Edge Collapse and Persistence of Flag Complexes

Jean-Daniel Boissonnat, Siddharth Pritam

Edge Collapse and Persistence of Flag Complexes*

Jean-Daniel Boissonnat
Université Côte d'Azur, INRIA, Sophia Antipolis, France
Jean-Daniel.Boissonnat@inria.fr

Siddharth Pritam
Université Côte d'Azur, INRIA, Sophia Antipolis, France
siddharth.pritam@inria.fr

Abstract

In this article, we extend the notions of dominated vertex and strong collapse of a simplicial complex as introduced by J. Barmak and E. Miniam. We say that a simplex (of any dimension) is dominated if its link is a simplicial cone. Domination of edges appear to be very powerful and we study it in the case of flag complexes in more detail. We show that edge collapse (removal of dominated edges) in a flag complex can be performed using only the 1-skeleton of the complex. Furthermore, the residual complex is a flag complex as well. Next we show that, similar to the case of strong collapses, we can use edge collapses to reduce a flag filtration $\mathcal{F}$ to a smaller flag filtration $\mathcal{F}^c$ with the same persistence. Here again, we only use the 1-skeletons of the complexes. The resulting method to compute $\mathcal{F}^c$ is simple and extremely efficient and, when used as a preprocessing for Persistence Computation, leads to gains of several orders of magnitude wrt the state-of-the-art methods (including our previous approach using strong collapse). The method is exact, irrespective of dimension, and improves performance of Persistence Computation even in low dimensions. This is demonstrated by numerous experiments on publicly available data.

2012 ACM Subject Classification Mathematics of computing, Topological Data Analysis, Computational geometry

Keywords and phrases Computational Topology, Topological Data Analysis, Strong Collapse, Persistent homology

Lines 500

1 Introduction

Improving the performance of computing persistent homology has been a central goal in Topological Data Analysis (TDA) since the early days of the field about 20 years ago. Very significant progress has been obtained on the two main components of the overall pipeline : the actual computation of persistence homology (PH) and the preprocessing of the sequence of complexes given as input. The first line of research led to improvement of the persistence algorithm and of its analysis, to efficient implementations and optimizations, and to a new generation of software [37, 8, 6, 45]. The other and complementary direction has been intensively explored with the goal of reducing the size of the complexes in the input sequence while preserving (or approximating in a controlled way) the persistent homology of the sequence [44, 30, 18, 13, 51, 41, 20, 27]. Among the most widely used complexes in TDA

* This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP/2007- 2013) / ERC Grant Agreement No. 339025 GUDHI (Algorithmic Foundations of Geometry Understanding in Higher Dimensions).

are the flag complexes and, in particular, the Vietoris-Rips complexes. These complexes are of great theoretical and practical interest since they are fully characterized by their graph (or 1-skeleton) and can thus be stored in a very compact way. Specific algorithms and very efficient codes have been developped for those complexes [6, 51]. Despite all these advances, further decisive progress was obtained very recently both for general simplicial complexes [12] and for flag complexes [11] using a special type of collapses, called strong collapses, introduced by J. Barmak and E. Miniam [5]. The basic idea is to simplify the complexes of the input sequence by using strong collapses and to compute the PH of an induced sequence of reduced simplicial complexes whose PH is the same or a close approximation of the PH of the initial sequence. In the case of flag complexes, the critical observation was that the construction of the reduced sequence can be done using only the 1-skeletons of the complexes, without constructing the full complexes, therefore saving time and space.

This paper further improves on these last results. Although the general philosophy is the same, there are some new key features that make the new method several orders of magnitude more efficient than all known methods.

1. Instead of strong collapses, we use the so-called edges collapses. In fact, we more generally define k-collapses that are identical to the *extended collapses* introduced in [4] (see also the early work of V. Welker [53]). When $k = 0$, we have strong collapses and when $k = 1$ edge collapses. Edge collapses share with strong collapses some important properties. Most notably, we can use edge collapses to reduce flag filtrations $\mathcal{F}$ to smaller flag filtrations $\mathcal{F}^c$ with the same persistence, using only the 1-skeletons of the complexes.

2. The reduction is exact and the PH of the reduced sequence is identical to the PH of the input sequence. Our algorithm thus computes the exact PH as does [6] but differs from [51, 11] where provably good approximations were computed.

3. In [12] and in [11], the reduced sequence associated to a filtration was usually a tower (a sequence of simplicial complexes connected through simplicial maps), and part of the computing time was devoted to transforming this tower in another equivalent filtration using ideas from [26, 40]. There is no such need in the algorithm presented in this paper, which is another main source of improvement.

4. The resulting method is simple and extremely efficient. On the theory side, we show that the edge collapse of a flag filtration can be computed in time $O(n\, n_c\, k^2)$, where n and n_c are the number of edges in the input and output 1-skeletons respectively and k is the maximal degree of a vertex in the input graph. The algorithm has been implemented. Numerous experiments on publicly available data show that the PH computation of flag complexes using edge collapse is much faster than with previous methods, and can even solve cases that were out of reach before. The code will be soon released in the Gudhi library [37].

An outline of this paper is as follows. Section 2 recalls some basic ideas and constructions related to simplicial complexes and simple collapses. We introduce k-collapse and then edge collapse in Section 3. In Section 4, we prove that simple collapse preserves persistence. In Section 5, we provide the main algorithm that reduces a flag filtration to another flag filtration using edge collapse. Experiments are discussed in Section 6.

2 Preliminaries

In this section we provide some background material. Readers can refer to [38] for a comprehensive introduction to these topics.

Simplex, simplicial complex and simplicial map. An **abstract simplicial complex** K is a collection of subsets of a non-empty finite set X, such that for every subset A in K, all the subsets of A are in K. From now on, we will call an *abstract simplicial complex* simply a *simplicial complex* or just a *complex*. An element of K is called a **simplex**. An element of cardinality $k+1$ is called a k-simplex and k is called its **dimension**. Given a simplicial complex K, we denote its geometric realization as $|K|$. A simplex is called **maximal** if it is not a proper subset of any other simplex in K. A sub-collection L of K is called a **subcomplex**, if it is a simplicial complex itself.

A map $\psi : K \rightarrow L$ between two simplicial complexes is called a **simplicial map**, if it always maps a simplex in K to a simplex in L. Simplicial maps are induced by vertex-to-vertex maps. A simplicial map $\psi : K \rightarrow L$ between two simplicial complexes K and L induces a continuous map $|\psi| : |K| \rightarrow |L|$ between the underlying geometric realizations. Any general simplicial map can be decomposed into more elementary simplicial maps, namely **elementary inclusions** (i.e., inclusions of a single simplex) and **elementary contractions** $\{\{u, v\} \mapsto u\}$ (where a vertex is mapped onto another vertex). The inverse operation of inclusion is called **simplicial removal** denoted as $K \hookleftarrow L$, where L is a subcomplex of K.

