 with *EE*. □

Proof 19 (Proof of Lemma 5.17) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions P_1, P_2, P_3, P_4 from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni$ such that P_1 witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2, \mathcal{C}_3$, P_2 witnesses $\mathcal{C}_2$ but not $\mathcal{C}_1, \mathcal{C}_3$, and P_2 witnesses $\mathcal{C}_3$ but not $\mathcal{C}_1, \mathcal{C}_2$. Further, P_4 only witnesses $\mathcal{T}$. Hence, P_1, P_2 and P_3 are labelled with *V* and P_4 with *EEE*. □

Proof 20 (Proof of Lemma 5.18) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions

$$P_1, P_2, P_3, P_4$$

from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni$ such that P_1 witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2, \mathcal{C}_3$, P_2 witnesses $\mathcal{C}_1, \mathcal{C}_2$ but not $\mathcal{C}_3$, and P_2 witnesses $\mathcal{C}_3$ but not $\mathcal{C}_1, \mathcal{C}_2$. Further, P_4 only witnesses $\mathcal{T}$. Hence, P_1, P_2 and P_3 are labelled with *V* and P_4 with *EEE*. □

Proof 21 (Proof of Lemma 5.19) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions P_1, P_2, P_3, P_4 from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni$ such that P_1 witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2, \mathcal{C}_3$, P_2 witnesses $\mathcal{C}_1, \mathcal{C}_2$ but not $\mathcal{C}_3$, and P_2 witnesses $\mathcal{C}_1, \mathcal{C}_3$ but not $\mathcal{C}_2$. Further, P_4 only witnesses $\mathcal{T}$. Hence, P_1 is labelled with *E*, P_2, P_3 with *V* and P_4 with *VEE*. □

Proof 22 (Proof of Lemma 5.20) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are four partitions P_1, P_2, P_3, P_4 from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni$ such that P_1 witnesses $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$, P_2 witnesses $\mathcal{C}_1, \mathcal{C}_2$ but not $\mathcal{C}_3$, and P_3 witnesses $\mathcal{C}_1, \mathcal{C}_3$ but not $\mathcal{C}_2$. Further, P_4 only witnesses $\mathcal{T}$. Hence, P_1 is labelled with *V*, P_2, P_3 with *VE*, and P_3 with *E*. □

Proof 23 (Proof of Lemma 5.21) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are four partitions P_1, P_2, P_3, P_4 from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni, \mathcal{C}_\ni$ such that P_1 witnesses $\mathcal{C}_1$ and not $\mathcal{C}_2, \mathcal{C}_3$, P_2 witnesses $\mathcal{C}_1$ and not $\mathcal{C}_2, \mathcal{C}_3$, and P_3 witnesses $\mathcal{C}_3$ but not $\mathcal{C}_1, \mathcal{C}_2$. Further, P_4 only witnesses $\mathcal{T}$. Hence, P_1, P_2, P_3 are labelled with *V*, and P_4 with *EEE*. □

Proof 24 (Proof of Lemma 5.22) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments, there are five partitions

$$P_1, P_2, P_3, P_4, P_5$$

from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni, \mathcal{C}_\ni, \mathcal{C}_\triangle$ such that P_1 witnesses $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\triangle$ and not $\mathcal{C}_3$, P_2 witnesses $\mathcal{C}_2$ and not $\mathcal{C}_1, \mathcal{C}_3, \mathcal{C}_4$, P_3 witnesses $\mathcal{C}_2, \mathcal{C}_3$ but not $\mathcal{C}_1, \mathcal{C}_4$, and P_4 witnesses $\mathcal{C}_4$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$. Further, P_5 only witnesses $\mathcal{T}$, and hence P_4 only witnesses $\mathcal{T}$. Hence, P_1, P_2 are labelled with *V*, P_3 with *VE*, and P_4, P_5 with *E*. □

