From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Mathematics

2015

Statistical Topology of Embedded Graphs

Schweinhart, Benjamin David

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01ks65hf592

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Statistical Topology of Embedded Graphs

Benjamin Schweinhart

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Mathematics
Advisor: Robert D. MacPherson

September 2015

© Copyright by Benjamin Schweinhart, 2015.

All Rights Reserved

Abstract

Complex systems with interesting topology and geometry abound in the physical and mathematical sciences. This thesis explores several methods for their analysis, namely swatches, persistent homology, and the knotting of embedded graphs. We begin by describing curvature flow on embedded graphs, and methods for simulating it in two and three dimensions. This complex system will serve as a case study for the remainder of the thesis.

The method of swatches characterizes the local topology of a regular cell complex in terms of probability distributions of local configurations. It is used to define a distance on cell complexes, which has many applications. Convergence in this distance is related to the concept of a Benjamini-Schramm graph limit, and we use it to formalize a universality conjecture about the long-term behavior of curvature-flow on graphs.

Persistent homology is a powerful tool for studying complex geometric structures. It establishes a correspondence between geometric features of a subset of a metric space and topological features of its ϵ-neighborhoods. This correspondence is represented by a scatter plot of feature points, which may be used to define many interesting statistics including an analogue of fractal dimension. We also describe a tree structure for the $(n-1)$-dimensional persistent homology of a subset of n-dimensional space, and apply it to compare two types of embedded graphs.

We conclude by proposing a definition for unknotted embedded graphs, and examining the relationship between curvature flow and knottedness in graphs.

Acknowledgements

I am deeply grateful to my advisor, Robert MacPherson, for his encouragement, advice, and support throughout my time at Princeton. I would also like to thank Jeremy Mason, Amit Patel, Jonathan Jaquette, Yasu Hiraoka, Kevin Emmett, Emerson Escolar, and Matthew Kahle for insightful discussions. Furthermore, I am grateful to Mark McConnell, Jeremy Mason, and Amit Patel for their feedback on the drafts of this thesis.

I would like to thank my mother, father, and sister for their love and encouragement, and express my heartfelt gratitude to all of my friends, particularly Ryan Peckner.

Finally, I would like to thank the NSF for providing financial support through the NSF Graduate Research Fellowship Program.

To my parents.

Contents

Abstract . . . iii

Acknowledgements . . . iv

1 Introduction **1**

2 Curvature Flow on Graphs **7**

2.1 Grain Growth . . . 8

2.2 Behavior of the Edges . . . 9

2.3 Behavior of the Vertices . . . 11

2.3.1 Axioms for Curvature Flow on Graphs . . . 12

2.3.2 Equations for the Vertices . . . 12

2.3.3 Scale and the Vertex Drag Equations . . . 13

2.4 Topological Changes . . . 14

2.5 Conjecture . . . 15

2.6 Computation . . . 17

2.6.1 Equations for Embedded Graphs in Two Dimensions . . . 17

2.6.2 Equations for Embedded Graphs in $\mathbb{R}^3$ . . . 19

2.6.3 Interpolation, and Other Technicalities . . . 20

2.6.4 Initial Conditions . . . 21

2.7 First Evidence for the Steady State Hypothesis . . . 23

3 The Method of Swatches **27**

3.1 Regular Cell Complexes 28

3.1.1 Regular Cell Complexes in Nature 29

3.2 Swatches and Clothes 30

3.2.1 Swatches 30

3.2.2 The Cloth 31

3.3 Similarity and Convergence of Cell Complexes 34

3.3.1 Convergence 34

3.3.2 Consequences of Convergence 36

3.4 Conjectures 37

3.5 Method of Computation 39

3.5.1 Choosing r 41

3.6 Applications to Embedded Graphs in Three Dimensions 42

3.6.1 Finite Size Effects 44

3.7 Applications to Embedded Graphs in Two Dimensions 46

3.7.1 Applications to Planar Graphs, Part 1 46

3.7.2 Applications to Planar Graphs, Part 2 48

3.7.3 Digging Deeper 52

4 Persistent Homology **54**

4.1 Introduction 54

4.2 Scatter Plots 56

4.2.1 Two Examples 57

4.3 Homology 58

4.3.1 A Brief Introduction to Homology 58

4.3.2 Regular Cell Complexes 59

4.3.3 Definition of Homology 60

4.3.4 Examples 61

4.3.5 Some Important Theory . . . 62
4.3.6 Alexander Duality . . . 63
4.3.7 Examples of Alexander Duality . . . 64
4.3.8 Alexander Duality for Subsets of the Torus . . . 65
4.4 Defining Persistent Homology . . . 66
4.5 Computation . . . 67
4.5.1 Filtrations . . . 67
4.5.2 The Alpha Complex . . . 68
4.5.3 Proof of the Correctness of the Alpha Complex . . . 69
4.5.4 Method of Computation . . . 71
4.6 Minimal Cycles . . . 72
4.6.1 What makes a cycle minimal? . . . 72
4.6.2 An example . . . 73
4.7 Persistence Trees . . . 73
4.7.1 Preliminary Definition of PH tree . . . 75
4.7.2 Definition . . . 75
4.7.3 Computing the PH tree . . . 76
4.7.4 Two Examples . . . 78
4.7.5 Displaying PH trees . . . 79
4.7.6 Extensions . . . 79
4.8 Recovering a 3D Cell Complex from its 1-Skeleton . . . 80
4.8.1 Recovered Cells . . . 81
4.8.2 Recovered Faces . . . 83
4.8.3 Testing Effectiveness . . . 83
4.8.4 Voluminous Cells . . . 84
4.8.5 Computational Results . . . 86

5 Persistent Homology Dimension and the Persistent Homology of Polymers 87

5.1 Introduction . . . 87

5.2 Dimension and Self-Similarity. . . . 89

5.2.1 Definition of *PH* Dimension. . . . 90

5.2.2 Definition of *PH* Self-Similarity. . . . 90

5.2.3 The Relation between *PH* Dimension and *PH* Self-Similarity. . . . 90

5.2.4 Relation between *PH* dimension and Hausdorff dimension. . . . 91

5.2.5 Probabilistically Determined Structures S. . . . 93

5.3 Three Probability Distributions of Polymers . . . 94

5.3.1 Branched Polymers. . . . 94

5.3.2 Brownian Trees. . . . 96

5.3.3 Self-Avoiding Walks. . . . 96

5.4 Computational Results. . . . 98

5.4.1 Numerical Determination of *PH* Statistical Self-Similarity and Estimation of *PH* Dimension. . . . 98

5.4.2 Three Dimensional Branched Polymers. . . . 98

5.4.3 Two Dimensional Branched Polymers. . . . 99

5.4.4 Two Dimensional Self-avoiding Walks. . . . 99

5.4.5 Brownian Trees. . . . 100

5.4.6 Shape Measures Involving the Aspect y. . . . 101

6 One-dimensional Features 104

6.1 Shortest Loops . . . 104

6.2 Homology . . . 105

6.3 Comparisons . . . 106

6.3.1 Table . . . 106

6.3.2 Histograms . . . 107

6.3.3 Loop Length Distribution . . . 107

6.4 Persistent Homology . . . 109
6.5 Neighboring Loop Distributions . . . 110

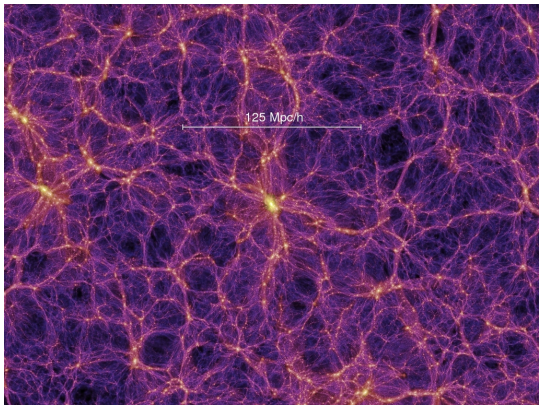
7 Knotting of Embedded Graphs **114**
7.1 The Knot Theory of the Circle . . . 114
7.2 Unknotted Graphs . . . 116
7.2.1 Previous Proposals . . . 118
7.3 A Proposed Definition . . . 119
7.3.1 Computation . . . 119
7.4 Conjecture . . . 120
7.5 Knotting of Shortest Loops . . . 120
7.5.1 Computational Methods . . . 121
7.5.2 Computational Results . . . 121

Chapter 1

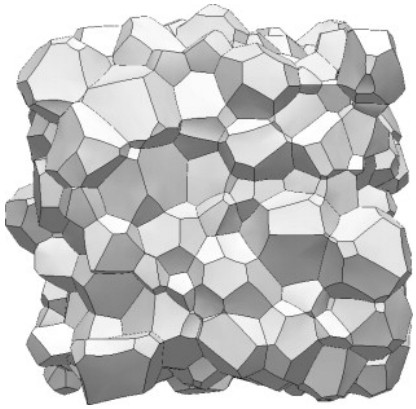
Introduction

Complex systems with interesting topology and geometry abound in the physical and mathematical sciences. Examples include the grain boundary network in a polycrystal, the network of covalent bonds in an oxide glass, mineral clusters formed by diffusion-limited aggregation, the distribution of matter in the universe, and the embedding of DNA inside the nucleus. Several of these are depicted in Figure 1.1. This thesis explores methods for the statistical analysis of these systems, namely swatches, persistent homology, and the knotting of embedded graphs. It was written with a broad audience of physicists, materials researchers, and other scientists in mind, so some topics are covered informally in order to keep the required background material to a minimum. We hope that a reader with basic knowledge of linear algebra, probability, and differential geometry will be able to follow the key ideas presented here. Some knowledge of algebraic topology would be helpful (such as presented in chapters 0-2 of Hatcher [22]), but is not required.

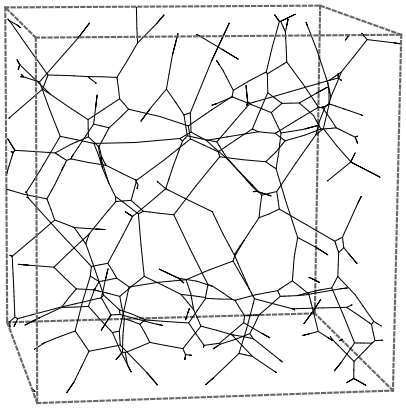
Graphs play an important role in this thesis, both as a tool and as an object of study. A graph is a set of vertices which are connected to their neighbors by edges. In Figure 1.2 a simple graph is shown with the vertices represented by black disks and the edges by straight line segments. (In an unfortunate divergence of terminology, this concept is referred to as a network in many contexts, with the word "graph" being used for plots). An embedded graph is a graph whose vertices and edges are embedded into an ambient space, whereas an abstract graph is one without a specified



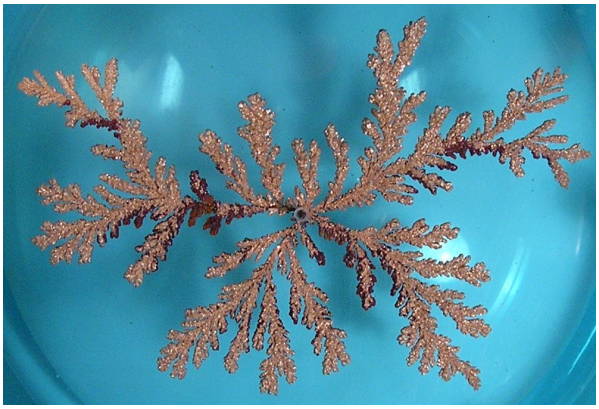
(a)



(b)



(c)



(d)

Figure 1.1: Complex systems with interesting topology and/or geometry: (a) the distribution of dark matter in a large simulation of the universe [66], (b) the grains of a simulated polycrystal [35], (c) an embedded graph in three dimensions resulting from curvature flow, and (d) a copper cluster grown by diffusion-limited aggregation [29].

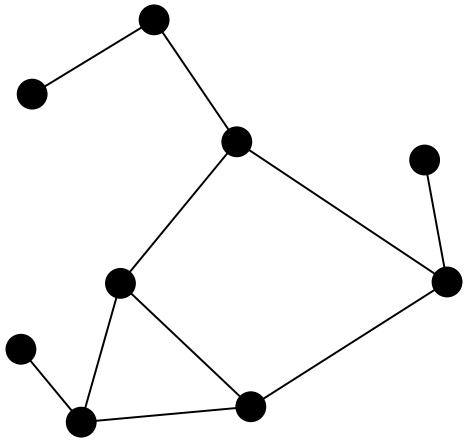


Figure 1.2: A graph with nine vertices (represented by black dots) and ten edges (represented by line segments).

embedding.

We begin the thesis in Chapter 2 by introducing curvature flow on embedded graphs. Under curvature flow, a graph embedded in $\mathbb{R}^n$ evolves to decrease the sum of the lengths of its edges. This process is depicted in two dimensions in Figure 1.3, and an example of an embedded graph in three dimensions resulting from curvature flow is shown in Figure 1.1c. This system will serve as a case study for the methods covered in the remainder of the thesis. After defining curvature flow on graphs, we describe how to simulate it in two and three dimensions. Due to the complexity of this system, most of our results will be computational rather than rigorous in nature. In later chapters, we state several conjectures about this system supported by computational experiments. In particular, we formalize a universality conjecture about the long-term behavior of graphs evolving under curvature flow: that for almost all initial conditions, all scale-invariant properties converge to universal values that do not depend on the initial state. We call this the "Steady State Hypothesis." We hope to approach this conjecture from a rigorous standpoint in the future. The main difficulty appears to be that of defining a topology on the configuration space of embedded graphs in $\mathbb{R}^n$.

The method of swatches, developed in Chapter 3, characterizes the local topology of a regular cell complex in terms of probability distributions of local configurations. A regular cell complex is a space that is built inductively by attaching cells to each other without twisting or tearing. A

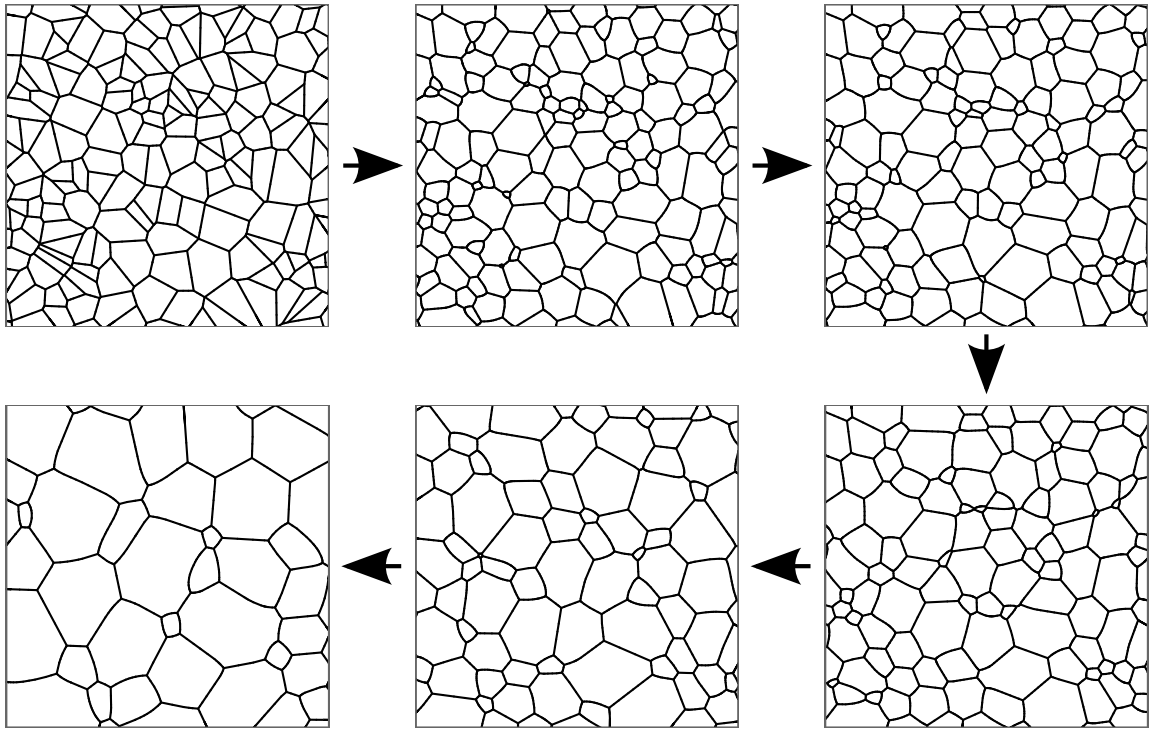


Figure 1.3: A planar graph evolving by curvature flow.

zero-cell is a point, a one-cell a line segment, a two-cell a disk, and so on. The method of swatches is useful for studying any complex system which has a natural representation as a regular cell complex, such as the polycrystal in Figure 1.1b or the embedded graph in Figure 1.1c. It is used to define a distance on cell complexes, which has many applications. Convergence in this distance is related to the concept of a Benjamini-Schramm graph limit, which formalizes the Steady State Hypothesis for embedded graphs evolving under curvature flow. At the end of the chapter, we present extensive computational evidence for this formalized hypothesis and other related conjectures.

In Chapters 4 and 5, we describe Persistent Homology, a powerful tool for studying complex geometric structures. Our approach to Persistent Homology is different from that presented elsewhere; we use it to study the geometry of a fixed object rather than the topology of a noisy data set. As such, it is useful for the study of complex systems which may not have interesting topology in themselves, such as the copper cluster shown in Figure 1.1d. The basic idea of Persistent Homology

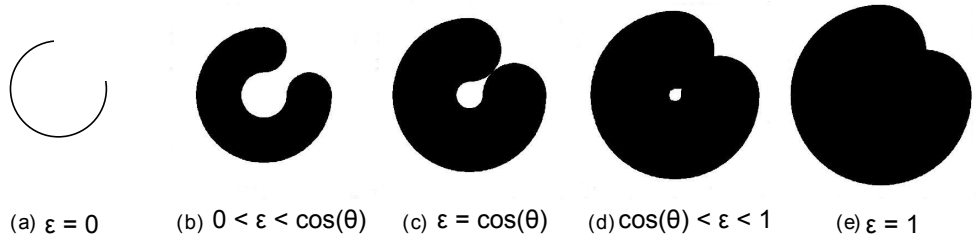
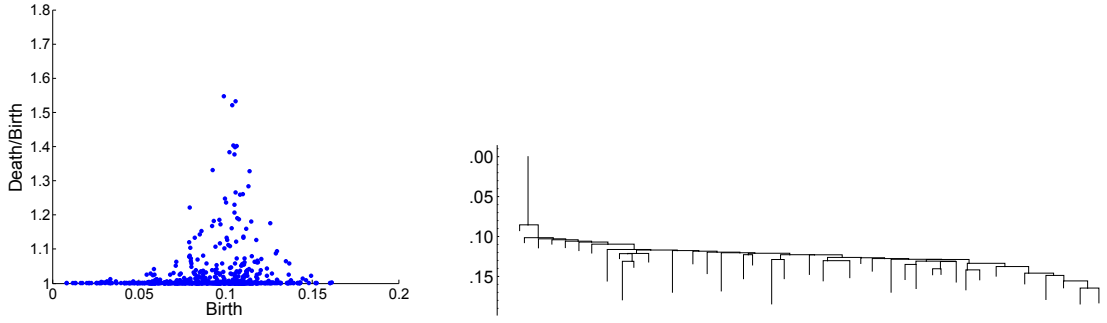


Figure 1.4: Neighborhoods of a circle of radius 1 with an arc removed. This object has one one-dimensional geometric feature, which gives rise to a one-dimensional topological feature (the hole) as it is expanded.

is that if X is a subset of $\mathbb{R}^n$, then the topological changes that occur when X is thickened inside the ambient space reflect its original geometry. For example, the arc of a circle in Figure 1.4 has a one-dimensional geometric feature (a gulf) which gives rise to a topological feature as it thickens (the hole that forms and disappears). In general, d-dimensional geometric features are represented in Persistent Homology by a scatter plot of feature points computed in terms of the $d-$dimensional topological features of the neighborhoods of X. An example scatter plot is shown in Figure 1.5a.

Chapter 4 provides a brief introduction to Homology before defining Persistent Homology and explaining its interpretation in terms of Alexander Duality. The chapter also includes an informal, intuitive explanation of these concepts. Afterward, we specialize to the case of the $(n-1)$-dimensional Persistent Homology of a subset X of n-dimensional space. The additional geometric structure in this case allows us to build a Persistence Tree representing the relationships between the $(n-1)$-dimensional features of X. An example tree is shown in Figure 1.5b. We apply this concept to compare the "Steady State" condition of an embedded graph in three-dimensional space evolving under curvature flow with the one-skeleton of a Voronoi tessellation of random points. We also explore how to recover a regular cell complex from its one-skeleton using Persistent Homology.

Our treatment of Persistent Homology is continued in Chapter 5, where we study statistics on the scatter plot of feature points. In particular, we define an analogue of fractal dimension called *PH* dimension which measures geometric complexity. We apply this to compare three different distributions on the space of polymers in two and three dimensions. The embedded graphs described in Chapter 2 do not have fractal properties, so we do not study them using this method.



(a) (b)

Figure 1.5: Persistent Homology data for a graph embedded in three dimensions: (a) a scatter plot of two-dimensional features and (b) its Persistence Tree.

Chapter 6 introduces several methods for studying the one-dimensional features of graphs embedded in three dimensions. These features are loops that are minimal in a certain sense. We generate four sets of such loops, and compare their properties, as well as their relation to one-dimensional Persistent Homology.

In the final chapter, Chapter 7, we define a notion of knottedness for embedded graphs, which generalizes the knot theory of the circle. We also state a conjecture that curvature flow "unknots" graphs, and provide heuristic evidence for it.

Chapter 2

Curvature Flow on Graphs

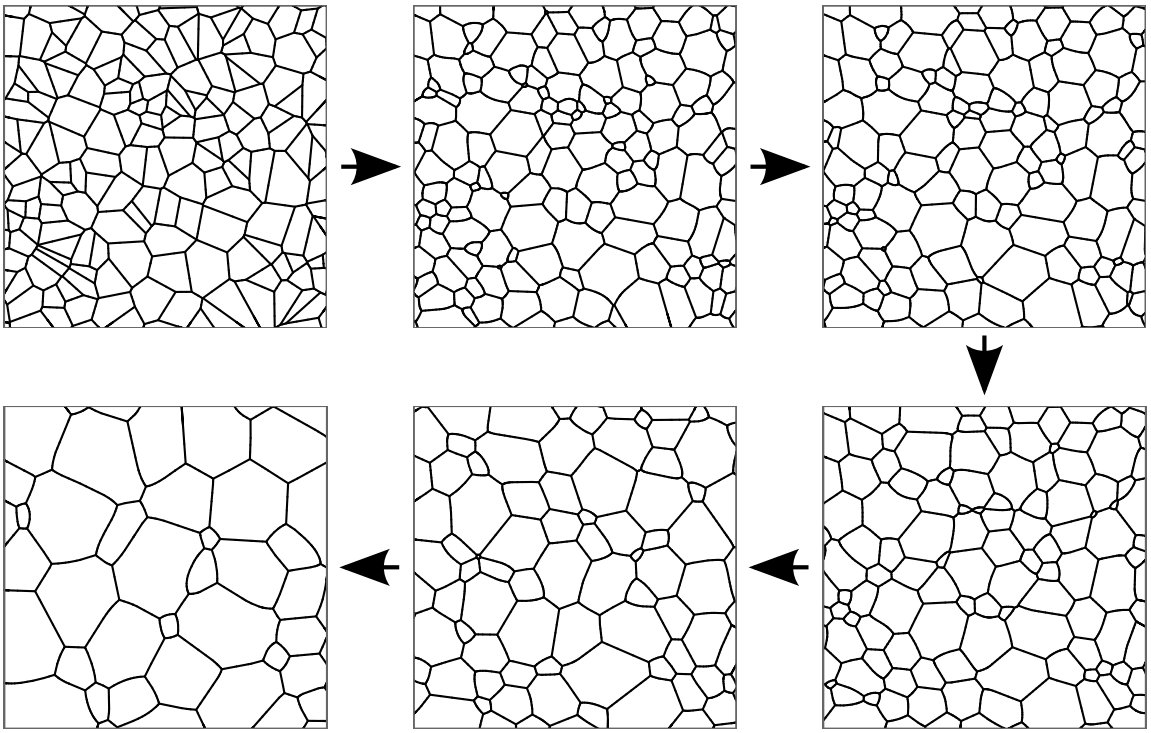


Figure 2.1: A planar graph evolving by curvature flow.

This chapter will introduce curvature-flow on graphs, which will serve as a case study for the

methods of statistical topology that are the core of this thesis. Let G be an embedded graph in $\mathbb{R}^n$, whose edges are smooth curves. We place no restrictions on the vertices *a priori*, but it will turn out that they will almost always be of degree three (where the degree of a vertex is the number of adjacent edges). We endow the edges of G with a constant energy per unit length, and let them evolve via energy minimization. This system is of intrinsic mathematical interest, and also has important physical applications. The evolution of a planar graph by curvature flow is a model for the important physical phenomenon called grain growth. A graph embedded in $\mathbb{R}^3$ can be interpreted as a very simple model of dislocations in a 3D polycrystalline material [24, 64].

The curvature-driven evolution of a planar graph is shown in Figure 2.1, which depicts the same small region as time proceeds. As the system evolves, the average edge length increases and the number of edges decreases. This process is called coarsening. However, in all but the first picture, the graphs appear very similar apart from scale. This observation leads to an important conjecture – the "Steady State Hypothesis" – first stated in Section 2.5, and formalized in Chapter 3. We shall study the geometric properties of the "Steady State" corresponding to embedded graphs in three dimensions using various topological techniques in Chapters 4 and 6. The computational method used to produce these images will be described in Section 2.6.

2.1 Grain Growth

The curvature-driven evolution of a planar graph is a model of an important physical phenomenon: the grain growth of polycrystalline materials. Many important materials are polycrystalline, including most metals. A crystalline material is one that can be described by a small basic configuration that is repeated at the points of a lattice. A polycrystalline material is one that is locally crystalline, but in which a crystalline orientation is not shared globally. Rather, the material is composed of grains, each with a different orientation. In two dimensions, the network of grain boundaries is a planar graph of degree three, and the energy present in these edges determines the evolution of the material in time. In general, the energy of an edge depends on the orientations of the neighboring grains, but physicists often make the simplifying assumption that it is constant throughout the ma-

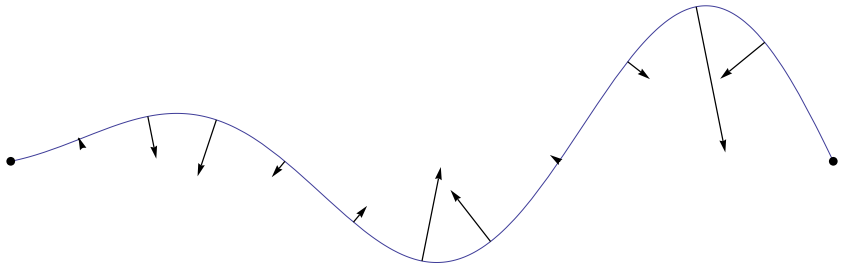


Figure 2.2: A planar curve with fixed endpoints, with curvature vectors drawn at several points. The curvature vector is in the direction of the unit normal with a magnitude equal to the curvature κ.

terial. Under this simplification, the boundary network evolves as a planar graph under curvature flow.

2.2 Behavior of the Edges

The behavior of the edges of a graph evolving under curvature flow is relatively well-studied and well-understood. Let E be an edge; at a given point in time there will be an embedding $\alpha_t : [0,1] \to \mathbb{R}^d$, where d is the dimension of the ambient space. In this section, we will take the behavior of the endpoints as given and derive an equation of motion that will be satisfied on the interior of the edge.

Before studying the behavior of α_t, let us review some concepts from differential geometry. Let $\beta : [0, L] \to \mathbb{R}^d$ be a smooth curve and suppose it is parametrized by arc length s. That is, $\|\beta'(s)\| = 1$ and the arc length of β in any subinterval (a, b) of $[0, L]$ is equal to $b - a$. Recall that the tangent vector T of β is the unit vector $T(s) = \beta'(s)$ and the normal vector $N(s) = \frac{T'(s)}{\|T'(s)\|}$. The curvature is $\kappa = \|T'(s)\|$ and the curvature vector is $T'(s) = \kappa N(s)$. A planar path with fixed endpoints and a closed curve in $\mathbb{R}^3$ are shown with their curvature vectors in Figures 2.2 and 2.3. Note that the curvature vectors are longest where the path is most curved, and "pulling" a point of the path in the direction of its curvature will straighten it out at that point. The curvature of a circle of radius r is a constant $1/r$ and the curvature of a straight line is zero. Note that a straight line is also the length-minimizing path between two points in $\mathbb{R}^n$. This hints at the relationship

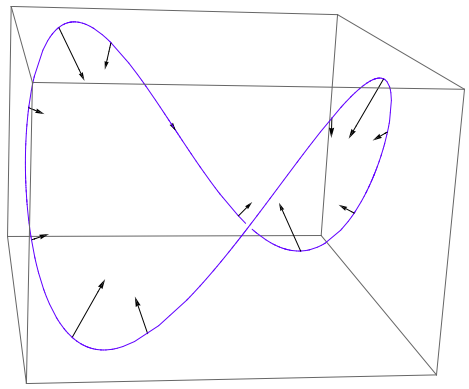


Figure 2.3: A closed curve in $\mathbb{R}^3$ with with curvature vectors drawn at several points.

between the minimization of curvature and the minimization of length.

Let us return to our consideration of the curve $\alpha_t(x)$ evolving by energy minimization, with a constant energy per unit length. A standard argument from the Calculus of Variations shows that if α_t satisfies

$$\frac{\partial \alpha_t}{\partial t} = \kappa N \tag{2.1}$$

where κ is the curvature and N is the normal vector then

$$\frac{\partial L(\alpha_t)}{\partial t} = -\int_0^1 \kappa^2 \, dt \tag{2.2}$$

and furthermore that this is the fastest that L can decrease for any family α_t normalized so that $\int_0^1 \left\| \frac{\partial \alpha_t}{\partial t} \right\| \, dt = C = \int_0^1 \kappa N \, dt$. [2]

Thus, we will use 2.1 as the equation of motion for the interior of the edges. For example, the curves in 2.2 and 2.3 will flow in the direction of the arrows at a speed proportional to their length. This law fully specifies the behavior of a closed curve evolving to minimize length, and in the plane it is known that all closed curves will even out to a circle before shrinking to a point. [20]. However, the situation here is more complicated as there are the vertices of the graph to consider.

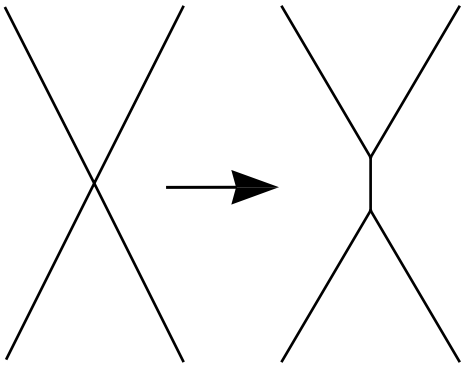


Figure 2.4: A vertex of degree four can always be replaced by a fifth edge to reduce the sum of the lengths of the neighboring edges.

2.3 Behavior of the Vertices

As a first step toward understanding the behavior of the vertices, let us note that that any vertex of degree four is unstable under energy minimization. It is always possible to split off two pairs of edges and add an infinitesimal edge between them to reduce the total edge length, as in Figure 2.4. The choice of pairing is determined by the angles between the tangent vectors - the one that yields the greatest length reduction will be the one that minimizes the sum of the angles opposite to the new, infinitesimal edge. A similar argument shows that any vertex of degree greater than 4 will also be unstable. Finally, a vertex of degree two is just a point along an edge, and the edge adjacent to a vertex of degree one will collapse very quickly. So here we will restrict our attention to to embedded graphs whose vertices all have degree three.

As a special case, consider a configuration in either $\mathbb{R}^2$ or $\mathbb{R}^3$ with one vertex V of degree three adjacent to edges e_1, e_2, and e_3. Suppose the other endpoints of the three edges are "pinned" to points p_1, p_2, and p_3 in space. A theorem, dating back to Fermat, states that there is a unique point p which minimizes the sum of the distances to the p_i, $d(p, p_1) + d(p, p_2) + d(p, p_3)$. This point is called the Fermat point of the points p_i. p lies in the plane formed by those three points, and if T_i are the unit vectors from the Fermat point to the p_i, then the angles between them equal $2\pi/3$. Equivalently, $T_1 + T_2 + T_3 = 0$. Any reasonable rule for the behavior of the vertex together with curvature flow on the edges will lead to an equilibrium where V is located at p and the three edges

are straight line segments.

2.3.1 Axioms for Curvature Flow on Graphs

The previous discussion may be summarized with the following axioms for the behavior of our system:

- G is composed of smooth curves meeting at vertices of degree three.
- The edges evolve by curvature flow.
- A system with one vertex of degree three whose edges' other vertices are pinned will reach an equilibrium where the tangent vectors at the vertex lie in a plane and meet in angles of $2\pi/3$.

2.3.2 Equations for the Vertices

Given these axioms, it turns out that any rule of vertex behavior of the form

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{T_1 + T_2 + T_3}{\lambda} \tag{2.3}$$

where the T_i are the unit tangent vectors of the adjacent edges pointing outward from the vertex and λ is a real number in $(0, \infty)$ is reasonably well-behaved[1]. [33]. λ can be viewed as a "vertex drag" parameter that specifies the speed of the vertices relative to the speed of the edges. The two limiting cases $\lambda \to \infty$ and $\lambda \to 0$ are also interesting if the overall speed of the system is scaled appropriately to keep the velocities finite. The latter is of particular importance, and is assumed more often than not in the planar case. It implies that the Herring Angle condition, $T_1+T_2+T_3 = 0$, always holds, and it will be briefly discussed in the next paragraph. In the other limit, $\lambda \to \infty$, the edges become straight and all of the action is at the vertices when time is appropriately rescaled. This is called a Vertex Model.[2]

[1] Kinderlehrer and Liu prove existence and uniqueness of solutions to these equations, together with curvature flow on the edges, for planar graphs whose vertices are "sufficiently close" to satisfying $T_1 + T_2 + T_3 = 0$. [33] Here, we will use these equations to model the behavior of general trivalent graphs, but will not tackle the difficult analytical questions of existence and uniqueness since it would distract from the focus on methods of statistical topology

[2] However, it is quite different than the Vertex Models used to simulate physical systems, such as in [23]. The model in that paper seeks to establish a vertex model to simulate the $\lambda \to 0$ limit.

When studying grain growth, the assumption is often made that $\lambda = 0$. ($\lambda = 0$ will be used as shorthand for the limiting case where $\lambda \to 0$.) That is, that the Herring Angle Condition always holds and the motion of the vertices is determined by the limiting curvatures of the incoming edges. [37] In this case, the change in area A of each grain is given solely in terms of the number of sides n by the von-Neumann Mullins Relation: [68, 54]

$$\frac{\mathrm{d}A}{\mathrm{d}t} = m\gamma\frac{\pi}{3}(n-6). \tag{2.4}$$

where m and γ are the mobility and energy per unit length, as above. In addition to being mathematically elegant, this situation is also the one that causes the total energy of the system to decrease most quickly and is viewed as most physically reasonable by many physicists. However, we will also be interested in cases where the vertices have non-zero drag. There is some evidence that vertices have non-zero drag in two-dimensional grain growth [18, 19].

