Some Heuristics for the Homological Simplification Problem

Erin W. Chambers * Tao Ju † David Letscher ‡ Mao Li § Christopher Topp ¶
Yajie Yan ‖

Abstract

In this paper, we consider heuristic approaches for solving the homological simplification problem. While NP-Hard in general, we propose an algorithm that in practice significantly reduces topological noise from large datasets, such as those from medical or biological imaging.

1 Introduction

In this paper, we will consider the homological simplification problem. Introduced in [2], this asks: given a simplicial pair (C, N) with $C \subset N$, can the persistent homology group of this pair be realized as the homology of some intermediate complex? This problem is one way to approach the problem of topologically accurate simplification, where the goal is to take a "noisy" shape and simplify it to reach some desired topological structure. Such algorithms are useful in a wide range of applications, as any surface or region reconstruction algorithm on scanned input data is apt to contain errors, and hence a postprocessing phase to simplify it is necessary.

For any surface, the homological simplification problem is solvable [8]; this work actually solves the more general problem of finding an ϵ-simplification in a filtration. However, such existence results do not hold in 3-dimensional manifolds where there a filtrations of manifolds that do not have ϵ-simplifications [8]. For the homological simplification problem, it is NP-Hard to determine if a simplification exists for 3-manifolds, even if the complex is embedded in $\mathbb{R}^3$ [2].

In this paper, we consider a 2-phase heuristic algorithm to simplify voxelized shapes. We first use a persistent homology-based algorithm to identify candidate simplications. This approach is inspired by prior work to find "nice" generators for homology groups on surfaces [5], but we expand to find not just the generators of homology group, but also representatives in the larger space that kill those generators. Phase 2 is then a validation: given such a candidate simplfication, we must check that it does result in a global simplification, since adding such things can introduce new topological features. We apply this algorithm to a number of types of input data, to assess how successful the heuristic is in practice. We find that our simplificaitons are able to remove over 99% of topological errors in several real-world data sets.

2 Related Work

Cubical complexes: A *cubical complex* is built from a collection of cells that are points, intervals, squares, cubes and higher dimensional analogs with the condition that the intersection of any two cells is also a cell of the cubical complex. Formally, the cells are build from products of intervals. An interval is a unit interval $[k, k+1]$ or a degenerate interval $[k, k]$; in d-dimensional space, a cube is a product of d elementary intervals. Given 2 cubes x and y, x is a face of y if $x \subseteq y$. A cubical complex of dimension d is a collection of cubes that is closed under taking faces such that the intersection of any two faces is a common subface. In this paper, we will use cubical complexes to represent the topology of our shapes.

*Department of Computer Science, Saint Louis University, echambe5@slu.edu.

†Department of Computer Science and Engineering, Washington University, St. Louis taoju@cse.wustl.edu.

‡Department of Computer Science and Engineering, Saint Louis University, letscher@slu.edu.

§Donald Danforth Plant Science Center, mli@danforthcenter.org .

¶Donald Danforth Plant Science Center, ctopp@danforthcenter.org .

‖Department of Computer Science and Engineering, Washington University, St. Louis yajieyan@wustl.edu.

Digital topology: In imaging and computer graphics, voxelizations are among the most common shape representatives. With Euclidean space divided using a cubical lattice, each individual cube is a single *voxel* and a voxelization of a shape is a set of these voxels. A common connectivity model of voxels is called 6-connectivity, which considers two voxels adjacent if they share a common 2-dimensional face [12]. This results in a different topology than the actual union of the voxels, which would typically connect anything that shares an edge or vertex as well. To be topologically consistent with this model, we will work in the dual shape where there is a vertex for each voxel, and edge whenever two voxels share a common face, a square or 2-cell for any 4 voxels around an edge, and a 3-cell or voxel whenever there are 8 voxels surrounding a common vertex. Note that this is still a cubical complex, but it is not a pure cubical complex.