Flag complex and Neighborhood. A complex K is a flag or a clique complex if, when a subset of its vertices has pairwise edges between them, they span a simplex. It follows that the full structure of K is determined by its 1-skeleton (or graph) we denote by G. For a vertex v in G, the **open neighborhood** $N_G(v)$ of v in G is defined as $N_G(v) := \{u \in G \mid [uv] \in E\}$, here E is the set of edges of G. The **closed neighborhood** $N_G[v]$ is $N_G[v] := N_G(v) \cup \{v\}$. Similarly we define the closed and open neighborhood of an edge $[xy] \in G$, $N_G[xy]$ and $N_G(xy)$ as $N_G[xy] := N[x] \cap N[y]$ and $N_G(xy) := N(x) \cap N(y)$, respectively. The above definitions can be extended to any k-clique $\sigma = [v_1, v_2, ..., v_k]$ of G; $N_G[\sigma] := \bigcap_{v_i \in \sigma} N[v_i]$ and $N_G(\sigma) := \bigcap_{v_i \in \sigma} N(v_i)$.

Star, Link and Simplicial Cone. Let σ be a simplex of a simplicial complex K, the **closed star** of σ in K, $st_K(\sigma)$ is a subcomplex of K which is defined as follows, $st_K(\sigma) := \{\tau \in K \mid \tau \cup \sigma \in K\}$. The **link** of σ in K, $lk_K(\sigma)$ is defined as the set of simplices in $st_K(\sigma)$ which do not intersect with σ, $lk_K(\sigma) := \{\tau \in st_K(\sigma) | \tau \cap \sigma = \emptyset\}$. The **open star** of σ in K, $st^o_K(\sigma)$ is defined as the set $st_K(\sigma) \setminus lk_K(\sigma)$. It is not a subcomplex of K.

Let L be a simplicial complex and a a vertex not in L. Then the simplicial cone aL is defined as $aL := \{a, \tau \mid \tau \in L \;\; or \;\; \tau = \sigma \cup a; \;\; \text{where} \;\; \sigma \in L\}$.

Sequences of complexes. A **sequence** of simplicial complexes $\mathcal{T} : \{K_1 \xrightarrow{f_1} K_2 \xrightarrow{f_2} K_3 \xrightarrow{f_3} \cdots \xrightarrow{f_{(m-1)}} K_m\}$, connected through simplicial maps f_i is called a **simplicial tower** or simply a *tower*. We call a tower a **flag tower** if all the simplicial complexes K_i are flag complexes. When all the simplicial maps the f_i are inclusions, then the tower is called a **filtration** and a flag tower is called a **flag filtration**.

Persistent homology. If we compute the homology classes of all the K_i, we get the sequence $\mathcal{P}(\mathcal{T}) : \{H_p(K_1) \xrightarrow{f_1^*} H_p(K_2) \xrightarrow{f_2^*} H_p(K_3) \xrightarrow{f_3^*} \cdots \xrightarrow{f_{(m-1)}^*} H_p(K_m)\}$. Here $H_p()$ denotes the homology class of dimension p with coefficients from a field $\mathbb{F}$ and f_i^* is the homomorphism induced from f_i. $\mathcal{P}(\mathcal{T})$ is a sequence of vector spaces connected through the f_i^* called a **persistence module**. More formally, a *persistence module* $\mathbb{V}$ is a sequence of vector spaces $\{V_1 \to V_2 \to V_3 \to \cdots \to V_m\}$ connected with homomorphisms $\{\to\}$ between them. A persistence module arising from a sequence of simplicial complexes captures the evolution of the topology of the sequence.

Any persistence module can be *decomposed* into a collection of intervals of the form $[i, j)$ [14]. The multiset of all the intervals $[i, j)$ in this decomposition is called the **persistence diagram** of the persistence module. An interval of the form $[i, j)$ in the persistence diagram of $\mathcal{P}(\mathcal{T})$ corresponds to a homological feature (a 'cycle') which appeared at i and disappeared at j. The persistence diagram (PD) completely characterizes the persistence module, that is, there is a bijective correspondence between the PD and the equivalence class of the persistence module [14, 58].

Two different persistence modules $\mathbb{V} : \{V_1 \to V_2 \to \cdots \to V_m\}$ and $\mathbb{W} : \{W_1 \to W_2 \to \cdots \to W_m\}$, connected through a set of homomorphisms $\phi_i : V_i \to W_i$ are **equivalent** if the ϕ_i are isomorphisms and the following diagram commutes [14, 24]. Equivalent persistence modules have the same interval decomposition, hence the same diagram.

$$
\begin{array}{ccccccccc}
V_1 & \longrightarrow & V_2 & & \cdots & & \longrightarrow & V_{m-1} & \longrightarrow & V_m \\
\downarrow{\scriptstyle \phi_1} & & \downarrow{\scriptstyle \phi_2} & & & & & \downarrow{\scriptstyle \phi_{m-1}} & & \downarrow{\scriptstyle \phi_m} \\
W_1 & \longrightarrow & W_2 & & \cdots & & \longrightarrow & W_{m-1} & \longrightarrow & W_m
\end{array}
$$

Simple collapse. Given a complex K, a simplex $\sigma \in K$ is called a **free simplex** if σ has a unique coface $\tau \in K$. The pair $\{\sigma, \tau\}$ is called a **free pair**. The action of removing a free pair: $K \to K \setminus \{\sigma, \tau\}$ is called an **elementary simple collapse.** A series of such elementary simple collapses is called a **simple collapse**. We denote it as $K \searrow L$. This operation preserves the homotopy type of the simplicial complex K, which we write $K \sim L$. In particular, there is a retraction map $|r| : |K| \to |L|$ between the underlying geometric realization of K and L which is a strong deformation retraction. A complex K' will be called **simply-minimal** if there is no free pair $\{\sigma, \tau\}$ in K'. A subcomplex K^{ec} of K is called **elementary core** of K if $K \searrow K^{ec}$ and K^{ec} is simply-minimal.

Removal of a simplex. We denote by $K \setminus \sigma$ the subcomplex of K obtained by removing σ, i.e. the complex that has all the simplices of K except the faces and the cofaces of σ.

3 Edge Collapse

In this section, we first extend the definition of a dominated vertex introduced in [5] to simplices of any dimension. Given a simplex $\sigma \in K$, we denote by Σ_σ the set of maximal (for the inclusion) simplices of K that contain σ. The intersection of all the maximal simplices in Σ_σ will be denoted as $\bigcap \Sigma_\sigma := \bigcap_{\tau \in \Sigma_\sigma} \tau$.