2.3.3 Scale and the Vertex Drag Equations

When an embedded graph evolving by curvature flow is dilated by a factor ρ the speed of the edges is multiplied by $\frac{1}{\rho}$. This is because curvature scales this way; for example, a circle of radius one has constant curvature equal to one but a circle of radius ρ has constant curvature equal to $\frac{1}{\rho}$. On the other hand, the equation for the vertices (2.3) is scale-invariant. This means that as the system evolves and the average edge length increases, the speed of the edges will decrease relative to the speed of the vertices and the system will approach the $\lambda = 0$ case. However, different long-term behavior is obtained by rescaling the velocity of the vertices continuously to keep the behavior of the system consistent across time. That is, if $G(t)$ is an embedded graph evolving with the time parameter t, d_ρ is dilation by ρ, and $\hat{G}(0) = \rho(G(0))$, then we want $\hat{G}(t) = \rho(G(\psi(t)))$ for some bijective, continuous function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$. There are many ways to achieve this, but here we will rescale the vertex equation by the average length of the edges of the configuration, $\bar{e}$:

$$\frac{dv}{dt} = \frac{T_1 + T_2 + T_3}{\bar{e}\lambda} \tag{2.5}$$

Note that no rescaling is necessary for the $\lambda = 0$ case.

A downside to this rescaled equation, is that unlike Equations 2.3 and 2.1 it is not local - it does not depend only on an infinitesimal neighborhood of a vertex. A different approach that yields similar long-term behavior but maintains the locality of the laws of motion is based on assigning non-zero masses to the vertices. At the beginning of the simulation, all vertices are given mass one. As the system evolves and coarsens, some vertices merge (as a result of the topological moves described in the next section), and the average vertex mass increases. We can define another equation of motion for the vertices based on their mass m:

$$\frac{dv}{dt} = \frac{T_1 + T_2 + T_3}{\sqrt{m}\lambda} \tag{2.6}$$

As we shall see later, the average vertex mass and average edge length both scale by the same power law in the long-term limit of the system. We also compare the systems resulting from equations 2.5 and 2.6 in Section 3.7.1 to show that they have the same long term behavior.

2.4 Topological Changes

Curvature flow on graphs does not only have singularities in space (vertices), but also in time. These singularities give rise to topological changes in the embedded graph. For example, consider what happens when an edge shrinks to a point as in Figure 2.5a. A vertex of degree four forms, which is unstable with respect to energy minimization. A new edge forms immediately after this singularity, with adjacent edges shuffled to minimize the sum of the two angles opposite to the new edge. This operation is called an edge flip. Another topological change occurs when the edges of a digon become infinitesimally close together, as in Figure 2.5b. These edges merge to form one edge. Note that the topological change that occurs when a polygon shrinks to a point can be expressed in terms of a series of the two previous operations.

Edge flips and digon deletions are the only topological changes that occur in planar graphs, but graphs embedded in three dimensions exhibit another singularity in time. Namely, two edges that

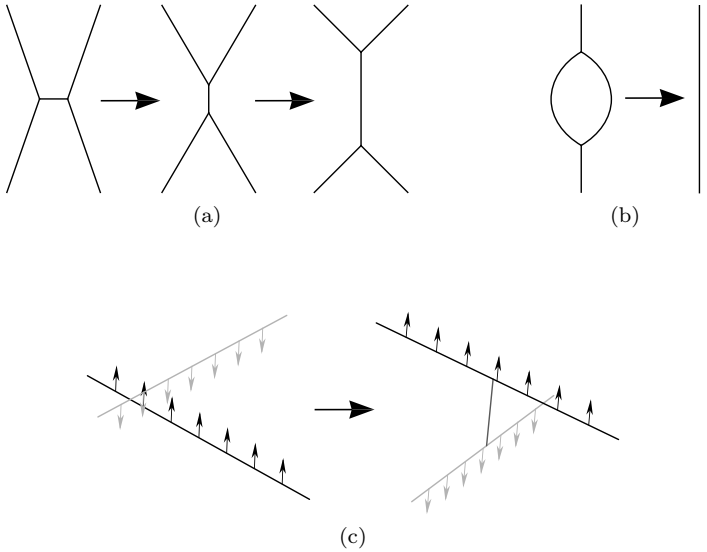


Figure 2.5: Topological operations that occur in the simulations include (a) edge flips, (b) digon deletions, and (c) edge intersections, resulting in joining edges.

do not share a vertex can move into a position where they suddenly intersect. It is ambigous what the correct behavior is at this kind of singularity. One possibility is that a new edge forms between two new vertices at the point of intersection on the existing edges as in Figure 2.5c. Alternatively, the edges could also pass freely through each other. As we will show in Section 3.6, this choice does not affect the long term statistical properties of the system.

2.5 Conjecture

Let G be the space of embedded graphs in $\mathbb{R}^n$. A scale-invariant property of embedded graphs is a function $S : G \to X$ that is invariant under dilations of $\mathbb{R}^n$. Examples of scale-free properties include the normalized standard deviation of the length of an edge (X is $\mathbb{R}$), the normalized probability distribution of edge lengths (X is the space of probability distributions on $\mathbb{R}$), and the percentage of edges contained in a triangle (X is $[0, 1]$).

We are now ready to state the main conjecture of the first half of this thesis:

Conjecture 1 (Steady State Hypothesis)**.** *Suppose* Ω *is the dynamical process on embedded graphs*

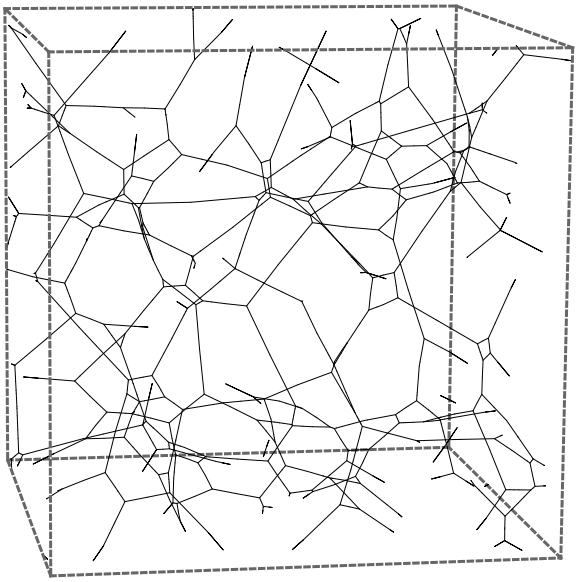
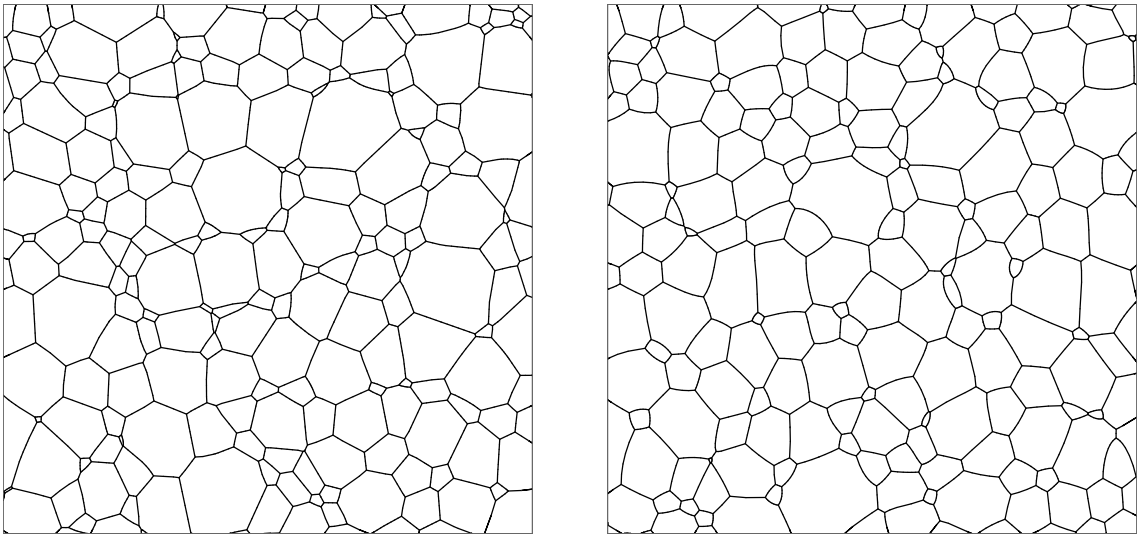


Figure 2.6: A small region in a steady state configuration in three dimensions.

in $\mathbb{R}^n$ in which the edges evolve by curvature flow and the vertices move by any of the rules defined in Section 2.3. For almost all infinite initial conditions G, the long-term behavior of $G(t)$ under the process Ω is such that all scale-invariant properties of the network converge. Furthermore, the limiting values are universal and do not depend on the initial state.

This conjecture can be easily adapted to context of curvature flow on finite graphs where the size of the initial conditions increases toward ∞. The statement of this conjecture will be formalized in Chapter 3, which introduces a topological metric on embedded graphs. We will also provide computational evidence for the Steady State Hypothesis in the cases of graphs embedded in two and three dimensions, beginning in Section 2.7 at the end of this chapter, and also in Chapter 3. This will strongly imply 2D and 3D configurations shown in Figures 2.7 and 2.6 are steady state configurations for their respective equations.



(a) (b)

Figure 2.7: Small regions in steady state configurations in two dimensions for the (a) Vertex Drag Equations with $\lambda = 1$ and (b) the von Neumann–Mullins equations.

2.6 Computation

Curvature flow on graphs was simulated in two and three-dimensional cubes with periodic boundary conditions (that is, the two- and three-dimensional tori $\mathbb{T}^2$ and $\mathbb{T}^3$). We model embedded graphs as networks of polygonal curves, where line segments composing a edge meet at nodes of degree two and edges meet at vertices of degree three. Several different sets of equations are used to model embedded graphs in two dimensions, and one set of equations is used for graphs in three dimensions.

2.6.1 Equations for Embedded Graphs in Two Dimensions

To model curvature flow of planar graphs, we use three different sets of equations. The first set, corresponding to values of vertex drag λ between 0 and ∞, are called the Vertex Drag Equations. The other two correspond to the limits as $\lambda \to \infty$ and $\lambda \to 0$ while rescaling time appropriately, and are called the Vertex Model and the von-Neumann Mullins equations, respectively.

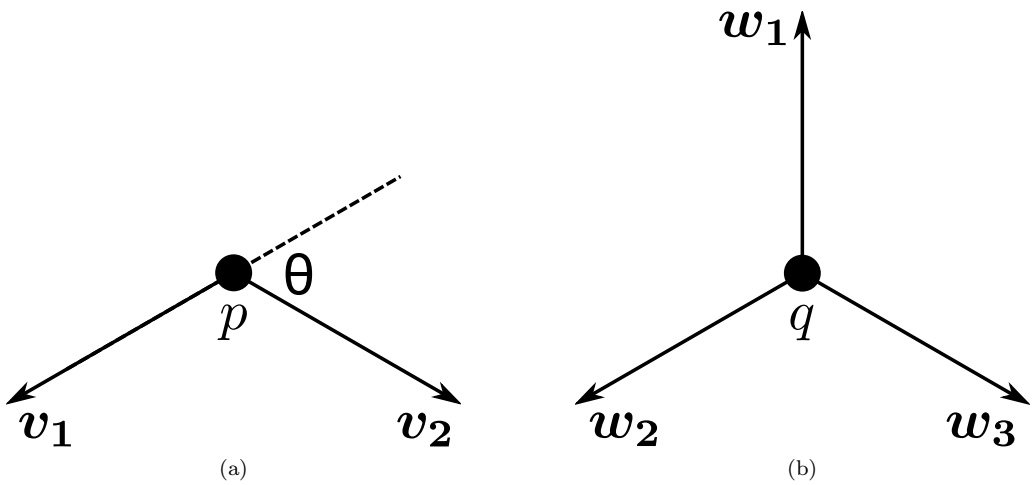


Figure 2.8: (a) A node p along an edge with turning angle θ. (b) A vertex q of degree three.

Vertex Drag Equations

In the Vertex Drag Equations, we use a discrete analogue of curvature flow to model the behavior of the edges. Consider the node in Figure 2.8a, with vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ to the adjacent nodes. In each timestep we move the node by

$$\frac{\Delta \boldsymbol{p}}{\Delta t} = \frac{2\gamma\theta}{\|\boldsymbol{v}_1\| + \|\boldsymbol{v}_2\|}\widehat{\boldsymbol{v}_1 + \boldsymbol{v}_2} \tag{2.7}$$

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are the vectors from p to the adjacent nodes, $\hat{\boldsymbol{v}}$ is the unit vector in the direction of $\boldsymbol{v}$, and θ is the turning angle at p. The expression $\theta/(\|\boldsymbol{v}_1\| + \|\boldsymbol{v}_2\|)$ is a well-known discrete approximation of curvature, which converges to the classically defined curvature as the mesh of a polygonal approximation of a curve goes to zero. [6] The accuracy of this approximation is highest when θ is low, so the polygonal curves in simulations are interpolated to keep the maximum turning angle below $\theta/15$ using the method described in Section 2.6.3. The vertex is moved in the direction of the midpoint of the two neighboring nodes, $\boldsymbol{v}_1 + \boldsymbol{v}_2$, rather than in the direction of the angle bisector, $\hat{\boldsymbol{v}}_1 + \hat{\boldsymbol{v}}_2$, in order to keep the spacing between nodes relatively uniform.

To model the behavior of the vertices, we directly port the Equations 2.3, 2.5, and 2.6 to the discrete case by interpreting the vectors $\boldsymbol{w}_i$ from a vertex to its adjacent nodes as tangent vectors

and multiplying them by γ. We call the three resulting equations the unrescaled Vertex Drag Equations, the rescaled Vertex Drag Equations, and the Vertex Mass Equations, respectively.

Vertex Model

The Vertex Model simulates the limit behavior of Equation 2.3 as $\lambda \to 0$ and the speed of the system is appropriately rescaled to keep the velocity of the vertices constant. In this case, the edges immediately flatten out and the behavior of the vertices determines the evolution of the system. As such, the Vertex Model is a special case of the Vertex Drag Equations where the edges are straight lines with no nodes in their interiors. The same equation for the vertices may be used, but note that the drag parameter λ does not affect the behavior of the system any differently than does the speed.

von Neumann-Mullins Equations

The von Neumann-Mullins Equations were developed in [36] as a model of two-dimensional grain growth in the case where the vertices experience no drag ($\lambda \to 0$) and always satisfy the Herring Angle Condition. That is, that the angles between the tangent vectors of the incoming edges at a vertex are always $2\pi/3$. In this situation, the von Neumann-Mullins relation 2.4 holds and the rate of change of the areas of each grain is determined by the number of edges of that grain. The movement of each vertex and node under the von-Neumann Mullins Equations is determined so that the resulting configuration will always satisfy their namesake relation up to numerical precision. Jeremy Mason's implementation of the von-Neumann Mullins Equations was used to produce the data used in this thesis [45, 36].

2.6.2 Equations for Embedded Graphs in $\mathbb{R}^3$

One set of equations of motion was used to simulate the behavior of embedded graphs in three dimensions under curvature flow. The equation for the nodes is the same as the one used in the two dimensional Vertex Drag Equations, Equation 2.7. However, as there is no analogue of the von Neumann-Mullins equations for embedded graphs in three dimensions, we designed a different

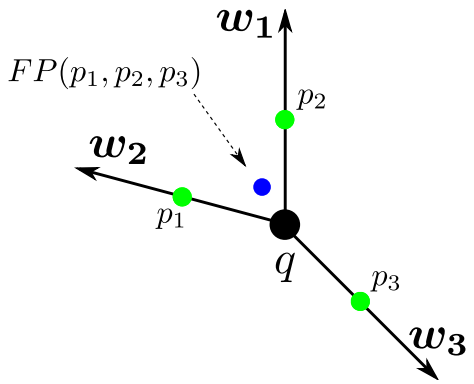


Figure 2.9: A vertex and the Fermat point to which it will move in the next timestep.

equation for the vertices that would approach the $\lambda \to 0$ limit as the system evolved. Recall that in this limit, the vertices experience no drag and move to minimize the length of their neighboring edges with no resistance. To simulate this, consider a vertex q such as the one illustrated in Figure 2.9. Select three points p_i along the line segments between q and its neighboring nodes so that the p_i are equidistant from p. In the next timestep, p is moved to the unique point that minimizes the sum of the distances to the p_i, the Fermat point of the p_i. This is illustrated in Figure 2.9. The lines segments from the p_i to the Fermat point meet at angles of $2\pi/3$. As the system evolves, the distance from the p_i to q is increased to speed to convergence to $\lambda \to 0$, but is always kept smaller than the smallest distance between nodes to prevent unusual behavior.

2.6.3 Interpolation, and Other Technicalities

The computational techniques described above were implemented in C++, and parallelized with OpenMP [8]. Thresholds were set for the topological moves in Section 2.4 based on the average edge length $\bar{e}$. If the maximum distance between two edges of a digon falls below $\frac{\bar{e}}{50}$, one of the two edges is deleted and the other is merged into the neighboring edges. If an edge's length is consistently decreasing and falls below $\frac{\bar{e}}{50}$, it is flipped. In some 3D simulations, if two edges that do not share a vertex come within $\frac{\bar{e}}{100}$, a new edge is added between the closest pair of nodes contained within them. The process for detecting these intersections is computationally costly, and a number of methods were implemented to make it more efficient.

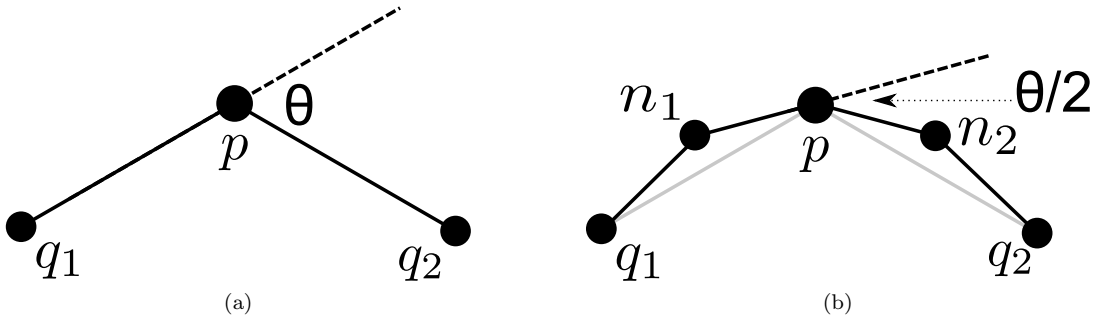


Figure 2.10: A node (a) before and (b) after interpolation. New nodes n_1 and n_2 are added at half the distance from q_1 and q_2 in a way such that the turning angle is halved. This maintains the curvature at p.

To maintain the accuracy of Equation 2.7, new nodes are added between a node and its neighbors if its turning angle falls below $\pi/15$. These nodes are added in a way to maintain the curvature at p while halving the turning angle, which is consistent with an interpretation of the polygonal curve as an approximation of a smooth curve with a given curvature at p. The procedure is depicted in Figure 2.10. Interpolation is not performed when it would add points within the threshold for minimum segment length, a fiftieth the length of an average edge. Nodes are deleted if they move within this very small threshold, or the turning angle falls below $\pi/400$.

2.6.4 Initial Conditions

Initial conditions for the simulations were generated from a set of random points by one procedure in two dimensions, and two procedures in three dimensions. In two dimensions, a Voronoi tessellation of a set of Poisson (uniformly) distributed points on the flat 2-torus (the square with periodic boundary conditions) was used as initial conditions. The Voronoi tesselation of a set of random points $\{p_i\}$ in $\mathbb{R}^n$ or the n-torus $\mathbb{T}^n$ is a decomposition of space into polyhedra, one for each point, such that the polyhedron containing p_i consists of all space closer to that point than to any of the other p_j. This results in a trivalent embedded graph in the plane, such as the one depicted in Figure 2.11. The software library CGAL was used to compute two-dimensional Voronoi tessellations [1].

The Voronoi tessellation of Poisson distributed points was also used to produce one of the types

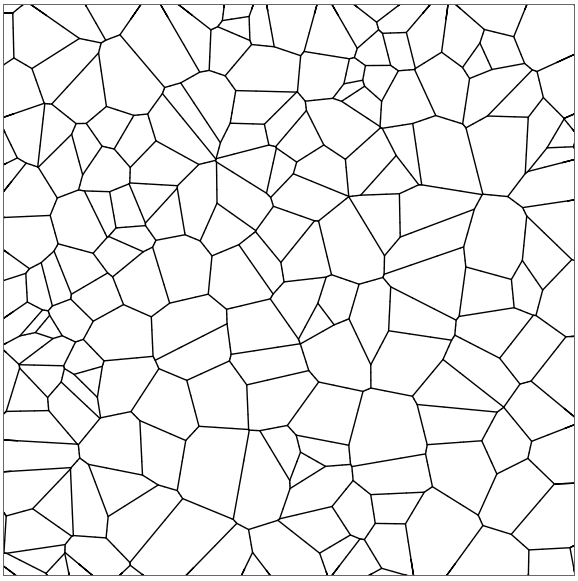
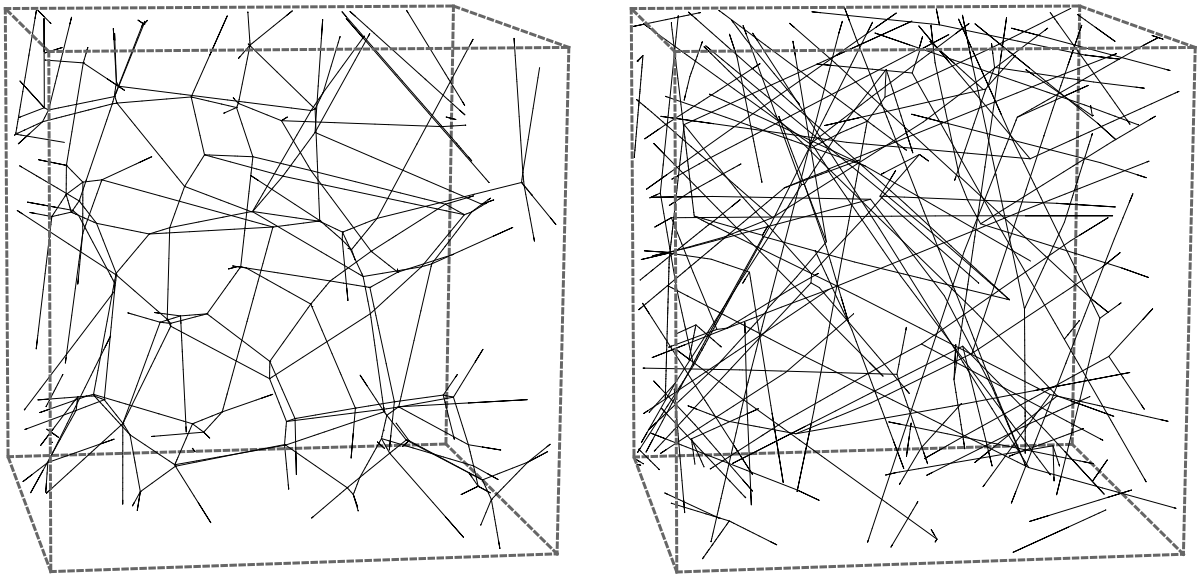


Figure 2.11: The initial condition of the 2D simulation - a Voronoi diagram of random points.

of initial conditions in three dimensions. This initial condition, called a Voronoi graph, is illustrated in Figure 2.12a and is constructed from the one skeleton of a Voronoi tesselation. Since the vertices of the one-skeleton have degree four, it has to be modified to make an admissible structure. Every vertex is replaced by an edge, and the four adjacent edges are assigned to the vertices to minimize the maximum angles opposite to the newly created edge. Ken Brakke's VOR3DSIM program was used to compute Voronoi tessellations in three dimensions [9].

For the second procedure in three dimensions, Poisson distributed points are again sampled on the flat three torus. Edges are placed between randomly selected points whose distance is below a threshold. If there are no points within the threshold around a certain point p, two cases are considered. If p had degree two, it is considered to be a node along the edge between its two neighbors. Otherwise, p is paired with the closest possible point. This process proceeds until the creation of edges is no longer possible. The resulting initial condition is called a random graph, and is depicted in Figure 2.12b. Note how different it looks than the Voronoi graph. In Section 3.6, the long term behavior under curvature flow of these two very different initial conditions is shown to be the same.



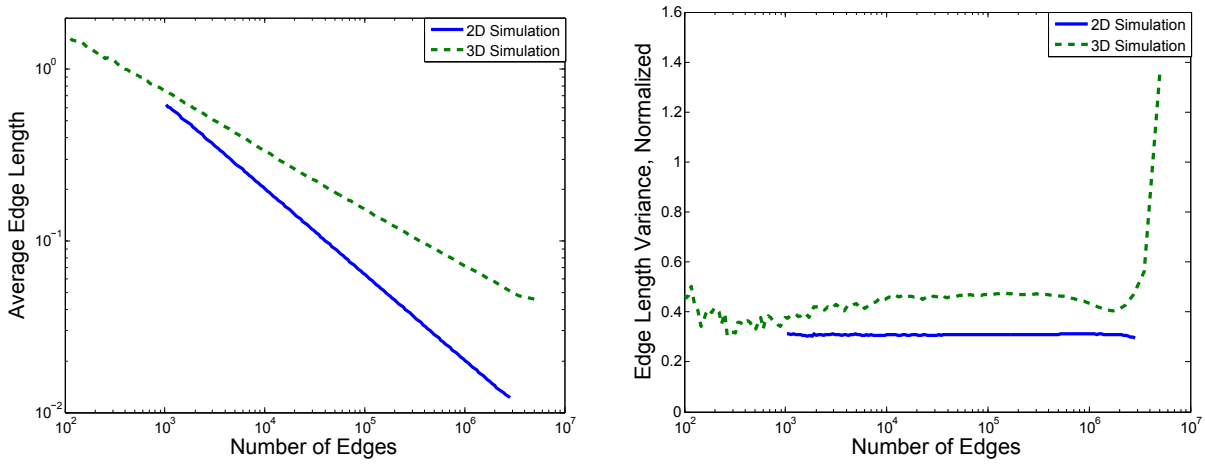
(a) (b)

Figure 2.12: Initial conditions of the 3D simulation include (a) Voronoi graphs and (b) random graphs, with the constructions given in the text. Note that (b) shows a smaller volume than (a), since edges are much denser in the random graph.

2.7 First Evidence for the Steady State Hypothesis

Recall the Steady State Hypothesis from Section 2.5, that if an embedded graph evolves by curvature flow it will eventually reach a steady state where all scale-free properties have converged, and take universal values that depend only on the dimension of the ambient space. Now that we have defined methods for the simulation of this evolution, let us examine some computational evidence for this conjecture. The statistics for 2D systems in this section come from a 2D system beginning with a Voronoi decomposition of one million random points on the flat 2-torus, and evolving via the Vertex Drag Equations with $\lambda = 1$. The 3D simulation began with a Voronoi graph with five million edges.

The average length of an edge is not a scale-free property (as is seen in Figure 2.13a), but the shape of the probability distribution of edge lengths is. Consider Figure 2.13b, which plots the normalized variance of edge length, the variance divided by the square of the mean, as time evolves for both a 2D and a 3D system. This is a scale-free property: if we scale the configuration by a factor ρ, the mean is multiplied by ρ and the variance by ρ^2. Thus, the Steady State Hypothesis

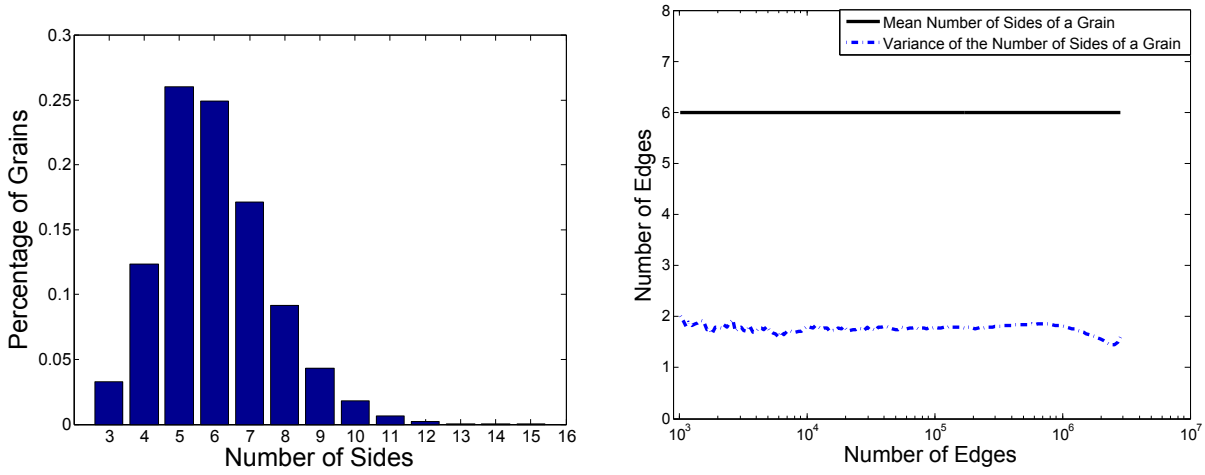


(a) (b)

Figure 2.13: The average length of an edge (a), and the normalized variance of edge length (b) plotted for 2D and 3D systems as they evolve. The evolution is parametrized by the number of edges, which generally decreases as time proceeds. Thus, these graphs should be read from right to left.

predicts that there will be a time interval in which this statistic is constant. This prediction is supported by the figure. Note that the number of edges is on the x-axis instead of time; this is to better enable the comparison of simulations that evolve at different rates (such as the 2D and 3D simulations shown). With a special exception discussed in Section 3.6, the number of edges decreases monotonically as the system evolves, so time goes forward as we move to the left in these plots. The deviation from the hypothesis at high and low numbers of edges is expected. When the number of edges is high, the system is still close to the initial conditions, not the "natural" condition resulting from sustained evolution under curvature flow. At the other end, when the number of edges decreases toward zero, statistical sampling error causes the plot to fluctuate. Furthermore, as the system decreases in size and the length of the edges approaches the same order of magnitude as the diameter of the ambient torus, it will begin to behave non-locally in strange ways.

Local topological statistics are a large class of scale-free properties. For example, if G is a planar graph, the distribution of the number of sides of each grain (component of the complement) is scale-free. The mean of this distribution is not an interesting statistic to us, as it is determined by the Euler characteristic of the ambient space and by the requirement that all vertices have degree



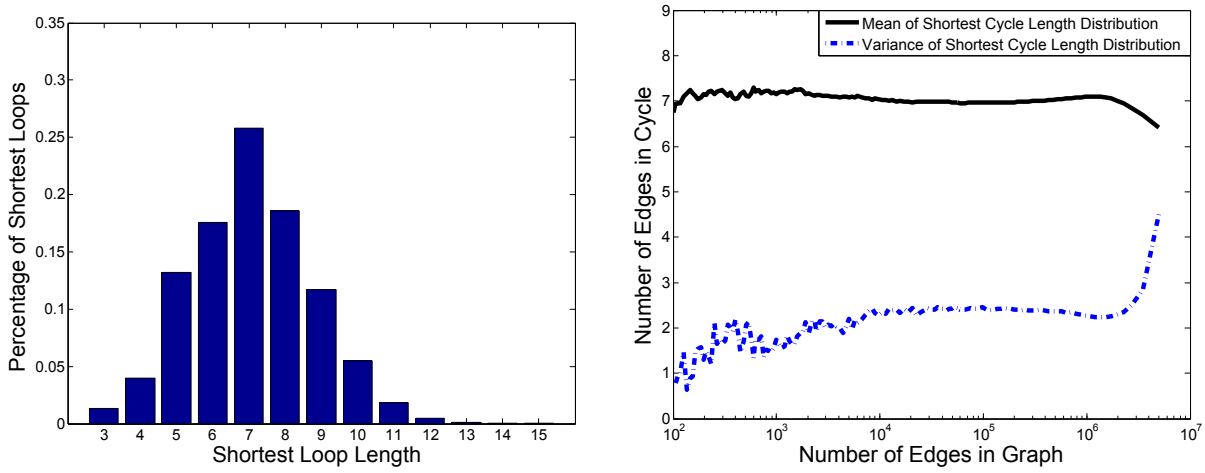
(a) (b)

Figure 2.14: The probability distribution of the number of sides of the grains in a planar graph is an important local topological statistic. A histogram of such a distribution is shown in (a), and how the mean and variance of that distribution evolve with time is shown in (b). The histogram in (a) is from the same system shown evolving in (b), when it has about 100,000 edges.

three. For example, if G is embedded in the two-torus, $\mathbb{T}^2$, $\chi(\mathbb{T}^2) = C - E + V = 0$ where C, E, and V are the number of grains, edges, and vertices respectively. Since every edge is adjacent to two vertices and every vertex is adjacent to three edges, $V = 2E/3$, so $C = E - V = E/3$. Because each edge is adjacent to two grains, this implies that the average number of edges of a grain is six. On the other hand, the shape of the distribution of edge lengths is an interesting scale-free property. A histogram showing this distribution is shown in Figure 2.14a, and the standard deviation is plotted in Figure 2.14b as the system evolves in time. As the Steady State Hypothesis predicts, there is a large time interval where this standard deviation remains essentially constant.

For a graph G embedded in three dimensions, there are no grains and only one component of the complement, so the statistics discussed in the previous paragraph do not make sense. Instead, we shall examine an analogue of them. For each edge, let $S(e)$ be the smallest number of edges in a cycle containing e.[3] The probability distribution of $S(e)$ is an interesting local topological statistic of G and an example distribution is shown in Figure 2.14a. Here, the mean of this distribution

[3]Note that, for graphs embedded in two dimensions, the shortest cycle containing an edge may not be the boundary of one of the two grains adjacent to it in general.



(a) (b)

Figure 2.15: The probability distribution of shortest cycle lengths for a 3D embedded graph is shown in (a), and the evolution of the mean and variance of that distribution over time is depicted in (b). The data for (a) is from the same system as (b) when it has around 100,000 edges.

is not determined by the ambient space and varies as the system evolves. Figure 2.15b shows the evolution of this statistic and the normalized variance of $S(e)$ as graph evolves under curvature flow. Again, the data is consistent with the Steady State Hypothesis. The properties of these shortest cycles will be discussed further in Chapter 6.

Traditionally, *ad hoc* measures of convergence such as the ones discussed in this section were used to claim the existence of a steady state. In the next section, we will put these claims on firmer ground by introducing a metric on embedded graphs, and regular cell complexes in general, that allows for the measurement of this convergence. This will also allow us to formalize the statement of the Steady State Hypothesis. The metric measures how similar the local topology of two configurations are. As one example convergence in it implies convergence of the cycle statistics introduced in the previous two paragraphs.

Chapter 3

The Method of Swatches

The method of swatches quantifies the statistical local topology of regular cell complexes. It was developed in collaboration with Jeremy Mason and Robert MacPherson, building on previous work of Mason and Lazar [44], and will appear in our paper "Topology of Random Cell Complexes, and Applications" [64]. The basic idea is to study probability distributions of local topological configurations in regular cell complexes, and to compare different regular cell complexes via metrics on these distributions. The method of swatches has many physical applications and can be used, for example, to quantify the similarity of regular cell complexes, to compare simulation with experiment, or to iteratively modify a structure to reach a desired state. Here, we will use it to formalize certain universality conjectures from physics and materials science.

Section 3.1 defines regular cell complexes, and gives several physical examples. The key concepts of a "swatch" and a "cloth" are introduced in Section 3.2. A metric on regular cell complexes is defined in Section 3.3 and the theory of convergence in this metric is described in Section 3.3.1. The Steady State Hypothesis is formalized using this language in Section 3.4, computational methods are discussed in Section 3.5, and applications to the embedded graphs from Chapter 2 are presented in Sections 3.6 and 3.7.

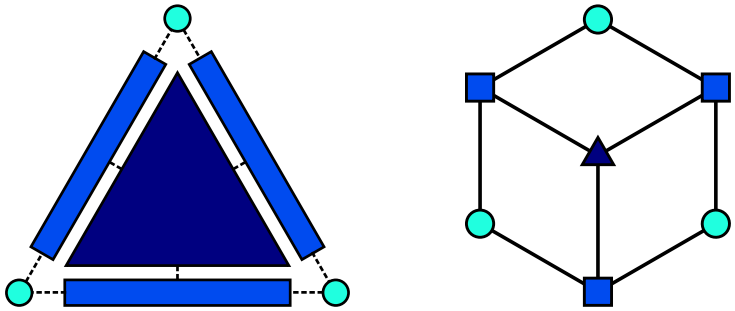


Figure 3.1: (a) Representation of a triangle as a cell complex by three 0-cells, three 1-cells, and one 2-cell. (b) Representation of a triangle as a graph with seven vertices, nine edges, and three vertex types.