Proof 25 (Proof of Lemma 5.23) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are six partitions

$$P_1, P_2, P_3, P_4, P_5, P_6$$

from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_5$ such that P_1 witnesses $\mathcal{C}_1, \mathcal{C}_4$ and not $\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5$, P_2 witnesses $\mathcal{C}_2, \mathcal{C}_4, \mathcal{C}_5$ and not $\mathcal{C}_1, \mathcal{C}_3$, P_3 witnesses $\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5$ but not $\mathcal{C}_1, \mathcal{C}_4$, P_4 witnesses $\mathcal{C}_4$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5$, and P_5 witnesses $\mathcal{C}_5$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$. Further, P_6 only witnesses $\mathcal{T}$, and hence P_1, P_2, P_3 are labelled with *V*, P_4, P_5, P_6 with *E*. □

Proof 26 (Proof of Lemma 5.24) Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are six partitions fig. 13.

$$P_1, P_2, P_3, P_4, P_5, P_6$$

from Algorithm 4 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_5, \mathcal{C}_6$ such that P_1 witnesses $\mathcal{C}_1, \mathcal{C}_4, \mathcal{C}_6$ and not $\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5$, P_2 witnesses $\mathcal{C}_2, \mathcal{C}_4, \mathcal{C}_5$ and not $\mathcal{C}_1, \mathcal{C}_3, \mathcal{C}_6$, P_3 witnesses $\mathcal{C}_3, \mathcal{C}_5, \mathcal{C}_6$ but not $\mathcal{C}_1, \mathcal{C}_2\mathcal{C}_4$, P_4 witnesses $\mathcal{C}_4$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5, \mathcal{C}_6$, and P_5 witnesses $\mathcal{C}_5$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_6$, and P_6 witnesses $\mathcal{C}_6$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_5$. Hence P_1, P_2, P_3 are labelled with *V*, P_4, P_5, P_6 with *E*. □