Dominated simplex. A simplex σ in K is called a **dominated simplex** if the link $lk_K(\sigma)$ of σ in K is a simplicial cone, i.e. if there exists a vertex $v' \notin \sigma$ and a subcomplex L of K, such that $lk_K(\sigma) = v'L$. We say that the vertex v' is *dominating* σ and that σ is *dominated* by v', which we denote as $\sigma \prec v'$.

k**-collapse.** Given a complex K, the action of removing a dominated k-simplex σ from K is called an **elementary** k**-collapse**, denoted as $K \searrow\searrow^k \{K \setminus \sigma\}$. A series of elementary k-collapses is called a k**-collapse**, denoted as $K \searrow\searrow^k L$. We further call a complex K k**-minimal** if it does not have any dominated k simplices. A subcomplex K^o of K is called a k**-core** if $K \searrow\searrow^k K^o$ and K^o is k-minimal.

The notion of k-collapse is the same as the notion of *extended collapse* introduced in [4]. We give it a different name to indicate the dependency on the dimension. A 0-collapse is a strong collapse as introduced in [5]. A 1-collapse will be called an **edge collapse**. It is not hard to see that an elementary simple collapse of a k-simplex σ is a k-collapse, as it is dominated by the vertex $v = \tau \setminus \sigma$, where τ is the unique coface containing σ.

The following lemma extends a result in [5] to general k-collapse. It shows that the domination of a simplex can be characterized in terms of maximal simplices.

▶ **Lemma 1.** *A simplex $\sigma \in K$ is dominated by a vertex $v' \in K$, $v' \notin \sigma$, if and only if all the maximal simplices of K that contain σ also contain v', i.e. $v' \in \bigcap \Sigma_\sigma$.*

Proof. Since $\sigma \prec v'$, $lk_K(\sigma) = v'L$ by definition. This implies that for any maximal simplex τ in $st_K(\sigma)$, $v' \in \tau$. Therefore, $v' \in \bigcap \Sigma_\sigma$. For the reverse direction, let $v' \in \bigcap \Sigma_\sigma$. Therefore, for any maximal simplex τ in $st_K(\sigma)$, we have $v' \in \tau$. Now since $v' \notin \sigma$, v' belong to all the simplices $\tau \setminus \sigma$, therefore $lk_K(\sigma) = v'L$. Hence $\sigma \prec v'$ if and only if $v' \in \bigcap \Sigma_\sigma$. ◀

Lemma 1 has important algorithmic consequences. To perform a k-collapse, one simply needs to store the adjacency matrix between the k-simplices and the maximal simplices of K.

Next we study the special case of a flag complex K and characterize the domination of a simplex σ of a flag complex K in terms of its neighborhood.

▶ **Lemma 2.** *Let σ be a simplex of a flag complex K. Then σ will be dominated by a vertex v' if and only if $N_G[\sigma] \subseteq N_G[v']$.*

Proof. Assume that $N_G[\sigma] \subseteq N_G[v']$ and let τ be a maximal simplex of K that contains σ. For a vertex $x \in \tau$ and for any vertex $v \in \sigma$, the edge $[x, v] \in \tau$. Therefore $x \in N_G[\sigma] \subseteq N_G[v']$. Every vertex in τ is thus linked by an edge to v' and, since K is a flag complex and τ is maximal, v' must be in τ. This implies that all the maximal simplices that contains σ also contain v'. Hence σ is dominated by v'.

Consider the other direction. If $\sigma \prec v'$, by Lemma 1, all the maximal simplices containing σ also contains v'. This implies $N_G[\sigma] \subseteq N_G[v']$. ◀

The above lemma is a generalisation of Lemma 1 in [11]. The next lemma, though trivial, is of crucial significance. Both lemmas show that edge collapses are well suited to flag complexes.

▶ **Lemma 3.** *Let K be a flag complex and let L be any subcomplex of K obtained by edge collapse. Then L is also a flag complex.*

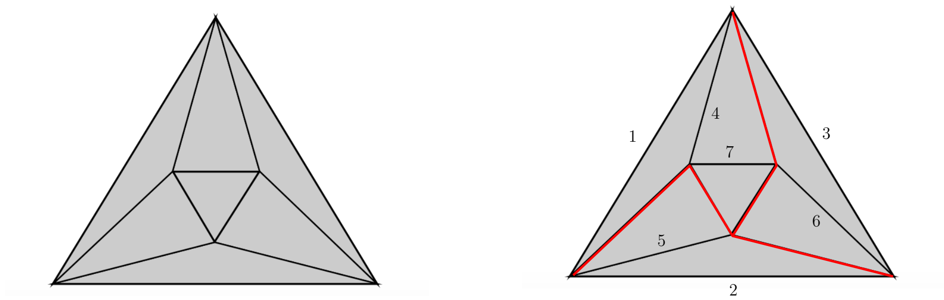

Figure 1 The above complex does not have any dominated vertex. However, by proceeding from the edges at the boundary one can edge collapse this complex to a 1-dimensional complex. The 1-core obtained in this way can be further reduced to a point using 0-collapse.

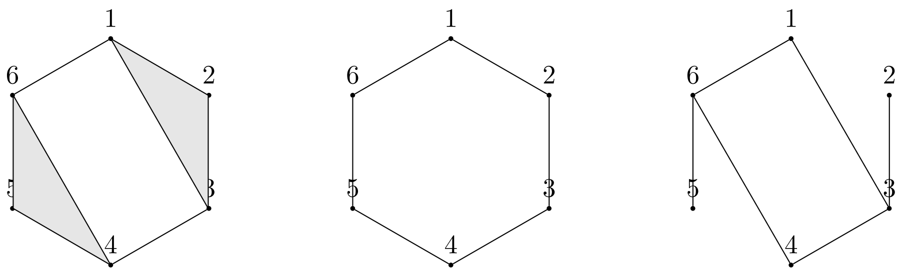

Figure 2 The complex in the left has two different 1-cores, the one in the middle is obtained after removing the inner edges $[1,3]$ and $[4,6]$, and the one in the right by removing outer edges $[1,2]$ and $[4,5]$. Note that the one in the right has further possibility of strong collapse (0-collapse).

Efficiency of reduction. In the practical cases we have considered (see Section 6), the ability of edge collapse to reduce the size of a complex is either comparable or superior than that of vertex collapse. This is llustrated in Figure 1, see also the **torus** example in Section 6.

NP Completeness. Unlike strong collapses (0-collapses), edge collapses (1-collapses) do not guarantee to have a unique core as illustrated in Figure 2. This leads to the natural optimization problem of computing an optimal sequence of edge collapses. However, this problem is difficult. More precisely, the following variant : Given a simplicial complex K, is it possible to compute a smallest 1-core (in terms of the number of edges), is strongly $\mathcal{NP}$-complete as shown below. To prove this, we will first recall a result of Egeciglu and Gonzalez [33]. Let K be a connected pure 2-dimensional simplicial complex that is embeddable in $\mathbf{R}^3$ and consider the following decision problem: given an integer k, does there exist a subset S of 2-simplices of K with $|S| \leq k$ such that $K \setminus S$ simply collapses to a 1-dimensional subcomplex of K. This problem is strongly $\mathcal{NP}$-complete [33].