3.1 Regular Cell Complexes

A regular cell complex is a space built inductively by attaching cells to each other, where an n-dimensional cell is homeomorphic to an n-dimensional ball. That is, a 0-cell is a point, a 1-cell is a line segment, and a 2-cell is a disk, and so on. The n-dimensional cells are attached via homeomorphisms from their boundaries (which are homeomorphic to $(n-1)$-dimensional spheres) into the structure built from the lower dimensional cells. In other words, they are attached without any twisting or tearing. Figure 3.1b shows a regular cell complex decomposition of a triangle.

The key property of regular cell complexes is that they can be represented by the adjacency graph of their cells, labeled by dimension. Edges are placed between adjacent cells of neighboring dimensions, as in as in Figure 3.1. This graph captures all topological properties of the cell complex. The combinatorial nature of this representation is easy to work with, and it is an essential building block in the method of swatches. More general definitions of cell complexes do not have this property, and therefore do not make sense for this framework. For example, the non-regular CW-complex decompositions for the 2-torus and Klein bottle depicted in Figure 3.2 give the same adjacency graph but different topologies.

As different regular cell complexes can have the same topology, it is important that the choice of decomposition be natural for the application at hand. All of the applications presented here have natural representations as regular cell complexes, so this issue does not arise in this thesis.

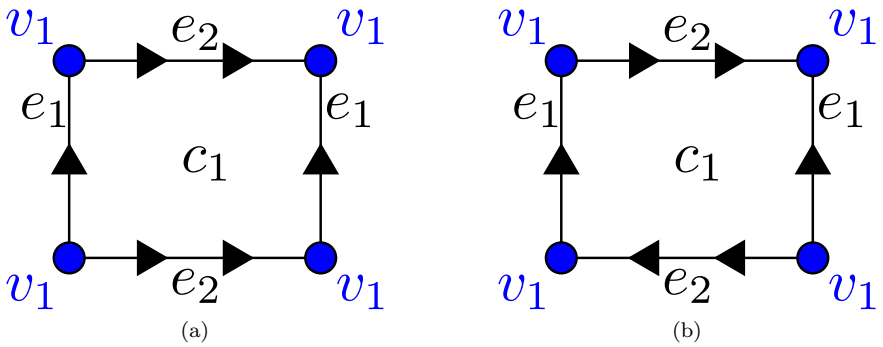
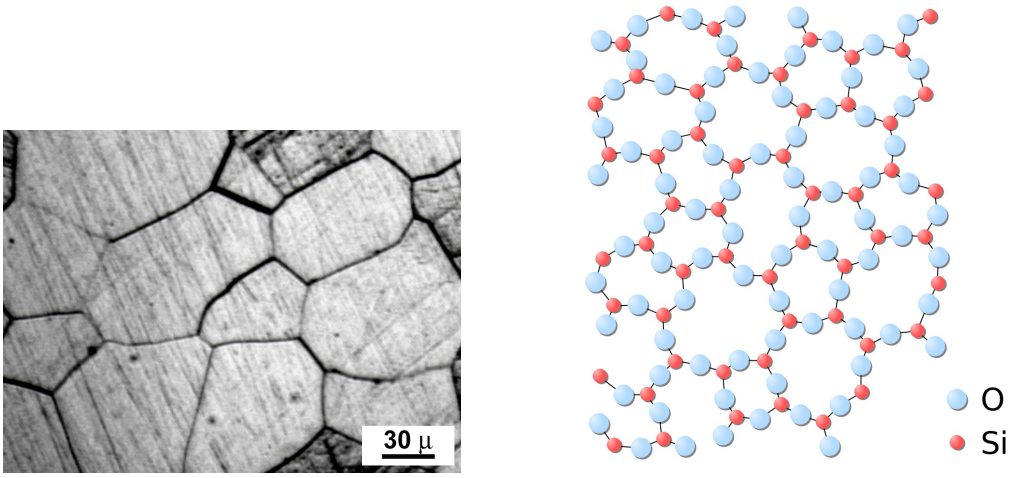


Figure 3.2: Non-regular cell complex decompositions of the (a) 2-torus and (b) Klein bottle with 1 0-cell, 2 1-cells and one 2-cell. Opposite edges are pasted together in the orientations given by the arrows. While these cell complexes have the same adjacency graph, they have different topologies.

3.1.1 Regular Cell Complexes in Nature

Regular cell complexes occur in many applications, with examples including the network of neural connections in the brain, the foam at the top of a mug of beer, and the graph structure of a social network. In this section, we will introduce two applications occurring in materials science: polycrystalline materials and oxide glass.

Recall from Section 2.1 that polycrystalline materials are composed of polygons (in two dimensions) or polyhedra (in three dimensions) each with its own crystalline orientation. Two dimensional polycrystalline materials can be represented as a regular cell complex in three different ways. The boundary network of the grains of such a material is a planar graph and therefore a one-dimensional cell complex. Adding the interior of the grains gives a two-dimensional cell complex but no additional topological information. Finally, the adjacency graph of these grains gives another one-dimensional cell complex that contains all local topological information about the material (the edges and vertices can be easily reconstructed from it). Three-dimensional polycrystalline materials also have three natural regular cell complex representations: their boundary network (a two-dimensional complex), the entire material (a three-dimensional complex), and the adjacency graph (a one-dimensional complex). Again, the information in all three representations is equivalent. A cross-section of 3D polycrystalline material is shown in Figure 3.3a. 3D polycrystalline materials also evolve by curvature flow in a process called 3D grain growth; for a treatment



(a) (b)

Figure 3.3: Regular cell complexes in nature: (a) A magnified photo of a cross section of a polycrystalline metal showing the grain boundary network [59]. (b) The network of covalent bonds in an oxide glass [27].

of that topic refer to [37].

Another important regular cell complex is the network of covalent bonds in an oxide glass, such as the one illustrated in Figure 3.3b. Like polycrystalline materials, this structure is disordered with no obvious symmetries. While the understanding of these materials is important both theoretically and in practice, the methods available to characterize them and their properties in terms of their local structure are surprisingly limited. We propose that the method of swatches be used to provide such a description for these and many other disordered materials.

3.2 Swatches and Clothes

3.2.1 Swatches

A swatch is a local topological configuration in a regular cell complex C. The swatch of radius r rooted at a root cell v is the ball of radius r centered at v in the adjacency graph of C, where every edge has length one. Recall that the adjacency graph has edges between adjacent cells of neighboring dimensions. The root cell is usually taken to be a 0-cell (vertex) of C, but may be a

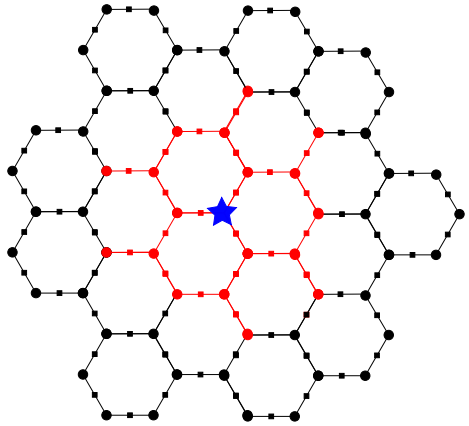


Figure 3.4: A swatch of radius 6 in a 1-dimensional regular cell complex.

cell of any dimension. A swatch of radius 6 in a one-dimensional cell complex is shown in red in Figure 3.4. The swatch is rooted at the 0-cell marked by the blue star. A subswatch of a swatch is a swatch of smaller radius at the same root vertex.

Two swatches have the same swatch type if they represent the same local topological configuration. This is equivalent to the existence of a graph isomorphism between them that fixes the root and respects the dimension of the cells. Two different swatch types from a one-dimensional cell complex are depicted in Figure 3.5. The one on the left is referred to as a free swatch, as there are no identifications between the cells in the neighborhood of the vertex.

It will be useful later to have a distance on swatches. Let the largest common subswatch of two swatches be the swatch of largest radius that is a subswatch of both. The distance between two swatches is defined as the reciprocal of the number of vertices in the largest common subswatch, or zero if the swatches are the same. For example, the largest radius for which the swatch in Figure 3.5b is free is $r = 3$, and the distance to the free swatch in Figure 3.5a is $1/13$.

3.2.2 The Cloth

Consider the probability distribution of swatch-types of radius r occurring at the 0-cells of a regular cell complex C. Letting r vary, we get a family of probability distributions called the cloth of C. The cloth characterizes the local topology of the cell complex in the sense of determining the probability

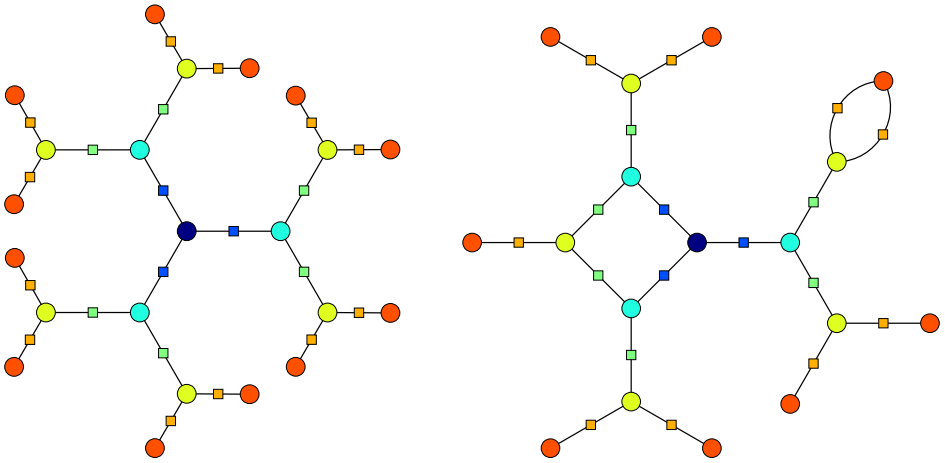


Figure 3.5: Swatches types of radius six in a cell complex containing only 0-and 1-cells. The vertex color indicates distance from the root, with the root colored dark blue. (a) A free swatch where there is a single path from the root to any given vertex. (b) A swatch that contains a cycle of length four and a cycle of length two.

at which any local configuration appears in the complex, as well as of prescribing all local topological properties of the regular cell complex (in the sense defined toward the end of Section 3.3.1).

The cloth is composed of levels, one for each value of the radius. Note that level r of the cloth contains strictly more information than all levels corresponding to lower radii, and the description of the cell complex becomes progressively more complete as the level increases. The relationship between different levels of the cloth is elucidated by a tree of swatch types, as in Figure 3.6. Level r of the tree is composed of all swatch types of radius r that occur in the type cell complex under consideration. An edge is placed between every swatch of radius r and its subswatch of radius $r-1$. Figure 3.6 is specialized to the case of the embedded graphs introduced in Chapter 2, where every vertex is adjacent to three edges, every edge is adjacent to two vertices, and the adjacency graph is 2-connected (meaning the removal of any one cell does not disconnect the adjacency graph).

The information in a cloth is equivalent to an assignment of swatch frequencies to each of the swatch types of the tree, subject to the conditions that the swatch frequencies of each level sum to one, and that the frequency of a swatch is equal to the sum of the frequencies of its immediate descendents.

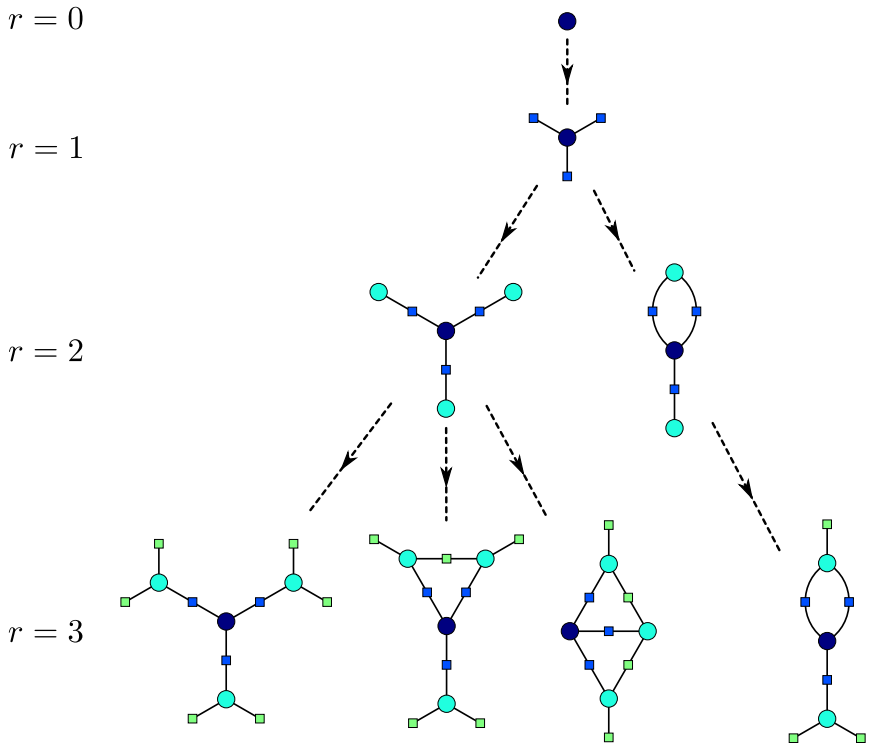


Figure 3.6: Levels 0 to 3 of the tree of swatch types, subject to the conditions that circle vertices be of degree three, square vertices be of degree two, and the graph be 2-connected.

3.3 Similarity and Convergence of Cell Complexes

We use the earth mover's distance on the probability distributions of local topological configurations at radius r to define a family of distances on cell complexes. The earth mover's distance is equal to the minimum cost of transforming one probability distribution on swatch types into a different probability distribution on the same swatch types. The transformation is performed by transferring probability mass between swatch types, with the overall cost given by the sum of the costs of the individual operations. The cost of an operation is the probability mass transferred times the distance between the two swatch types [52, 62]. Given two cell complexes C_1 and C_2, let $d_r(C_1, C_2)$ equal the earth mover's distance between probability distributions on swatches of radius r induced by the two cell complexes. Note that d_r is uniformly bounded and non-decreasing in r, and that it stabilizes for some finite r if the cell complexes are finite. The limit distance on cell complexes is defined as the limit of d_r with increasing r, or

$$d(C_1, C_2) = \lim_{r \to \infty} d_r(C_1, C_2)$$

d is impractical computationally, but is important for theoretical reasons. Note that these are only a few of the many possible distances that can be defined between regular cell complexes using the method of swatches. One feature of this method is its modularity; different distances can be introduced via the same framework based on the application at hand, by either modifying the probability distribution distance or the distance on swatches themselves.

3.3.1 Convergence

Consider a sequence of regular cell complexes $C_1, C_2, \ldots = \{C_i\}$, and suppose that it is a Cauchy sequence in the limit distance d. That is, the elements of the sequence all become arbitrarily close above some sufficiently large i. This condition is equivalent to the convergence of all swatch frequencies, and implies the convergence of other important properties as well. A key mathematical result of Benjamini and Schramm [7, 39] is that the sequence $\{C_i\}$ may be associated with a universal limit object σ. The limit object is not a cell complex itself, but is instead a probability

distribution on the space $\mathcal{C}^\bullet$ of countably infinite, connected regular cell complexes with a specified root cell (a root must be specified because swatches are inherently rooted). Sampling from σ may be viewed as sampling a random configuration from the universal state that the cell complex sequence approaches. Note that swatch frequencies for any radius r, and therefore the distance d, may be extended to such distributions: the frequencies are the probabilities that swatches of radius r appear at the root of a random cell complex drawn from σ. This allows the distance from a finite cell complex to σ to be computed, and makes the notion of a sequence of cell complexes converging to the probability distribution σ sensible.

The limit distribution σ is constructed by assigning probabilities to a collection of subsets of $\mathcal{C}^\bullet$ defined by swatches. Suppose that S is a swatch of radius r, and E_S is the set of all cell complexes in $\mathcal{C}^\bullet$ where S appears at the root vertex. These sets form a basis for a topology on $\mathcal{C}^\bullet$, and this is the same topology induced by the distance d. Thus, to define a probability distribution on $\mathcal{C}^\bullet$, it is sufficient to give consistent probabilities for each of these sets. Here, the probability of E_S is the limiting value of the swatch frequency of S in the sequence C_i as $i \to \infty$. By construction, C_i converges to σ in the sense that the distance d between C_i and σ becomes arbitrarily small for sufficiently large i.

The previous discussion is summarized by this theorem:

Theorem 1. *Let $C_1, C_2, \ldots = \{C_i\}$ be a Cauchy sequence of regular cell complexes in d, then there is a unique probability distribution σ on $\mathcal{C}^\bullet$ such that $\{C_i\}$ converges to σ in d. The value of σ on a basis set E_S is given by $\sigma(E_S) = \sigma_r(S)$ where r is the radius of the swatch S and σ_r is the limiting distribution on the set of swatches of radius r.*

The limit distribution has two important and interesting properties. The first is that the probability of E_S is equal to the sum of the probabilities of E_{T_i}, where T_i ranges over the swatch extensions of S of radius one greater than that of S. The second is called involution invariance, and is related to the fact that swatch frequencies at different radii are far from independent – given a swatch of radius r, the swatches rooted at neighboring vertices are determined up to radius $r - 1$. This property is formally defined using operations that swap a regular cell complex with a given

root cell with with one one centered at one of its immediate neighbors. To state it, consider the space $\mathcal{C}^\dagger$ of regular cell complexes with both a distinguished root cell and a distinguished edge incident to that root. Define a probability distribution on this space:

$$\sigma^\dagger(E_S^e) = \frac{\sigma(S)}{\overline{\deg}},$$

where E_S^e is the set of extensions of a swatch S with specified root edge e, and $\overline{\deg}$ is the average degree of all root cells sampled from σ. If the involution on $\mathcal{C}^\dagger$ given by swapping the root of a regular cell complex with an edge specified to the other cell adjacent to that edge preserves this probability measure, σ is said to be involution-invariant. All distributions on $\mathcal{C}^\dagger$ arising from convergent sequences have this property. While the definition of involution invariance requires that we define swatches at root cells of any dimension, it implies that a similar consistency property holds if we restrict our consideration to regular cell complexes rooted at 0-cells (no two 0-cells are adjacent). Also, note that a sequence of regular cell complexes converges in the distance d based on swatch frequencies rooted at 0-cells if and only if it converges in a different distance based on swatches rooted at any cells. This is true because the frequency of any swatch rooted at a d-cell can be expressed in terms of larger swatches rooted at nearby 0-cells.

3.3.2 Consequences of Convergence

The convergence of a regular cell complex sequence to a limit implies the convergence of all local topological properties of that sequence as well. For example, the probability than an edge belongs to a cycle of length four will converge. More precisely, let H be the labeled adjacency graph of a square, let G_i be the labeled adjacency graph of the cell complex C_i, and let $\text{inj}(H, G_i)$ be the number of times H appears in G_i. Although both $v(G_i)$ (the number of vertices of G_i) and $\text{inj}(H, G_i)$ will usually diverge with increasing i, if $\{G_i\}$ converges, the normalized quantity $\text{inj}(H, G_i)/v(G_i)$ will converge as well. [39]

More generally, the normalized number of adjacency preserving maps from H to G_i converges for any labeled graph H. [39] This may be used to find, e.g., the probability that a 0-cell is adjacent to

a specified number of 1-cells, the probability that a 1-cell is connected to 0-cells of specified degree, the joint probability of adjacent 2-cells being incident to specified numbers of 1-cells (as in the Aboav-Weaire relation [70, 4] for a 2D microstructure), or the probability of a 1-cell participating in a cycle of specified length. In this sense, the cloth provides a complete description of the local topology of the underlying cell complex, as initially claimed at the beginning of this chapter.

Convergence in d also implies the convergence of some, but not all, global properties of the regular cell complexes. By this we mean properties that can be expressed in terms of adjacency-preserving maps from the adjacency graph of G_n into a graph H. We should not expect all global properties to converge with this metric, nor should we want them to as the local topology of a graph should not depend on whether the ambient space is $\mathbb{R}^n$ or $\mathbb{T}^n$, for instance. An example of a convergent global graph property is the number q-colorings of C_n for sufficiently high q, i.e., the number of different ways that q colors may be assigned to the cells of C_n such that no vertices connected by an edge have the same color. For a more thorough treatment of this and other topics related to Benjamini-Schramm convergence, the interested reader is encouraged to refer to chapters 17-22 of the book "Large Networks and Graph Limits" by to the book by Lovàsz [39].

3.4 Conjectures

Now, we may formalize the Steady State Hypothesis first stated in Section 2.5, and introduce several other universality conjectures. We will state them here for the case of infinite graphs, but they can be easily adapted to context of curvature flow on finite graphs where the size of the initial conditions increases toward ∞.

Conjecture 2 (Steady State Hypothesis, Formalized)**.** *Suppose Ω is the dynamical process on embedded graphs in $\mathbb{R}^n$ in which the edges evolve by curvature-flow and the vertices move by any of the rules defined in Section 2.3. There exists a unique probability distribution σ_Ω on the space of abstract one-dimensional regular cell complexes with a root vertex specified such that for almost all infinite graphs $G \subset \mathbb{R}^n$ and any sequence $\{t_i\}$ of positive real numbers with $t_i \to \infty$, the sequence $G(t_1), G(t_2) \dots G(t_i), \dots$ converges to σ_Ω in the limit distance d.*

At first glance, this conjecture [formalized steady state hypothesis] appears weaker than Conjecture 1 [informal steady state hypothesis], since it requires only the convergence of topological properties. One could define many alternative distances on regular cell complexes that take into account metric information as well as topological information[1]. For convergence in any distance $\hat{d}$ to make sense in this context, it must be scale-free, statistical (since initial conditions are generated stochastically), and local (since the global topology should not matter). We believe that the local topological properties of graphs evolving under curvature flow will not stabilize unless all scale-free properties of the graph have stabilized. Formally,

Conjecture 3. *Let Ω be as above. Let G be an infinite, embedded graph such that for any sequence $\{t_i\}$ of positive real numbers with $t_i \to \infty$, $G(t_1), G(t_2) \dots G(t_i), \dots$ is a convergent sequence of regular cell complexes. Then, $G(t_1), G(t_2) \dots G(t_i), \dots$ is a Cauchy sequence in any distance $\hat{d}$ that is scale-free, statistical, and local.*

Finding appropriate definitions for the "scale-free," "statistical," and "local" is the first challenge to approaching this conjecture. Note that the converse does not hold unless some "unboundedness" condition is put on $\hat{d}$: it must be able to see beyond neighborhoods of any fixed size. This is to rule out distances like d_0 and d_1 that converge trivially in this case.

Recall that three different sets of equations of motion were defined for planar graphs evolving under curvature-flow, the von–Neumann Mullins Equations, the Vertex Drag Equations with vertex drag parameter λ, and the Vertex Model. We conjecture:

Conjecture 4 (Interchange of Limits)**.** *There exist distinct limit distributions for the von-Neumann–Mullins Equations, the Vertex Drag Equations for each λ, and the Vertex Model. Call these σ_0, σ_λ, and σ_∞, respectively. Then, as the notation suggests, $lim_{\lambda \to 0} \sigma_\lambda = \sigma_0$ and $lim_{\lambda \to \infty} \sigma_\lambda = \sigma_\infty$.*

The second half of this conjecture is effectively a statement about the interchange of limit operations, as the Vertex Model is a limit model of the Vertex Drag Equations. The first half is perhaps deeper, as the von-Neumann–Mullins Equations have a fundamentally different form

[1] We believe such metrics will be very useful in other applications.

than the Vertex Drag Equations, but are derived based on the Herring Angle Condition which is a consequence of the limit $\sigma \to 0$. Evidence for this conjecture will be discussed in Section 3.7.2.

In three dimensions, recall that two different rules were proposed for what happens when two edges that do not share a vertex intersect.

Conjecture 5. *Let Ω_1 be curvature flow on graphs embedded in three-dimensions where edges pass freely through each other when they intersect, and let Ω_2 be the same except that a new edge is added at the point of intersection as in Figure 2.5c. Then σ_{Ω_1} and σ_{Ω_2} are equal as probability distributions.*

Finally, we state a similar conjecture for the grain growth of 3D polycrystalline materials:

Conjecture 6. *Let Ξ be the dynamical process on three-dimensional regular cell complexes in $\mathbb{R}^3$ defined by curvature flow on the faces and an appropriate, consistent set of rules on the edges and vertices. There exists a unique probability distribution σ_Ξ on the space of three-dimensional regular cell complexes with a root vertex specified such that for almost all admissisable initial conditions D and any sequence $\{t_i\}$ of positive real numbers with $t_i \to \infty$, the sequence $D(t_1), D(t_2) \dots D(t_i), \dots$ converges to σ_Ξ.*

Unfortunately, the distribution of local topological configurations occurring in 3D grain growth is far too broad for the computational methods presented here to provide evidence for this conjecture, even at low radii. We hope to investigate it by other means in the future.

3.5 Method of Computation

To test the conjectures stated in the previous section, we compute cloths for a variety of different embedded graphs throughout their evolution via curvature flow (simulated using the methods described in Section 2.6). To do this, we represent the graphs as one-dimensional regular cell complexes, then sample swatches at each of the vertices for radii less than a certain value r. We use the program *Nauty* to compute canonical forms for the adjacency matrices of these swatches, and thus

are able to classify them up to swatch type. While the graph isomorphism problem is thought to be difficult in theory, *Nauty* works very quickly for all of the examples in this computation.[2] [47]

To compute the distance d_r between two cloths C_1 and C_2 defined in Section 3.3, we represent the set of swatches appearing in them as a tree of swatches as in Section 3.2.2. With this data structure, the distance is computed inductively as follows. The distance d_0 is always 0. Knowing d_{r-1}, we calculate the additional contribution from all swatches that split at radius r. Suppose a swatch S of radius $r-1$ splits into $T_1, \ldots T_n$ at radius r, and that they occur with frequencies $\phi^1_{r-1}(S)$ and $\phi^1_r(T_i)$ in C_1 and $\phi^2_{r-1}(S)$ and $\phi^2_r(T_i)$ in C_2. As the difference between $\phi^1_{r-1}(S)$ and $\phi^2_{r-1}(S)$ is already accounted for by d_{r-1}, the additional amount of probability mass that must be moved to transform one into the other is

$$m_S := \frac{-\left|\phi^1_{r-1}(S) - \phi^2_{r-1}(S)\right| + \sum_{i=1}^{n}\left|\phi^1_r(T_i) - \phi^2_r(T_i)\right|}{2}$$

As these swatches all have S as a subswatch of radius $r-1$, the total cost contributed by this split is $m_S/|S|$, where $|S|$ is the number of cells of S. To get d_r, add d_{r-1} to the sum of the contributions of all swatches that split at radius r. This approach only works when the distance between two swatches is a function of their largest common subswatch.

Separate implementations of these algorithms were programmed by both the author and Jeremy Mason, and the author plans to post his software online by the date of completion of this thesis. The two programs work for general regular cell complexes, not just those of dimension one.

To test the conjectures stated the previous section, we found candidate steady state configurations for which all measured properties had converged. Then, we tracked the distance from an embedded graph to one or more of these reference conditions as it evolves. If the Steady State Hypothesis holds and the reference condition is in fact from the steady state, this distance should decrease toward zero and stabilize for a significant period of time.

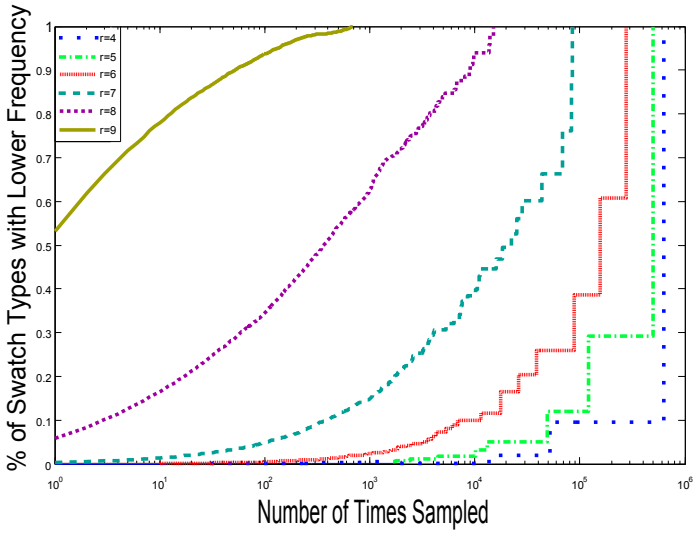


Figure 3.7: The percentage of swatch types that occur with a frequency below a certain value for several different radii r. The x-axis is the number of times a swatch has been sampled in a configuration, and the y-axis is the percentage of all swatch types that occur less than this many times. The data from this figure was computed from a 3D embedded graph with one million edges resulting from a simulation beginning with a Voronoi graph with about twenty million edges.

3.5.1 Choosing r

Since the number of swatch types increases dramatically as a function of radius, the sample size required to accurately compute the cloth increases dramatically as well. To see this, examine Figure 3.7 which shows the cumulative percentage of swatch types that appear at below a certain frequency for several different values of r. The data from this figure was computed from a 3D embedded graph with one million edges resulting from a simulation beginning with a Voronoi graph with about twenty million edges. As shown in the figure, between 50% and 60% of the vertices in the configuration have swatch types shared with no other vertex at radius r=9, and about 6% have unique swatches at radius r=8. This indicates that radius $r = 7$ is the highest one for which the swatch frequencies have been sufficiently sampled. Nevertheless, there are so many swatch types for this radius that the cloth still offers a very detailed description, and we use the distance d_r with $r = 7$ in the following.

[2]The Graph Isomorphism problem is in NP but it is unknown whether it is NP-complete.

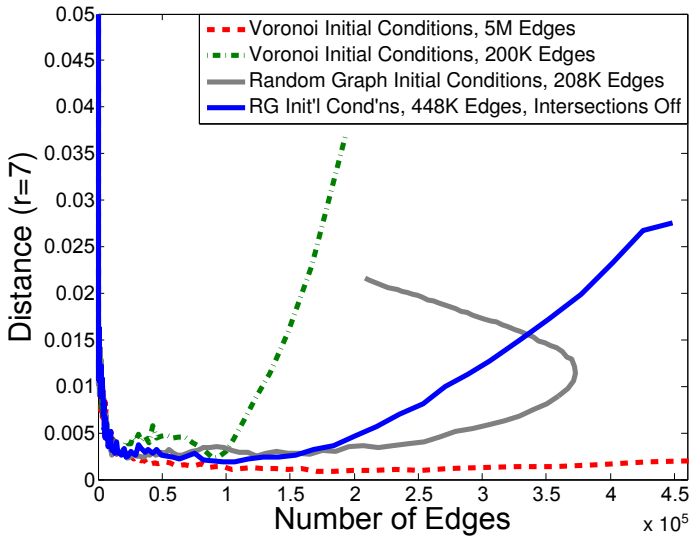


Figure 3.8: Distances of several simulations to the reference condition.

The curves in the Figure 3.7 become blocky toward the right of the figure for lower values of r because some swatch types are very common in the configuration. Unless r is sufficiently high, the most common swatch type is the free swatch of radius r.

3.6 Applications to Embedded Graphs in Three Dimensions

While the mathematics of graphs embedded three dimensions is in general more complicated than for planar graphs, we will begin our discussion of applications of the method of swatches with that case. This is because we have only one set of equations of motion in three dimensions, as described in Section 2.6.2, and therefore only one conjectured universal limit distribution. We will provide evidence for the existence of such a distribution, as well as Conjecture 5.

We used the regular cell complex distance d_7 to compare the simulated embedded graphs to a candidate steady state throughout their evolution. This reference condition was an embedded graph with 10^6 edges resulting from a simulation that started as a Voronoi graph with $6.1 \cdot 10^7$ edges, and for which all measured properties had converged.

Figure 3.8 shows the distance to the reference condition for four simulations. For all four, the

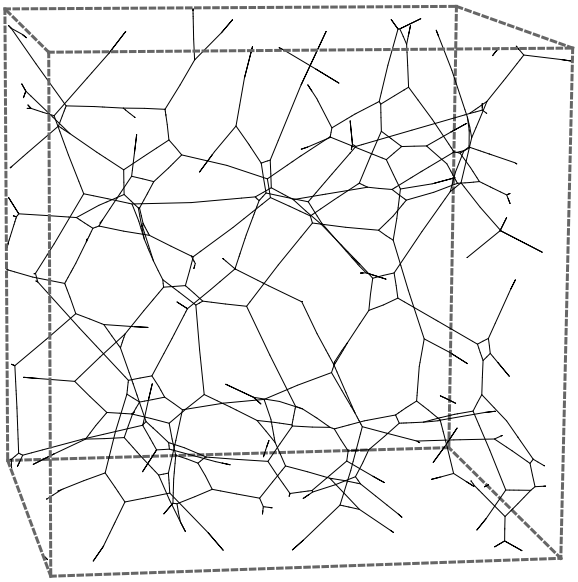


Figure 3.9: A small region of the reference condition for 3D embedded graphs.

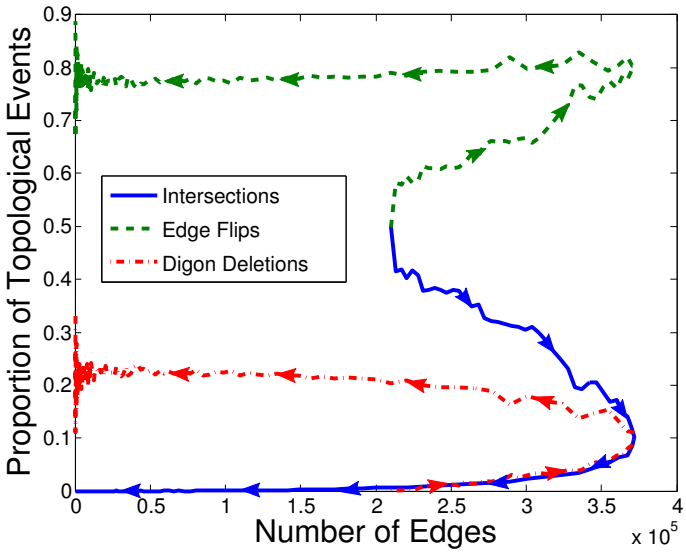


Figure 3.10: Relative rates of topological changes throughout a simulation with a random graph initial condition. As before, the evolution is parametrized by the number of edges which generally decreases as time proceeds (except for the system plotted in solid gray).

distance to the reference condition decreased rapidly as they evolved, indicating convergence toward the candidate steady state depicted in Figure 3.9. The evolution of the systems is parametrized by the number of edges in the system, which generally decreases with time but can increase if there are frequent edge intersections. For example, the simulation for one of the random graph initial conditions (shown by the solid gray line in Figure 3.8) initially experienced many intersections, leading to a transient where the number of edges increased. As shown in Figure 2.5c though, the number of intersections as a fraction of all topological changes declined as the simulation progresses, eventually decreasing to almost none in the steady state. This suggests that the steady state may be insensitive to the detection of edge intersections. As further evidence for this, the remaining three simulations in Figure 3.8 allowed edges to pass freely through each other, and yet converged to the same candidate steady state. This provides strong evidence for Conjectures 2 and 5, and allows the simulation to be made more efficient by neglecting the computationally expensive edge intersections.