Homology: Homology groups and persistent homology on filtrations are commonly used tools to find topological features in spaces; we refer the reader to recent books covering the topic [6, 14]. Most relevant to our setting, cubical persistent homology has been considered in some prior work [18], including optimized data structures to do such calculations. The p^{th} Betti number of a space X, $\beta_p(X)$ is defined to be the rank of the p^{th} homology groups. Given map $f : C \to N$, we can define a persistent notion of Betti number, where the p^{th} Betti number of f is the rank of the induced map on homology: $\beta_p(C \to N) = rk(f^*(H_p(C)))$. Extending this to filtrations of more than 2 spaces precisely gives the notion of persistent homology groups (or their ranks).

Homological simplification: Consider two spaces $C \subset N \subset \mathbb{R}^3$. The homological simplification problem is generally phrased in terms of Betti numbers: we wish to find a space X such that the p-dimensional Betti number, $\beta_p(X)$ is equal to the Betti number of the inclusions $C \to N$, $\beta_p(C \to N)$. In three dimensions, this problem has been shown to be NP-Hard [2].

Topological repair: Various methods have been developed in different research communities for removing topological errors of surfaces in $\mathbb{R}^3$. In computer graphics, algorithms exist for modifying a given surface to either remove features smaller than a given size or achieve a specific topology (e.g., a single connected component with a prescribed genus) [19, 21, 3]. These methods make decisions of where and how to modify the surface solely based on the shape of the given surface, whereas homological simplification bases its decision on the persistence of topological features between two spaces. In medical image analysis, there has been active research on ractiying the topology of reconstructions of biological structures, such as the cortical surface in the human brain [16, 10, 15, 20]. However most of these methods are specialized for removing redundant handles and cannot deal with other types of topological noises such as disconnected components or cavities. Finally, in scientific visualization, a line of research aims at simpliying the topological structure (e.g., the Morse-Smale complex) of a scalar function [4, 8, 9, 17]. While these methods effectively remove topological noises on *all* level sets of the function, they are unncessarily expensive if the goal is to fix the topology of one level set. In addition, methods designed to work on functions in $\mathbb{R}^3$ are limited to reducing the set of maxima or minima, and hence they are less useful for removing topological handles on the level sets.

3 Heuristic Shape Simplification

Since we will use voxelized representatives of our input shapes, we will focus on the restriction of the homology simplification problem to voxelized shapes. We restate the problem as follows where we refer to C as the core, and N as the neighborhood, and we wish to find a space X which is somehow "in between" them with homology equal to the persistent homology of $C \to N$.

The voxelized homological simplification problem: Suppose $C \subset N \subset \mathbb{R}^3$ are voxel regions. Determine if there exists X such that $C \subset X \subset N$ and $\beta_p(X) = \beta_p(C \to N)$ for all p.

Proposition 1 *The voxelized homological simplification problem is NP-hard.*

Proof. [Proof sketch] We omit details due to space constraints, but just briefly note that the proof for simplicial complexes embedded in $\mathbb{R}^3$ is a reduction from 3SAT [2]. They construct gadgets that are easily modified to be built from voxels on an $O(m)$ by