We will show that, for a 2-dimensional complex, elementary edge collapses and elementary simple collapses are equivalent. It will then follow that finding an optimal edge collapse is $\mathcal{NP}$-complete as well.

▶ **Lemma 4.** *Let e be an edge of a 2-dimensional complex K. Then e is dominated in K if and only if it is free, i.e. if it has a unique coface.*

Proof. Let $e = [xy]$ be a dominated edge of K. By Lemma 1, there exists a vertex $v' \notin e$ of

K such that $v' \in \bigcap \Sigma_e$. Hence, the 2-simplex $[x, y, v'] \in \bigcap \Sigma_e$. Now, as K is 2 dimensional, the maximal simplices are 2-dimensional and we must have $[x, y, v'] = \Sigma_e$. This implies that e is a free edge. The other direction is obvious since a free edge is always dominated. ◀

Using the above lemma and the result by Egeciglu and Gonzalez [33], we get:

▶ **Theorem 5.** *Let K be a simplicial complex that is connected, pure, 2-dimensional and embeddable in $\mathbf{R}^3$, and let k be an integer. It is strongly $\mathcal{NP}$-complete to decide whether there exists a subset S of 2-simplices of K, with $|S| \leq k$, such that there is an edge collapse from $K \setminus S$ to a 1-dimensional subcomplex of K.*

4 Simple Collapse and Persistence

In this section we provide one of the main result of this article. This can be seen as a generalization of Theorem 2 of [12].

▶ **Theorem 6.** *Let $f : K \to L$ be a simplicial map between two complexes K and L and let K^{ec} and L^{ec} be the associated elementary cores. Then there exists a map $f^{ec} : K^{ec} \leftrightarrow L^{ec}$, induced by f, such that the persistence of $f^* : H_p(K) \to H_p(L)$ and $f^{ec^*} : H_p(K^{ec}) \leftrightarrow H_p(L^{ec})$ are the same for any integer $p \geq 0$. The induced map f^{ec} may not be simplicial. Nevertheless, it can be expressed as a combination of inclusions, contractions and removals of simplices.*

Proof. Let us consider the following diagram between the geometric realizations of the complex $|K|$, $|L|$, $|K^{ec}|$ and $|L^{ec}|$.

$$\begin{array}{ccc} |K| & \xrightarrow{|f|} & |L| \\ {\scriptstyle |i_k|}\uparrow\downarrow{\scriptstyle |r_k|} & & {\scriptstyle |i_l|}\uparrow\downarrow{\scriptstyle |r_l|} \\ |K^{ec}| & \xrightarrow{|f^{ec}|} & |L^{ec}| \end{array}$$

and the associated diagram after computing the p-th singular homology groups

$$\begin{array}{ccc} H_p^o(|K|) & \xrightarrow{|f|^*} & H_p^o(|L|) \\ {\scriptstyle |i_k|^*}\uparrow\downarrow{\scriptstyle |r_k|^*} & & {\scriptstyle |i_l|^*}\uparrow\downarrow{\scriptstyle |r_l|^*} \\ H_p^o(|K^{ec}|) & \xrightarrow{|f^{ec}|^*} & H_p^o(|L^{ec}|) \end{array}$$

Here $|r_k|$ and $|r_l|$ are the deformation retractions on the geometric realizations associated with the simple collapse and $|i_k|$ and $|i_l|$ are the inclusion maps. $H_p^o()$ denotes the singular homology and $*$ is the induced homomorphisms by the corresponding continuous maps. The map $|f^{ec}|$ is defined as $|f^{ec}| := |r_l||f||i_k|$. Now by definition $|f^{ec}||r_k| = |r_l||f||i_k||r_k|$. And $|r_l||f||i_k||r_k| \sim |r_l||f|$ (homotopic) since $|r_k|$ is a deformation retraction, therefore $|i_k||r_k|$ is homotopic to the identity over $|K|$. Since homotopic maps induce identical homomorphisms on the corresponding homology groups, we have $|f^{ec}|^*|r_k|^* = |r_l|^*|f|^*$ (commutativity) [38, Proposition 2.19]. Also, since $|r_k|^*$, $|r_l|^*$ are induced by deformation retractions, they are isomorphisms on their respective singular homology groups. This proves that the two maps $|f| : |K| \to |L|$ and $|f^{ec}| : |K^{ec}| \to |L^{ec}|$ have the same singular persistent homology. Now for simplicial complexes, singular homology is isomorphic to simplicial homology [38, Theorem

2.27]. Since $|f^{ec}| = |r_l||f||i_k|$ and the inclusion i_k and f are simplicial except r_l which is removal of simplices, f^{ec} can be expressed as composition of inclusions, contractions and removals of simplices. Therefore, we deduce that the persistent simplicial homologies $f^* : H_p(K) \to H_p(L)$ and $f^{ec^*} : H_p(K^{ec}) \leftrightarrow H_p(L^{ec})$ are equivalent. ◀

The use of singular homology in the proof is due to the lack of a simplicial map associated with the retraction ($|r|$) of simple collapse. Due to the same reason, the induced map $f^{ec} : K^{ec} \leftrightarrow L^{ec}$ between the elementary cores may not be necessarily simplicial. Nevertheless, the proof explicitly constructs this map and shows that it can be expressed as a combination of inclusions, contractions and removals of simplices. When a sequence of simplicial complexes contains removals of simplices, it is called a zigzag sequence. There are algorithms [45, 42] to compute zigzag persistence but they are not as efficient as the usual algorithms for filtrations and towers.

In the next section, we consider the case of flag filtrations and show that we can restrict the way the edge collapses are performed so that the reduced filtration is also a flag filtration.

5 Edge collapse of a flag filtration

In Section 3, we have introduced edge collapse for general simplicial complexes and provided an easy criterion for edge-domination in a flag complex using only the 1-skeleton of the complex. In this section, we provide an algorithm to simplify a flag filtration using edge collapse and again using only the 1-skeleton of the complex.

The correctness of the following algorithm rely on the notion of **removable edge**. Let G be a graph and K be the associated flag complex. We say that an edge e in a graph G is removable either if it is dominated in K or if there exists a sequence of edge collapses in K such that e is dominated in the reduced complex K^c. Let $G^c \subseteq G$ be the 1-skeleton of K^c. The flag complexes K and K^c are homotopy equivalent and we say that G and G^c are **edge-equivalent**, which we denote as $G \sim_e G^c$.