3.6.1 Finite Size Effects

While Figure 3.8 indicates that the distance to the reference condition generally decreases as the simulation proceeds, the distance to the reference condition visibly increases for small numbers of edges. This is explained by the decrease in the sample size increasing the statistical error in the swatch frequencies and increasing the apparent distance to the reference condition. A rigorous test of convergence to the steady state should account for this source of error. Let R_n be a set of representative subsamples of the candidate steady state. A measured distance may be compared to the distribution of distances from elements of R_n to the reference condition; if the measured distance falls within one standard deviation of the mean of this distribution, then the network being considered is likely in the same statistical state as the candidate steady state. Practically, the set R_n contains random subsamples of the reference condition. A single subsample is constructed from the vertices and edges within some radius of a randomly selected vertex to attain the desired number of edges, with vertices on the boundary excluded.

This procedure is used to evaluate the convergence of several simulations to the steady state in

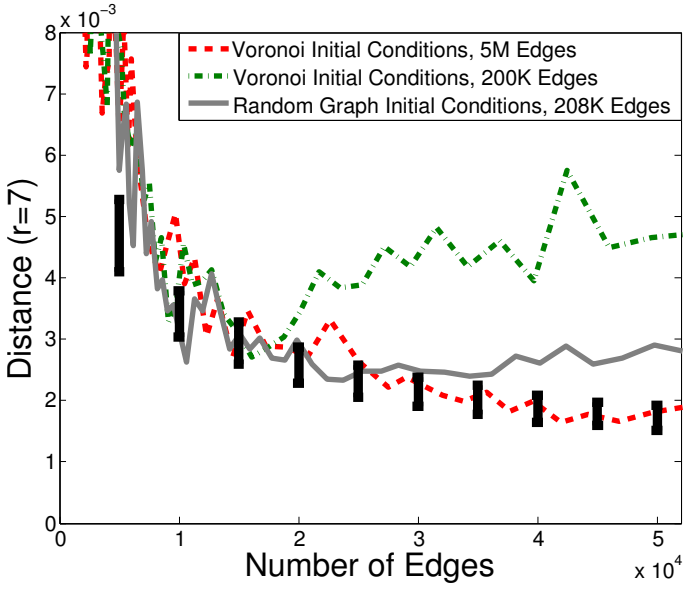


Figure 3.11: Distances of several simulations to the reference condition. Error bars show one standard deviation of the distance between a steady state-configuration with the indicated number of edges and the reference condition.

Figure 3.11. The error bars in the figure extend one standard deviation above and below the mean distance from the elements of R_n to the reference condition. Note that the simulation starting from a random graph (the solid line) is within one standard deviation of the subsamples throughout most of the interval between $25,000$ and $10,000$ edges, and that the same is true for the simulation starting from the larger Voronoi graph (the evenly dashed line) between $50,000$ and $10,000$ edges. This offers strong evidence that they have converged to the same statistical state as the reference condition in the indicated intervals. By contrast, the simulation starting from the smaller Voronoi graph (the unevenly dashed line) is within one standard deviation of the subsamples only for a short interval before the effects of the periodic boundary conditions overwhelm those of the limited sampling. Generally, we find that simulations with Voronoi graph initial conditions do not converge until there remains only around one-tenth of the initial number of edges.

3.7 Applications to Embedded Graphs in Two Dimensions

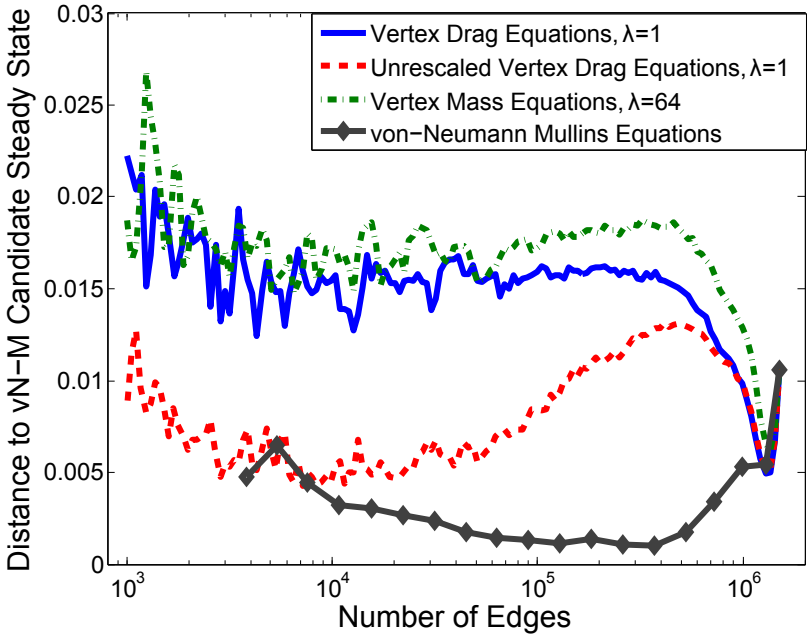
Recall that several different equations of motion were introduced for embedded graphs in two dimensions: the Vertex Model, the von Neumann–Mullins Equations, the Vertex Drag Equations, the Unrescaled Vertex Drag Equations, and the Vertex Mass Equations. The latter three all have a positive, real parameter λ. First, we will use the method of swatches to compare the last four sets of equations for a fixed value of λ, and then we will see how varying λ affects the long term behavior of systems evolving under the Vertex Drag Equations. We will provide strong evidence for Conjecture 2 (the Steady State Hypothesis) and Conjecture 4 (which holds that the Vertex Model and von Neumann–Mullins Equations produce systems that are limiting cases of the Vertex Drag Equations as λ approaches ∞ or 0, respectively).

3.7.1 Applications to Planar Graphs, Part 1

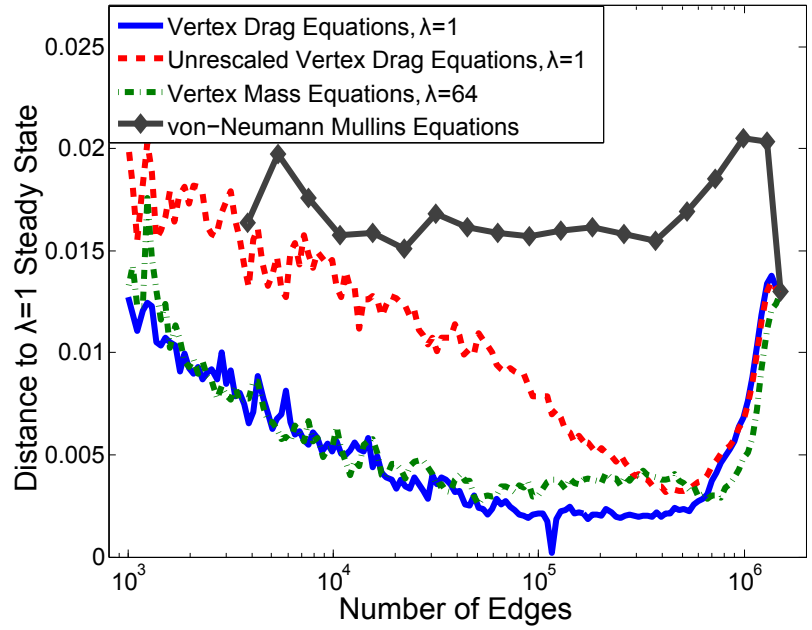
Consider Figure 3.12, which shows the distance from several graphs evolving by curvature flow to two candidate steady states. These systems all began with the same Voronoi decomposition of 500,000 Poisson distributed points on the 2-torus. The reference condition in Figure 3.12a evolved from a different Voronoi decomposition of the same size under the von Neumann–Mullins Equations until it had 184551 edges remaining. The candidate steady state in Figure 3.12b is taken from the system evolving by the Vertex Drag Equations shown by the solid blue line at the point when it had 115233 edges. This is why the solid blue line goes to zero at this point in the evolution[3]. All measured scale-free properties had converged for the two reference conditions.

First let us examine the behavior of the system evolving under the Unrescaled Vertex Drag Equations with $\lambda = 1$. This is shown by the evenly dashed red line in both figures. Recall that the vertex drag parameter is rescaled in the Vertex Drag Equations to keep the movement of the vertices and edges consistent as the size of the average edge increases. On the other hand, in the Unscaled Vertex Drag equations, the effective value of vertex drag decreases toward zero as the system coarsens and the edges slow down. As predicted in Section 2.3.3, this causes the system to

[3]Note how quickly the distance bounces back.



(a)



(b)

Figure 3.12: The distance from several simulations to candidate steady states resulting from (a) the von Neumann–Mullins Equations and (b) the Vertex Drag Equations with $\lambda = 1$. As before, the evolution of the systems is parametrized by the number of edges, which in this case is monotonically decreasing with time.

drift slowly toward the von-Neumann Mullins reference state.

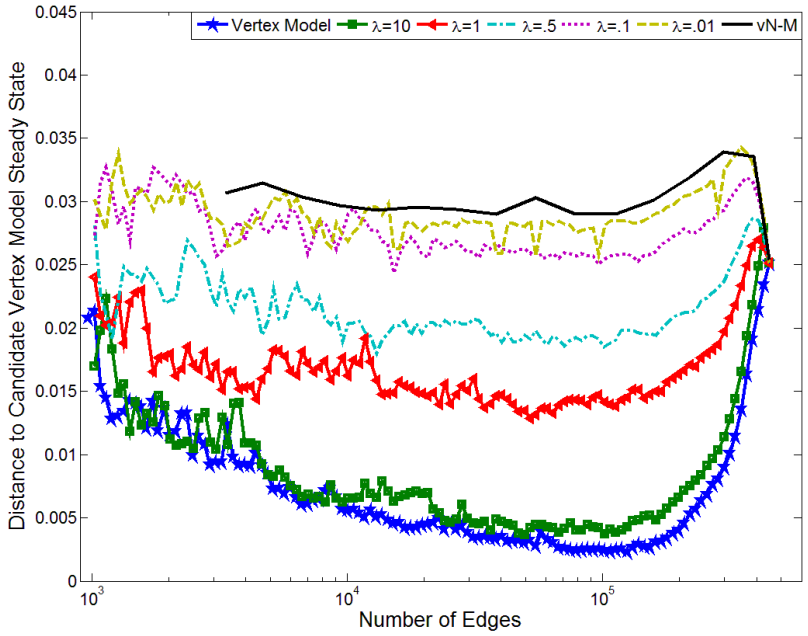
The Vertex Mass Equations were introduced in Section 2.3.3 as an alternative to the Vertex Drag Equations for which the behavior would be locally defined but still consistent across time. A system evolving by these equations is shown by the unevenly dashed green line in the above figures. As time proceeds, this configuration approaches the candidate steady state for the Vertex Drag Equations. The value $\lambda = 64$ for this simulation was chosen in order to produce this result, but by varying λ one should get the same range of behavior exhibited by the Vertex Drag Equations in the next section.

3.7.2 Applications to Planar Graphs, Part 2

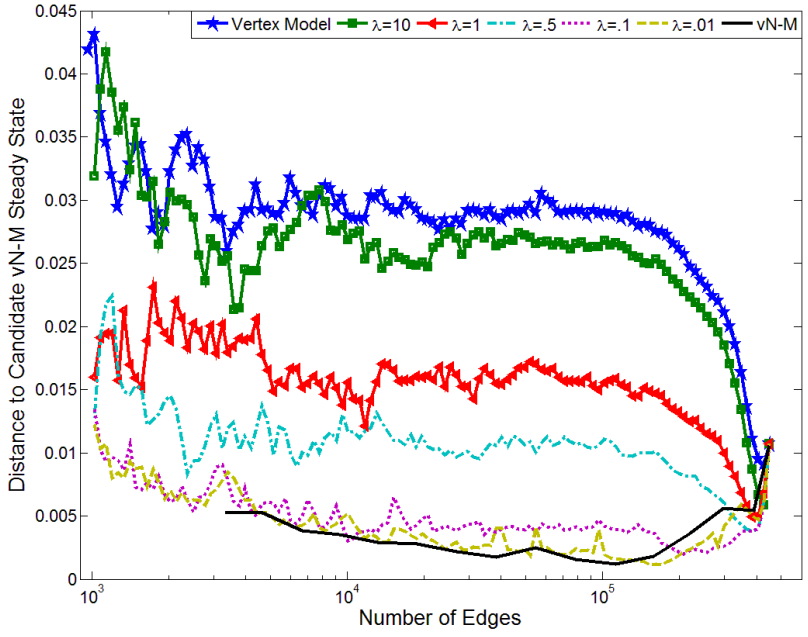
Conjecture 4 states that the Vertex Drag Equations give rise to a distinct steady state for each value of λ, and that the Vertex Model and von Neumann–Mullins Equations yield steady states that are the limiting cases as $\lambda \to \infty$ and $\lambda \to 0$. Figures 3.13 and 3.14 provide strong evidence for this conjecture.

Figures 3.13a and 3.13b track the distance of several simulations to candidate steady states for the Vertex Model and von Neumann–Mullins Equations, respectively. The reference conditions were generated from two simulations beginning with the same Voronoi decomposition of 500,000 Poisson distributed points in the 2-torus, and the other systems began with a Voronoi decomposition of 150,000 Poisson distributed points. In Figure 3.13a, the systems approach the Vertex Model reference condition as λ increases, whereas in Figure 3.13b they approach the von Neumann–Nullins reference condition as $\lambda \to 0$. The data in these figures is summarized by Figure 3.14 which shows the maximum distance of the simulations to the two reference conditions when they have between 30,000 and 100,000 edges. These three figures provide strong computational evidence for Conjecture 4.

A set of candidate steady state conditions for several values of λ are shown in Figure 3.15. The images show the same region in simulations begun with the same set of initial conditions, when they had reached 50,000 edges.



(a)



(b)

Figure 3.13: The distance from several simulations to candidate steady states resulting from (a) the Vertex Model and (b) the von Neumann–Mullins Equations. Note that the Vertex Drag Equations approach the Vertex Model reference states as $\lambda \to \infty$ and the von Neumann–Mullins reference states as $\lambda \to 0$.

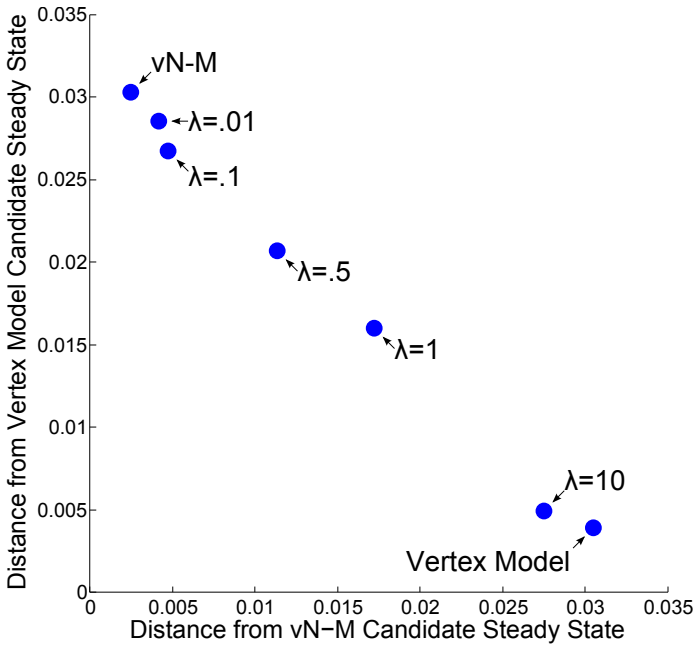


Figure 3.14: The distance from several simulations to candidate steady states for the Vertex Model and von Neumann–Mullins Equations, during a time interval in which the distances appear to have stabilized (between 30,000 and 100,000 edges). As the distances fluctuate due to finite-size effects, the maximum distances in this interval are displayed. This figure shows evidence for a one-parameter family of steady states.

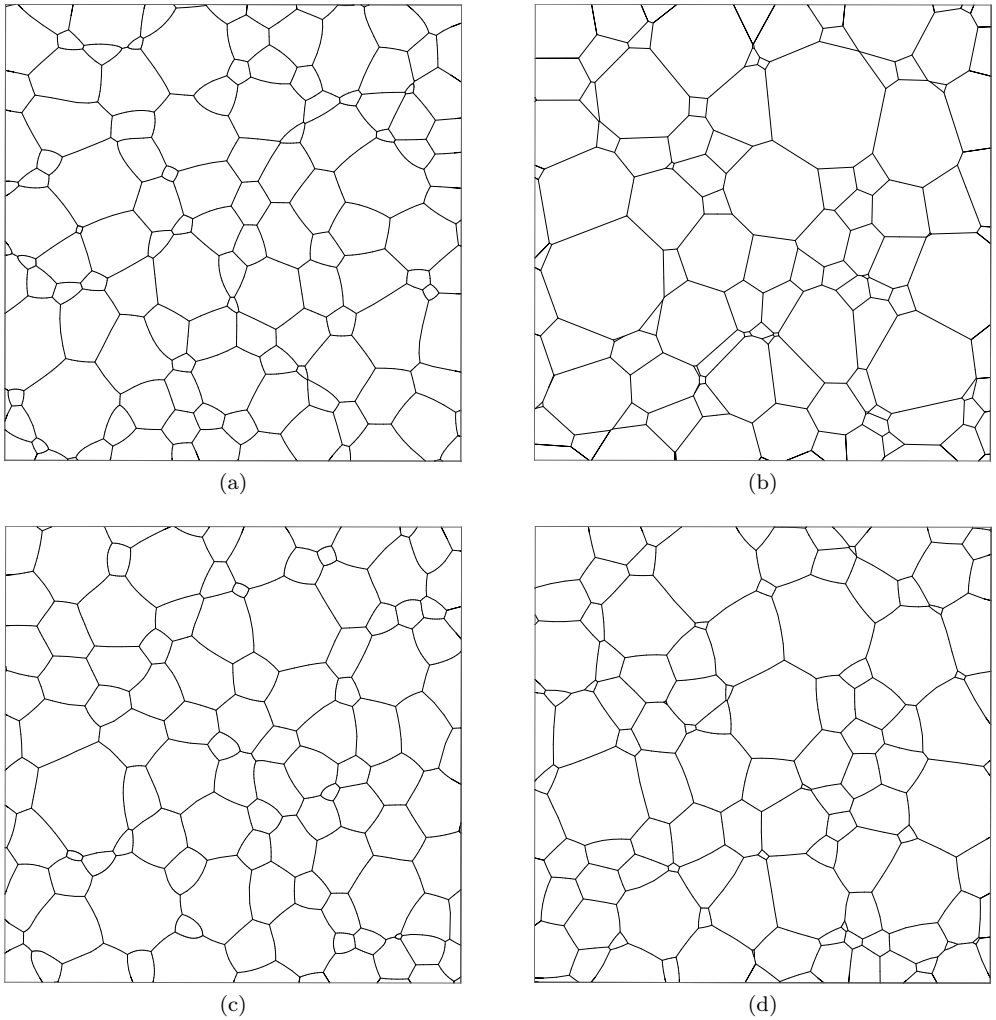


Figure 3.15: Candidate steady state conditions for (a) the von-Neumann Mullins Equations, (b) the Vertex Model, and the Vertex Drag Equations for (c) $\lambda = .01$ and (d) $\lambda = 1$. The pictures show the same region of simulations started from the same initial conditions.

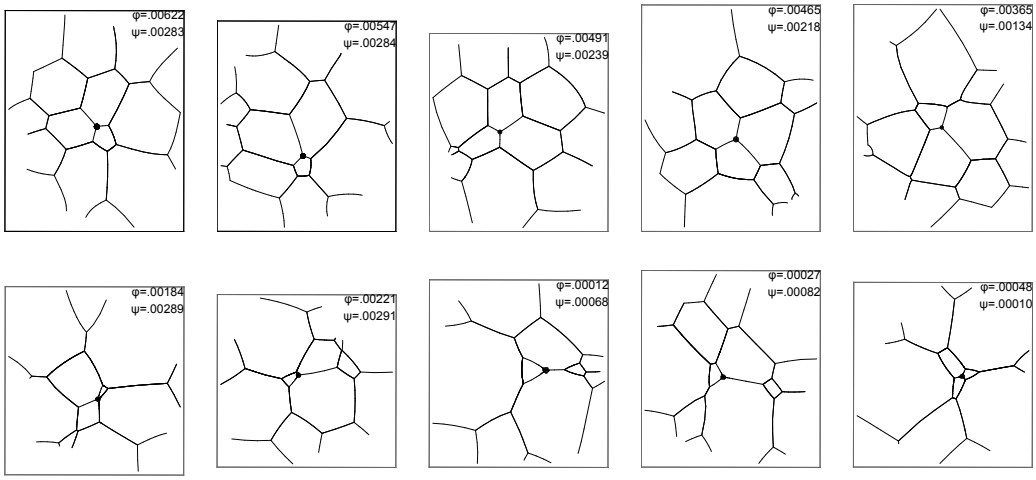
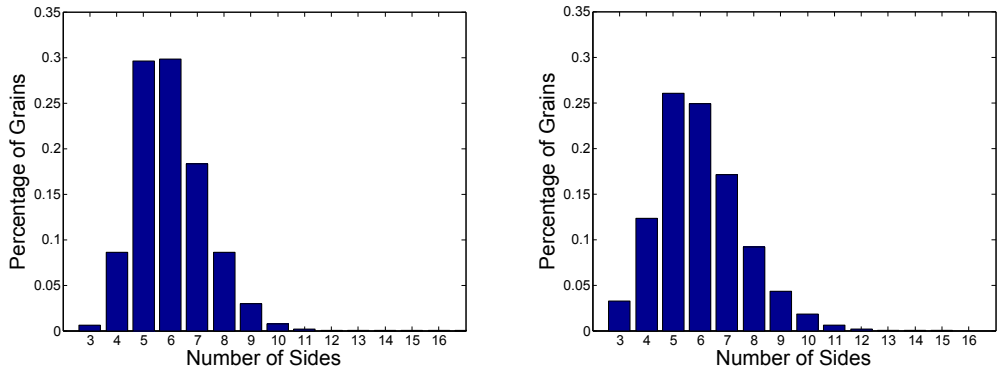


Figure 3.16: Ten example swatches from candidate steady state configurations resulting from the von Neumann–Mullins Equations (first row) and the Vertex Drag Equations with $\lambda = 1$ (second row). The frequencies of each swatch type are given by φ (for the von-Neumann Mullins configuration) and ψ (for the Vertex Drag configuration). The swatches in each row are those which have the highest frequency of occurring in that configuration minus the probability that they occur in the other configuration. The root vertices are marked in each swatch.

3.7.3 Digging Deeper

In this section, we will use the method of swatches to examine why two regular cell complexes are different. Specifically, we will compare a configuration C_1 resulting from the von Neumann–Mullins Equations such as in Figure 3.15a and a configuration C_2 that evolved by the Vertex Drag Equations with $\lambda = 1$ such as the one in Figure 3.15d. Let the swatch frequency of S in C_1 be $\varphi(S)$ and the frequency of S in C_2 be $\psi(S)$. Ten example swatches of radius $r = 8$ from the two configurations are shown in Figure 3.16. The five swatches in the top row are the ones for which $\varphi - \psi$ is greatest, and the ones in the bottom row maximize $\psi - \varphi$. These swatches can be viewed as the ones that contribute most to the distance between the two graphs.

The first thing to notice is that the values of both ψ and ϕ are much larger in the top row than in the bottom one. In fact, three of the five swatch types in the top row are in the top five most common swatches in the von Neumann–Mullins configuration, and the first two in that row are among the top five most common in the vertex drag configuration. That is, the swatches in C_1



(a) (b)

Figure 3.17: Histogram of the number of sides in each grain occuring in configurations resulting from (a) the von Neumann–Mullins Equations and (b) the Vertex Drag Equations with $\lambda = 1$.

which contribute most to the distance between it and C_2 are among those that are most common in both C_1 and C_2, and the swatch frequencies of C_1 are heavily concentrated in fewer swatch types. The distribution of local topological configurations in C_2 is much broader - not only are the most common swatch types less frequent than in C_1 but there are also almost twice as many distinct swatch types (15176 as opposed to 8383).

Also, note that all of the swatches in the second row have at least one grain with three or four sides, whereas none of the swatches in the first row have grains with this number of sides. Similarly, two of them have grains adjacant to the root with more than 8 edges while none of the swatches in the first row do. This suggests that a difference in the distribution of the number of sides of each grain contributes greatly to the swatch distance between the two configurations. These distributions are shown in Figure 3.17, and the one corresponding to the Vertex Drag Equations is broader as expected.

The swatch distance also picks up on more subtle correlations between the number of sides of neighboring grains, but these simple effects are the ones most apparent when looking at only a few example swatches.

Chapter 4

Persistent Homology

4.1 Introduction

Persistent Homology (abbreviated PH) is a powerful tool for studying complex geometric structures. Suppose X is a geometric structure embedded in a metric space such as n-dimensional Euclidean space, $\mathbb{R}^n$, or the flat n-dimensional torus, $\mathbb{T}^n$. Let X_ϵ be the ϵ-neighborhood of X: the set of points within distance ϵ of it. PH tracks how the topology of X_ϵ changes as ϵ increases - intuitively, as X thickens. PH is indexed by a dimension $i < n$; the i-th dimensional PH, PH_i, corresponds to i-dimensional geometric features of X. For example, consider Figure 4.1, showing the arc of a circle of length $\pi + 2\theta$. As ϵ increases, a hole forms and disappears. This topological change is captured

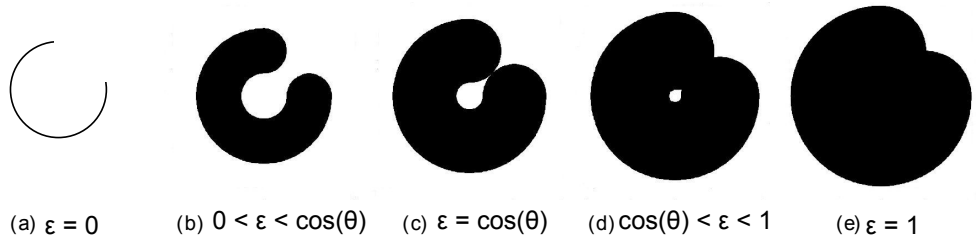


Figure 4.1: Neighborhoods of a circle of radius 1 with an arc removed. This object has one PH_1 class, whose birth is $\epsilon = \cos(\theta)$, and death is one.

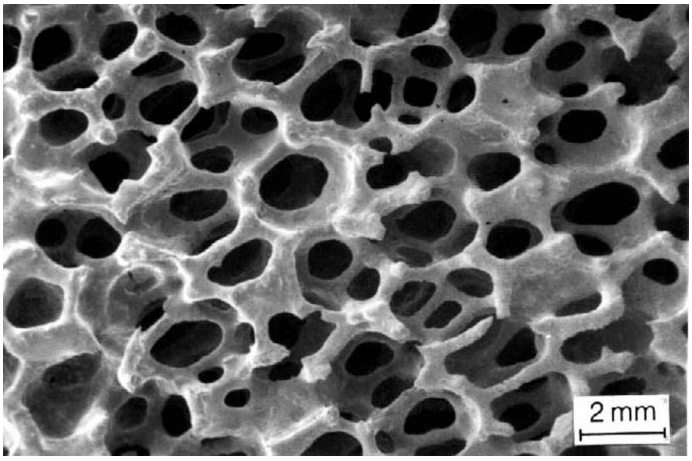


Figure 4.2: An open-cell foam. [21] One would like to recover a 3D cell complex from this materials in order to study its properties.

by the 1-dimensional PH, PH_1. More generally, let $X \subset \mathbb{R}^2$ and consider the holes that form and disappear in X_ϵ as ϵ increases. Each hole is represented by a PH_1 interval (b_i, d_i), where the birth b_i is the ϵ at which the hole appears and the death d_i is the ϵ at which it disappears. Each interval corresponds to a geometric "feature" of X that causes a hole to form, and whose size and shape are reflected in the birth and death times. Intuitively, this can be seen as a "void" or a "gulf" largely surrounded by X as shown in Figure 4.1.

This geometric intuition can be extended to higher dimensions, with i-dimensional PH classes corresponding to i-dimensional features of the object. For example, consider the open cell-foam $Y \subset \mathbb{R}^3$ in Figure 4.2. In this case, we are interested in the second dimensional PH, PH_2. It tracks 3D holes that form and disappear as Y_ϵ is expanded; these holes correspond to 3D voids in Y (the "cells" in the open-cell foam). PH_1 of Y is a bit harder to visualize - each PH_1 interval corresponds to a feature of Y that can trap ropes with radii between the birth and death values, such as the gaps between voids.

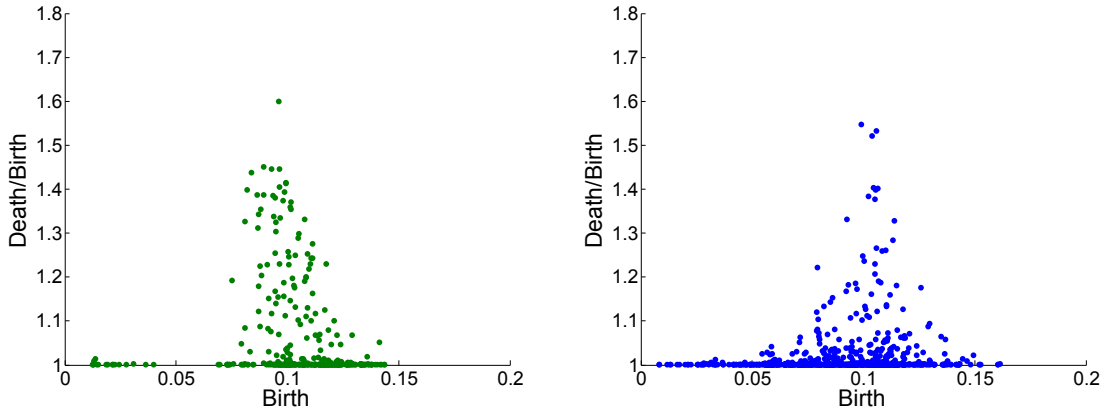
PH was originally developed by Edelsbrunner et al. [15], and was extended by many including Carlsson and Zomorodian [71]. In collaboration with MacPherson, the author introduced the philosophy described here - using PH to study the geometry of a fixed object, rather than the topology of noisy measurements - in the paper, "Measuring Shape with Topology" [40]. The main results

of that paper will be discussed in Chapter 5. In this chapter, we will introduce PH and several techniques developed by the author to extract additional geometric information from it. We will also apply them to the case study presented in Chapter 2.

The next section will introduce the Scatter Plot of Feature Points, which is the main way to display PH data. The intuitive explanation of Persistent Homology presented in the previous few paragraphs will be enough to understand that section, and most of the other data presented below. However, the next few sections will give a brief technical description of Persistent Homology, starting with Homology in Section 4.3, and then proceeding to Persistent Homology in Section 4.4. Section 4.3 also includes a discussion of Alexander duality, which underlies much of the intuition presented above. After Persistent Homology has been defined, we will briefly outline the methods used for computing persistence in Section 4.5, and then discuss the author's original work in applying Persistent Homology and developing new methods to analyze the data resulting from it in Sections 4.6 through 4.8. The proofs in Sections 4.4 and 4.5 are more technical than the rest of the thesis, and may be skipped. We should note that Persisent Homology can be defined more generally for filtrations of topological spaces (we'll define filtration later) and functions from manifolds to the real line, but here we restrict consideration to the $\epsilon-$expansion case that is important in these applications.

4.2 Scatter Plots

The Scatter Plot of Feature Points is one of the main ways to visualize and interpret Persistent Homology data. Recall from the previous section that the i-th dimensional Persistent Homology PH_i of a subset X of a metric space gives rise to a set of intervals I_i corresponding to i-dimensional geometric features of X. The start point of the interval is the birth of the feature - the first ϵ at which it causes an i-dimensional homology class to form in X_ϵ. Similarly, the end point of the interval is the death of the feature. Both the birth and the death scale with the size of X, and measure the size of geometric features of X. For example, the birth of a PH_1 class of an object in the plane corresponds to the radius of the smallest disk that is trapped by the corresponding



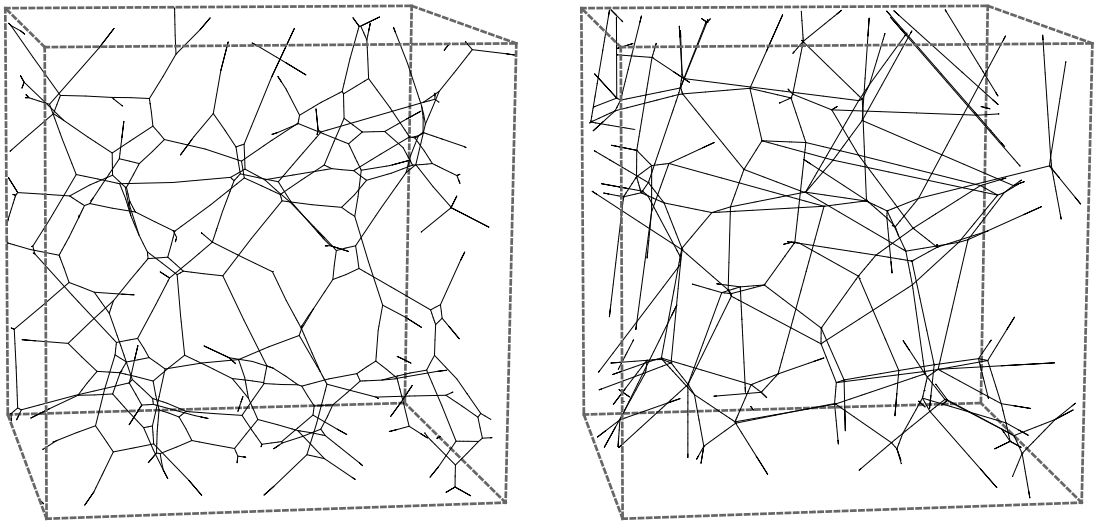
(a) (b)

Figure 4.3: PH_2 Scatter Plots for (a) the one-skeleton of a Voronoi tesselation of 100 Poisson distributed points in $\mathbb{T}^3$ and (b) an embedded graph that evolved by curvature flow until it had 1600 edges.

gulf, and the death corresponds to the radius of the largest disk that can fit inside the gulf. Some authors plot the birth versus the death in a scatter plot, but here we will different coordinates: $(\text{birth}, \text{death}/\text{birth})$. We prefer this choice because the ratio of death to birth is scale-invariant and measures something about the shape of features rather than the size. For example, in the case of PH_1 of a planar set, the ratio of the death to the birth measures how enclosed gulfs are.

4.2.1 Two Examples

PH_2 Scatter Plots for the one-skeleton of a Voronoi tessellation of 100 Poisson distributed points in $\mathbb{T}^3$ and an embedded graph with 1600 edges resulting from curvature flow are shown in Figure 4.3 (we will refer to the latter as a "CDE configuration" for the rest of the chapter, short for "curvature-driven evolution configuration"). The configurations were chosen to have comparable size, and a small region of each is shown in Figure 4.4. The two distributions differ in shape, but both have a peak centered near birth = .1 of similar height. This indicates that the two configurations have voids of similar size. This is surprising; it is expected that the Voronoi graph will have many large voids corresponding to the interiors of Voronoi cells, but there is no reason to suspect *a priori* that the other configuration will have similarly sized voids. The equations of motion that resulted in this



(a) (b)

Figure 4.4: Small regions in (a) a CDE configuration and (b) a Voronoi graph.

state are local and contain no explicit volume exclusion. We will develop a method for visualizing these unexpected voids in Section 4.6. It will also turn out that while the two configurations have voids on similar size scales, the relationships between neighboring voids in the two systems are quite different. This topic will be discussed in Section 4.7. An additional application will be presented in Chapter 5, where we will use the statistical distribution of feature points to study three different probability distributions on polymers.