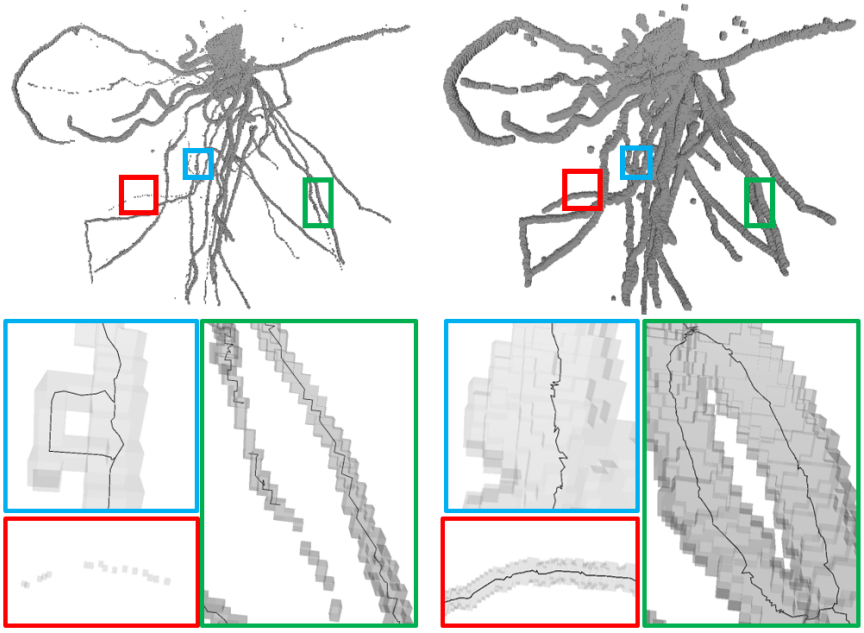

Figure 1: The core and neighborhood for the root system of corn. Note that in the inclusion of core into the neighborhood some loops are filled (blue inset), components connect (red inset) and new loops form (green inset). A solution to the homological simplification problem will accept the first two types of modifiations and reject the creation of new loops and voids.

$O(n)$ grid, where n is the number of booleans variables and m is the number of clauses in the 3SAT instance. □

It is typically impractical to solve the homological simpflication problem; however, we are concerned with simplifying a shape as much as possible. In particular, we will try to find a shape X where $C \subset X \subset N$ with each of the Betti numbers β_p as close as possible to the persistent Betti numbers $\beta_p(C \to N)$. If they are equal, we have a solution to the homological simplification problem. Even if we do not achieve maximal simplification, we will typically simplify our shape significantly.

Our simplification procedure will be a two phase process. It starts with the core C, which our algorithm will always include in the result, and then tries to expand to remove topological noise. First, a modified version of the standard persistence algorithm [7] is used to find sets of voxels whose addition would remove topologcial features that are not in the neighborhood. However, these candidates might add new features that are not desired. The second phase examines the modified shape and checks if the persistence Betti numbers have been reduced. If so, the changes are kept. This process of candidate generation and validation is repeated until no more valid additions can be found. There are no guarantees that this will remove all of the undesired topological features; however, our experiments in Section 5 show that in real world examples the accuracy of the simplification is very high.

Types of topological errors Consider $C \subset N \subset \mathbb{R}^3$ and the maps induced by inclusion $f_p^* : H_p(C) \to H_p(N)$ for $p = 0, \ldots, 3$. If every f_p^* is an injection then C is already a solution to the homological simplification problem. Any extraneous topological noise will appear in the kernel of some f_p^*. Consider a cycle $\alpha \neq 0$, where $f_p(\alpha) = 0$. α might connect two components, fill a void or loop in the large space. To interpret α, we assume α is *irreducible*, that is it cannot be written as a non-trivial sum of other elements of the kernel. Depending on the dimension p, we have different interpretations of the type of noise we observe. Each of these cases can be observed in the root data in Figure 1.

$\alpha \in \ker f_0^*$**:** Here α is represented by two points lying in different components of C. These components can be connected by a path γ in N that connects the two points. In this case, γ is a one dimensional chain in N. By adding γ to C we obtain a new space $C \cup \gamma$ where the kernel of $H_0(C \cup \gamma) \to H_0(N)$ has rank one less than $H_0(C) \to H_0(N)$.

$\alpha \in \ker f_1^*$**:** In this case α is a curve that follows a non-trivial topological feature in C, such as going around a handle, but which'' bounds some surface in N that kills that homological feature. If this surface is a disk D then $H_1(C \cup D) \to H_1(N)$ has rank one less than $H_1(C) \to H_1(N)$. In other words, we can add the disk D to remove the handle that is present in C. However, if the curve bounds a surface with genus, then if we were to add the surface in, a generator of the kernel is removed but two or more generators are added (since this surface has its own handles). Hence the resulting object is not simpler in terms of homology, and we cannot use this chain to simplify the space.