Algorithm. Let $\mathcal{F} : K_1 \hookrightarrow K_2 \hookrightarrow \cdots \hookrightarrow K_n$ be a flag filtration and $\mathcal{G}_{\mathcal{F}} : G_1 \hookrightarrow G_2 \hookrightarrow \cdots \hookrightarrow G_n$ be the associated sequence of 1-skeletons. We further assume that $G_i \hookrightarrow G_{i+1}$ is an elementary inclusion, namely the inclusion of a single edge we name e_{i+1}. The edges in $E := \{e_1, ..., e_n\}$ are thus indexed by their order in the filtration and we denote by G_i the subset $\{e_1, ..., e_i\}$. Our algorithm computes a subset of edges $E^c \subseteq E$ and attach to each edge in E^c a new index. We thus obtain a new sequence of flag complexes $\mathcal{F}^c$, we call the *core sequence*. The construction of E^c and of the new indices is done so that $\mathcal{F}^c$ has the same persistence diagram as $\mathcal{F}$.

We now explain how to compute E^c. See [Algorithm 1] for the pseudo-code. The main **for** loop on line 6 (called the forward loop) iterates over the edges in the filtration $\mathcal{F}$ by increasing filtration values, i.e. in the *forward direction*, and check whether or not the current edge e_i is dominated in the graph G_i. If *not*, we insert e_i in the set E^c and assign i as the new index of e_i (i.e. we keep the original index). Note that we check the domination of e_i in G_i, not in the final graph G_n. The non-domination of e_i in G_i implies that G_i and G_{i-1} are not edge equivalent and therefore the status of some edges that were dominated in G_{i-1} can change to non-dominated. This is why, after the insertion of edge e_i in E^c, we trigger another search in G_i by decreasing filtration values, i.e. in the *reverse direction* ([Line 9-26]), called the backward loop).

If $e = [u, v]$, we define the edge-neighborhood of an edge $e \in G$ as $NEIGHBORS(e, G) = \{[x, y], x \in \{u, v\}, y \in N_G([uv])\}$. Notice that the only edges that can change their status are in the edge-neighbourhood of an edge that has been inserted in E^c (Lemma 8). To benefit from this fact and to restrict the search, we assign G_i to a temporary graph G, and we assign the edge-neighborhood of e_i in the graph G_i to E^{nbd} [Line 9-10]. Thereafter, we iterate through the edges of G_i in decreasing order of their indices [Line 12-26]. Specifically, we proceed as follows. If an edge $e_j \notin E^c$ and $e_j \notin E^{nbd}$ [Line 13-14], e_j is still dominated and we remove it from G [Line 22]. If $e_j \notin E^c$ and $e_j \in E^{nbd}$, then we check whether it is dominated or not. If e_j is dominated, we remove it from G [Line 19]. Otherwise, we insert e_j in E^c and assign to it the ***new*** *index* i, i.e. the index of the edge e_i that has triggered the backward search in G_i. Next we enlarge the edge-neighborhood E^{nbd} by inserting the edge-neighbors of e_j in G. We then repeat this process [Line 12-26] until the last index $(j = 1)$ in G_i. Upon termination of the forward loop [Line 6-30], we output E^c as the final set.

We now prove the correctness of the above algorithm after some more definitions.

Critical Edges: Edges in E^c are called **critical** while edges in $E \setminus E^c$ are called **non-critical**. All edges have an original index i given by the insertion order in the input filtration $\mathcal{F}$. The critical edges received a second index j, called their **critical index**, when they are inserted in E^c. By convention, if an edge is not critical and thus has never been inserted in E^c, we will set its critical index to be ∞. Hence, at the end of Algorithm 1, each edge $e \in E$ has two indices, an original and a critical index. To make this explicit, we denote e as e_i^j. Clearly $i \leq j$. We further distinguish the cases $i = j$ and $i < j$. If $i = j$, e_i has been put in E^c during a forward move (forward loop) and we call e_i a **primary critical edge**. If $i < j$, e_i has been put in E^c during a backward move (backward loop) and we call it a **secondary critical edge**.

For $i = 1, ..., n$, we define the **critical graph** at index i, denoted G_i^c, whose edges are the edges in E^c with a critical index at most i. We denote the associated flag complex as K_i^c.

Correctness. We now prove some lemmas to certify the correctness of our algorithm. The following simple lemma justifies the fact that the search for new critical edges during the backward loop of Algorithm 1 is restricted to the neighborhood of already found critical edges.

▶ **Lemma 7.** *Let e be an edge in a graph G and let e' be a new edge. If e is dominated in G and does not belong to $EN_{G'}(e')$, then it is still dominated in $G' = G \cup e'$.*

The following lemma characterizes non-critical and critical edges in terms of being dominated or removable in certain specific graphs G_i. It essentially says that a non-critical edge is always removable and that a critical edge is removable until it becomes critical.

▶ **Lemma 8.** *Let e_i^j be an edge with $i < j$, then it is removable in all G_t, $i \leq t < \min(n+1, j)$.*

Proof. According to the algorithm, if $i < j$, e_i^j is dominated in G_i (j being finite or not).

Algorithm 1 Core flag filtration algorithm

```
procedure CORE-FLAG-FILTRATION(E)
    input : set of edges E of G_F sorted by filtration value.
    E^c ← ∅; i ← 1;
    E^nbd ← ∅
    G ← ∅
    for e_i ∈ E do                                  ▷ For i = 1, ..., n in increasing order
        if e_i is non-dominated in G_i then
            Insert {e_i, i} in E^c.
            G ← G_i
            E^nbd ← NEIGHBORS(e_i, G_i)
            j ← i − 1
            for e_j in G_i do                       ▷ For j = (i − 1), ..., 1 in decreasing order
                if e_j ∉ E^c then
                    if e_j ∈ E^nbd then
                        if e_j is non-dominated in G then
                            Insert {e_j, i} in E^c.
                            E^nbd ← E^nbd ∪ NEIGHBORS(e_j, G)
                        else
                            G ← G \ e_j
                        end if
                    else
                        G ← G \ e_j
                    end if
                end if
                j ← j − 1
            end for
        end if
        G ← ∅
        i ← i + 1
    end for
    return E^c                                      ▷ E^c is the 1-skeleton of the core flag filtration.
end procedure
```

Let us first consider the case $i < j = \infty$. Note that e_i^∞ is non-critical and let j_i be the smallest primary critical index greater than i. If no such index exists, set $j_i = n + 1$. We show by induction that e_i^∞ remains removable in all G_t, $i \geq t < n + 1$. As shown above, it is true for $t = i$ since e_i^j is dominated in G_i. So assume that e_i^j is removable in G_{t-1} and consider the insertion of e_t in G_t, for some $t < j_i$. By definition of j_i, e_t is dominated in G_t, which implies that e_i^j is removable in G_t.

Consider now $t = j_i$. Since e_{j_i} is a primary critical edge, it is non-dominated in G_{j_i}. According to the algorithm, a backward loop has been triggered at j_i. During this backward loop, e_i^∞ has *not* been inserted in E_c since its second critical index is ∞, This is only possible because e_i^∞ has been found to be dominated in G. Since G is initialized as G_{j_i}, it follows that e_i^∞ is removable in G_{j_i}. We can now proceed in a similar way for all $t, j_i < t < n + 1$.