4.3 Homology

4.3.1 A Brief Introduction to Homology

To define and understand Persistent Homology, we must first introduce homology. Homology was introduced by Poincaré in 1895 [60]. The homology groups $H_i(X)$ are vector spaces indexed by a non-negative integer i that detect i-dimensional topological features of X[1]. For example, the rank of $H_0(X)$ is the number of connected components of X. $H_1(X)$ measures the number of independent

[1]What we use here is homology with coefficients in the field $\mathbb{Z}_2$. Homology may be defined over a general field or group, though it will not always be a vector space in the latter case.

loops in X. H_1 of the circle has rank one, indicating that all of the different ways to draw a loop around the circle are generated by a single element (that is, the only way to draw loops around the circle are to go around it n times). On the other hand, H_1 of the two torus $\mathbb{T}^2$ has rank two; the meridian and the equator give independent ways of drawing loops around it. $H_1(S^2) = 0$, because there is no way to draw a loop on the 2-sphere that cannot be collapsed to a point. H_2 of a subset X of $\mathbb{R}^3$ counts the number of independent voids in X. For example, if $X = S^2$, then $H_2(S^2)$ has rank one; the 2-sphere has one hole. Similarly, the rank of $H_2(\mathbb{T}^2)$ is one, but $H_2(S^1) = 0$ because S^1 is one-dimensional and has no three-dimensional holes.

In general, elements of $H_i(X)$ correspond to formal sums of topological features of X. However, in many cases it is possible to choose a basis for this vector space such that each basis vector corresponds to a single topological feature of X. For example, a basis of $H_0(X)$ is given by a vertex for each connected component of X. It is important to note that this basis, and the correspondence between topological features and basis elements, may not be unique for higher i.

4.3.2 Regular Cell Complexes

Homology can be defined for general topological spaces, but here we will introduce it for the special case of a regular cell complex. As introduced in Section 3.1, a regular cell complex is a space built inductively by attaching cells in each dimension, where an n-dimensional cell is homeomorphic to an n-dimensional ball. That is, a 0-cell is a point, a 1-cell is a line segment, a 2-cell is a disk, and so on. The n-dimensional cells are attached via homeomorphisms from their boundaries (which are homeomorphic to $(n-1)$-dimensional spheres) into the structure built from the lower dimensional cells. In Figure 4.5 regular cell complex decompositions are shown for a triangle and a 2-torus. Most reasonable topological spaces can be represented as a regular cell complex, though there are examples of otherwise nice four-dimensional spaces which are homeomorphic to no regular cell complex. Homology works for them, too, but the development of the theory is slightly more complicated.

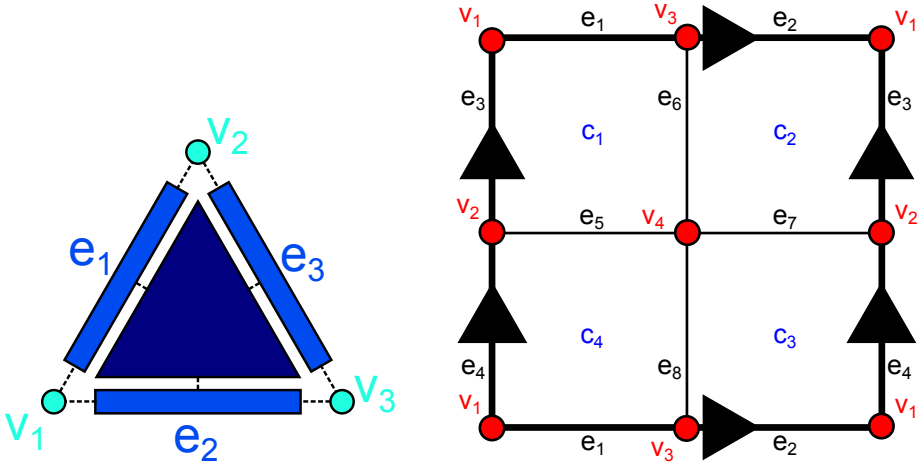


Figure 4.5: Regular cell complex decompositions for (a) a triangle and (b) $\mathbb{T}^2$. The triangle has one two-cell, three one-cells, and three zero-cells. Opposite edges on the boundary of the two-torus are identified in the directions that the arrows are pointing.

4.3.3 Definition of Homology

Suppose X is a regular cell complex. The chain vector space $C_n(X)$ is the vector space on the n-cells of X over the field $\mathbb{Z}_2$. That is, if σ^n is an n-cell of X, then $\sigma_i^n + \sigma_i^n = 0$. The boundary homomorphism $\partial_n : C_n(X) \to C_{n-1}(X)$ is a linear transformation which sends an n-cell to the sum of the $(n-1)$-cells adjacent to it, and extends linearly to all of $C_n(X)$. For example, the boundary of the two-cell depicted in Figure 4.5a is the sum of its three edges, and the boundary of each of the edges is the sum of the two vertices at its endpoints. The kernel of ∂_n is called the vector space of n-cycles of X, $\mathbb{Z}_n(X)$, and the image of ∂_n is the vector space of $(n-1)$-boundaries $B_{n-1}(X)$. A simple but fundamental computation shows that $\partial_{n-1} \circ \partial_n = 0$; that is, boundaries are cycles, meaning $B_n(X) \subset Z_n(X)$. We define the n-dimensional homology to be $H_n(X) = Z_n(X)/B_n(X)$. This vector space is usually referred to as the n-dimensional cellular homology group with coefficients in $\mathbb{Z}_2$. Homology may be defined for general topological spaces over any group, but this will suffice for the applications presented here[2].

[2]If the coefficient group is not a field, the homology will be a group but not a vector space.

4.3.4 Examples

Before proceeding, let's compute a couple of examples. For the triangle in Figure 4.5a, $C_0(T)$ is generated by the basis $\{v_1, v_2, v_3\}$, $C_1(T)$ is generated by $\{e_1, e_2, e_3\}$ and $C_2(T)$ is generated by the sole 2-cell t. The boundary of t is $\partial_2(t) = e_1 + e_2 + e_3$, so in matrix form the homomorphism $\partial_2 : C_2(T) \to C_1(T)$ is $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$, so $Z_2(T) = 0$ (this map only sends zero to zero) and therefore $H_2(T)$ is the trivial group 0. Also, $B_1(T) \cong \mathbb{Z}_2$ and is generated by the sum of the three basis elements. The boundaries of each of the edges are the sum of the adjacent vertices, so the matrix of ∂_1 is

$$\partial_1 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

It follows that $Z_1(T) \cong \mathbb{Z}_2$ and is generated by the sum of the three edges, $e_1 + e_2 + e_3$ (this is also easily seen from the fact that a one-cycle is necessarily a closed path, or a sum of them). This same element is the generator of $B_1(T)$, so $H_1(T) = \mathbb{Z}_2/\mathbb{Z}_2 = 0$ is trivial. Finally, the boundary of a 0-cell is zero, so $C_0(T) \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. Quotienting out by the image of ∂_1 identifies each of the vertices with the others, so $H_0(T)$ is $\mathbb{Z}_2$. In summary, T only has nontrivial 0-dimensional homology. This makes sense, as T can be squished to a point, so it only has one 0-dimensional topological feature and no higher dimensional topology. It turns out that any space that can be compressed to a point (called a contractible space) has these same homology groups, $H_0 \cong \mathbb{Z}_2$ and $H_i = 0$ for all $i > 0$.

Let us proceed to a more interesting example, the 2-torus in Figure 4.5b, $\mathbb{T}^2$. With chain bases ordered by the indices in the figure, a computation yields:

$$\partial_2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \qquad \partial_1 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

We can compute $H_0(\mathbb{T}^2)$ easily by noting that quotienting out by the image of ∂_1 simply identifies all of the vertices since every row has exactly two ones. Thus $H_0(\mathbb{T}^2) \cong \mathbb{Z}_2$, confirming that the 2-torus has only one component. Gauss-Jordan elimination shows that the dimension of the kernel of ∂_2 is one, so $H_2(\mathbb{T}^2)$ is $\mathbb{Z}_2$ - the 2-torus has one three-dimensional hole. The computation of $H_1(\mathbb{T}^2)$ is a bit more complicated - the standard algorithm is to find bases for the $C_1(\mathbb{T}^2)$ and $C_2(\mathbb{T}_2)$ that make the answer obvious [71]. We won't do the computation here, but the answer is $H_1(\mathbb{T}^2) = \mathbb{Z}_2 \times \mathbb{Z}_2$. There are other ways to define homology so that this computation is easier to do by hand [22], and it is also very easy for a computer.

4.3.5 Some Important Theory

Now, we will state some important theoretical results. First, homology groups do not depend on the choice of regular cell complex decomposition. That is, if Y is a topological space that is homeomorphic to the regular cell complexes X_1 and X_2, the homology groups computed from X_1 and X_2 are the same. In fact, a much stronger statement is true: homology groups are a homotopy invariant of the space Y. Two continuous maps $f : Y \to Y$ and $g : Y \to Y$ are *homotopic* if there is a continuous map $H : Y \times [0, 1] \to Y$ such that $H(y, 0) = f(y)$ and $H(y, 1) = g(y)$. That means that f can be "deformed continuously" into g. Two spaces X and Y are *homotopy equivalent* and are said to have the same *homotopy type* if there exist continuous functions $h : X \to Y$ and $k : Y \to X$ such that the compositions $k \circ h$ and $h \circ k$ are homotopic to the identity maps of X and

Y, respectively. The statement that homology is a homotopy invariant means that if X and Y are homotopy equivalent, then they have isomorphic homology groups.

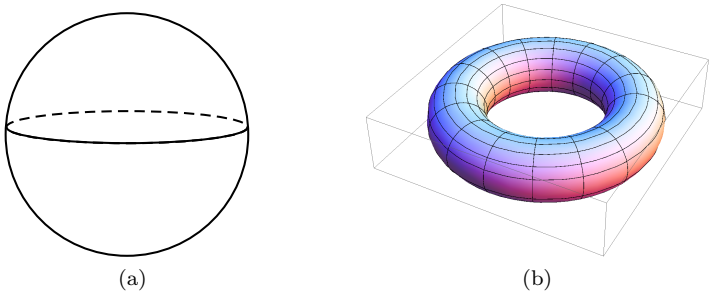
Homology is also functorial. By that, we mean that if there are two topological spaces X and Y and a function $f : X \to Y$, then f induces a homomorphism $f_* : H_i(X) \to H_i(Y)$ for each i. Furthermore these induced homomorphisms behave well with respect to composition: if Z is another topological space and there is a function $g : Y \to Z$ then $(g \circ f)_* = g_* \circ f_*$. Finally, the identity map induces the identity homomorphism. The most important induced homomorphisms in the remainder of this chapter will be those induced by inclusion maps $X \hookrightarrow Y$ where $X \subset Y$. In this case, we will just say that the induced homomorphism is the inclusion homomorphism. As an example, consider the ϵ-expansion of the arc of a circle in Figure 4.1. As ϵ increases, each neighborhood includes into the next. Let the neighborhoods be C^1 through C^5, and the inclusion maps be i^1 through i^4. As each C_j is connected, the induced maps on H_0, $i^j_* : H_0(C^j) \to H_0(C^{j+1})$ are isomorphisms. The maps on H_1 are somewhat more interesting. i^1 and i^2 are the trivial map, because $H_1(C^1) = H_1(C^2) = 0$. $i^3 : H_1(C^3) \to H_1(C^4)$ is an isomorphism sending $\mathbb{Z}_2$ to $\mathbb{Z}_2$. Finally, $i^4 : H_1(C^4) \to H_1(C^5)$ is the zero map, since $H_1(C^5) = 0$. Persistent Homology is a way to understand the structure of these inclusion maps in more complex cases.

4.3.6 Alexander Duality

Alexander Duality is a beautiful property of homology that provides theoretical backing for most of the intuition for Persistent Homology. It states that if X is a subset[3] of S^n, then there is a strong relationship between the homology of X and the homology of the complement of X, denoted $S^n - X : H_i(X) \cong H_{n-i-1}(S^n - X)$ if $i \neq 0$ and $i \neq n-1$, and the ranks of $H_0(X)$ and $H_0(S^n - X)$ are one less than the ranks of $H_{n-1}(S^n - X)$ and $H_{n-1}(X)$, respectively.[4] This is not just an algebraic equivalence; Alexander duality gives a geometric correspondence between i-dimensional topological features of X and $(n - i - 1)$-dimensional topological features of its complement. For

[3]Subject to some conditions that are satisfied by all of the physical examples considered later in the paper. Namely, X is required to be compact and locally contractible.

[4]Convention holds that $H_{-i}(X)$ is always zero. If the reader is familiar with reduced homology, then the duality is more elegantly stated as $\tilde{H}_i(X) = \tilde{H}_{n-i-1}(S^n - X)$. Reduced homology is defined by modding out by a base component; it decreases the rank of H_0 by one, and does not change the other homologies.



(a) (b)

Figure 4.6: (a) A circle embedded in S^2 and (b) a torus embedded in $\mathbb{R}^3$. The circle in (a) can be seen as a subset of $\mathbb{R}^2$ by removing any other point of S^2.

example, homology classes in $H_n(X)$ correspond to (sums of) components of the complement of $S^n - X$, and vice versa. This will be seen in some of the examples below.

4.3.7 Examples of Alexander Duality

As a first example, consider $S^1 \subset S^2$, in Figure 4.6a. The complement of S^1 is two disks (the hemispheres separated by the equator), so the rank of $H_0(S^2 - S^1)$ is two. The rank of $H_1(S^1)$ is one less than this, and it is equal to $\mathbb{Z}_2$. Similarly, the rank of $H_1(S^2 - S^1)$ is zero (each disk is contractible), and the rank of $H_0(S^1)$ is one greater than this and so equals $\mathbb{Z}_2$.

Now, consider the 2-torus $\mathbb{T}^2 \subset S^3$ in Figure 4.6b. The complement of $\mathbb{T}^2$ inside of S^3 is equal to two disjoint solid donuts. By Alexander Duality, $\text{rank}(H_0(\mathbb{T}^2)) = 1 + \text{rank}(H_2(S^3 - \mathbb{T}^2)) = 1 + 0 = 1$ and $\text{rank}(H_2(\mathbb{T}^2)) = \text{rank}(H_0(S^3 - \mathbb{T}^2)) - 1 = 2 - 1 = 1$. Now, Alexander Duality also holds that $H_1(\mathbb{T}^2) \cong H_1(S^3 - \mathbb{T}^2)$. Each solid donut in the complement of $\mathbb{T}^2$ is homotopy equivalent to a circle, and the homology of a disjoint union is equal to the sum of the homologies, so Alexander Duality tells us$H_1(\mathbb{T}^2) = H_1(S^3 - T^2) \cong H_1(S^1) \times H_1(S_1) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$. This is consistent with the answer stated in Section 4.3.4. This correspondence can be seen geometrically: the H_1 generator going around the equator corresponds to the inner doughnut which wraps around in the same way, and the H_1 generator going around the meridian matches the outer doughnut.

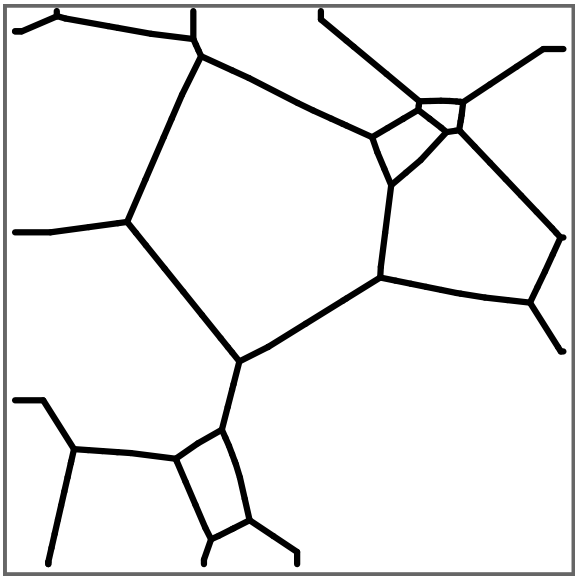


Figure 4.7: An embedded graph in $\mathbb{T}^2$. Its H_1 has rank 11 and its complement has 10 components. There is a basis of H_1 such that all members of the basis that do not wrap around the torus correspond to components of the complement.

4.3.8 Alexander Duality for Subsets of the Torus

A version of Alexander Duality holds for subsets of general n-dimensional manifolds, as well, but the algebra is more complicated. If $Y \subset M^n$, then $H_i(Y) \cong \check{H}^{n-i-1}(M^n - Y)$, where $\check{H}^{n-i-1}(M^n - Y)$ is a group called the "compactly supported cohomology" of $M^n - Y$. The definition of this group is beyond the scope of this thesis (refer to [22] or [51] for a full treatment), but we will explore what it means in the special case of $\mathbb{T}^2$. Let Y be the embedded graph in Figure 4.7, where the boundaries of the square are periodic so the graph is in fact embedded in $\mathbb{T}^2$ (notice that the edges that go out one side come back on the other). Computation shows that $H_1(Y)$ has rank 11, but the number of components of its complement is just 10, not the 12 we would expect for a subset of S^2. The reason is that $H_1(\mathbb{T}^2) \cong \mathbb{Z}_2 \times \mathbb{Z}_2$, and Y contains both of these "global" homology classes. If we restrict our attention to local subsets, the correspondence is as expected. That is, if a topological feature of Y is contained in a subset of $\mathbb{T}^2$ homeomorphic to a disk, then any topological features of it will correspond to a topological feature of the complement. In the applications discussed later in this

chapter, the global topological properties are just artifacts of the periodic boundary conditions, and we are only interested in the local topological features for which Alexander duality behaves as it does for a subset of the sphere or Euclidean space.

4.4 Defining Persistent Homology

The intuitive definition of Persistent Homology given in the introduction can be formalized as follows. As before, let X be a subset of $\mathbb{R}^n$ or a more general metric space M. Consider the product $\prod_{\mathbb{R}\geq 0} H_i(X_\epsilon)$, together with the maps $i_{\epsilon_1,\epsilon_2} : H_i(X_{\epsilon_1}) \hookrightarrow H_i(X_{\epsilon_2})$ induced by inclusion for all $0 \leq \epsilon_1 \leq \epsilon_2$. This is called the Persistenct Homology $PH_i(X)$, and at first glance it looks very complicated. However, under the right hypotheses, the structure of this object is easy to understand and provides much useful information on the geometry of X. This was first seen by Carlsson and Zomorodian in [71], and the next paragraph is based on their work.

Theorem 2. *Suppose X is a subset of a metric space and the rank of $H_i(X_\epsilon)$ is finite for all ϵ. Then, the structure of $PH_i(X)$ can be represented by a set of intervals with endpoints in the non-negative real numbers, each representing a homology class that is born at the beginning of the interval and dies at its end. This result can be more formally stated in terms of Equation 4.1 below.*

Sketch of Proof. To show this, we will first construct an auxiliary space A as a subspace of $X \times \mathbb{R}_{\geq 0}$. Suppose $p : X \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is projection onto the second factor, and define A to be such that $p^{-1}(\epsilon) = X_\epsilon$. The condition that $H_i(X_\epsilon)$ has finite rank for all ϵ implies that that $\mathbb{R}_{\geq 0}$ breaks up into a finite set of intervals (possibly single points) $S_0, S_1, \ldots, S_k, \ldots, S_n$ such that $H_i(p^{-1}(x))$ is constant in each S_k. Thus, the structure of $PH_i(x)$ is captured by the simpler product $\widehat{PH}_i(X) = \prod_{k=0}^{n} H_i(S_k)$ together with the appropriate inclusion maps:

$$H_i(S_0) \to H_i(S_1) \to \ldots \to H_i(S_n) = H_i(M).$$

This product has the structure of a graded $\mathbb{R}[t]$ module, where the grading is given by k, and

t acts by sending a homology class in $H_i(S_{k-1})$ to its image in $H_i(S_k)$ (except when $k = n$, where it acts trivially). As each $H_i(S_i)$ is finitely generated, the product is as well, and the structure theorem for finitely generated modules over a principal ideal domain applies [5]. This implies that

$$\widehat{PH}_i(X) = \bigoplus_{j=0}^{a} \mathbb{Z}_2[t]^{w_j} \oplus \bigoplus_{j=0}^{b} \mathbb{Z}_2[t]^{v_j} / t^{n_j} \tag{4.1}$$

where the exponents w_j and v_j give the grading, that is the first k such that the generating class appears in $H_i(S_k)$. The decomposition in this canonical form is unique, and is interpreted as follows. Each term in the decomposition gives an interval, with birth time given by the grading, w_j or v_j (that is, the birth is equal to the leftmost point in the interval S_{w_j} or S_{v_j}). If the term is isomorphic to $\mathbb{R}[T]$, the interval extends to ∞, and the corresponding persistent homology class never dies. However, if the term is of the form $\mathbb{R}[t]/t^{n_j}$, applying t n_j times to the generator t kills it, so the death time is equal to the rightmost point in the interval S_{v_j}.

Before proceeding, we shall note that Persistent Homology and its structure theorem are perhaps more naturally understood in the context of cosheaf theory rather than module theory. A cosheaf $\mathscr{F}$ over the real line is an object that assigns to every interval I a group $\mathscr{F}(I)$, subject to certain consistency requirements. Persistent Homology may be viewed as a cosheaf $\mathscr{PH}$ that assigns to each interval I the group $\mathscr{PH}(I) = H_i(p^{-1}(I))$. A structure theorem for constructible cosheaves yields the same classification as above. A full treatment of (co)sheaf theory is outside of the scope of this thesis, and interested readers should refer to Justin Curry's PhD thesis for a treatment of these topics. [12]

4.5 Computation

4.5.1 Filtrations

The computation of Persistent Homology starts with an object called a filtration of simplicial complexes, which matches or approximates the topology of X and its ϵ-neighborhoods. A simplicial complex is a special type of regular cell complex, where an n-dimensional cell is an n-simplex. An

n-simplex is the n-dimensional convex hull of $n+1$ vertices. A 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, and so on. Any regular cell complex can be subdivided to create a simplicial complex. A filtration is a sequence of simplicial complexes, each mapping into the next:

$$K_0 \to K_1 \to K_2 \to \ldots \to K_j \to \ldots \to K_n.$$

The Persistent Homology of a filtration is then the product $\prod_{j=0}^{n} H_i(K_j)$ together with the usual map[5]. It is easily computed using a standard matrix algebra algorithm, the technical details of which are presented in [71]. Many programs implement this algorithm, including JavaPlex [25], Perseus [55], and Dionysus [53]. JavaPlex and its predecessor jPlex [65] were used for the computations in this thesis.

4.5.2 The Alpha Complex

The Alpha Complex, originally discovered by Edelsbrunner and Mücke [16], is probably the best tool for computing Persistent Homology of subsets of $\mathbb{R}^2, \mathbb{R}^3$, and flat tori of corresponding dimensions. Given a set of points[6] S in $\mathbb{R}^n$ or $\mathbb{T}^n$, the Alpha Complex of radius r is a simplicial complex homotopy equivalent to the union of balls of radius r at those points (recall from Section 4.3 that homotopy equivalent spaces have the same homology groups). It is a subset of the Delaunay triangulation on S, which is the unique triangulation with vertex set equal to S all of whose top-dimensional simplices have empty circumspheres[7]. That is, for each n-simplex in the Delaunay triangulation, the unique sphere containing the $n+1$ vertices of the simplex has no other points of S in its interior. The Delaunay triangulation is equivalently defined as the dual graph of the Voronoi tessellation introduced in Section 2.6.4. The Delaunay triangulation is very well studied, and there are many implementations of several different algorithms to compute it. The author used the Computational

[5]This definition can be used to generalize Persistent Homology; the maps in a filtration can be general continuous maps, not just inclusions as they are in the examples presented here.

[6]perhaps approximating a more complicated space, as described in Section 4.5.4

[7]This definition only specifies a unique triangulation if the points are in general position. That is, no more than $n+1$ points are contained in any n-plane.

Geometry Algorithms Library (CGAL), which has implementations of the Delaunay triangulation for both Euclidean space and the flat torus in two and three dimensions [1].

The Alpha Complex defines a filtration of simplicial complexes by assigning an alpha value to each simplex of the Delaunay triangulation that specifies when it enters the complex. The alpha value of a simplex is equal to the radius of the smallest sphere that contains the vertices of the simplex on its surface, but no other points of S in its interior. For a top-dimensional simplex, this is simply the radius of the circumsphere, by definition of the Delaunay triangulation. However, the smallest sphere containing the vertices of a lower-dimensional simplex does not always have an empty interior. In this case, the alpha value turns out to be the smallest alpha value of the higher dimensional simplices containing it. The alpha value of a 0-simplex is always zero. CGAL [1] includes a program to compute the Alpha Complex, and the author also programmed his own software for that purpose.

4.5.3 Proof of the Correctness of the Alpha Complex

To prove that the Alpha Complex on a set of points S has the same Persistent Homology as S, we require a fundamental result of combinatorial topology called the Nerve Lemma. If a subspace of a metric space Y (or, more generally, a subspace of a paracompact topological space) is covered by the open sets $\{U_i\}$, the nerve of the cover $\{U_i\}$ is an abstract simplicial complex with $\{U_i\}$ as its vertex set. A finite subset of $\{U_i\}$ forms a simplex in the nerve if and only if the intersection of all of its members is nonempty. The Nerve Lemma states that if $\{U_i\}$ is a cover such for every subset of its members whose intersection is nonempty, that intersection is contractible (homotopy equivalent to a point), then the nerve of $\{U_i\}$ is homotopy equivalent to S. In particular, it has the same homology. For a proof of the Nerve Lemma, refer to chapter 4G of [22].

We will now exhibit a cover of S_ϵ, the ϵ-neighborhood of S, that satisfies the hypotheses of the Nerve Lemma. S_ϵ is a union of balls centered at the points of S, $\{p_i\}$. Let V_i be the intersection of the ball of radius ϵ centered at p_i with the Voronoi cell corresponding to that point. Since both Voronoi cells and r-balls are convex, all of the elements of $\{V_i\}$ are convex, as are any nonempty intersections of its members. Convex sets are contractible, so the cover $\{V_i\}$ satisfies the hypotheses

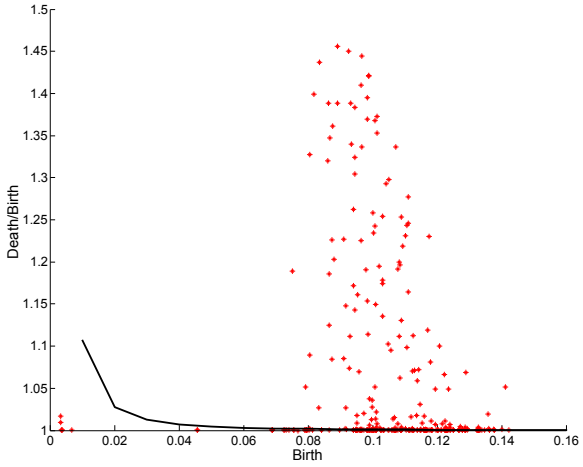


Figure 4.8: A scatter plot for the PH_2 of a Voronoi tesselation of 100 Poisson distributed points on the 3-torus. All of the noise due to the approximation of the skeleton by a point cloud is below the black curve.

of the Nerve Lemma and its nerve has the same homology as S_ϵ. Next, we shall show that the Alpha Complex gives the same simplicial complex as the nerve of $\{V_i\}$.

The Alpha Complex at radius ϵ has a simplex with vertices $\{q_j\}$ if and only if there is a sphere of radius less than or equal to ϵ that contains $\{q_j\}$ on its surface, and no other points of S in its interior. This is equivalent to the existence of a point $\tilde{q}$ that is (1) within distance ϵ of each of the q_j and (2) and is equally distant from each of the q_j and closer to them than to any other point of S. Property (1) implies that $\tilde{q}$ is contained in the ball of radius ϵ at each of the q_j, and property (2) implies that $\tilde{q}$ is contained in the Voronoi cells of each of the q_j. Therefore, the open sets in $\{V_i\}$ corresponding to $\{q_j\}$ intersect nontrivially and there is a simplex in the nerve of $\{V_i\}$ with vertices $\{q_j\}$. Thus we see that every simplex in the Alpha Complex corresponds to a simplex in the nerve of $\{V_i\}$, and the converse follows from the same reasoning. Therefore, the Alpha Complex at radius ϵ gives the same simplicial complex as the nerve of $\{V_i\}$, and has the same homology as S_ϵ. Since this is true for all ϵ, the Persistent Homology of the Alpha Complex filtration is the same as that of S, as claimed.

4.5.4 Method of Computation

To compute the Persistent Homology of a subset X of $\mathbb{R}^2$ or $\mathbb{R}^3$, approximate it by a point cloud and compute the Alpha Complex. The method for generating the point cloud depends on the application at hand. For the embedded graphs, we interpolate a set of points along the edges. In some cases, it may be necessary to perturb the points slightly to be in general position.

Since the Alpha Complex has the same Persistent Homology as the neighborhoods of the point cloud, the accuracy of the measurement depends on the choice of points. Let P be the point cloud chosen to approximate X, and let d be the maximum radius of a ball contained within $X-P$. Suppose a j-dimensional persistence class β of P is born at time b, and β not correspond to any persistence class of X. That means that the image of β is zero under the inclusion $i : H_j(P_b) \to H_j(X_b)$. However, since X_b includes into P_{b+d} via a map l, this implies that the image of β under the inclusion map $k : H_j(P_b) \to H_j(P_{b+d}) = l \circ i$ is zero, so its death time is less than $b + d$. Therefore, in the scatter plot of the Persistent Homology of X, all classes due to approximation error are below the line death = birth $+ d$. For example, in the Scatter Plot for the one-skeleton of a Voronoi tesselation of 100 Poisson distributed points in $\mathbb{T}^3$ shown in Figure 4.8, all of the noise is below the black curve. However, the existence of signal below the curve depends on the geometry of X and cannot be ruled out no matter how small the mesh.

The Alpha Complex produces accurate results, and is efficient in two and three dimensions, the most important dimensions for physical applications. In the worst case, it has $O(n^{\lceil \frac{d}{2} \rceil})$ simplices, where n is the number of points and d is dimension, and in practice the number is often quite a bit smaller than that [48, 58]. In the computations presented here, neither the Alpha Complex nor the Persistence algorithm were bottlenecks for speed or memory. Other filtrations have been used by other authors for different applications in higher dimensions, such as the Vietoris-Rips complex and Witness complex. They can be much faster, but there is a substantial tradeoff in accuracy.

4.6 Minimal Cycles

The Scatter Plot of Feature Points contains much useful information about the geometry of a subspace X of a metric space. However, the points themselves do not directly correspond to specific geometric structures. In order to further study the geometry of X, we will develop methods to extract more geometric information from persistence. This section will introduce a method for finding minimal cycle representatives for PH_i classes.

As before, we begin with a point cloud $\{p_i\}$ approximating X. The discussion here will assume that that it is possible to perturb $\{p_i\}$ slightly so each PH_i birth time is unique. This makes the problem of computing minimal PH cycles much more computationally tractable than that of computing a minimal homology basis of a space whose i-dimensional homology has rank greater than one. Instead of having to find a minimal basis, one simply needs to find a minimal cycle at each step.

First, slightly perturb the points if necessary so all of the i-dimensional simplices have unique alpha values. This way, each PH_i class can be uniquely matched with an i-simplex whose alpha value equals the birth time, and whose entry into the complex causes the PH_i class to form. Sort the PH classes by birth value, with the lowest birth value first. Iterate through this list, keeping a list of current PH_i classes, adding new ones when they are discovered and removing old ones after their death. Also, keep a list of boundaries so that homology may be computed. At each step in the list, find the unique i-simplex corresponding to the class. Now, the task is to find an i-chain containing that simplex that is homologically independent from the current class list, and is minimal in a certain sense. Of course, a definition of "minimal" is required.

4.6.1 What makes a cycle minimal?

What is a minimal cycle containing a given simplex? For $i = 0$, the choice is relatively obvious: the whole component containing the point in question[8]. Alexander Duality makes the choice easy for $i = n - 1$, as well. In that case, there are two components to either side of the unique $(n-1)$-

[8] A single point in the component could represent this homology class. This is more "minimal" in a certain sense, but also more "complex" as it requires an arbitrary choice point.

dimensional simplex corresponding to the PH class. Compute the boundaries of the components containing each of these cells. If neither is homologically independent from the current list of PH classes, then the current class is a "global" one corresponding to the topology of the ambient space. If both are independent, then the boundary of the union of the two components is an already-detected PH class. The algebraic machinery of persistent homology chooses one of these two components to continue representing the union, and the other to represent the new class. This ambiguity is unsatisfying, and will be dealt with in the next section. If i does not equal $n-1$ or 0, there are many more reasonable ways to define a minimal cycle. For instance, one could choose the new cycle with the fewest simplices, or the one whose simplices have the lowest volume. The choice of definition will ultimately depend on the application at hand. However, note that most reasonable definitions will not in general uniquely specify a minimal cycle.

4.6.2 An example

As an application, consider again the CDE configuration shown in Figure 4.4a. In Section 4.2.1, we saw that this structure has surprisingly many relatively large PH_2 classes, corresponding to voids in the structure. Using the method described in the previous paragraph, minimal representatives were computed to visualize these voids. One such void is shown in Figure 4.9. The picture highlights all edges of the original embedded graph containing a vertex of a two-simplex in the minimal cycle corresponding to the void[9].

4.7 Persistence Trees

Let $X \subset \mathbb{R}^n$ and consider its $(n-1)$-dimensional Persistent Homology, $PH_{n-1}(X)$ (the $n-1$ subscript will be suppressed in the remainder of this section). With the definition of a minimal cycle, it becomes apparent that a PH class can be born in two different ways: it can form by splitting off from a previously existing PH cycle, or it can enclose volume previously in the unbounded

[9]Note that there are two sets of vertices here - the vertices of the original graph, and the vertices of the alpha complex which are the $\{p_i\}$. Here, we mean the latter.

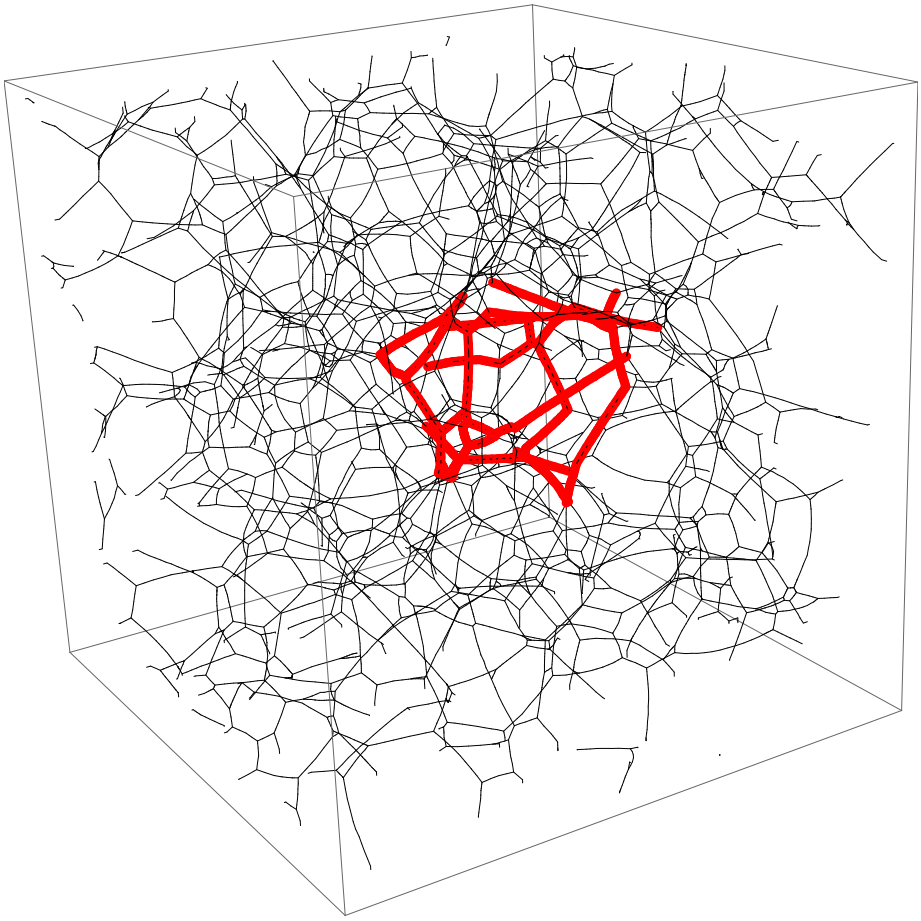


Figure 4.9: A minimal cycle representation of a PH_2 class of a CDE configuration, shown in red.