$\alpha \in \ker f_2^*$**:** Here α is a surface bounding a hollow region (a "void") in C, but which is not hollow in N. This void could be a simple ball, but could also contain topology. In either case, we can fill this in to simplify the topology. If the surface is a sphere, Alexander's theorem implies that it will bound a ball B in N [11]. In this case, it is easy to simplify since $H_2(C \cup B) \to H_2(N)$ has rank one less than $H_2(C) \to H_2(N)$. If the surface does not bound a ball, then the simplification may also simplify $\ker f_1$, as some loops will disappear when the void disappears.

The commonality of all of these cases is finding a $(k+1)$-dimensional structure in N whose boundary is in the kernel of f_k^* and whose addition simplifies the topological structure. Our techniques are based on this observation, repeatedly adding simplifications until no more can be found. Ideally, after this we obtain a space X with $C \subset X \subset N$ and $H_k(C) \to H_k(X)$ is a surjection and $H_k(X) \to H_k(N)$ is an injection for all k; in this case, we would actually have a solution to the homological simplification problem. Our algorithm will only simplify the shape, although we have no guarantees of optimality or approximation ratio. In Section 5, we will discuss our implementation and show that it works well on several data sets.

Candidate generation Consider $C \subset X \subset N$ where X is initially equal to C but is extended to removed topological features that are not shared by C and N. We will build a boundary matrix, ∂_p, for calculating the p-dimensional persistent homology of the filtration $C \subset X \subset N$. There is a column of ∂_p for each $(p+1)$-cell of X and a non-zero entry in that column for each p-cell that is a face of the $(p+1)$-cell. The rows and columns are ordered so that cells of C occur first, then cells of X and finally cells of N. The standard persistence algorithm [7] adds columns of ones to their right until a canonical reduced form is obtained. Many improvements to the algorithm have been made to make this calculation more efficient, but we are unable to utilize them as they discard information about generators along this way, which is the main data our approach requires.

After the reductions, the columns of ∂_p represent cycles involving the cells with non-zero entries. If that column corresponds to cells of N and the non-zero entries involved include d-cells of C, then this cycle, z, is a feature of C that does not persist in N and should be removed. To remove this feature we need to find a $(d+1)$ chain, b, in X such that $\partial_p b = z$. There are potentially many candidates for z; we find one by creating a copy of the identity matrix M with an entry for each column of ∂_p and performing the same column operations as we did in the reduction. An invariant of these column operations is that if v is the i-th column of M, then $\partial_p v$ is the i-th column of ∂_p. So we choose b to be the union of cells with non-zero entries in the column of M corresponding to the cycle z; recall that b is actually a chain in the dual cubical complex. We build a candidate simplification S which is the set of voxels that contain b. When S is added to X, the cycle z will become trivial; however, additional features could be added, so we cannot add S and guarantee reduction of Betti numbers.

Candidate validation To determine if the addition of S will actually simplify the shape, recall that our algorithm tries to minimize the sum of Betti numbers $B(X) = \sum_{p=0}^{3} \beta_p(X)$, with the goal of reducing this complexity to be equal to $\sum_{p=0}^{3} \beta_p(C \to N)$. If S is set of voxels that form a candidate simplification, we will calculate $B(X \cup S)$. If $B(X \cup S) < B(X)$ and $\beta_p(X \cup S) \leq \beta_p(X)$ for all p, then we will consider this a *valid simplification*. We cannot be sure that the addition of S has not