Consider now the case $i < j \leq n$. The proof is very similar to the previous case. As e_i^j has not been inserted in E^c until the backward loop triggered at index j, e_i^j remains removable in all G_t, $i \leq t < j$. ◀

Note that our statement does not imply that a critical edge e_i^j, $i < j \leq n$, can never be removable in G_t, $t \geq j$. It just means that we are sure that it will remain removable until the point it becomes critical. By definition, $G_i \setminus G_i^c = \{e_t^m | \ t \leq i, m \geq i\}$ is the set of edges whose critical index $m \geq i$, including the non-critical edges ($m = \infty$). Using Lemma 8,

▶ **Lemma 9.** *For each i, $G_i^c \sim_e G_i$.*

The proof of the following theorem certifying the correctness of our algorithm follows directly through the application of Lemma 9 and Theorem 6.

▶ **Theorem 10.** *Let $\mathcal{F} : K_1 \hookrightarrow K_2 \hookrightarrow \cdots \hookrightarrow K_n$ be a flag filtration and $\mathcal{G}_{\mathcal{F}} : G_1 \hookrightarrow G_2 \hookrightarrow \cdots \hookrightarrow G_n$ be the associated sequence of 1-skeletons, such that $G_i \hookrightarrow G_{i+1}$ is an elementary inclusion of an edge e_{i+1}. Let G_i^c be the critical graph and K_i^c be its flag complex as defined before. Then the associated flag filtration of the critical edges, $\mathcal{F}^c : K_1^c \hookrightarrow K_2^c \hookrightarrow \cdots \hookrightarrow K_n^c$ is equivalent to $\mathcal{F}$.*

Proof. Let us consider the following diagram of the flag complexes for any $i \in \{1, ..., n\}$, where K_i^c is the flag complex of the critical graph G_i^c.

$$\begin{array}{ccc} K_i & \hookrightarrow & K_{i+1} \\ \uparrow\downarrow r_i & & \uparrow\downarrow r_{i+1} \\ K_i^c & \hookrightarrow & K_{i+1}^c \end{array}$$

Using Lemma 9, K_i is homotopic to K_i^c. And r_i is a deformation retraction induced by the corresponding edge collapse. Now let us consider the following diagram after computing the homology groups.

$$\begin{array}{ccc} H_p(K_i) & \hookrightarrow & H_p(K_{i+1}) \\ \uparrow\downarrow {r_i}^* & & \uparrow\downarrow {r_{i+1}}^* \\ H_p(K_i^c) & \hookrightarrow & H_p(K_{i+1}^c) \end{array}$$

The equivalence of the persistence then follows directly from the application of Theorem 6. ◀

Complexity: Write n_v for the total number of vertices, n for the total number of edges and k for the maximum degree of a vertex in G_n. We represent each graph G_i as an adjacency list, where every vertex stores a *sorted* list of at most k adjacent vertices. Additionally, we store the set of edges (E and E^c) as a separate data structure.

The cost of inserting and removing an edge from such an adjacency list is $\mathcal{O}(k)$. Since the size of $N_G[v]$ is at most k for any vertex v, the cost of computing $N_G[e]$ for an edge e is $\mathcal{O}(k)$. Checking if an edge e is dominated by a vertex $v \in N_G[e]$ reduces to checking whether $N_G[e] \subseteq N_G[v]$. Since all the lists are sorted, this operation takes $\mathcal{O}(k)$ time per vertex v, hence $\mathcal{O}(k^2)$ time in total.

Let us now analyze the worst-case time complexity of Algorithm 1. At each step i of the forward loop [Line 6], either e_i is dominated (which can be checked in $O(k^2)$ time) or an backward loop is triggered [Line 12]. The backward loop will consider all edges with (original) index at most i and check whether they are dominated or not. Writing n_c for the number of primary critical edges, the worst-case time complexity is $nk^2 + \sum_{i=1}^{n_c} n\, k^2 = \mathcal{O}(nn_ck^2)$. The space complexity is $\mathcal{O}(n)$. In practice, n_c is a small fraction of n (see Table 1).

6 Computational Experiments

Our algorithm [Algorithm 1] has been implemented for VR filtrations as a C++ module named EdgeCollapser. Our previous preprocessing method described in [11] to simplify VR filtrations using strong collapse is called the VertexCollapser (previously called the RipsCollapser). Both EdgeCollapser and VertexCollapser take as input a VR filtration and return the reduced flag filtration according to their respective algorithms.

We present results on five datasets **netw-sc**, **senate**, **eleg**, **HIV** and **torus**. The first four datasets are publicly available [22] and are given as the interpoint distance matrix of the points. The last dataset **torus** has 2000 points sampled in a spiraled fashion on a torus embedded in a 3-sphere of $\mathbf{R}^4$ [39]. The reported time includes the time of EdgeCollapser/VertexCollapser and the time to compute the persistent diagram (PD) using the Gudhi library [37].

The code has been compiled using the compiler 'clang-900.0.38' and all computations were performed on a '2.8 GHz Intel Core i5' machine with 16 GB of available RAM. Both EdgeCollapser and VertexCollapser work irrespective of the dimension of the complexes associated to the input datasets. However, the size of the complexes in the reduced filtration, even if much smaller than in the original filtration, might exceed the capacities of the PD computation algorithm. For this reason, we introduced, as in Ripser (a state of the art software to compute PH of VR complex [6]), a parameter *dim* and restricts the expansion of the flag complexes to a maximal dimension *dim*.

The experimental results using EdgeCollapser are summarized in Table 1. Observe that the reduction in the number of edges done by EdgeCollapser is quite significant. The ratio between the number of initial edges and the number of critical edges is approximately 20. If the number of edges in a graph is $|E|$ then the size of the $(k+1)$-cliques $\mathcal{O}(|E|^k)$. Therefore the reduction in the size of k-simplices can be as large as $\mathcal{O}(20^k)$. This is verified experimentally too, as the reduced complexes are small and of low dimension (column Size/Dim) compared to the input VR-complexes which are of dimensions respectively 57, 54 and 105 for the first three datasets **netw-sc**, **senate** and **eleg**. [1]

Comparison with VertexCollapser. The same set of experimental results using VertexCollapser are summarized in Table 2. VertexCollapser can be used in two modes: in the exact mode (step=0), the output filtration has the same PD as the input filtration while, in the approximate mode (step>0), a certified approximation is returned. For appropriate comparison, we use VertexCollapser in exact mode. It can be seen that EdgeCollapser is faster than VertexCollapser by approximately two orders of magnitude. The main reason for this is the efficient preprocessing algorithm behind EdgeCollapser. As it can be noticed in some

[1] The sizes of the complexes are so big that we could not compute the exact number of simplices.

cases, the reduction obtained using by VertexCollapser is better than using EdgeCollapser, but even in those cases EdgeCollapser is faster than VertexCollapser.