Appendix C Graphs of functions

See Figs. 12, 13, 14

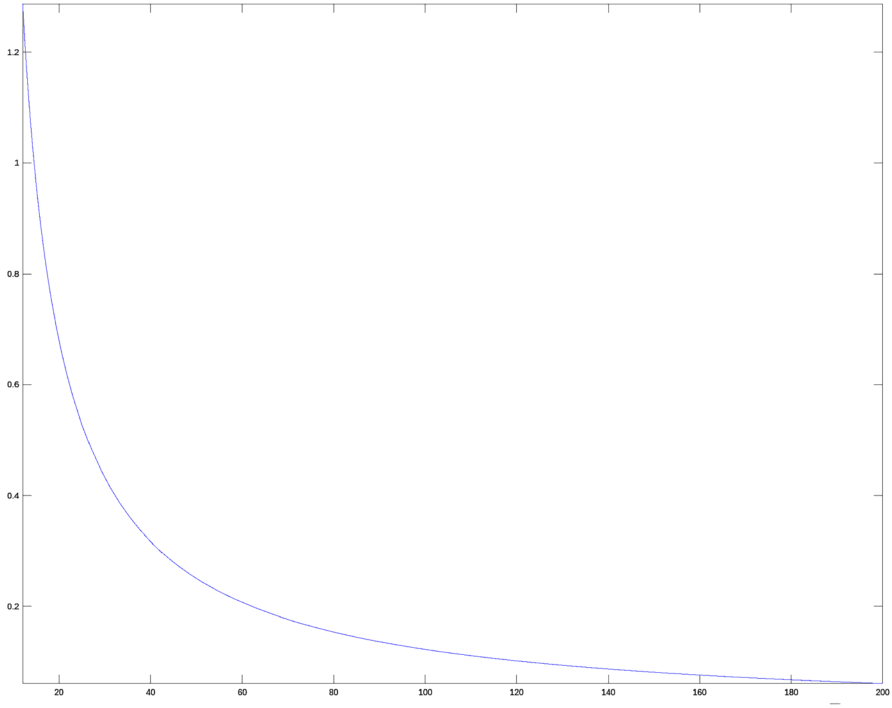


Fig. 12 Graph of $\Psi_1\left(1, \frac{R}{\varepsilon}\right)$

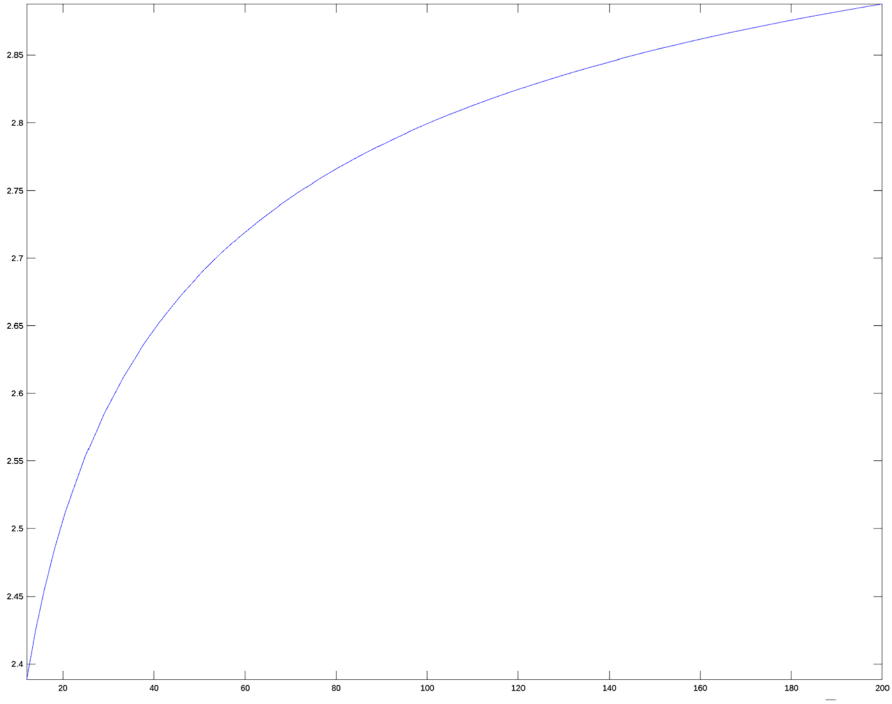


Fig. 13 Graph of $\Psi_2\left(1, \frac{R}{\varepsilon}\right)$

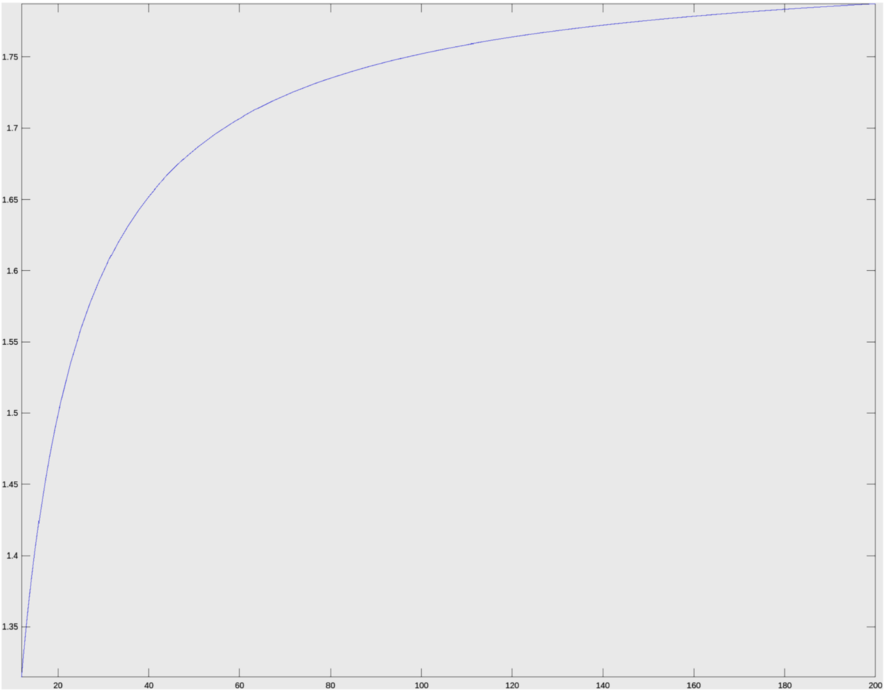


Fig. 14 Graph of $\Psi_3\left(1, \frac{R}{\varepsilon}\right)$

Appendix D Pseudocode of potential algorithm

Algorithm 1 $\Delta_{\varepsilon,R}(p)$

Data: An ε-dense sample P of an embedded 2-complex $|X|$, a point $p \in P$.
Result: -1 if $p \in P_{NLM}$, 0 if $p \in P_{LM,0}$, 1 if $p \in P_{LM,1}$, 2 if $p \in P_{LM,2}$.
begin
  $\mathcal{G}_p \longleftarrow \{q \in P \mid \|p-q\| \leq R+\varepsilon\}$;
  connect $q, q' \in \mathcal{G}_p$ if $\|q-q'\| \leq 3\varepsilon$;
  $\mathcal{C}_p \longleftarrow \{q \in \mathcal{G}_p \mid q$ is path connected to p in $\mathcal{G}_p\}$;
  remove $q \in \mathcal{C}_p$ if $\|p-q\| \geq R-\varepsilon$;
  if $\mathcal{RK}_0 = 0$ *and* $\mathcal{RK}_1 = 0$ **then**
    return *0*
  else if $\mathcal{RK}_0 = 1$ *and* $\mathcal{RK}_1 \neq 1$ **then**
    return *-1*
  else if $\mathcal{RK}_0 = 1$ *and* $\mathcal{RK}_1 = 1$ **then**
    if $\forall q_1, q_2 \in C_p$, $\exists q_0$ *such that* $\|q_1-q_0\|, \|q_2-q_0\|, \|q_2-q_1\| \in [\sqrt{3(R^2-\varepsilon^2)}, \sqrt{3}R]$ **then**