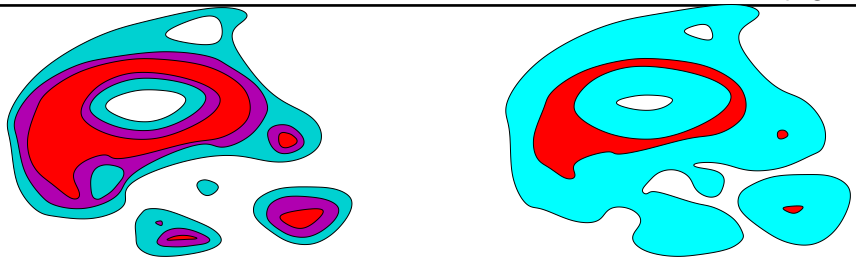

Figure 2: (a) Three intensity threshold of input: $t-\epsilon, t, t+\epsilon$. (b) Corresponding core and neighborhood.

created new topological feature, but our check does guarantee that more features are removed than were introduced.

Our algorithm We will repeat the process of finding candidate simplifications using the modified standard persistence algorithm and adding them one at a time if they are valid simplifications. This process repeats until no valid simplifications are found.

4 Removing Topological Errors in 3D Imaging

In a 3D imaging technology such as CT or MRI, the output is a grid of voxels each with an intensity value. A typical representation would be a map $f : \{0, \ldots, k-1\}^3 \to \mathbb{R}$. Shapes of interest, in theory, can be extracted by selecting voxels in a given intensity range. A typical segmentation might try to extract all voxels with intensity over a specified threshold, e.g. $X = f^{-1}([t, \infty))$. However, there can be a variety of errors in this segmentation.

There is a natural two dimensional filtration of the region where $X_{t,\delta} = N_\delta\left(f^{-1}([t,\infty))\right)$. Where t specifies the intensity threshold and IN_δ is in fact a morphological dilation when $\delta > 0$ or erosion when $\delta < 0$, using the ball of radius $|\delta|$ as the structuring element[13]. Formally, N_δ is a neighborhood of a set for positive δ and points with a $|\delta|$ neighborhood contained in the set for negative δ defined by

$$N_\delta(X) = \begin{cases} \cup_{x\in X} B_\delta(x) & \delta \geq 0 \\ \{x \mid B_{-\delta(x)} \subset X\} & \delta < 0 \end{cases}$$

If t is a threshold for an initial segmentation $X = f^{-1}([t,\infty))$, we can give two parameters, ϵ and δ, representing noise levels on the threshold value and geometric scale, respectively. We will define the core $C = X_{t+\epsilon,-\delta}$ and the neighborhood $N = X_{t-\epsilon,\delta}$. This collapses the two dimensional filtration into a pair of spaces $C \subset N$. In this case the core C represents voxels that we want to include in the final shape, they all have neighborhoods meeting a higher intensity threshold. And the neighborhood N constains all the voxels near some voxel with a lower intensity threshold. See Figure 2 for an example.

A good solution to the homological simplification with this C and N would have features that do not appear or disappear with a small change in threshold, expanding or contraction of the shape, or a combination of these changes. In practice, a partial solution to the homological simplification problem might only remove some of the noise; this still resuts in an improvement over simple thresholding techniques. In the following section, we will discuss our observations of a reduction in more than 99% of the noise in real world datasets.

5 Experimental Results

We experimented on three different collections of data: CT scans of corn root systems, synthetic root systems and brain volumes reconstructed from histological sections. The corn data was from a single variety of corn, with three different scans each viewed at two different resolutions. The synthetic root was designed to roughly resemble a root systems and was studied at nine different resolutions. The brain scans were of the BigBrain dataset [1] downsampled at ten different resolutions. Figure 3 gives some information about the datasets. The largest regions had close to 400 million voxels. This forced some additional techniques described below to handle the scale of the data sets. The largest core in these experiments has 2.5 million voxels and the largest neighborhood had over 4 million voxels. The most complicated example had approximately