In terms of size reduction, EdgeCollapser either outperforms VertexCollapser by a big amount or is comparable. Some intuition can be gained from the case of *torus*. This is a well distributed point sets sampled from a manifold without boundary. The fact that there is no boundary implies that there are only few dominated vertices, which dramatically reduces the capacity of VertexCollapser to collapse. To better grasp this fact, one can play with examples of well distributed points on a circle or a sphere (without boundary) and on a disk (with boundary). Remarkably, EdgeCollapser does not face this problem. in this case.

EdgeCollapser computes the exact PD of the input filtration while VertexCollapser has an exact and an approximate modes, Results in Table 2 are obtained using the exact mode of VertexCollapser, while results in Table 1 [11] are obtained using the approximate mode. In both cases, EdgeCollapser performs much better than VertexCollapser. It would be easy to implement an approximate version of EdgeCollapser similarly to what has been done for VertexCollapser. Instead of triggering the backward loop of the algorithm [Line12-26] at each primary critical edge we find, we can trigger the backward loop at certain snapshot values only. See Section 5 of [11] for more details on the approximate methodology and description of snapshot.

Data	Pnt	Thrsld	EdgeCollapser +PD				
			Edge(I)/Edge(C)	Size/Dim	*dim*	Pre-Time	Tot-Time
netw-sc	379	5.5	8.4K/417	1K/6	∞	0.62	0.73
senate	103	0.415	2.7K/234	663/4	∞	0.21	0.24
eleg	297	0.3	9.8K/562	1.8K/6	∞	1.6	1.7
HIV	1088	1050	182K/6.9K	86.9M/?	6	491	2789
torus	2000	1.5	428K/14K	44K/3	∞	288	289

Table 1 The columns are, from left to right: dataset (Data), number of points (Pnt), maximum value of the scale parameter (Thrsld), Initial number of edges/Critical (final) number of edges *Edge(I)/Edge(C)*, number of simplices (Size) and dimension of the final filtration (Dim), parameter (*dim*), time (in seconds) taken by Edge-Collapser and total time (in seconds) including PD computation (Tot-Time).

Data	Pnt	Thrsld	VertexCollapser +PD					
			Size/Dim	*dim*	Pre-Time	Tot-Time	*Step*	Snaps
netw-sc	379	5.5	175/3	∞	366.46	366.56	0	8420
senate	103	0.415	417/4	∞	15.96	15.98	0	2728
eleg	297	0.3	835K/16	∞	518.36	540.40	0	9850
HIV	1088	1050	127.3M/?	4	660	3,955	4	184
torus	2000	1.5		4	∞*	∞	0	428K

Table 2 The columns are, from left to right: dataset (Data), number of points (Pnt), maximum value of the scale parameter (Thrsld), number of simplices (Size) and dimension of the final filtration (Dim), parameter (*dim*), time (in seconds) taken by VertexCollapser, total time (in seconds) including PD computation (Tot-Time), parameter *Step* (linear approximation factor) and the number of snapshots used (Snaps). *The last experiment (torus) could not finish (>12hrs) the preprocessing due to large number of snapshots and the size of the complex.

Comparison with Ripser. Ripser [6] computes the *exact* PD associated to the input filtration up to dimension *dim*. EdgeCollapser (as well as VertexCollapser) are not really competitors of Ripser since they act more as a preprocessing of the input filtration and do not compute Persistence Homology. Hence they can be associated to any software computing flag filtrations. Nevertheless, we run Ripser[2] on the same datasets as in Table 1 to demonstrate the benefit of using EdgeCollapser. Results are presented in Table 3. The main observation is that, in most of the cases, EdgeCollapser computes PD in all dimensions and outperforms Ripser, even when we restrict the dimension of the input filtration given to Ripser.

Data	Pnt	Threshold	Val		Val		Val	
			dim	Time	*dim*	Time	*dim*	Time
netw-sc	379	5.5	4	25.3	5	231.2	6	∞
senate	103	0.415	3	0.52	4	5.9	5	52.3
”	”	”	6	406.8	7	∞		
eleg	297	0.3	3	8.9	4	217	5	∞
HIV	1088	1050	2	31.35	3	∞		
torus	2000	1.5	2	193	3	∞		

Table 3 Time is the total time (in seconds) taken by Ripser. ∞ means that the experiment ran longer than 12 hours or crashed due to memory overload.

Acknowledgements. We want to thank Marc Glisse for useful discussions and Vincent Rouvreau for his help with Gudhi.

References

1 M. Adamaszek and J. Stacho. Complexity of simplicial homology and independence complexes of chordal graphs. *Computational Geometry: Theory and Applications*, 57:8–18, 2016.

2 D. Attali, A. Lieutier, and D. Salinas. Efficient data structure for representing and simplifying simplicial complexes in high dimensions. *International Journal of Computational Geometry and Applications (IJCGA)*, 22:279–303, 2012.

3 Dominique Attali and André Lieutier. Geometry-driven collapses for converting a čech complex into a triangulation of a nicely triangulable shape. *Discrete & Computational Geometry*, 54(4):798–825, 2015.

4 Dominique Attali, André Lieutier, and David Salinas. Vietoris-rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry*, 46(4):448–465, 2013.

5 J. A. Barmak and E. G. Minian. *Discrete and Computational Geometry*, pages 301–328.

6 U. Bauer. Ripser. URL: `https://github.com/Ripser/ripser`.

7 U. Bauer, M. Kerber, and J. Reininghaus. Clear and compress: Computing persistent homology in chunks. In *Topological Methods in Data Analysis and Visualization III, Mathematics and Visualization*, pages 103–117. 2014.

8 U. Bauer, M. Kerber, J. Reininghaus, and H. Wagner. PHAT – persistent homology algorithms toolbox. *Journal of Symbolic Computation*, 78, 2017.

9 J-D. Boissonnat and C. S. Karthik. An efficient representation for filtrations of simplicial complexes. In *ACM Transactions on Algorithms*, 2018.

[2] We used the command <./ripser inputData –format distances –threshold inputTh –dim inputDim >.

10 J-D. Boissonnat, C. S. Karthik, and S. Tavenas. Building efficient and compact data structures for simplicial complexes. *Algorithmica*, 79:530–567, 2017.

11 J-D. Boissonnat and S. Pritam. Computing persistent homology of flag complexes via strong collapses. *International Symposium on Computational Geometry (SoCG)*, 2019.

12 J-D. Boissonnat, S.Pritam, and D. Pareek. Strong Collapse for Persistence. In *26th Annual European Symposium on Algorithms (ESA 2018)*, volume 112, 2018.