      return *2*
    else
      return *-1*
  else if $\mathcal{RK}_0 = 2$ *and* $\mathcal{RK}_1 = 0$ **then**
    if $\max\{\mathcal{D}(c_1), \mathcal{D}(c_2)\} \leq 5\varepsilon$ **then**
      if $\langle q_1-p, q_2-p\rangle > -R^2+2R\varepsilon-7\varepsilon^2$ **then**
        return *1*
      else
        return *-1*
    else
      return *-1*
  else if $\mathcal{RK}_0 = n$, $n \neq 0,1,2$ *and* $\mathcal{RK}_1 = 0$ **then**
    return *-1*

Algorithm 2 Spanning triangle components

Data: Parameters ε, R and $P_{LM,1}$.
Result: The set of triangle spanning components.
begin
  Initialise empty set T;
  Let C be the set of connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$;
  for $\mathcal{T} \in C$ **do**
    if $\exists p \in \mathcal{C}$ *such that* $B_{R/2+\varepsilon}(p) \cap P \subset P_{LM,2}$ **then**
      Add $\mathcal{T}$ to T;
  return T

Algorithm 3 Spanning edge components

Data: Parameters ε, R and $P_{LM,1}$.
Result: The set of triangle spanning components.
begin
  Initialise empty set E;
  Let C be the set of connected components of $\check{\mathcal{C}}_{\frac{5\varepsilon}{2}}(P_{LM,2})$;
  for $\mathcal{E} \in C$ **do**
    if $\mathcal{D}(\mathcal{T}) \geq \frac{3R}{2} - 2\varepsilon$ **then**
      Add $\mathcal{E}$ to E;
  return E