	Region size	Core size	Neighborhood size	Complexity
Corn	$2.3 \times 10^6 - 1.5 \times 10^8$	$1.6 \times 10^3 - 2.4 \times 10^5$	$2.4 \times 10^3 - 4.6 \times 10^6$	$109 - 6,713$
Synthetic	$3.9 \times 10^6 - 3.8 \times 10^8$	$1.8 \times 10^3 - 1.7 \times 10^5$	$2.3 \times 10^3 - 2.7 \times 10^6$	$208 - 26,605$
Brain	$3.9 \times 10^5 - 1.7 \times 10^7$	$5.1 \times 10^4 - 2.5 \times 10^6$	$6.1 \times 10^5 - 3.6 \times 10^6$	$388 - 15,393$

Figure 3: Characteristics of the data analyzed, including the number of voxels in the region examined, core and neighborhood and the initial complexity of the shapes.

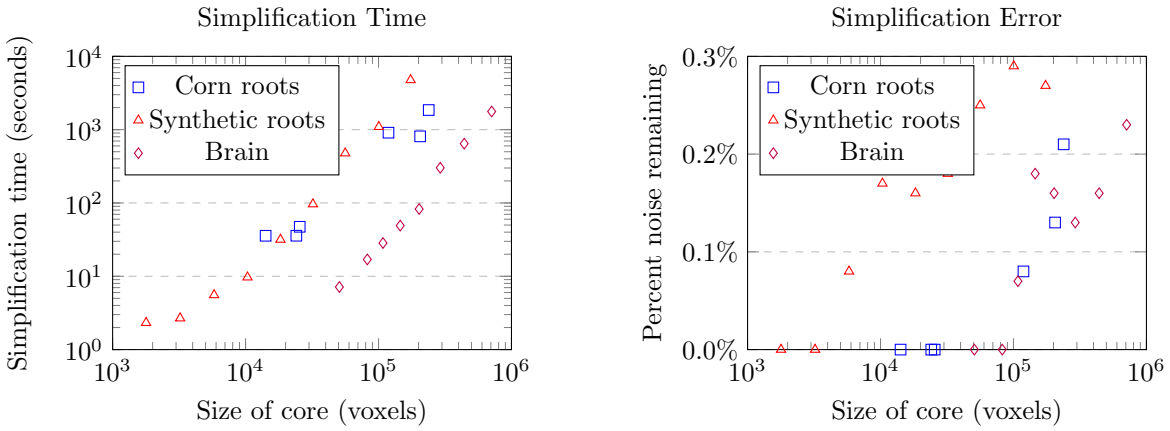

Figure 4: (a) Experimental runtime and (b) error rates on three datasets.

8 thousand components, 17 thousand loops and one thousand voids. All but approximately 100 of those where noise features. In general, the complexity of the initial shapes was very high and there were very few features that were shared between the core and the neighborhood.

Practical concerns The shapes that we have considered have regions as large as several hundred million voxels. As a result, we utilize a sparse representation for our shapes, so that the neighborhoods have about 6 million voxels at most. While the fastest persistent homology implementations can handle inputs of this size, the standard persistence algorithm has trouble on this size of inputs. However these faster algorithms did not allow us to generate candidate simplifications. We choose to do this calculation on windows of data at most 250^3 voxels in size. These windows were overlapped to cover the shape. There is the thorectical possibility that this would fail to repair some very large features, but it made the algorithms computational feasible. This allowed us to perform the caluclations on the largest shapes in under five hours on a Linux desktop with a 2.20 GHz processor I5 processor and 64 GB RAM; see Figure 4.

Accuracy For all of the test shapes, over 99.7% of the topological errors were removed by our method. In several cases, our algorithm was able to find solutions to the homological simplification problem. See Figure 4 for the error rates in the experiments. We note that it is computationally infeasible to determine in the other cases if solutions to the homological simplification problem exist.

6 Future Work

There are several natural directions to purse next. First, we would like to continue scaling the algorithms to larger datasets through optimizations and potential parallelization. We also plan to consider a more robust set of candidate simplifications that might be able to reduce the errors that are not repaired; at the same time, we would like to restrict to simplifications that have nice geometric properties as well as topological properties. Finally, more on the theoretical side, it is interesting to consider the notion of hardness of approximation or any approximation guarantees of heuristics such as our greedy approach.