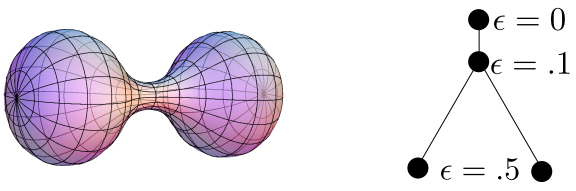


Figure 4.10: A surface with its PH_2 tree

component of the complement. For example, if the pinched sphere in Figure 4.10 is expanded, a persistent homology class will split. This motivates a preliminary definition of the PH tree.

4.7.1 Preliminary Definition of PH tree

The PH tree is composed of edges and nodes in the plane that capture geometric information about the voids of X. The $y-$coordinates of the edges and nodes contain the same information as the birth and death times of the PH intervals; the x-coordinates do not contain information about X and are used to display the tree nicely in the plane. Each PH birth corresponds to a node in the tree, whose $y-$coordinate equals the birth. If the birth corresponds to the formation of a cycle enclosing volume previously contained in the unbounded component of the complement, the node has no parents (no edges coming into it) and one child edge. The child edge starts with the same $y-$coordinate as the node, and ends when the corresponding PH class either splits or dies (whichever comes first). If a node corresponds to a splitting event, one parent edge comes in from above and two child edges split out below. Note the difference between this and the interval representation of PH: the classes involved in a splitting event are represented by two intervals, but three edges. The curvature flow simulations were performed on the three torus $\mathbb{T}^3$, for which there is no privileged unbounded component. As such, a different definition of the PH tree is required. It turns out that the correct way to do this is to build the tree using $H_0(\mathbb{T}^3 - X_\epsilon)$, not $H_2(X_\epsilon)$. To see this, we should first note that Alexander Duality does not imply that the second dimensional homology of a subset X_ϵ of$\mathbb{T}^3$ is isomorphic to the zero-dimensional homology of $X_\epsilon - \mathbb{T}^3$, as $H_2(\mathbb{T}^3) = \mathbb{Z}^3 \neq 0$. For sufficiently large ϵ, $H_2(X_\epsilon)$ will contain one or more of this generators, depending on how X_ϵ wraps around the torus. However, these generators contain information about the global topology of the ambient space, not the local geometry of X_ϵ. As we are generally interested in the local properties of X, we lose only uninteresting information by looking at $H_0(\mathbb{T}^3 - X_\epsilon)$ instead of $H_2(X_\epsilon)$.

4.7.2 Definition

The general definition of the PH tree is as follows. Let X be a subset of an $n-$dimensional manifold M, such as $\mathbb{R}^n$, S^n, or $\mathbb{T}^n$. For each component of $M - X$, the PH tree contains a node

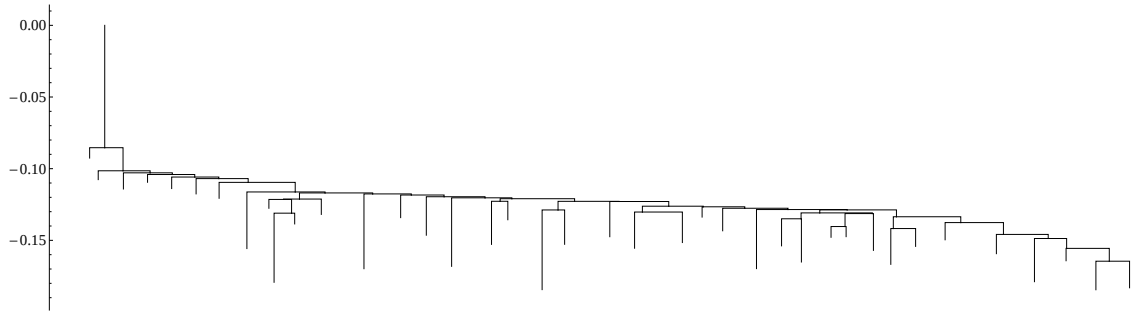


Figure 4.11: The PH tree of a CDE configuration with 800 edges.

with $y-$coordinate 0. Each of these nodes is adjacent to one edge, which ends at a node with $y-$coordinate equal to the time when the component first splits or dies. When a component splits, a new node is added with $y-$coordinate equal to the splitting time. Two child edges are adjacent to the new node, corresponding to the two split components. If $M = \mathbb{R}^n$, the resulting structure will be slightly different than the one defined above but the information will be equivalent. Specifically, assuming $H_n(X) = 0$, there will be an additional edge corresponding to the unbounded component, making the tree connected.

4.7.3 Computing the PH tree

To compute the PH tree in practice, a simple data structure is used: a component class with birth and death values, and pointers to a parent component and zero or two child components. The component class also includes a list of top dimensional cells that are contained in the component at its birth time. To begin the computation, suppose $X \subset \mathbb{R}^n$ or $X \subset \mathbb{T}^n$. As in Section 4.5, approximate X by a set of points and compute their Delaunay triangulation. Recall that the Alpha Complex is a filtration on the simplices of the Delaunay triangulation, where the alpha values of the simplices are determined by their circumradii, and the circumradii of their faces. To start, compute the components of the complement X and a list of the top-dimensional cells of each. Two lists of components will be kept in memory: the list of components at the beginning (stored to maintain access to the data structure) and a list of current components. Sort the list of $(n-1)$-dimensional simplices of the Delaunay triangulation by their alpha value, from low to high. Iterate through

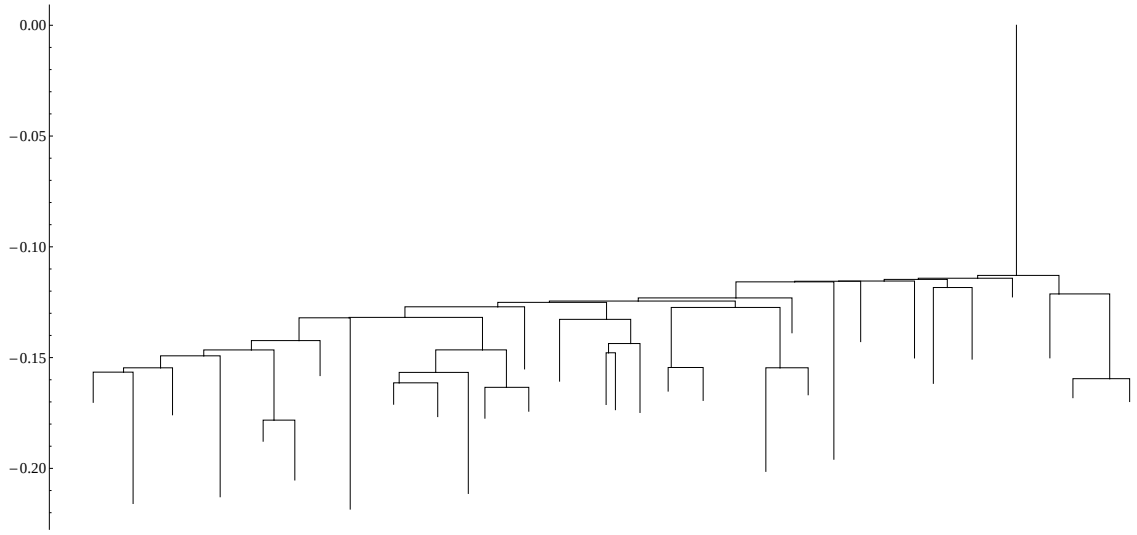


Figure 4.12: The PH tree of the one-skeleton of a Voronoi tesselation of 40 Poisson distributed random points.

this list; at each $(n-1)$-dimensional simplex, check to see if it separates the two $n-$simplices it is adjacent to into two different components. In that case, find the current component containing both of these simplices, and remove it from the current component list. Add two new components to the current component list, and set their parent pointer to the component just removed from the list. Their birth time and their parents' death time is equal to the alpha value of the current $(n-1)$-simplex. After finishing the list of $(n-1)$simplices, iterate through the current component list and assign death values to each equal to the maximum circumradius of its top dimensional simplices.

The PH tree is now easy to compute: create a node for each component born at 0, each splitting event, and each death event, with the y coordinates being 0, the splitting time, and the death time respectively. Connect these nodes with edges given by the parent and child relationships of the corresponding components. This process does not give x-coordinates to any edge or node, and they can be assigned in a way that facilitates visualization of the tree.

If the PH of X has already been computed, there is another way to compute the PH tree. As in the previous section, it is assumed that the points have been perturbed if necessary so that all

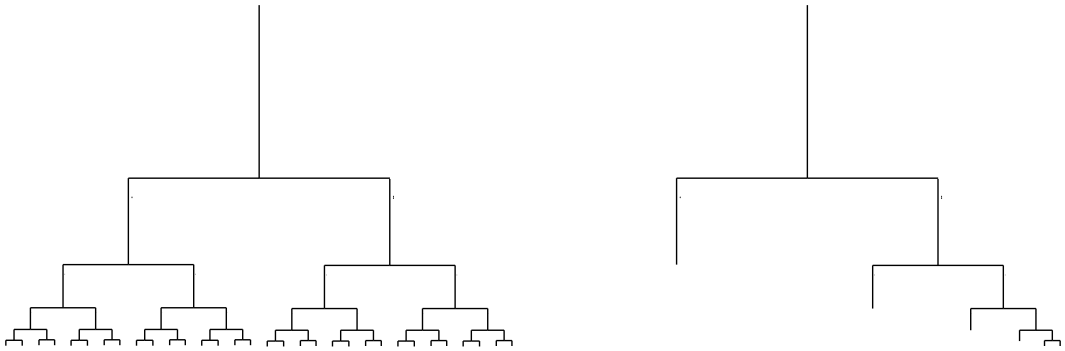


Figure 4.13: (a) A binary tree with maximal branching. (b) A binary tree with minimal branching.

of the alpha values of the $(n-1)$-simplices of the faces are unique. Thus every PH birth can be matched with a single $(n-1)$-simplex whose addition to the alpha complex creates the new PH class. Then, the same method as in the last paragraph is used except that only the $(n-1)$-simplices matched with PH intervals need to be considered.

4.7.4 Two Examples

PH trees for a CDE configuration and the one-skeleton of a Voronoi decomposition of the three torus are shown in Figures 4.11 and 4.12, respectively. The differences between the two are striking, especially with reference to the surprisingly similar scatter plots in Figures 4.3b and 4.3a. There is far more branching in the Voronoi tree than in the CDE configuration tree, reflecting that the voids in the Voronoi skeleton are related; many are adjacent to each other. To quantify this, let $g(T)$ equal the percentage of nodes in a tree T with grandchildren that have grandchildren through both of their children. $g(T)$ quantifies how branched a binary tree is, with the tree depicted in Figure 4.13b having $g(T) = 0$ and the one in Figure 4.13a having $g(T) = 1$ (note that the infinite analogues of these trees have the same values for g). The g values for Figures 4.11 and 4.12 are .233 and .363, and trees from larger configurations of those types give similar values.

4.7.5 Displaying PH trees

To display the tree, each edge is depicted as going out horizontally from its parent node until it reaches its x-coordinate, and then down vertically until it reaches the y coordinate of its child node(s). The x-coordinates are determined by giving the tree a potential, and modifying it iteratively to reduce the potential. Each edge is given a quadratic repulsive potential with the other (non-parent) edges in the tree. The coefficient of this potential is smaller if the edges are siblings. Finally, an exponential attractive potential is given between the x-coordinate of an edge's beginning and the x-coordinate of its end to prevent the tree from spreading out too much. The total potential of the tree is the sum of these potentials. The potentials and their coefficients were determined by trial-and-error to create the most informative embeddings of the PH trees depicted in the figures above.

4.7.6 Extensions

The PH_{n-1} tree is very useful for understanding the geometry of a subset X of an $n-$dimensional manifold. In the future, the author plans to quantify the properties of complex physical systems using statistics on PH trees. It should be noted that it is not possible to define a similar structure with nice properties for PH_i with i not equal to 0 or $n-1$. This is because there is no well-defined notion of splitting or merging for any definition of minimal cycle in these dimensions. For example, consider the set $X \subset \mathbb{R}^3$ depicted in Figure 4.14 and consider its PH_1. If a minimal PH_1 cycle is defined as one with shortest length, then as ϵ increases, the sum of two minimal cycles merges with another minimal cycle. This cannot be depicted by a tree. This is why groups are necessary to describe the homology of spaces, rather than simpler combinatorial notions.

A structure similar to the PH_{n-1} tree can be defined for general filtrations

$$\ldots \to X_0 \to X_1 \ldots \to X_i \to X_{i+1} \to \ldots$$

where all of the X_i are subsets of a fixed n-manifold and the maps are not necessarily inclusions. The process for constructing this object is the same, but it should be called a PH_{n-1} graph rather

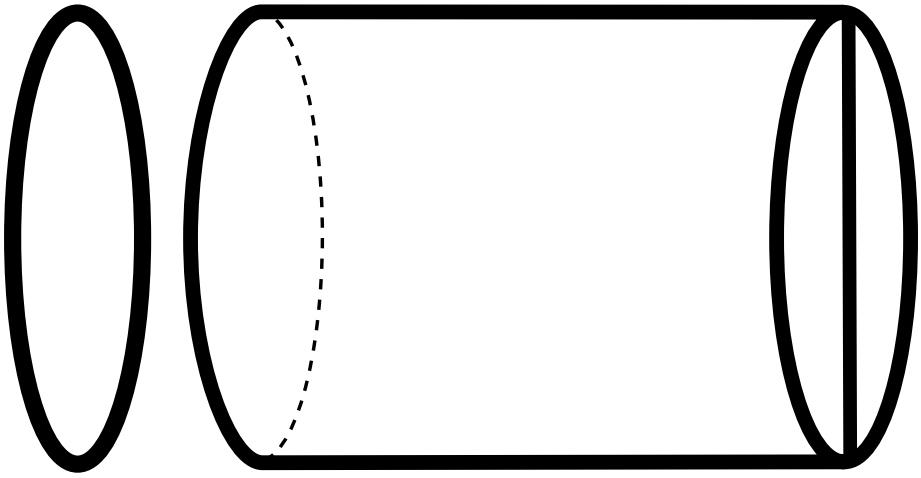


Figure 4.14: A cylinder adjacent to a circle. The end of the cylinder opposite to the circle has a cord drawn across its diameter. If this object were expanded, the homology class represented by the circle would eventually merge with the sum of the homology classes represented by the two loops on the opposite end of the cylinder.

than a PH_{n-1} tree because it may have cycles.

4.8 Recovering a 3D Cell Complex from its 1-Skeleton

Let C be a three-dimensional regular cell complex, and S its one-skeleton[10]. In some cases, it is possible to recover many of the cells and faces of C using the Persistent Homology of S. As motivation, consider the open-cell foam depicted in Figure 4.15. Open-cell foams are very useful materials, whose large surface areas make them excellent substrates for catalyzing chemical reactions. For example, the catalytic converter in our automobiles is a platinum foam that facilitates the removal of dangerous chemicals from exhaust. To model the effectiveness of an open-cell foam, one would like to construct a dual graph of the material and study the max-flow problem for this network. The existence of such a dual graph is not obvious in all cases; it is related to the knottedness of the foam (see Chapter 7). However, if the open-cell foam is the one-skeleton of a three-dimensional cell complex, the adjacancy graph of its cells would be the dual graph. Persistent homology can be used

[10]Note: we do not assume that the 3-cells of C are convex.

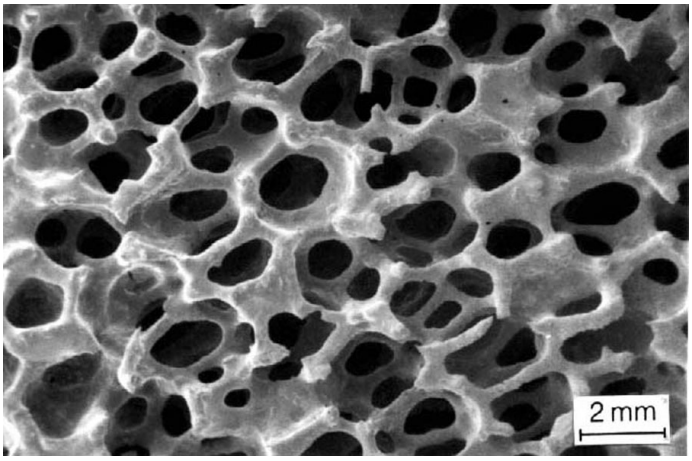


Figure 4.15: An open-cell foam [21]. One would like to recover a 3D cell complex from this material in order to study its properties.

to recover much of this information. In this section, we will explain how a 3D cell complex may be recovered from its one skeleton, and provide some computational results for the case of a Voronoi decomposition. The author intends to pursue the application to open-cell foams in the future.

4.8.1 Recovered Cells

To recover information about C from S, compute the Persistent Homology of S via the method explained in Section 4.5. That is, place points along each of the edges in the one-skeleton S at a fine mesh and compute the Alpha Complex of those points. Consider the resulting Persistence Tree. Let $\{c_i\}$ be the set of representative cycles for the persistence classes at the leaves of the tree at their birth times. Each c_i is a union of three-simplices, with vertices being points along the edges of the original graph S. We call c_i a recovered cell, and the edges containing vertices of the three-simplices of c_i are the recovered edges. For example, consider the one-skeleton of a Voronoi cell depicted in Figure 4.16. The black edges are the original edges of the one-skeleton, and the blue spheres are the points contained in the representative cycle of its persistence class. The recovered edges are precisely those black edges with blue spheres along them. Note that not all edges are recovered, and some faces have no recovered edges.

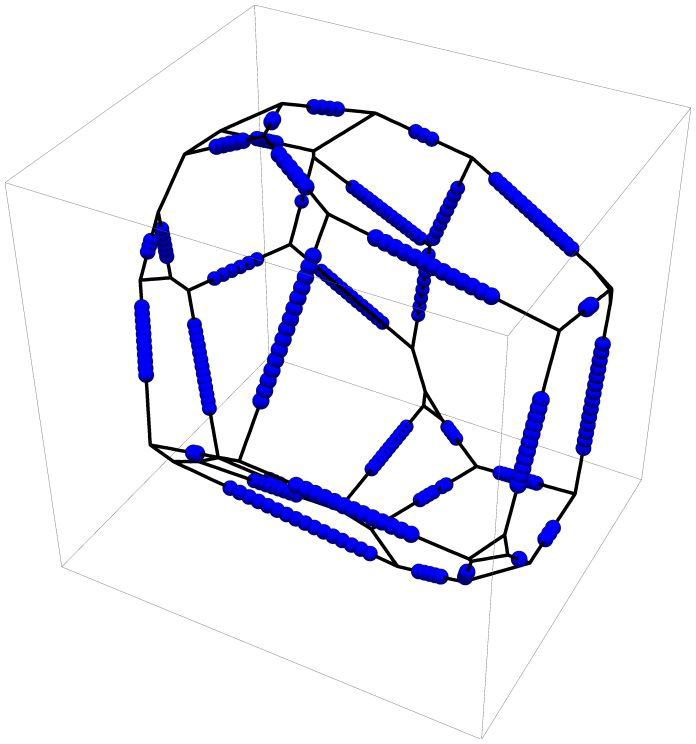


Figure 4.16: The one-skeleton of a Voronoi cell (black edges), and the points of its corresponding persistent homology representative (blue spheres).

4.8.2 Recovered Faces

The recovered faces of C are determined via the adjacency of the recovered cells, just as the faces of the original complex are determined by the adjacency of the original cells. That is, two recovered cells share a recovered face if they share more than one edge. In general, not all of the faces of the original complex are recovered in this way. Small faces are often lost because their points are not included in the minimal cycle, as in Figure 4.16. As will be seen below, faces that are large relative to the size of the cells containing them are sometimes missed as well.

4.8.3 Testing Effectiveness

To compare the recovered complex with the original, we must establish a correspondence between the original cells and the recovered cells of C. This correspondence is not necessarily injective in either direction. In the case of a Voronoi tesselation, each face is contained within a single plane, so if a recovered cell contains two edges of an original cell that are not coplanar, the recovered cell and original cell correspond to each other. In general, one must use heuristic methods to establish this correspondence, i.e. those based on the percentage of edges of the cell recovered by the face (note that these methods will only be heuristic, as a single face can contain most of the edges of a cell).

Not all cells of C give rise to a unique recovered cell c shared by no other original cells. The first of four possible cases is shown in Figure 4.16, where a recovered cell corresponds to a single cell, and that cell corresponds only to the one recovered cell. Cells may also fracture: if they are "thin" in a particular direction, they can give rise to multiple persistence classes, and therefore multiple recovered cells. An example of this is depicted in Figure 4.17. The blue and orange spheres are from two different recovered cells, but come from a single original cell. The third case occurs when a recovered cell corresponds to more than one cell of the original cell complex. This is often occurs when two original cells share a face that is large relative to their sizes. When the union of cells expands, only one component appears in the complement. This case is shown in Figure 4.18. In the potential application to open-cell foams, this loss of information does not matter much as the

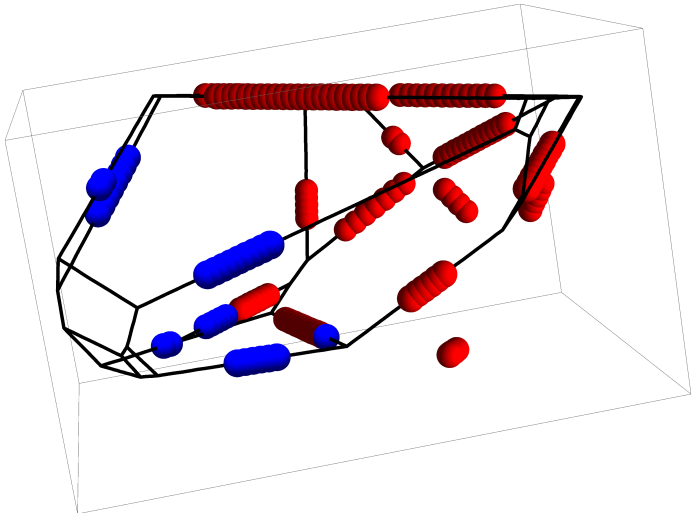


Figure 4.17: An original cell that fractures into two recovered cells, one shown by red spheres and the other by blue. Note that the red recovered cell also corresponds to two original cells (the other is not shown).

recovered graph will have the same homotopy type. Finally, some original cells do not give rise to any corresponding recovered cells.

4.8.4 Voluminous Cells

However, a special kind of cell always gives rise to a recovered cell. We say that a convex, piecewise linear polyhedron is voluminous if the radius of the largest inscribed sphere is greater than the radius of the largest circle that can be inscribed on any of its faces. Note the similarity between this definition and that of the alpha value in Section 4.5. It is easily seen that a voluminous cell corresponds to at least one recovered cell, and that that recovered cell does not correspond to any other cell in the original cell complex. Note that voluminous cells may fracture, and the converse is false even for convex polyhedra (there exist non-voluminous convex polyhedra that give rise to a PH_2 class).

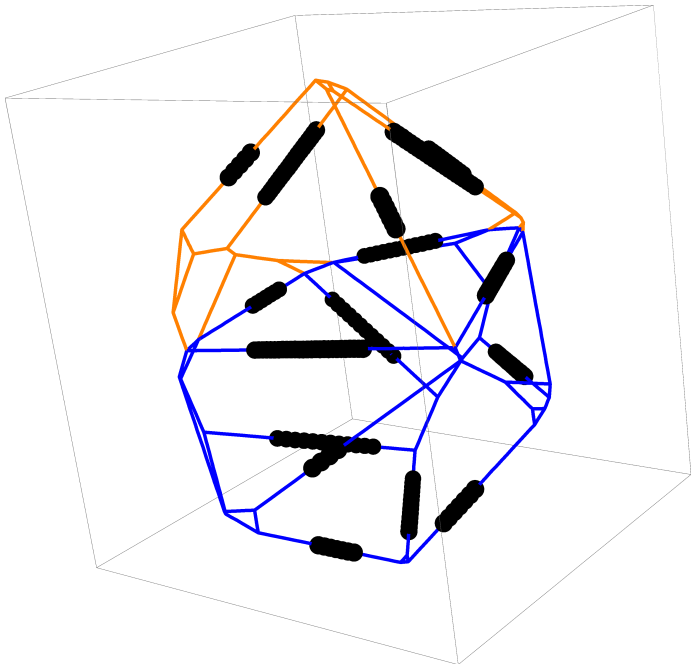


Figure 4.18: A recovered cell (black points) corresponding to two different original cells, one shown with orange edges the other blue.

4.8.5 Computational Results

This method recovers most of the cells of a Voronoi tesselation from its one-skeleton. For a Voronoi tesselation of 120 Poisson distributed points on the three-torus, 118 of 120 original cells correspond to a recovered cell. However, 4 original cells fracture to correspond to more than one recovered cell, and 38 of the recovered cells correspond to more than one original cells. Of the faces, 298 out of 947 are recovered. This is not a very good result, though many of the missing faces are quite small. For physical applications, it will be important to develop heuristic methods to recover more faces, perhaps based on the relative locations of the recovered cells.

Chapter 5

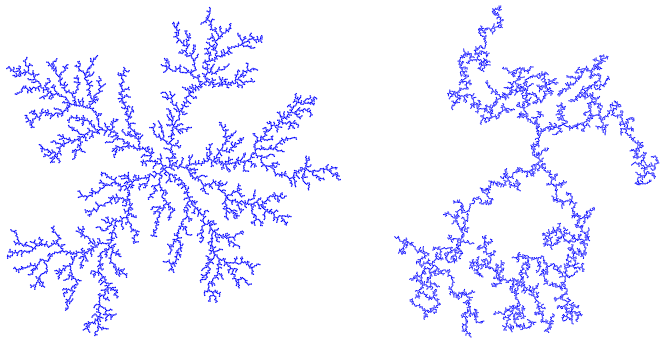
Persistent Homology Dimension and the Persistent Homology of Polymers

5.1 Introduction

Consider the complex geometric structures in Figure 5.1. While their local structure and global topology (both are contractible) are similar, their global geometry is quite different. Both have "gulfs"; the ones in Brownian Tree in Figure 5.1a tend to be long and think, while those in the Branched Polymer in Figure 5.1b are more varied in their structure. Also, the Brownian Tree has a distinct center, whereas the Branched Polymer looks uniform at different spatial locations and scales. In this section, we will use Persistent Homology to quantify these differences. The results here were first presented in [40], and the exposition here will follow the treatment in that paper.

The objects in Figure 5.1 are examples of objects called polymers; an m-dimensional polymer of order n is a connected subset of $\mathbb{R}^m$ that is the union of n distinct m-balls of the same radius r. We assume that the balls are non-overlapping, i.e. two m-balls are either disjoint, or else tangent.

In Section 5.3 we will introduce three different probability distributions on the space of polymers. Each is a model for a physical process of interest, and they have been studied by mathematical physicists as representatives of interesting universality classes. Later in the section, we will discuss computational results pertaining to all three distributions in the plane, and two of them in three-space.



(a) (b)

Figure 5.1: A (a) Brownian Tree and a (b) Branched Polymer of order 5568.

We will use Persistent Homology to study these three distributions of polymers. An example of a PH_1 Scatter Plot for a single large 2D Branched Polymer is shown in Figure 5.2. Note that the y-coordinate has been rescaled by taking the arc-secant; this choice allows us to better visualize the whole range of possible behavior from death/birth = 0 (gulfs that are very narrow) to death/$textbirth$ >> 0 (gulfs that are very wide). We call the rescaled y-coordinate the aspect. There is an incredible amount of information in this plot, and that of any other individual polymer. As such, while we were interested in visualizing representatives of individual PH classes and studying their relationships in the previous chapter, here we will study statistics of entire PH scatter plots.

The three probability distributions on the space of polymers induce probability distributions on the corresponding Persistent Homology Scatter Plots introduced in Section 4.2. The remainder of this chapter will introduce and apply methods for studying these induced distributions. In particular, Section 5.2 will define a notion of fractal dimension based on the Persistent Homology, called the PH-dimension. Computations will show that it agrees with Hausdorff dimension in

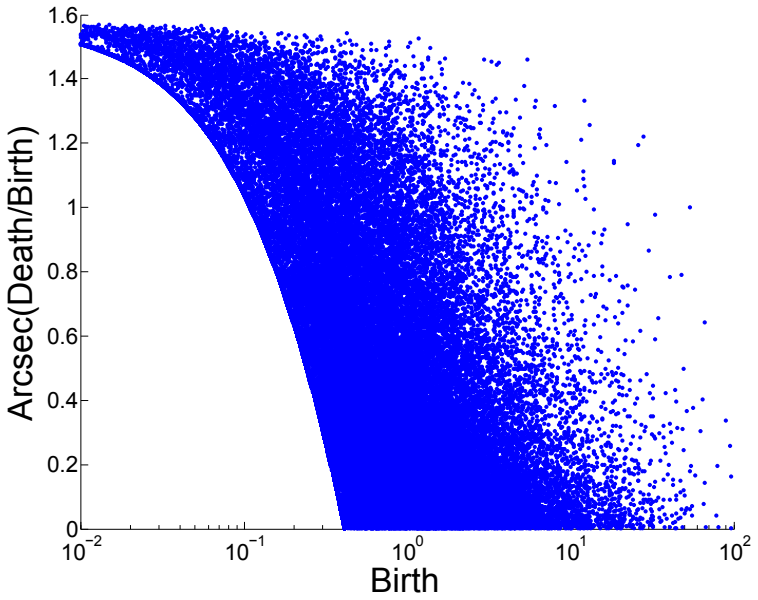


Figure 5.2: PH_1 Scatter Plot for a Large 2D Branched Polymer. The sharp cutoff on the left is due to the geometry of small configurations of balls, and is not of interest to us.

certain cases, and with conjectural values of the fractal dimension of the polymers.

5.2 Dimension and Self-Similarity.

In this section, we introduce three notions that are derived from the scatter plot of feature points: *PH* dimension and *PH* self-similarity, and *PH* statistical self-similarity. These ideas are particularly useful when S is fractal. They are Persistent Homology analogues of Hausdorff dimension and ordinary self-similarity. It will be useful to refer to the points in the *PH* Scatter Plot as *PH* points. Also, many self-similar objects will have infinitely many *PH* points of arbitrarily small size. Despite this, the *PH* classification theorem in Section 4.4 still holds if there are finitely many topological changes for ϵ in the range (a, ∞) for all $a > 0$. The classifications are consistent for all a, so if we take the limit as $a \to 0$, we get a classification of the *PH* for the positive real line.

Let $F(x)$ be the number of *PH* points (x_j, y_j) of S such that $x_j \geq x$. Since F is essentially an integral from the right with respect to x, it is a nowhere increasing function.

5.2.1 Definition of PH Dimension.

We say that S has PH dimension d if

$$\text{for all} \quad c < d, \quad \lim_{x \to 0} x^c F(x) = \infty \quad \text{and} \tag{5.1}$$

$$\text{for all} \quad c > d, \quad \lim_{x \to 0} x^c F(x) = 0 \tag{5.2}$$

i.e. if $F(x)$ grows roughly like x^{-d} as x decreases to 0.

5.2.2 Definition of PH Self-Similarity.

Let ρ be a number, $0 < \rho < 1$, and let ℓ be an integer, $\ell > 0$. We say that S is PH *self-similar* with similarity ratio ρ and replication number ℓ if for small enough x, the number of PH points with coordinates $(\rho x, y)$ is ℓ times the number of PH points with coordinates (x, y). (To rule out trivial cases, we assume S has PH points for arbitrarily small x coordinate.)

5.2.3 The Relation between PH Dimension and PH Self-Similarity.

The following proposition is an analogue of the relationship between self-similarity and Hausdorff dimension.

Proposition.

If S is PH self-similar with similarity ratio ρ and replication number ℓ, then its PH dimension is

$$d = \frac{\log(\ell)}{\log(\rho^{-1})} = -\frac{\log \ell}{\log \rho}$$

(To rule out trivial cases, we assume S has PH points for arbitrarily small x coordinate)

Proof.

Suppose that $x \leq c$ is "small enough x" in the definition of PH self-similar. It is enough to show that for $x \leq c$, $A_1 + B_1 x^{-d} \leq F(x) \leq A_2 + B_2 x^{-d}$ for some A_i, B_i and for $d = -\log \ell / \log \rho$.

For positive integers k,

$$F(\rho^k c) = F(c) + \ell + \ell^2 + \cdots + \ell^k = A + B(\rho^k)^{-d}$$

where $A = F(c) + \ell/(1-\ell)$, and $B = \ell/(1-\ell)$. Now suppose $\rho^{k+1}c \leq x \leq \rho^k c$. Then

$$F(x) \geq F(\rho^k c) = A + B(\rho^k c)^{-d} \geq A + B(x/\rho)^{-d} = A + B\rho^d x^{-d} \equiv A_1 + B_1 x^{-d}$$

$$F(x) \leq F(\rho^{k+1} c) = A + B(\rho^{k+1} c)^{-d} \leq A + B(x\rho)^{-d} = A + B\rho^{-d} x^{-d} \equiv A_2 + B_2 x^{-d}$$

5.2.4 Relation between *PH* dimension and Hausdorff dimension.

The formula $d = -\log \ell / \log \rho$ shows that *PH* dimension bears the same relation to *PH* self-similarity that Hausdorff dimension bears to ordinary self-similarity. Nevertheless, *PH* dimension and Hausdorff dimension in general measure different things. For example, d-dimensional Euclidean space has Hausdorff dimension d but it has *PH* dimension 0 since its only non-trivial *PH* class will be given by the one component of $\mathbb{R}^d$. In this case and in general, *PH* dimension always measures the complexity of a shape, whereas Hausdorff dimension measures some combination of complexity and topological dimension.

Now, we will construct another example showing that *PH* dimension can exceed Hausdorff dimension. For any ρ and ℓ, let the set $S_{\rho,\ell}$ be the set of real numbers

$$0 = x_{1,0}, x_{1,1}, \ldots, x_{1,\ell} = x_{2,0}, x_{2,1}, \ldots, x_{2,\ell^2} = x_{3,0}, x_{3,1} \ldots, x_{3,\ell^3} = x_{4,0}, \ldots$$

where $x_{k,j+1} - x_{k,j} = 2\rho^k$. The set $S_{\rho,\ell}$ has *PH* points only for $i = 0$, where the points are all of the form $(\rho^k, \pi/2)$ which occurs ℓ^k times for each $k = 1, 2, \ldots$ (representing a degree 0 persistent homology class that is born at $\epsilon = 0$ and dies at $\epsilon = 2\rho^k$). So it is *PH* self-similar with similarity ratio ρ and replication number ℓ. Therefore, it has *PH* dimension $-\log \ell / \log \rho$ if $i = 0$ and I contains $\pi/2$. However, it has Hausdorff dimension 0 and it is not self-similar in the usual sense.

Finally, consider the fractal S depicted in Figure 5.3 with several of its ϵ-neighborhoods. It is

an example of a structure that is both both Self-Similar and *PH* Self-Similar, and whose Hausdorff dimension and *PH* dimension agree. The *PH* points of S can be computed exactly.

S is a generalized version of the Sierpinski triangle. To construct it, let S^1 be an equilateral triangle with sides of length one and vertices $(0,0),(1,0)$, and $(1/2,\sqrt{3}/2)$. Let f_1, f_2, and f_3 be contractions of the plane with shared factor $\rho \leq 1/2$ that fix $(0,0)$, $(1,0)$, and $(1/2,\sqrt{3}/2)$, respectively. Let $S^k = f_1(S^{k-1}) \cup f_2(S^{k-1}) \cup f_3(S^{k-1})$ for $k > 1$. Then S is defined to be $\cap_k S^k$. If $\rho = 1/2$, S is the Sierpinski triangle.