Algorithm 4 Partitioning P_{NLM}

Data: An ε-dense sample P of an embedded 2-complex $|X|$, partitioned into P_{NLM}, $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$.
Result: A partition $\{P_i\}$ of P_{NLM}, and for each P_i, two sets $S_E(P_i)$, $S_T(P_i)$.
begin
  For each $p \in P_{NLM}$, find all the edge spanning components $\mathcal{E}$ such that $\mathcal{E} \cap B_{(R+2\varepsilon)\kappa}(p) \neq \emptyset$, and place them in $S_E(p)$;
  Find all the triangle spanning components $\mathcal{T}$ such that $\mathcal{T} \cap B_{(R+2\varepsilon)\kappa}(p) \neq \emptyset$, and place them in $S_T(p)$;
  Partition P_{NLM} into $\{ P_i \}$ such that for each $p, q \in P_i$, $S_E(p) = S_E(q)$ and $S_T(p) = S_T(q)$;
  Assign $S_E(P_i)$ and $S_T(P_i)$ to each P_i;
  for P_i *and* P_j *with* $S_E(P_j) \subseteq S_E(P_i)$ *and* $S_T(P_j) \subseteq S_T(P_i)$ **do**
    if $S_E(P_j), S_T(P_j) \neq \emptyset$ **then**
      Merge P_j into P_i with labels $S_E(P_i)$, $S_T(P_i)$;
    else if $|S_T(P_j)| \geq 2$ *and* $\forall p \in P_j$ *such that* $\mathrm{Sig}_{\varepsilon,R}(p) = (n, 0)$, $n \in \mathbb{Z}_{\geq 0}$ **then**
      Merge P_j into P_i with labels $S_E(P_i)$, $S_T(P_i)$;
  return $\{P_i\}$, and $S_E(P_i)$, $S_T(P_i)$ for each P_i

Algorithm 5 Order $\{P_i\}$

```
Data: An ε-dense sample P of an embedded 2-complex |X|, partition {P_i} of
      P_NLM with two sets S_E(P_i), S_T(P_i) for each P_i and partitions of
      P_LM,0, P_LM,1, P_LM,2.
Result: Two sets P^1, P^2 ⊂ { P_i }.
begin
    Initialise empty P^1 and P^2;
    for P_i ∈ {P_i} do
        if S_E(P_i) ≠ ∅ then
            Add P_i to P^1
        else if ∃p ∈ P_i such that Sig_R(p) ≠ (1, n) then
            Add P_i to P^1
        else if |S_T(P_i)| ≠ 1 then
            Add P_i to P^1
        else
            Add P_i to P^2
    return P^1, P^2
```

Algorithm 6 Classification of P^1

Data: An ε-dense sample P of an embedded 2-complex $|X|$, P^1, and partitions of P_{NLM}, $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$.

Result: A labeled list C, where the label for P_i is -1 if P_i corresponds to 2 vertices, 0 if P_i corresponds to a vertex, 1 if P_i corresponds to a vertex and an edge, 2 if P_i corresponds to two vertices and an edge, 3 if P_i corresponds to just an edge.

begin
- Initialise empty list C;
- **for** $P_i \in P^1$ **do**
  - **if** $|S_E(P_i)|= 1$ *and* $S_T(P_i) = \emptyset$ **then**
    - **if** $\mathcal{E} \notin S_E(P_j) \forall P_j \neq P_i$ **then**
      - Add P_i to C with label VV;
    - **else if** $\exists P_j \neq P_i$ *such that* $\mathcal{E} \in S_E(P_j)$ **then**
      - Add P_i to C with label 0;
  - **else if** $S_E(P_i) \neq \emptyset$ **then**
    - Add P_i to C with label 0;
  - **else**
    - **for** $\mathcal{T} \in S_T(P_i)$ **do**
      - Let $LN(\mathcal{T}) = \{P_k \mid \mathcal{T} \in S_T(P_k)\}$
    - Let $N(P_i) = \bigcap_{\mathcal{T} \in S_T(P_i)} LN(\mathcal{T})$;
    - **if** $N(P_i) = \{P_i, P_k\}$ **then**
      - Add P_i to C with label VE;
      - Add P_k to C with label V, unless P_k is already in C;
    - **else if** $N(P_i) = \{P_i, P_k, P_l\}$ **then**
      - Add P_i to C with label E;
      - Add P_k to C with label V, unless P_k is already in C;
      - Add P_l to C with label V, unless P_l is already in C;
- **if** $\exists P_i \in P^1 \setminus C$ **then**
  - Add P_i to C with label VVE;
- **return** C

Algorithm 7 Classification of P^2

Data: An ε-dense sample P of an embedded 2-complex $|X|$, P^2, and partitions of P_{NLM}, $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$, a labelled list C obtained from Algorithm 6.