References

[1] Katrin Amunts, Alan Evans, and Karl Zilles. Bigbrain dataset.

[2] Dominique Attali, Ulrich Bauer, Olivier Devillers, Marc Glisse, and André Lieutier. Homological reconstruction and simplification in r3. In *Proceedings of the Twenty-ninth Annual Symposium on Computational Geometry*, SoCG '13, pages 117–126, New York, NY, USA, 2013. ACM.

[3] Marco Attene, Marcel Campen, and Leif Kobbelt. Polygon mesh repairing: An application perspective. *ACM Computing Surveys (CSUR)*, 45(2):15, 2013.

[4] P-T Bremer, Bernd Hamann, Herbert Edelsbrunner, and Valerio Pascucci. A topological hierarchy for functions on triangulated surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 10(4):385–396, 2004.

[5] T. K. Dey, K. Li, and J. Sun. On computing handle and tunnel loops. In *Cyberworlds, 2007. CW '07. International Conference on*, pages 357–366, Oct 2007.

[6] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. AMS Press, 2009.

[7] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 454–463. IEEE, 2000.

[8] Herbert Edelsbrunner, Dmitriy Morozov, and Valerio Pascucci. Persistence-sensitive simplification functions on 2-manifolds. In *Proceedings of the twenty-second annual symposium on Computational geometry*, pages 127–134. ACM, 2006.

[9] David Günther, Alec Jacobson, Jan Reininghaus, Hans-Peter Seidel, Olga Sorkine-Hornung, and Tino Weinkauf. Fast and memory-efficienty topological denoising of 2d and 3d scalar fields. *IEEE transactions on visualization and computer graphics*, 20(12):2585–2594, 2014.

[10] Xiao Han, Chenyang Xu, Ulisses Braga-Neto, and Jerry L Prince. Topology correction in brain cortex segmentation using a multiscale, graph-based algorithm. *IEEE Transactions on Medical Imaging*, 21(2):109–121, 2002.

[11] John Hempel. *3-manifolds*, volume 349. American Mathematical Soc., 2004.

[12] Reinhard Klette and Azriel Rosenfeld. *Digital geometry: Geometric methods for digital picture analysis*. Elsevier, 2004.

[13] Laurent Najman and Hugues Talbot. *Mathematical morphology: from theory to applications*. John Wiley & Sons, 2013.

[14] Steve Y. Oudot. *Persistence Theory: From Quiver Representations to Data Analysis*, volume 209 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2015.

[15] Florent Ségonne, Jenni Pacheco, and Bruce Fischl. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE transactions on medical imaging*, 26(4):518–529, 2007.

[16] David W Shattuck and Richard M Leahy. Automated graph-based analysis and correction of cortical volume topology. *IEEE transactions on medical imaging*, 20(11):1167–1177, 2001.

[17] Maxime Soler, Melanie Plainchault, Bruno Conche, and Julien Tierny. Topologically controlled lossy compression. *arXiv preprint arXiv:1802.02731*, 2018.

[18] Hubert Wagner, Chao Chen, and Erald Vuçini. *Efficient Computation of Persistent Homology for Cubical Data*, pages 91–106. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[19] Zoë Wood, Hugues Hoppe, Mathieu Desbrun, and Peter Schröder. Removing excess topology from isosurfaces. *ACM Transactions on Graphics (TOG)*, 23(2):190–208, 2004.

[20] Rachel Aine Yotter, Robert Dahnke, Paul M Thompson, and Christian Gaser. Topological correction of brain surface meshes using spherical harmonics. *Human brain mapping*, 32(7):1109–1124, 2011.

[21] Qian-Yi Zhou, Tao Ju, and Shi-Min Hu. Topology repair of solid models using skeletons. *IEEE Transactions on Visualization and Computer Graphics*, 13(4), 2007.