13 M. Botnan and G. Spreemann. Approximating persistent homology in euclidean space through collapses. *In: Applicable Algebra in Engineering, Communication and Computing*, 26:73–101.

14 G. Carlsson and V. de Silva. Zigzag persistence. *Found Comput Math*, 10, 2010.

15 G. Carlsson, V. de Silva, and D. Morozov. Zigzag persistent homology and real-valued functions. *International Symposium on Computational Geometry (SoCG)*, pages 247–256, 2009.

16 G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian. On the local behavior of spaces of natural images. *In: International Journal of Computer Vision*, 76:1–12, 2008.

17 J. M. Chan, G. Carlsson, and R. Rabadan. Topology of viral evolution. *In: Proceedings of the National Academy of Sciences*, 110:18566–18571, 2013.

18 F. Chazal and S. Oudot. Towards persistence-based reconstruction in Euclidean spaces. *International Symposium on Computational Geometry (SoCG)*, 2008.

19 C. Chen and M. Kerber. Persistent homology computation with a twist. *In European Workshop on Computational Geometry (EuroCG)*, pages 197–200, 2011.

20 A. Choudhary, M. Kerber, and S. Raghvendra:. In *Polynomial-Sized Topological Approximations Using The Permutahedron*. International Symposium on Computational Geometry (SoCG), 2016.

21 H. Edelsbrunner D. Cohen-Steiner and J. Harer. Stability of persistence diagrams. *Discrete and Compututaional Geometry*, 37:103–120, 2007.

22 Datasets. URL: `https://github.com/n-otter/PH-roadmap/''`.

23 V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *In: Algebraic and Geometric Topology*, 7:339 – 358, 2007.

24 H. Derksen and J. Weyman. Quiver representations. *Notices of the American Mathematical Society*, 52(2):200–206, February 2005.

25 T. K. Dey, H. Edelsbrunner, S. Guha, and D. Nekhayev. Topology preserving edge contraction. *Publications de l'Institut Mathematique (Beograd)*, 60:23–45, 1999.

26 T. K. Dey, F. Fan, and Y. Wang. Computing topological persistence for simplicial maps. In *International Symposium on Computational Geometry (SoCG)*, pages 345–354, 2014.

27 T. K. Dey, D. Shi, and Y. Wang. *SimBa: An efficient tool for approximating Rips-filtration persistence via Simplicial Batch-collapse*. In European Symp. on Algorithms (ESA), pages 35:1–35:16, 2016.

28 T. K. Dey and R. Slechta. Filtration simplification for persistent homology via edge contraction. *International Conference on Discrete Geometry for Computer Imagery*, 2019.

29 C. H. Dowker. Homology groups of relations. *The Annals of Mathematics*, 56:84–95, 1952.

30 P. Dłotko and H. Wagner. Simplification of complexes for persistent homology computations,. *Homology, Homotopy and Applications*, 16:49–63, 2014.

31 H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010.

32 H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Compututational Geometry*, 28:511–533, 2002.

33 Omer Egecioglu and Teofilo F. Gonzalez. A computationally intractable problem on simplicial complexes. *Computational Geometry*, 6:85–98, 1996.

34 B. T. Fasy, J. Kim, F. Lecci, and C. Maria:. Introduction to the R-package tda. *CoRR abs/1411.1830*, 2014.

35 E. Fieux and J. Lacaze. Foldings in graphs and relations with simplicial complexes and posets. *Discrete Mathematics*, 312(17):2639 – 2651, 2012.

36 F. Le Gall. Powers of tensors and fast matrix multiplication. *ISSAC '*, 14:296–303, 2014.

37 Gudhi: Geometry understanding in higher dimensions. URL: `http://gudhi.gforge.inria.fr/`.

38 A. Hatcher. *Algebraic Topology.* Univ. Press Cambridge, 2001.

39 Benoît Hudson, Gary L. Miller, Steve Oudot, and Donald R. Sheehy. Topological inference via meshing. *International Symposium on Computational Geometry (SoCG)*, 2010.

40 M. Kerber and H. Schreiber:. Barcodes of towers and a streaming algorithm for persistent homology. *International Symposium on Computational Geometry (SoCG)*, 2017. `arXiv:1701.02208`.

41 M. Kerber and R. Sharathkumar. Approximate Čech complex in low and high dimensions. In *Algorithms and Computation*, pages 666–676. by Leizhen Cai, Siu-Wing Cheng, and Tak-Wah Lam. Vol. 8283. Lecture Notes in Computer Science, 2013.

42 C. Maria and S. Oudot. Zigzag persistence via reflections and transpositions. In *Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA) pp.* 181–199, January 2015.

43 N. Milosavljevic, D. Morozov, and P. Skraba. Zigzag persistent homology in matrix multiplication time. In *International Symposium on Computational Geometry (SoCG)*, 2011.

44 K. Mischaikow and V. Nanda. Morse theory for filtrations and efficient computation of persistent homology. *Discrete and Computational Geometry*, 50:330–353, September 2013.

45 D. Mozozov. Dionysus. URL: `http://www.mrzv.org/software/dionysus/`.

46 J. Munkres. *Elements of Algebraic Topology.* Perseus Publishing, 1984.

47 N. Otter, M. Porter, U. Tillmann, P. Grindrod, and H. Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science, Springer Nature*, page 6:17, 2017.

48 Steve Y. Oudot and Donald R. Sheehy. Zigzag zoology: Rips zigzags for homology inference. *Foundations of Computational Mathematics*, 15, 2015.

49 J. Perea and G. Carlsson. A Klein-bottle-based dictionary for texture representation. *In: International Journal of Computer Vision*, 107:75–97, 2014.

50 H. Schreiber. Sophia. URL: `https://bitbucket.org/schreiberh/sophia/`.

51 D. Sheehy. Linear-size approximations to the Vietoris–Rips filtration. *Discrete and Computational Geometry*, 49:778–796, 2013.

52 M. Tancer. Recognition of collapsible complexes is NP-complete. *Discrete and Computational Geometry*, 55:21–38, 2016.

53 Volkmar Welker. Constructions preserving evasiveness and collapsibility. *Discrete Mathematics*, 207(1):243 – 255, 1999.

54 J. H. C Whitehead. Simplicial spaces nuclei and m-groups. *Proc. London Math. Soc*, 45:243–327, 1939.

55 A. C. Wilkerson, H. Chintakunta, and H. Krim. Computing persistent features in big data: A distributed dimension reduction approach. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 11–15, 2014.

56 A. C. Wilkerson, T. J. Moore, A. Swami, and A. H. Krim. Simplifying the homology of networks via strong collapses. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 11–15, 2013.

57 A. Zomorodian. The tidy set: A minimal simplicial set for computing homology of clique complexes. In *International Symposium on Computational Geometry (SoCG)*, pages 257–266, 2010.

58 A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 2005.