Result: A labelled list C.

begin
- **for** $P_i \in P^2$ **do**
  - **if** $P_i \notin C$ **then**
    - Let $LN = \{P_k \mid \mathcal{T} \in S_T(P_k)\}$;
    - **if** $LN \cap P^2 = \{P_i, P_k, P_l\}$ **then**
      - Add P_i, P_k, P_l to C with label E;
    - **else if** $LN \cap P^2 = \{P_i, P_k\}$ **then**
      - Add P_i to C with label E;
      - Add P_l to C with label VEE;
    - **else if** $LN \cap P^2 = \{P_i\}$ **then**
      - **if** $LN = \{P_i\}$ **then**
        - Add P_i to C with label $VVVEEE$;
      - **else if** $LN = \{P_i, P_k\}$ *and* P_k *has label* V **then**
        - Add P_i to C with label $VVEEE$;
      - **else if** $LN = \{P_i, P_k\}$ *and* P_k *has label* VVE **then**
        - Add P_i to C with label VEE;
      - **else if** $LN = \{P_i, P_k, P_l\}$ *and* P_k *has label* V, P_l *label* VE **then**
        - Add P_i to C with label VEE;
      - **else if** $LN = \{P_i, P_k, P_l\}$ *and* P_k *has label* VE, P_l *label* VVE **then**
        - Add P_i to C with label E;
      - **else if** $LN = \{P_i, P_k, P_l\}$ *and* P_k *has label* V, P_l *label* V **then**
        - Add P_i to C with label $VEEE$;
      - **else if** $LN = \{P_i, P_k, P_l, P_j\}$ *and* P_k, P_l, P_j *have label* V **then**
        - Add P_i to C with label 8;
      - **else if** $LN = \{P_i, P_k, P_l, P_j, P_m\}$ *and* P_k, P_l, P_j *have label* V *and* P_m *has label* E **then**
        - Add P_i to C with label 9;
- **return** C

Algorithm 8 Number of triangles, edges and vertices.

Data: An ε-dense sample P of an embedded 2-complex $|X|$, partitions of P_{NLM}, $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$ and the labelled list C from Algorithm 7.
Result: The triangles, edges, and vertices in X.
begin
  Initialise an empty weighted graph B;
  $\forall$ spanning components $\mathcal{T}$ of $P_{LM,2}$, add weight 2 node to B, labelled with $\mathcal{T}$;
  $\forall$ spanning components $\mathcal{E}$ of $P_{LM,1}$, add weight 1 node to B, labelled with $\mathcal{E}$;
  $\forall$ components $\mathcal{V}$ of $P_{LM,0}$, add weight 0 node to B, labelled with $\mathcal{V}$;
  for $P_i \in C$ **do**
    if P_i *has label* VV **then**
      Add 2 weight 0 nodes to B, labelled with P_i;
    else if P_i *has label* V **then**
      Add weight 0 node to B, labelled with P_i;
    else if P_i *has label* VE **then**
      Add 2 weight 0 nodes to B, labelled with P_i;
      Add weight 1 node to B, labelled with P_i;
    else if P_i *has label* VVE **then**
      Add weight 0 node to B, labelled with P_i;
      Add weight 1 node to B, labelled with P_i;
    else if P_i *has label* E **then**
      Add two weight 0 nodes to B, labelled with P_i;
      Add weight 1 node to B, labelled with P_i;
    else if P_i *has label* VEE **then**
      Add weight 1 node to B, labelled with P_i;
    else if P_i *has label* $VVEEE$ **then**
      Add weight 0 node to B, labelled with P_i;
      Add two weight 1 nodes to B, labelled with P_i;
    else if P_i *has label* $VEEE$ **then**
      Add two weight 0 nodes to B, labelled with P_i;
      Add three weight 1 nodes to B, labelled with P_i;
    else if P_i *has label* $VVVEEE$ **then**
      Add three weight 0 nodes to B, labelled with P_i;
      Add three weight 1 nodes to B, labelled with P_i;