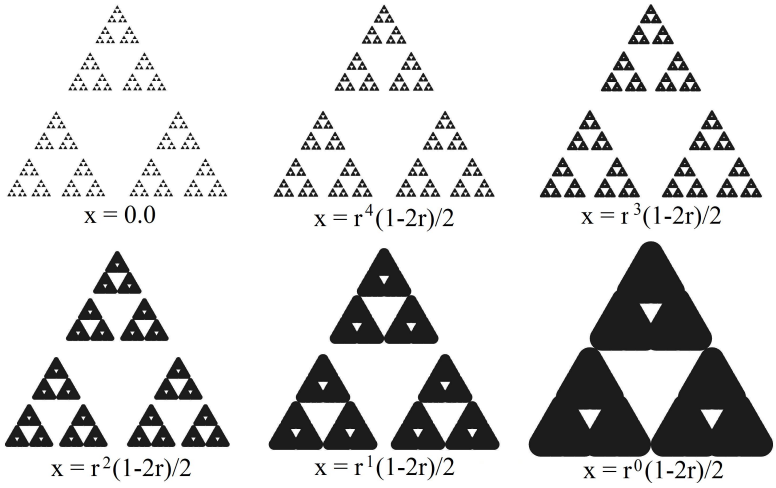


Figure 5.3: A self-similar fractal and its ϵ-neighborhoods

Proposition.

S is *PH* self-similar with similarity ratio ρ and replication number 3. Therefore its *PH* dimension is $-\log 3/\log \rho$ (provided that either $i = 1$ and the interval I contains the number $\tilde{y}$ defined below, or else $i = 0$ and the interval I contains the number $\pi/2$). Furthermore, it is self-similar in the usual sense with similarity ratio ρ and replication number 3, so its Hausdorff dimension is also $-\log 3/\log \rho$.

Proof.

We calculate all of the *PH* points of S exactly.

First, consider $i = 1$. For every integer $k \geq 0$, there are 3^k independent cycles in $H_1(S_\epsilon)$ which are born when $\epsilon = \rho^k(1/2 - \rho) \equiv b_k$. These are the values of ϵ illustrated in Figure 5.3, where the cycles are visible as newly formed holes in S_ϵ. These cycles die when $\epsilon = \rho^k\sqrt{1/3 - \rho + \rho^2 - \rho^3 + \rho^4} \equiv d_k$. Therefore S has 3^k *PH* points (x_k, y_k) where $x_k = (b_k + d_k)/2 = \rho^k(b_0 + d_0)/2 = \rho^k x_0$ and $y_k = \text{arcsec}(d_k/b_k) = \text{arcsec}(d_0/b_0) \equiv \tilde{y}$ is independent of k.

Next consider $i = 0$. For every integer $k \geq 0$, there are $2 \cdot 3^k$ cycles that die when $\epsilon = \rho^k(1/2 - \rho) \equiv d_k$. A canonical basis of $H_0(S_\epsilon)$ is indexed by the connected components of S_ϵ. So we can see the homology classes dying since connected components are fused together at these values of ϵ, Fig. 5.3. Every homology class is born at $\epsilon = 0$, so we take $b_k = 0$ for all k. Therefore S has $2 \cdot 3^k$ *PH* points (x_k, y_k) where where $x_k = (b_k + d_k)/2 = d_k/2 = \rho^k(d_0)/2 = \rho^k x_0$ and $y_k = \text{arcsec}(d_k/b_k) = \text{arcsec}(\infty) = \pi/2$, again independent of k.

5.2.5 Probabilistically Determined Structures S.

Unlike the example above, the physically interesting structures from Section refsec:polymers we consider are determined by a probabilistic process. In this section, we will define versions of self-similarity and dimension that are adapted to this situation.

Suppose S is a probabilistic process. The probability distribution from which S was generated induces a probability distribution on its *PH* points. That is, its *PH* points are drawn from a distribution with a *PH density function* $f(x, y)$ which measures the expected number of *PH* points per unit area at (x, y). The density function f depends only on the probabilistic process giving rise to S, unlike the individual *PH* points which depend on a particular realization S of that process. In this case, the expected value $\mathbb{E}[F(z)]$ of the function $F(z)$ will be

$$\mathbb{E}[F(z)] = \int_{y \text{ in } I} \int_{x=z}^{\infty} f(x, y)\, \mathrm{d}x\, \mathrm{d}y$$

We call the process giving rise to S *PH* statistically self-similar with self-similarity dimension d

if the *PH* points in the plane are sampled from a spatial Poisson process with *PH* density function of the form

$$f(x, y) \approx x^{-(d+1)} \cdot g(y)$$

for some function $g(y)$, at least for x small compared to the diameter δ of S (at which point $f(x, y)$ will vanish). Here $g(y)$, which is a function of the aspect y alone, encodes additional information about the shape of S unrelated to the dimension. In this case,

$$\mathbb{E}\left[F(z)\right] \approx z^{-d}$$

up to a multiplicative constant, independent of the interval I, and d will be the *PH* dimension almost surely.

We hypothesize that many physically interesting statistical structures S that are statistically self-similar in the usual sense with similarity dimension d are also *PH* statistically self-similar with self-similarity dimension d. We give numerical evidence for this hypothesis in Section 5.4 in the case of 2D self-avoiding walks and 3D branched polymers by calculating the *PH* points of individual realizations S of that process (averaging to improve accuracy) and estimating F from this data.

5.3 Three Probability Distributions of Polymers

5.3.1 Branched Polymers.

Branched polymers are drawn from a distribution that can be viewed as the uniform distribution on the space of polymers. Given a polymer composed of n unit balls $\{B_1, \ldots, B_n\}$ each of radius r, let $x_1, \ldots, x_n$ be their centers. We define the tree-type of the polymer to be the tree with vertices $\{1, \ldots, n\}$ and edges $\{i, j\}$ such that B_i and B_j are tangent. There is a measure Ω_T on the space of branched polymers whose tree-type is T :

$$\Omega_T = \prod_{\substack{ij \in T \\ i<j}} \Omega(x_i - x_j)$$

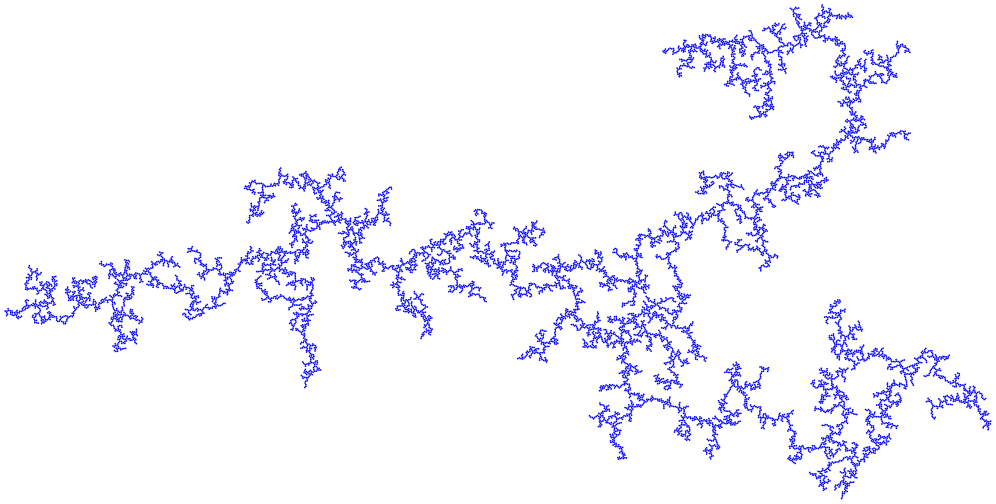


Figure 5.4: A Branched Polymer of order 10,000.

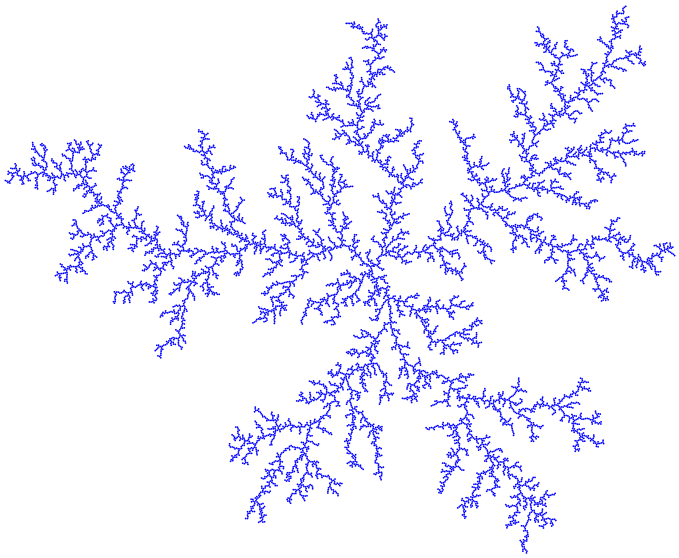


Figure 5.5: A Brownian Tree of order 10,000.

where $\Omega(x_i - x_j)$ is the area form of the unit sphere of vectors $x_i - x_j$. The probability measure on the space of all branched polymers is defined as the sum of this measure over each tree type, normalized to have total volume one.

We used an algorithm created by Kenyon and Winkler to inductively construct branched polymers in two and three dimensions [32]. A two-dimensional branched polymer of order 10,000 is depicted in Fig. 5.4.

5.3.2 Brownian Trees.

Another distribution on the space of polymers samples objects called Brownian trees. They are built inductively – given a Brownian tree with $n-1$ balls, another ball is placed on a circle at some distance away from it. It then moves via Brownian motion until it reaches the existing structure and sticks. This process is called diffusion limited aggregation. A Brownian tree composed of 10,000 balls is depicted in Figure 5.5.

To create Brownian trees, we used a program written by Mark J. Stock [67].

5.3.3 Self-Avoiding Walks.

The third distribution we consider is that of self-avoiding (random) walks on a square lattice, which are a model for linear polymers. A self-avoiding walk is a subset of the edges of a lattice that forms a non-intersecting path. They are given the uniform distribution: if there are ℓ self-avoiding walks of length n, each is given weight $1/\ell$. We realize self-avoiding walks as polymers by viewing each edge as three touching balls. An example of a self-avoiding walk composed of 500,000 lattice edges is shown in figure 5.6.

We used the fast-pivot algorithm created and implemented by Tom Kennedy to generate self-avoiding walks [30] [31].

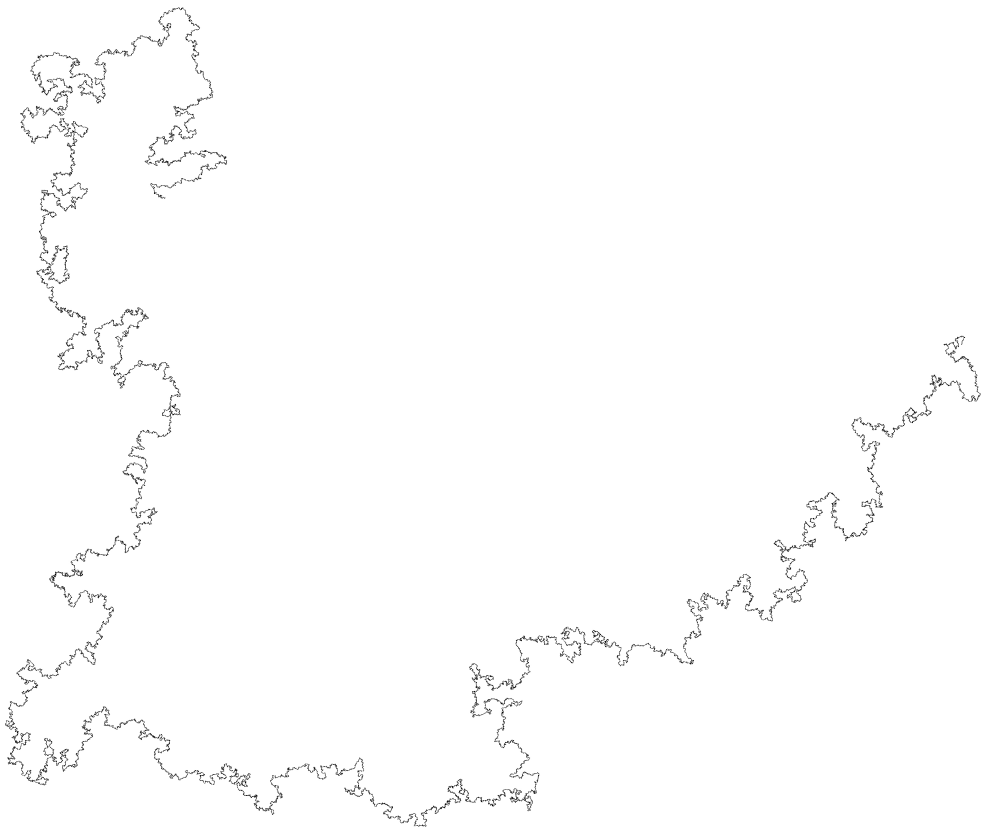


Figure 5.6: A Self-avoiding Random Walk with 500,000 lattice edges.

5.4 Computational Results.

5.4.1 Numerical Determination of PH Statistical Self-Similarity and Estimation of PH Dimension.

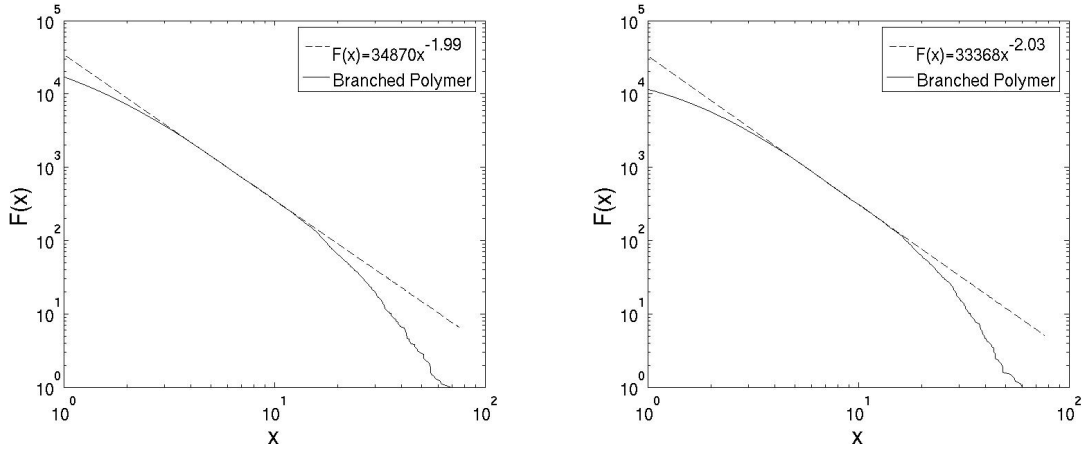
We consider 2D self-avoiding walks, 2D and 3D branched polymers, and 2D and 3D Brownian trees. Using the methods described in Section 4.5, we compute scatter plots of PH points for each polymer. Then, we calculate the function $F(x)$ defined in Section 5.2 and observe whether or not it appears to follow a power law for some interval on the x axis. If so, we use linear regression (taking the log of both axes so that a power law $F(x) = cx^{-d}$ becomes a linear relation) to estimate the exponent, which will be an estimate of the PH dimension.

Even if the probabilistic process we are modeling is PH statistically self-similar, we do not expect $F(x)$ to follow a power law for either very high values or very low values of x in our individual simulations. Each simulation has two natural length scales: the diameter δ of the polymer S itself, and the diameter r of an individual ball in the polymer. We can only expect that our estimate of $F(x)$ is of physical interest if x is well above r and well below δ. The function $F(x)$ is identically zero when $x > \delta$ so it cannot follow a power law. Increasing the number n while fixing r would increase the upper end of the range of accuracy. When x is close to r, the features reflect the local geometry of small collections of balls, rather than the limiting object. Note that power law behavior begins at a small multiple of the component ball size in Figures 5.7 and 5.8 ($r = 1$ for branched polymers and Brownian trees, while it is $1/2$ for self-avoiding walks which have additional local structure as they are constructed on a lattice).

Our finite polymers do not have a non-zero PH dimension as defined in Section 5.2.1, because that takes the limit as $x \to 0$.

5.4.2 Three Dimensional Branched Polymers.

Plots of $F(x)$ for degree $i = 1$ and degree $i = 2$ of three-dimensional branched polymers is shown in Figure 5.7 on log-log paper. The data were gathered from ten branched polymers of order 30,000. For both $i = 1$ and $i = 2$, the plot appears to be linear from about size ~ 4 to size ~ 15, so



(a) (b)

Figure 5.7: Log-log plots of $F(x)$ for 3D branched polymers for (a) PH_1 and (b) PH_2.

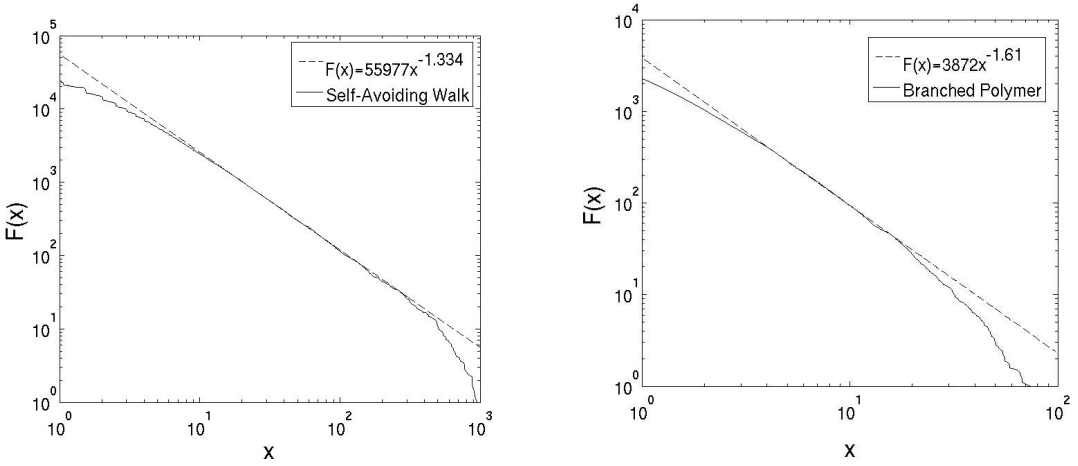
the process appears to be PH statistically self-similar. Linear regression yields exponents -1.99 and -2.03, respectively. Thus the PH dimension of 3D branched polymers is about 2, which is consistent with the scaling exponent of $1/2$ that Brydges and Imbrie proved for three-dimensional branched polymers in [10].

5.4.3 Two Dimensional Branched Polymers.

Here, our results are inconclusive. The function $F(x)$ for 2D branched polymers appears to follow a power law with exponent -1.61 for a certain range. It was computed from a data set of 10 branched polymers of order $n = 10,000$ and it is plotted in figure 5.8b. However, the fit is not as convincing as it is for 3D branched polymers and 2D self-avoiding walks. Furthermore, the rigorous methods of Brydges and Imbrie [10] do not work in two dimensions and do not provide a value for the scaling exponent of 2D branched polymer.

5.4.4 Two Dimensional Self-avoiding Walks.

A similar plot for the 2-dimensional self-avoiding walks and degree $i = 1$ is shown in figure 5.8a. The data were gathered from five self-avoiding walks composed of 500,000 lattice edges each. The



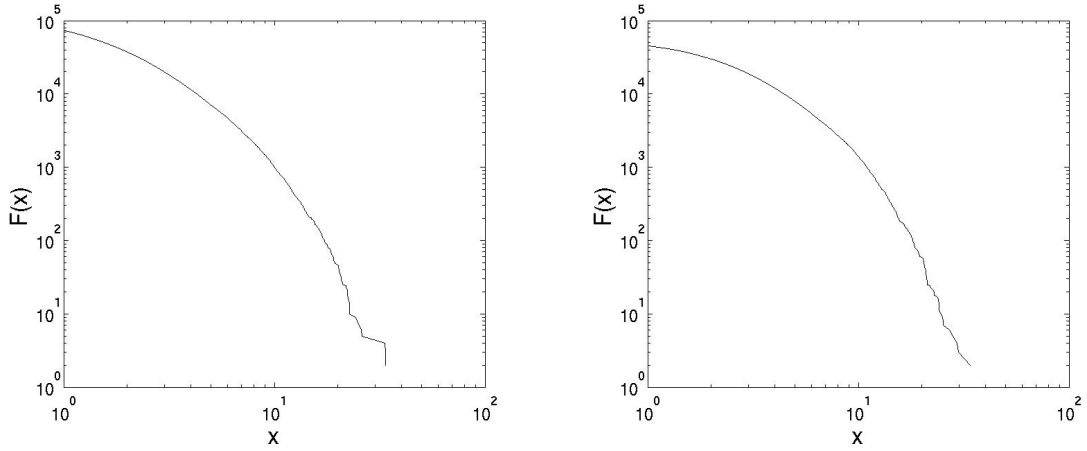
(a) (b)

Figure 5.8: Log-log plots of $F(x)$ for (a) 2D S.A.W. using PH_1 and (b) 2D B.P. using PH_1.

plot appears to follow a power law for a very long range: from size ~ 10 to size ~ 150, which is consistent with PH statistical self-similarity. Least squares regression yields an estimate of -1.334 for the exponent and therefore a value of 1.334 for the PH dimension. This is very close to the conjectured value of 4/3 for the Hausdorff dimension, which was first hypothesized by Chemist P. Flory and has been proven rigorously for an object believed to be in the same universality class as two-dimensional self-avoiding walks by B. Nienhuis [57].

5.4.5 Brownian Trees.

Unlike in the previous cases, the function $F(x)$ for Brownian trees does not come close to following a power law, so we conclude that they are not PH statistically self-similar. The plots for one- and two-dimensional PH for 3D Brownian trees are shown in figures 5.9a and 5.9b (a plot for 2D Brownian trees is similar). The data were gathered from three Brownian trees of order 30,000. Note that the plot is significantly bowed out, indicating that the function on a log-log plot is more concave than linear. This result is expected, as Mandelbrot et. al. have shown with high statistical certainty that Brownian trees are not self-similar [42] [41].



(a) (b)

Figure 5.9: Log-log plots of $F(x)$ for 3D Brownian trees using (a) PH_1 and (b) PH_2.

5.4.6 Shape Measures Involving the Aspect y.

The function $F(x)$ is not the only way to extract useful information from the Persistent Homology of polymers. Consider the bar chart in Figure 5.10. For each interval I of possible aspects listed on the horizontal axis($0-.2$, $.2-4$, etc.), the height of the bar is the ratio $P_{BP}(I)/P_{BT}(I)$, where $P_{BP}(I)$ is the proportion of two-dimensional branched polymer features ($i = 1$) whose aspect y is in the interval I and $P_{BT}(I)$ is the proportion of two-dimensional Brownian tree features ($i = 1$) whose aspect y is in the interval I. Recall that the aspect y of a persistent homology class is equal to $\text{arcsec}(d/b)$, where d and b are the death point and birth point of the class, respectively. This effectively measures the angular opening at the edge of a gulf in the structure.

The proportion of features with low aspect measurements appearing in the branched polymers and the Brownian trees is similar. However, features with aspect greater than 1.2 are more common in branched polymers and those with aspect greater than 1.4 are much more abundant. This makes sense when one examines the pictures of the branched polymer and the Brownian tree shown in Figure 5.1. Features with a large aspect value are gulfs with the property that a large ball can fit inside them but only a very small ball can enter them. Intuitively, this difference seems to capture the fact that most features of the Brownian trees have a similar shape. These gaps between the

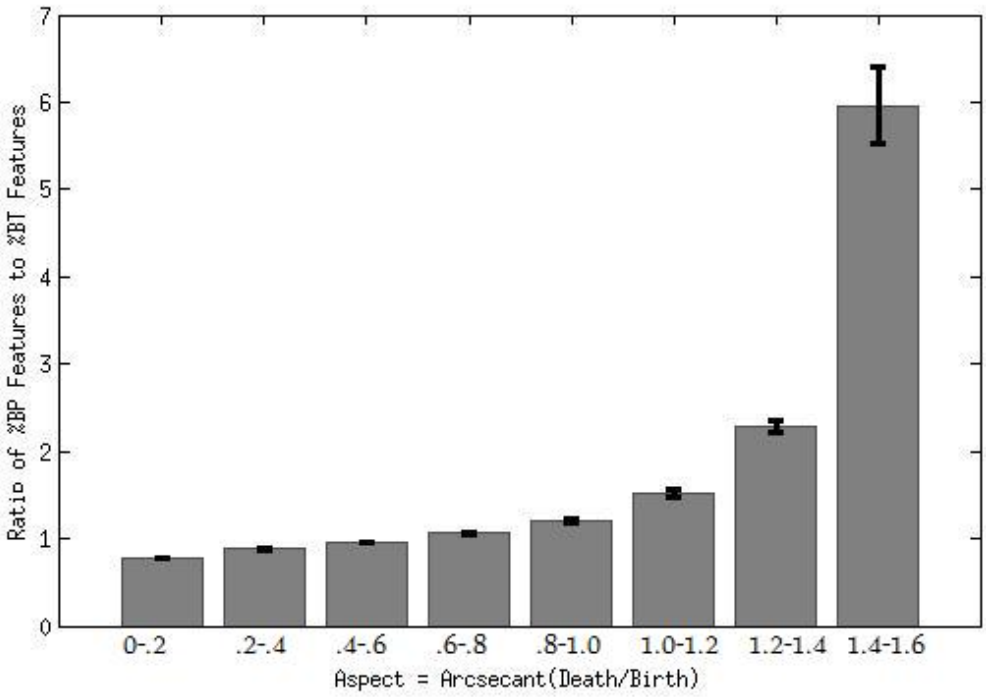


Figure 5.10: Ratio of branched polymer features to Brownian tree features for various aspects

branches of the Brownian tree can trap balls only slightly larger than those that can enter them. On the other hand, branched polymers have more varied gulfs.

Chapter 6

One-dimensional Features

In this chapter, we study the one-dimensional features of steady state configurations for curvature flow on embedded graphs in three dimensions (we call these "CDE configurations"). These features are loops that that are minimal in a certain sense. We generate sets of minimal loops in two different ways: by computing loops of shortest length containing containing each edge of the graph, and by computing minimal homology bases. These methods are described in Sections 6.1 and 6.2, respectively, and we compare them in Section 6.3. We also examine the first-dimensional Persistent Homology of a CDE configuration in Section 6.4.

6.1 Shortest Loops

There are several ways to define a shortest loop containing an edge. Here, we study two notions: the loop of shortest length in terms of the Euclidean distance, and the loop (or loops) with the fewest number of edges. These are often different, as in the example in Figure 6.1.

Simple algorithms were used to compute these loops. Fix an edge e with vertices v_1 and v_2. For shortest loops based on the Euclidean distance, we built paths starting at v_1 and extended them until we reached v_2. These paths were stored in a priority queue sorted by length, and at each step the shortest path was removed and extensions of it added back into the queue. The process

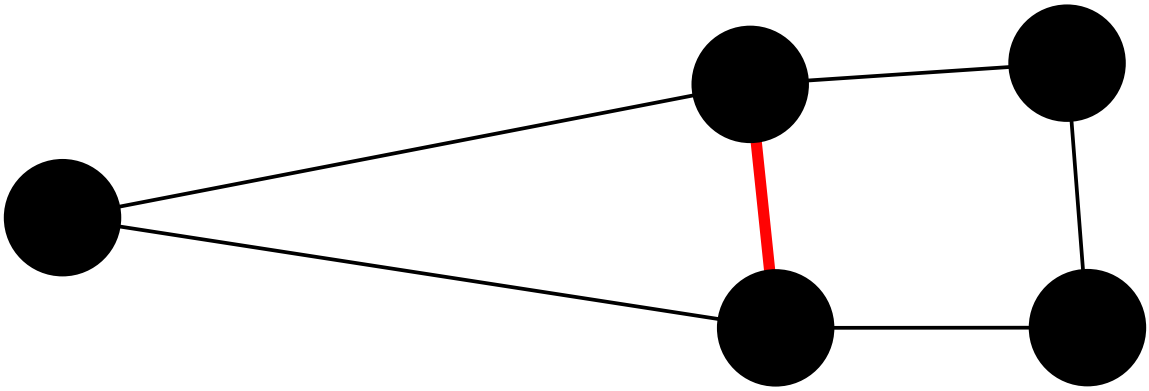


Figure 6.1: The two notions of shortest cycle differ for the thick edge. The rectangle is the loop with shortest Euclidean length and the triangle the one with the fewest edges.

continued until a loop was found, and all of the paths in the queue were longer than that loop. For shortest loops based on the number of edges, we used a slightly different data algorithm. Starting with $d = 1$, we iteratively computed lists of paths with d edges until we found a loop (or loops).

The properties of these loops are examples of what are called ring statistics in materials science. Ring statistics are a good first step toward studying the statistical topology of embedded graphs. [61]

6.2 Homology

Another way to generate a set of loops is to find a basis for the one-dimensional homology that is minimal in a certain sense. As described in Section 4.3, a basis for the one-dimensional homology of an embedded graph is an algebraically independent set that spans the space of loops in the graph. Note that a set of shortest cycles at each edge is not necessarily algebraically independent, and may not span the space of loops, so these notions are quite different.

We used Mark McConnnell's code for computing minimal homology bases [46] in terms of both the number of edges and the Euclidean distance. The program takes as input a set of edges with their lengths (for us, either 1 or the Euclidean length), and uses a greedy algorithm to compute a homology basis of small total length. At each step, the shortest algebraically independent loop is added to the basis.

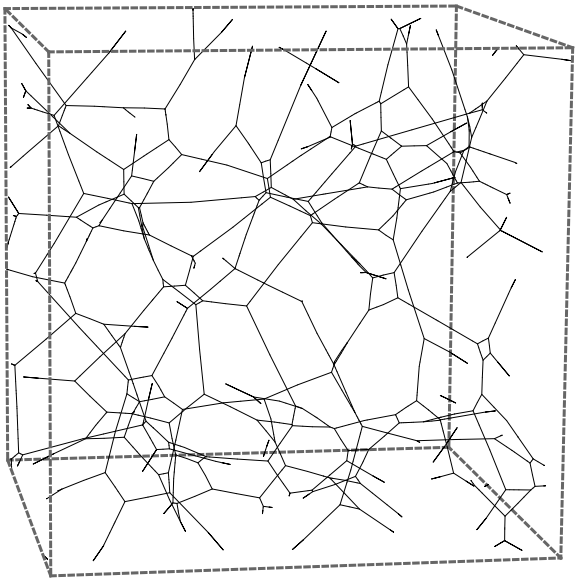


Figure 6.2: A small region in a CDE configuration.

6.3 Comparisons

In the previous section, we defined four minimal sets of loops for an embedded graph. Here, we compare them in the case of a CDE configuration with around 500,000 edges.

6.3.1 Table

The number of loops in the four sets and their pairwise intersections is displayed in Table 6.1. The names are displayed in shorthand, with "Loops by Distance" meaning the set of shortest loops at the edges of the graph in terms of the Euclidean distance, for example. The diagonal elements of the table are the sizes of the four sets. The dimension of the first-dimensional homology of the graph is 161038, so we should find 161035 loops plus three "global" generators that wrap around the ambient torus for a basis. The minimal homology basis determined by distance had only 161034 elements because the program missed one loop whose Euclidean length was unusually long.

The homology bases are larger than each of the shortest loop sets, indicating that some one-dimensional features are missed by looking only at shortest loops. Also, the bases do not contain

	Loops by Edges	Loops by Distance	H_1 by Edges	H_1 by Distance
Loops by Edges	137925	115787	129688	121547
Loops by Distance	115787	138953	122045	138952
H_1 by Edges	129688	122045	161035	134901
H_1 by Distance	121547	138952	134901	161034

Table 6.1: The order of the pairwise intersections of four sets of loops in a CDE configuration with about 500,000 edges. Each row and column corresponds to a set of loops.

all of the shortest loops, so some of the shortest loops are algebraically redundant. Despite these differences, the percentage overlap between the four sets is quite high.

6.3.2 Histograms

Consider the histograms in Figure 6.3 showing the distribution of number of edges per loop in the four sets. While loops in the sets determined by distance (6.3b and 6.3d) tend to have slightly more edges than those determined by edges (6.3a and 6.3c), the difference is not as much as one might expect. This indicates that the number of edges in a loop correlates strongly with its length in the Euclidean distance. The difference between the sets of shortest cycles (6.3a and 6.3b) and the minimal homology bases (6.3c and 6.3d) is far more noticeable - longer loops are needed to get a complete basis.

6.3.3 Loop Length Distribution

Now, we compare the distributions of Euclidean lengths for the four minimal sets of loops. The functions in Figure 6.4 plot a function g where $g(x)$ is the percentage of loops of greater length than x. This function is obtained from integrating the probability density of lengths from the right. The shape of the function is similar in all four sets, but decreases faster for the shortest loop sets than for the homology bases, indicating that longer loops are needed for a complete basis.

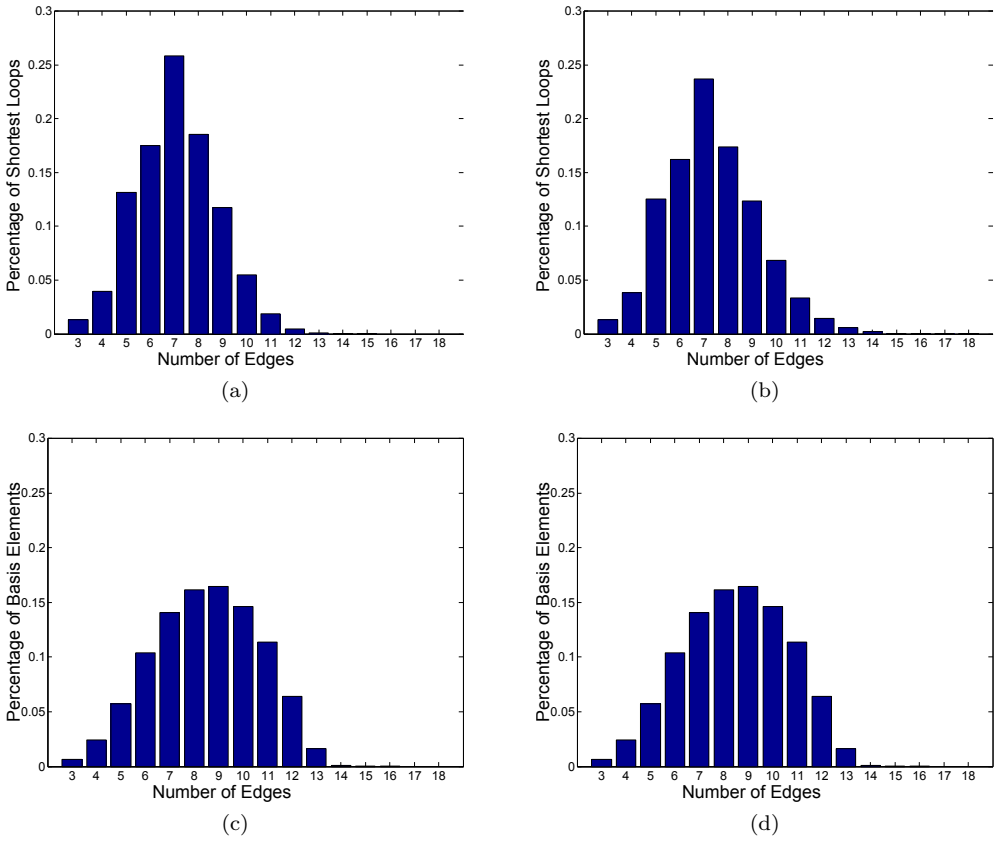


Figure 6.3: Histograms showing the distribution of the number of edges per loop in the four sets: shortest cycles for each edge in terms of (a) number of edges and (b) the Euclidean distance, and minimal homology bases in terms of (c) number of edges and (d) the Euclidean distance.

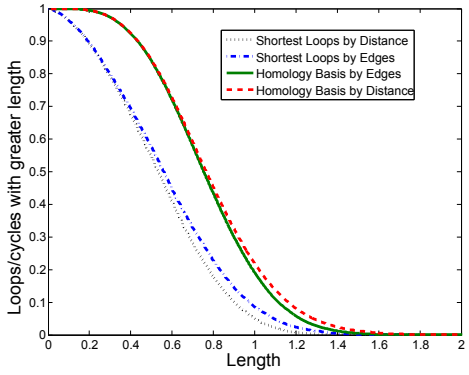


Figure 6.4: Functions representing the distribution of loop lengths in terms of the Euclidean distance. The length of a cycle is plotted against the percentage of loops with greater length.

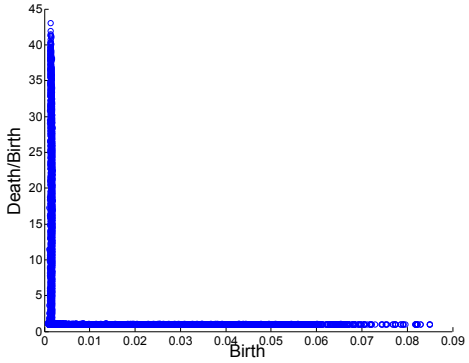
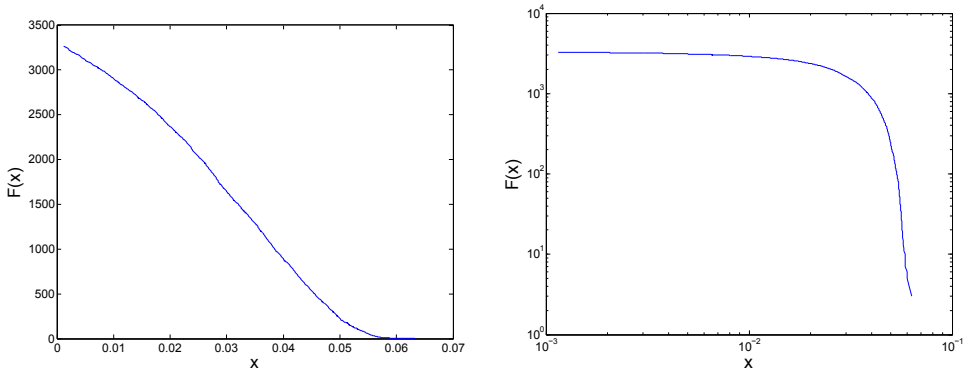


Figure 6.5: A Scatter Plot of Feature Points for the first-dimensional Persistent Homology of a CDE configuration.

6.4 Persistent Homology

The first-dimensional Persistent Homology of an embedded graph provides a measure of its one-dimensional geometric features. For the CDE configuration we are studying, all of these features are loops in the graph. This can be seen from the Scatter Plot of Feature Points in Figure 6.5. The feature points are concentrated along the axis where death equals birth (noise), and in a spike at very small birth values. Recall that Persistent Homology is computed by interpolating points along the graph and taking the alpha complex of those points. The birth values of the spike fall in the range of the mesh of this interpolation, and decreasing the mesh moves the spike to the left by a proportional amount. Thus, all interesting feature points correspond to topological features (loops) of the original graph.

The distribution of PH_1 death times measures the size of loops in the configuration. A death event occurs when increasing ϵ causes a loop to be filled in in the ϵ-neighborhood of the configuration. Thus, roughly speaking, the deaths are the radii of a set of loops that form a homology basis for the graph. Figure 6.6a shows the distribution of these death times, in terms of a function plotting the death time versus the percentage of feature points with later deaths. We excluded feature points with a death-to-birth ratio less than 1.1 to reduce the effect of noise. Note that this is very similar to the function defined in the previous section, and also to the function $F(x)$ used in Section 5.2 to define PH dimension. For completeness, we have plotted it on log-log paper in Figure 6.6b.



(a) (b)

Figure 6.6: Plots showing the distribution of death times for the PH_1 of a CDE configuration. Both plots show the same function, which plots death values against the percentage of feature points with higher death values.

This function is not a power law, so the CDE configuration is not PH self-similar. This is not unexpected, as it does not exhibit fractal properties.

The shapes of the functions plotted in Figures 6.4 and 6.6a are very similar. In fact, the distributions match very closely after a suitable linear reparametrization of the size axis for the PH_1 distribution. This is shown in Figure 6.7. The behavior is different for low death values due to the effect of noise in the PH_1 computation. Both functions measure the distribution of loop sizes occurring in a CDE configuration, but are computed in very different ways. As the length of the cycles in the minimal basis measures the circumference of loops, and Persistent Homology roughly measures the radius of loops, this matching could be interpreted as saying that the length of loops in the configuration is proportional to their radii. That is, that the loops tend to be "round."

6.5 Neighboring Loop Distributions

When studying grain growth (curvature flow on graphs in two dimensions), one of the most important statistics is the distribution of the number of sides of a grain. This has no direct analogue for graphs embedded in three dimensions, but we can substitute the lengths of the shortest loops defined in 6.1. Note, however, that a grain boundary is not necessarily a shortest loop in terms of

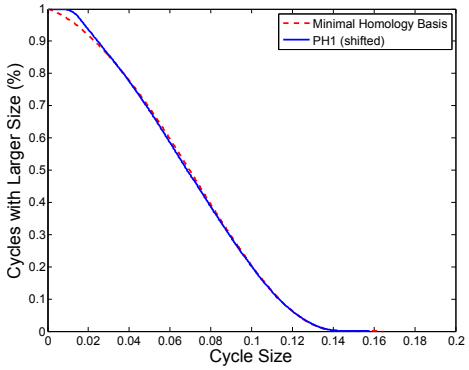


Figure 6.7: Functions measuring the size distribution of loops in a CDE configuration based on PH_1 (solid blue) and a minimal homology basis computed in terms of the Euclidean distance (dashed red). The size axis of the former was linearly reparametrized to highlight the similarity in the shape of the functions.

distance or number of edges for any of its edges.

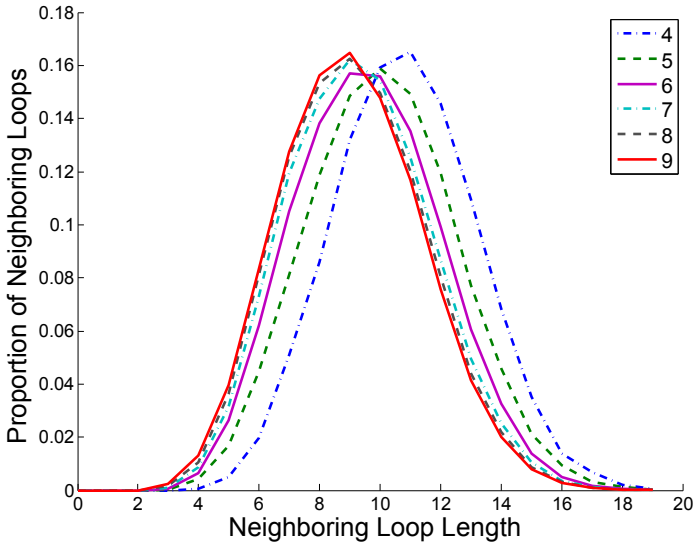
For grain growth with zero vertex drag, materials scientists observed that grains with more than the mean number of sides tended to have neighbors with fewer edges, and vice versa. Specifically, the average number of sides of a neighbor of a grain with n edges is approximately

$$s_n \approx \frac{6+\sigma}{n}$$

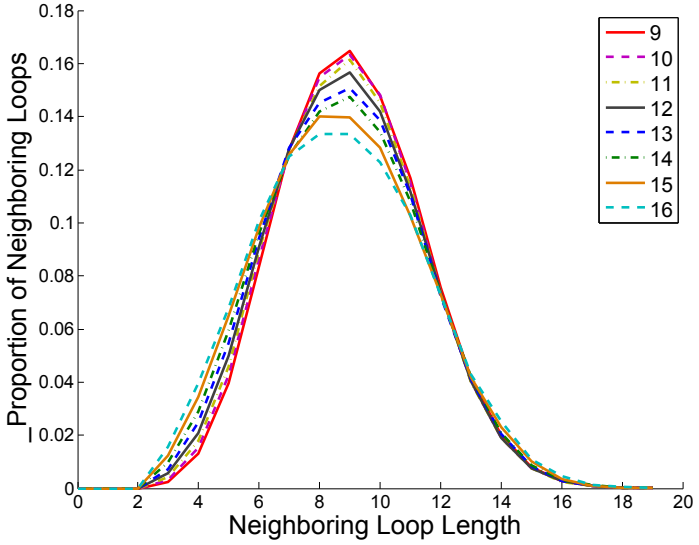
where σ is the variance of the distribution of the number of sides of a grain. Due to the work of Mason, Ehrenborg, and Lazar, it is known that such a formula is true up to an error term proportional to the total curvatures of sides of the grain, which tend to be small. [43] Their result holds for any planar graph with the property that edges meet at a vertex at an angle of $2\pi/3$.

Here, we present computational evidence in the case of a CDE configuration (which is, as always, embedded in three dimensions) that the average length of a shortest loop adjacent to a shortest loop of length n is inversely proportional to n. That is $\mathbb{E}(\rho_n)$ is inversely proportional to n, where ρ_n is the conditional probability distribution of the number of edges in a shortest loop adjacent to another shortest loop of length n, and the shortest loops are computed in terms of the Euclidean distance. Figure 6.8 shows ρ_n for several values of n. Note that as n increases, the probability

distributions shift to the left.



(a)



(b)

Figure 6.8: Conditional probability distributions of the number of edges in a shortest cycle adjacent to a shortest cycle with n edges. To reduce clutter, these are split into two figures with $n = 9$ in both. The shortest cycles were computed in terms of the Euclidean distance.

Chapter 7

Knotting of Embedded Graphs

The content of this chapter is motivated by the observation that the steady state condition of an embedded graph in three dimensions evolving under curvature flow appears to be unknotted. This leads to an obvious question: what does it mean for an embedded graph to be knotted? We will propose a definition in Section 7.3. First, we will review the knot theory of the circle.

7.1 The Knot Theory of the Circle

The knot theory of the circle is is the study of embeddings of the circle, S^1 into the three-sphere S^3 (which is $\mathbb{R}^3$ with a point added at infinity). Two embeddings K_1 and K_2 are considered equivalent if one can be deformed continuously into the other inside the ambient space. That is, there is a continuous map $\phi : S^1 \times [0,1] \to S^3$ such that $\phi(\cdot, t)$ is a homeomorphism onto its image for each $t \in [0,1]$, and $\phi(\cdot, 0) = K_1$ and $\phi(\cdot, 1) = K_2$. This transformation from one embedding to another is called an isotopy. The central problem of knot theory is the classification of knots: what are all possible embeddings of the circle into three-space, and how to tell them apart. The simplest knot is the standard embedding of the unit circle in three-space, and is called the unknot. Knots are usually depicted via their projections into the plane, with the knot shown be to crossing under or over wherever the projection is not injective (the projection assumed to be 1-1 or 2-1 everywhere,

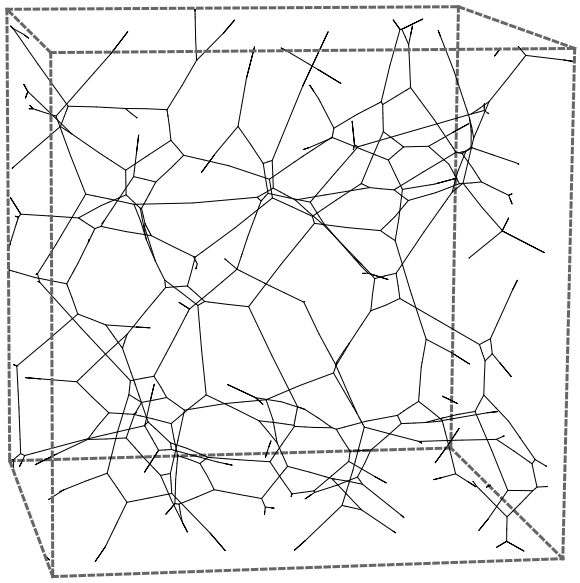


Figure 7.1: A small region in a steady state configuration in three dimensions.

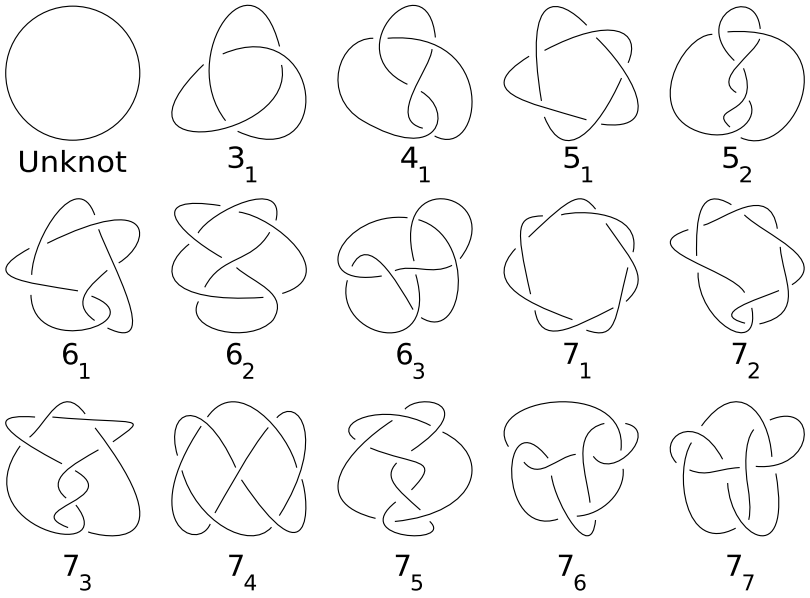


Figure 7.2: A table of knots indexed by crossing number [28]. Each diagram represents a distinct isotopy class of embeddings of the circle into S^3.

or is perturbed slightly so that this is the case). These projections are called knot diagrams, and diagrams for some of the simplest distinct knots are shown in Figure 7.2.

Knot theorists have developed many different invariants in order to aid in the classification of knots. The crossing number is perhaps the simplest of these, and is equal to the minimum number of crossings among all diagrams representing the knot. The diagrams in Figure 7.2 realize these crossing numbers, and those numbers are listed in the figure. For example, the trefoil knot is the only one with crossing number 3, and there are three distinct knots with crossing number 6. Another useful knot invariant is the Jones Polynomial, which will be discussed in Section 7.5.1.

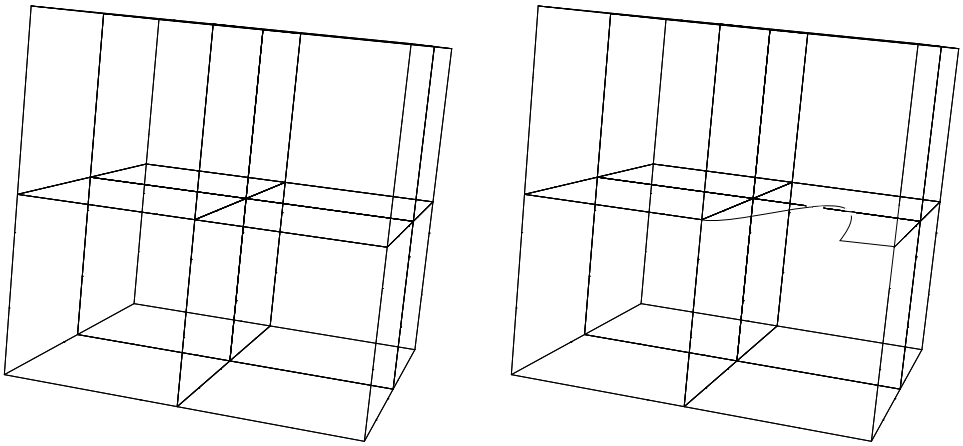
Knots are determined by the topology of their complement. For example, the unknot is the only knot whose complement is homeomorphic to a solid torus. Furthermore, this complement topology is determined by an invariant called the fundamental group. Roughly speaking, the fundamental group of a knot complement keeps track of the different ways to draw loops around the knot. The unknot is the only knot whose complement has fundamental group equal to the integers $\mathbb{Z}$.čite1997lickhorish

7.2 Unknotted Graphs

In analogy with the knot theory of the circle, we are interested in studying isotopy classes of embeddings of a graph G into three-dimensional space. Note that this is a strictly more general subject, as the circle can be realized as a graph with one edge and one vertex. Two questions are of particular interest: given a graph G, which embedding (or embeddings) should we distinguish as unknotted (if any), and how to determine the degree of knottedness of a graph. In this thesis, we will propose an answer to the first question. First, we will examine some examples.

Consider the cubic lattice graph in Figure 7.3a. This is the simplest embedding of the corresponding abstract graph, with no local complexity. Our intuition is that this graph should be unknotted. On the other hand, the graph in Figure 7.3b obtained from it by looping on edge around another edge in the graph should be knotted.

If the 2×2 cubic lattice graph is unknotted, so should the 3×3 cubic lattice graph in Figure 7.4.



(a) (b)

Figure 7.3: Two embeddings of a graph in three-space: (a) a subset of a standard cubic lattice and (b) a knotted embedding.

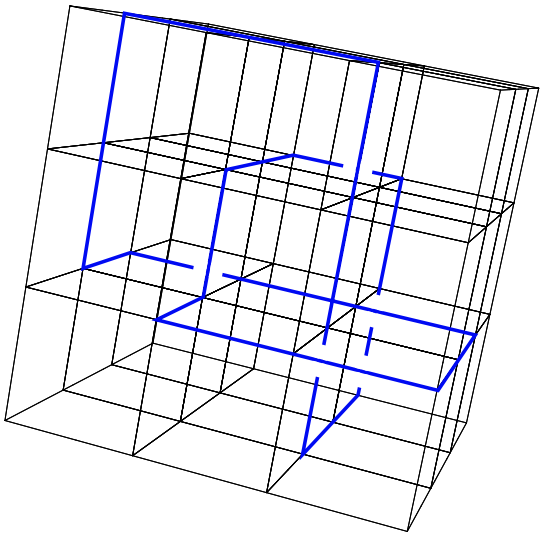


Figure 7.4: A larger subset of the standard cubic lattice graph with a trefoil knot highlighted in blue. This is the knot of minimal length appearing in the cubic lattice graph [14].

However, this lattice graph contains a knot itself. A trefoil knot is highlighted in blue in the figure. In fact, all possible knotted circles appear as subsets of an infinite cubic lattice graph. As such, a definition of unknotted graphs should not require that arbitrary subsets of a graph be unknotted.

7.2.1 Previous Proposals

Several definitions of knotted graphs appear in the literature. All of them are generalize the knot theory of the circle, are interesting mathematical concepts, and are good definitions in their own context. However, none of them is the correct definition for the present context. For example, in [63], Scharlemann and Thompson call an embedded graph unknotted if and only if it is planar (that is, it has an embedding in $\mathbb{R}^2$). This would mean that the cubic lattice graph is knotted, for example. Another very interesting concept is that of an intrinsically knotted graph, a graph which has a knotted cycle in every possible embedding [17]. In this thesis, however, we are interested in studying different embeddings of an abstract graph.

In [26], Hyde and Friedrichs seek to find a distinguished embedding for every graph, and call this the untangled embedding of that graph. For finite graphs, they first restrict their attention to the set of embeddings of a graph G that have minimal genus - that is, the embedding of the graph that can be embedded in a surface of minimal possible genus. This does not uniquely specify an embedding, so they then propose a method for determining the simplest one by measuring the average edge length in the universal cover of the minimal genus surface. For infinite graphs, Hyde and Friedrichs propose that the barycentric embedding is the unique untangled one - that is, the embedding where each vertex is at the barycenter of its neighbors. It is important to note that this criterion is not just topological, but geometric. Finding an embedding of minimal complexity is an interesting problem, but we believe that some graphs should have no unknotted embedding, such as a complete graph on a sufficiently large number of vertices. We would also like to allow for the possibility that a graph has multiple topologically inequivalent unknotted embeddings.

7.3 A Proposed Definition

Recall that an embedded circle is the unknot if and only if its complement also has the homotopy type of a circle. In analogy, we propose that unknotted graphs be the ones whose complement is as simple as possible:

Definition 1. *An embedded graph is* ***unknotted*** *if and only if its complement has the homotopy type of a graph.*

That is, the complement of the graph can itself be continuously deformed into a graph. As in the case of the circle, this is equivalent to a statement about the fundamental group of the complement:

Theorem 3. *An embedded connected graph is unknotted if and only if the fundamental group of its complement is free.*

This theorem follows from a result of Waldhausen which states that a large class of three-manifolds are determined by their fundamental group [69]. Specifically, a 3-manifold (subject to some niceness hypotheses that always hold in this case) is a genus g handlebody if and only if its fundamental group is free. The genus g handlebody deformation retracts onto a wedges of g circles, which is a graph. These are the only 3-manifolds homotopy equivalent to graphs because the fundamental group of a graph is free.[1]

The cubic lattice graph is clearly unknotted by this definition, and it is dual to itself. A large class of examples is given by the one-skeletons of three-dimensional regular cell complexes. In this case, the complement is homotopy equivalent to the adjacency graph of the three-cells of the complex.

7.3.1 Computation

Given a presentation of the fundamental group of a graph complement, one needs to determine whether it is free in order to ascertain whether the graph is unknotted. This is a very difficult problem in theory; in fact, whether a finitely presented group is free is not recursively recognizable.

[1] I would like to thank David Futer and Herman Gluck for the reference on this result.

That is, there is no one algorithm that will provide a correct answer in all cases. [49] In practice, however, it may not be impossible. By computing the fundamental group of the complement, the program SnapPea correctly identified the unknotted circle in many (sometimes complicated) cases in spite of the same theoretical difficulties [56].

7.4 Conjecture

Conjecture 7. *Suppose Ω is the dynamical process on embedded graphs in $\mathbb{R}^3$ in which the edges evolve by curvature-flow and the vertices move by any of the rules defined in Section 2.3. For an embedded graph G, let $G(t)$ denote the evolution of G under Ω. For almost all infinite, trivalent graphs G embedded in $\mathbb{R}^3$ there exists a $T > 0$ such that $G(t)$ is unknotted with probability one for all $t > T$.*

In the future, we hope to study this conjecture using computational and theoretical methods. For now, we will present some heuristic evidence.

7.5 Knotting of Shortest Loops

The shortest loops of an embedded graph are the smallest loops containing each edge of the graph with respect to the number of edges or the total length in the Euclidean distance, as defined in the previous chapter. One might expect that the knottedness of these loops would correlate with the knottedness of the graph. It turns out that the presence of knotted shortest loops is neither a necessary not a sufficient condition for a graph to be knotted. This can be seen from the cubic lattice graph in Figure 7.4. Sufficiently subdividing all the edges but those in the blue trefoil makes it a knotted shortest loop with respect to the number of edges, and contracting only its edges yields a knotted shortest loop with respect to length. In the other direction, a knotted graph none of whose edges are knotted can be used to produce a knotted graph with no knotted shortest loops by replacing each edge with an unknotted digon. Still, the knotting of shortest loops is a useful heuristic in cases where the edge length distribution is "homogeneous", such as that of curvature

flow on embedded graphs.

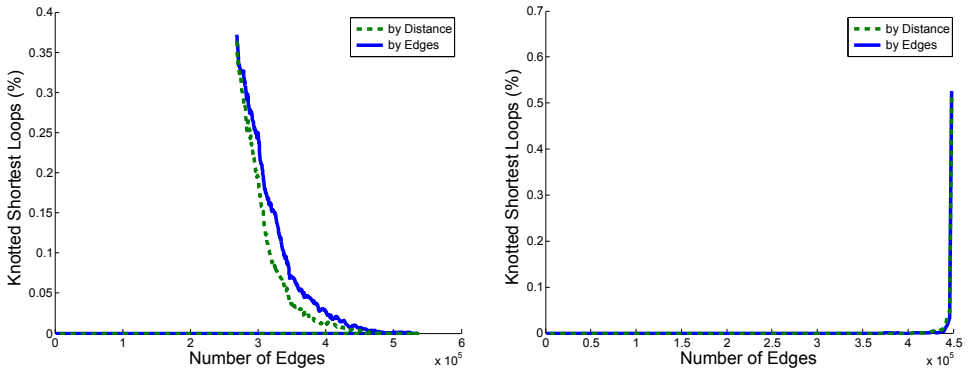
7.5.1 Computational Methods

In this section, we examine how the percentage of knotted shortest loops changes as a graph evolves by curvature flow. We start with random graph initial conditions, which were designed to be heavily knotted. We compute shortest loops of an embeddded graph at a series of timesteps as it evolves, and determine whether these loops are knotted. First, we apply the Fary-Milnor Theorem which states that if a closed curve has total curvature less than 4Π than it must be the unknot. For smooth curves, the total curvature is the integral of its curvature, but for the discretized curves coming from the computer simulation it is simply the sum of the exterior angles. [50] For most shortest loops in the case at hand, the total curvature is less than 4Π so they are unknotted.

The knottedness of the remaining loops is determined using the Jones Polynomial. The Jones Polynomial of a knot is a Laurent polynomial with integer coefficients that is invariant under isotopy. For a definition of this polynomial, refer to a standard text on knot theory such as Lickorish [38]. Here, all we require is that the Jones Polynomial of the unknot is equal to one, and that it is the only knot with a diagram with fewer than 18 crossings with this property [13]. It is unknown whether the Jones Polynomial "detects the unknot" - that is, that the only knot with trivial Jones Polynomial is the unknot. Fortunately, all of the knots encountered in these computations had fewer than 18 crossings. In other applications, it may be necessary to use other knot invariants which are known to detect the unknot, such as Khovanov homology [34]. We used the Mathematica KnotTheory‘ package to compute Jones Polynomials [3].

7.5.2 Computational Results

Figure 7.5a shows the percentage of shortest loops as time proceeds for an graph embedded in three dimensions evolving by curvature flow. The initial condition was a random graph with about 270,000 edges, and new edges were added at each intersection event. As before, the evolution is parametrized by the number of edges in the system which increases initially before decreasing. The number of knotted loops decreases very rapidly whether the shortest loops are computed based on



(a) (b)

Figure 7.5: The percentage of knotted shortest loops in embedded graphs evolving under curvature flow for which (a) a new edge is created at every intersection and (b) intersections are ignored. The evolution of the system is parametrized by the number of edges, as before. Shortest loops are computed either in terms of the number of edges in a loop (solid blue) or the geometric length of the loop (dashed green). In all cases, the percentage of knotted shortest loops falls very quickly.

the number of edges in a loop (plotted in solid blue) or the Euclidean length of the loop (plotted in dashed green). Thus, curvature flow unknots the shortest loops in an embedded graph. The proportion of knotted loops decreases even more quickly if intersections are ignored, as shown in Figure 7.5b.

Bibliography

[1] Cgal, computational geometry algorithms library. http://www.cgal.org.

[2] Curvature flows and grayson's theorem. *http* : *//smp.uq.edu.au/sites/smp.uq.edu.au/files/GeometricFlowsBook_C*

[3] Mathematica knottheory' package. http://katlas.org/wiki/KnotTheory.

[4] D. A. Aboav. The arrangement of cells in a net. *Metallography*, 13(1):43–58, 1980.

[5] Michael Artin. *Algebra*. Pearson, 1991.

[6] Alexander G. Belyaev. A note on invariant three-point curvature approximations. *Surikaisekikenkyusho Kokyuroku (Singularity Theory and Differential Equations)*, 1999.

[7] I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electronic Journal of Probability*, 6:1–13, 2001.

[8] OpenMP Architecture Review Board. Openmp 3.1, July 2011. http://www.openmp.org/mp-documents/OpenMP3.1.pdf.

[9] K. Brakke. VOR3DSIM. Personal Communication.

[10] David C. Brydges and John Z. Imbrie. Branched polymers and dimensional reduction. *Annals of Mathematics*, 158:1019–1039, 2003.

[11] Toen Castle, Myfanwy E. Evansa, and S. T. Hyde. Ravels: knot-free but not free. novel entanglements of graphs in 3-space. *New Journal of Chemistry*, 2008.

[12] Justin Curry. *Sheaves, Cosheaves, and Applications.* PhD thesis, University of Pennsylvania, 2014.

[13] Oliver T. Dasbach and Stefan Hougardy. Does the jones polynomial detect unknottedness? *Experimental Mathematics*, 1997.

[14] Yuanan Diao. Minimal knotted polygons on the cubic lattice. *Journal of Knot Theory and Its Ramifications*, 1993.

[15] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplificitaion. *Discrete and Computational Geometry*, 2002.

[16] H. Edelsbrunner and E. P. Mücke. Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13(1):43–72, 1994.

[17] Joel Foisy. Intrinsically knotted graphs. *Journal of Graph Theory*, 2001.

[18] G Gottstein, A H King, and L S Shvindlerman. The effect of triple-junction drag on grain growth. *Acta materialia*, 48(2):397–403, 2000.

[19] G Gottstein and L S Shvindlerman. Triple junction drag and grain growth in 2d polycrystals. *Acta materialia*, 50(4):703–713, 2002.

[20] Matthew A. Grayson. The heat equation shrinks embedded plane curves to round points. *Journal of Differential Geometry*, 1987.

[21] D. J. Green. Image of an open-cell ceramic foam. http://www.ems.psu.edu/ green/564-535.html.

[22] Allen Hatcher. *Algebraic Topology.* Cambridge University Press, 2002.

[23] Reiner Henseler, Barbara Niethammer, and Felix Otto. *Free Boundary Problems*, chapter A Reduced Model for Simulating Grain Growth, pages 177–187. Birkhuser Basel, 2003.

[24] John P Hirth and Jens Lothe. *Theory of dislocations.* John Wiley & Sons, 1982.

[25] Han Hong and Chee Yap, editors. *JavaPlex: A research software package for persistent (co)homology*, Lecture Notes in Computer Science 8592, 2014. http://appliedtopology.github.io/javaplex/.

[26] S. T. Hyde and O. D. Friedrichs. From untangled nets and graphs to tangled materials. *Solid State Physics*, 2010.

[27] Jdrewitt. The amorphous structure of glassy silica in two-dimensions (figure), May 2008. http://commons.wikimedia.org/wiki/File:Silica.svg.

[28] Jkasd. Knot table.svg. Wikimedia Commons. https://commons.wikimedia.org/wiki/File

[29] Kevin R. Johnson. A dla cluster grown from a copper sulfate solution in an electrodeposition cell. Licensed under CC BY 2.5 via Wikimedia Commons. *https* : *//commons.wikimedia.org/wiki/File* : *DLA_Cluster.JPG*.

[30] Tom Kennedy. Code for simulating self-avoiding walks.

[31] Tom Kennedy. A faster implementation of the pivot algorithm for self-avoiding walks. *J. Statistical Physics*, 106:407–429, 2002.

[32] Richard Kenyon and Peter Winkler. Branched polymers. *American Mathematical Monthly*, 116(7):612–628, 2009.

[33] David Kinderlehrer and Chun and Liu. Evolution of grain boundaries. *Mathematical Models & Methods in Applied Sciences*, 2001.

[34] P. B. Kronheimer and T. S. Mrowka. Khovanov homology is an unknot-detector. *Publications mathmatiques de l'IHS*, 2011.

[35] E. Lazar, J. K. Mason, Robert D. MacPherson, and D. Srolovitz. A more accurate three-dimensional grain growth algorithm. *Acta Materialia*, 2011.

[36] E.A. Lazar, R.D. MacPherson, and D.J. Srolovitz. A more accurate two-dimensional grain growth algorithm. *Acta Materialia*, 58(2):364–372, January 2010.

[37] Menachem Lazar. *The Evolution of Cellular Structures via Curvature Flow*. PhD thesis, Princeton University, 2012.

[38] W. B. Raymond Lickorish. *An Introduction to Knot Theory*. Springer, 1997.

[39] L. Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[40] R. D. MacPherson and B. Schweinhart. Measuring shape with topology. *Journal of Mathematical Physics*, 53(7):073516, 2012. http://arxiv.org/pdf/1011.2258v2.pdf.

[41] B.B. Mandelbrot, H. Kaufman, A. Vespignani, I. Yekutieli, and C.-H. Lam. Deviations from self-similarity in plane dla and the infinite drift scenario. *Europhysics Letters*, 29(8):599–604, 1995.

[42] B.B. Mandelbrot, A. Vespignani, and H. Kaufman. Crosscut analysis of large radial dla: Departures from self-similarity and lacunarity effects. *Europhysics Letters*, 32(3):199–204, 1995.

[43] J K Mason, R Ehrenborg, and E A Lazar. A geometric formulation of the law of aboavweaire in two and three dimensions. *Journal of Physics A*, 2012.

[44] J. K. Mason, E. A. Lazar, R. D. MacPherson, and D. J. Srolovitz. Statistical topology of cellular networks in two and three dimensions. *Physical Review E*, 86(5):051128, 2012.

[45] Jeremy Mason. 2d grain growth simulation. Personal Communication.

[46] Mark McConnell. Minimal homology basis program. Personal communication.

[47] Brendan D McKay and Adolfo Piperno. Practical graph isomorphism, ii. *Journal of Symbolic Computation*, 60:94–112, 2014.

[48] P. McMullen. The maximum number of faces of a convex polytope. *Mathematika*, 1970.

[49] C. F. Miller. Decision problems for groups - survey and reflections. In *Algorithms and Classification in Combinatorial Group Theory*. Springer-Verlag, 1992.

[50] J. W. Milnor. On the total curvature of knots. *Annals of Mathematics*, 1950.

[51] Jesper M. Møller. From singular chains to alexander duality. http://www.math.ku.dk/ moller/f03/algtop/notes/homology.pdf.

[52] G. Monge. Mémoire sur la théorie des déblais et remblais, mémoires acad. *Royale Sci. Paris*, 3, 1781.

[53] Dmitriy Morozov. Dionysus. http://www.mrzv.org/software/dionysus/.

[54] W.W. Mullins. Two-Dimensional Motion of Idealized Grain Boundaries. *Journal of Applied Physics*, 27:900–904, March 1956.

[55] Vidit Nanda. Perseus, the persistent homology software. http://www.sas.upenn.edu/ vnanda/perseus/index.html.

[56] Sam Nead. can you fool snappea? MathOverflow, November 2009. URL:http://mathoverflow.net/q/4948 (version: 2009-11-11).

[57] Bernard Nienhuis. Exact critical point and critical exponents of o(n) models in 2 dimensions. *Phys. Rev. Lett.*, 49:1062, 1982.

[58] Olivier Devillers. Nina Amenta, Dominique Attali. Complexity of delaunay triangulation for points on lower-dimensional polyhedra. Technical report, INRIA, 2006.

[59] Edward Pleshakov. Micrograph of a polycrystalline metal; grain boundaries evidenced by acid etching., March 2008. $http://en.wikipedia.org/wiki/Grain_boundary/media/File:CrystalGrain.jpg$.

[60] Henri Poincare. Analysis situs. *Journal de l'cole Polytechnique*, 1895.

[61] Sebastein Roy and Philippe Jund. Ring statistics analysis of topological networks: New approach and application to amorphous ges2 and sio2 systems. *Computational Materials Science*, 2010.

[62] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *Computer Vision, 1998. Sixth International Conference on*, pages 59–66. IEEE, 1998.

[63] M. Scharlemann and A. Thompson. Detecting unknotted graphs in 3-space. *Journal of Differential Geometry*, 1991.

[64] B. Schweinhart, J. K. Mason, and R. D. MacPherson. Topological similarity of random cell complexes, and applications to dislocation configurations. 2014. http://arxiv.org/pdf/1407.6989v1.pdf.

[65] Harlan Sexton and Mikael Vejdemo-Johansson. jplex, December 2008. http://www.math.duke.edu/ hadams/jplex/index.html.

[66] Volker Springel, Simon D. M. White, Adrian Jenkins, Carlos S. Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, John A. Peacock, Shaun Cole, Peter Thomas, Hugh Couchman, August Evrard, Joerg Colberg, and Frazer Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 2005. http://www.mpa-garching.mpg.de/galform/virgo/millennium/.

[67] Mark J. Stock. Dla-nd.

[68] J. von Neumann. Written Discussion. In *Metal Interfaces*, pages 108–110, Cleveland, Ohio, 1952. American Society for Metals.

[69] Friedhelm Waldhausen. On irreducible 3-manifolds which are sufficiently large. *Annals of Mathematics*, 1968.

[70] D. Weaire. Some remarks on the arrangement of grains in a polycrystal. *Metallography*, 7(2):157–160, 1974.

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