Acknowledgements The author would like to thank Kate Turner, Chris Williams, Jonathan Spreer, Stephan Tillmann, Vanessa Robins, Vigleik Angeltveit, Martin Helmer, and James Morgan for very helpful discussions; and thanks Sara Kališnik Hintz and Paul Bendich for comments on an earlier version. Additonally, the author would like to thank both reviewers for their very insightful and helpful comments, without which the paper would be infinitely less coherent than it currently is.

Author Contributions Not applicable, single-author paper.

Funding Open access funding provided by Institute of Science and Technology (IST Austria). The work in this paper was supported by an Australian Federal Government Grant, 2019-2022, Stratified Space Learning.

Data Availability Not applicable.

Declarations

Ethical Approval Not applicable.

Conflict of interest Not applicable.

References

1. McInnes, L., Healy, J., Melville, J.: UMAP: uniform manifold approximation and projection for dimension Reduction. arXiv:1802.03426 (2020)
2. Kileel, J., Moscovich, A., Zelesko, N., Singer, A.: Manifold learning with arbitrary norms. J. Fourier Anal. Appl. **27**(5), 82 (2021). https://doi.org/10.1007/s00041-021-09879-2
3. Cheng, S., Dey, T.K., Ramos, E.: Manifold reconstruction from point samples. Proc. Sixteenth Ann. ACM-SIAM Symp. Discrete Algorithms (2005). https://doi.org/10.5555/1070432.1070579
4. Dey, T.K.: Curve and Surface Reconstruction. Cambridge University Press, Cambridge (2007). https://doi.org/10.1017/CBO9780511546860
5. Dey, T.K., Wang, Y.: Dimension detection with local homology. In: Proceedings of the 26th Canadian Conference on Computational Geometry (2014)
6. Bokor, Y., Turner, K., Williams, C.: Reconstructing linearly embedded graphs: a first step to stratified space learning. Found. Data Sci. **4**(4), 537–561 (2021). https://doi.org/10.3934/fods.2021026
7. Bendich, P., Wang, B., Mukherjee, S.: Towards stratification learning through homology inference. AAAI Fall Symposium on Manifold Learning and its Applications (2010)
8. Stolz, B., Tanner, J., Harrington, H., Nanda, V.: Geometric anomaly detection in data. Proc. Natl. Acad. Sci. **117**(33), 19664–19669 (2020). https://doi.org/10.1073/pnas.2001741117
9. Aanjaneya, M., Chazal, F., Chen, D., Glisse, M., Guibas, L.J., Morozov, D.: Metric graph reconstruction from noisy data. Int. J. Comput. Geom. Appl. **22**, 205–325 (2012). https://doi.org/10.1142/S0218195912600072
10. Chazal, F., Cohen-Steiner, D., Lieutier, A.: A sampling theory for compact sets in euclidean space. Discrete & Comput. Geom. **41**(3), 461–479 (2009). https://doi.org/10.1007/s00454-009-9144-8
11. Bendich, P., Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Morozov, D.: Inferring local homology from sampled stratified spaces. In: 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), pp. 536–546 (2007) https://doi.org/10.1109/FOCS.2007.45
12. Carlsson, G.: Topological pattern recognition for point cloud data. Acta Numer. **23**, 289–368 (2014). https://doi.org/10.1017/S0962492914000051
13. Chazal, F., Lieutier, A.: Stability and computation of topological invariants of solids in $\mathbb{R}^n$. Discrete & Comput. Geom. **37**(4), 601–617 (2007). https://doi.org/10.1007/s00454-007-1309-8
14. Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2000)