

Fractal Dimension Estimation with Persistent Homology

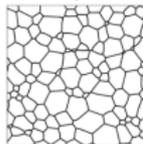
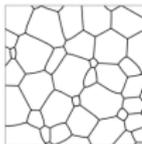
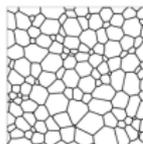
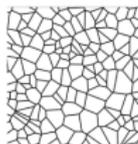
Benjamin Schweinhart

October 31, 2019

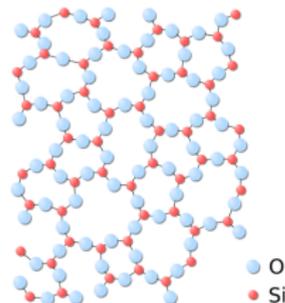
Overview



Fractal Dimension and
Persistent Homology



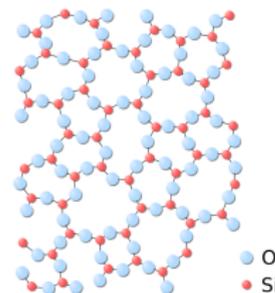
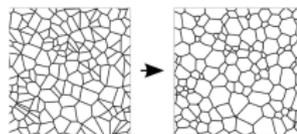
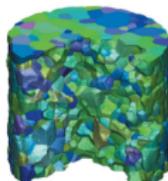
Cellular Structures
and Grain Growth



Local Structure of
Bond Networks

I study the topology and geometry of complex geometric objects. Broadly speaking, this falls under the purview of topological data analysis. I'm interested in both theory and applications, especially to materials science.

Topology and Geometry of Complex Geometric Objects



Fractal Dimension and
Persistent Homology

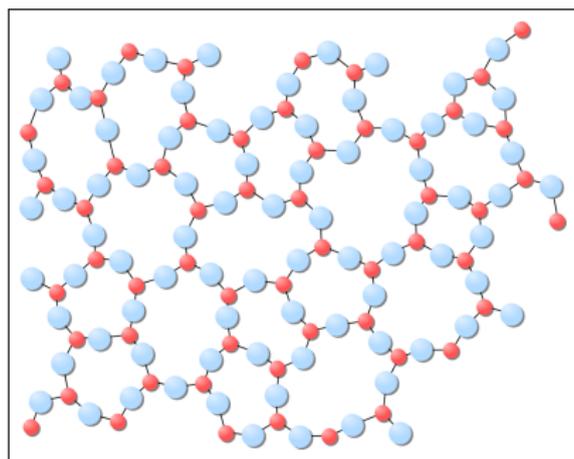
Cellular Structures
and Grain Growth

Local Structure of
Bond Networks

Presentation Outline

- 1 Brief Aside: The Structure of Glass
- 2 Background: Fractal Dimension and Persistent Homology
- 3 Previous Work and Definitions
- 4 Results for Minimum Spanning Trees (PH_0)
- 5 Computational Results
- 6 Proof Sketch

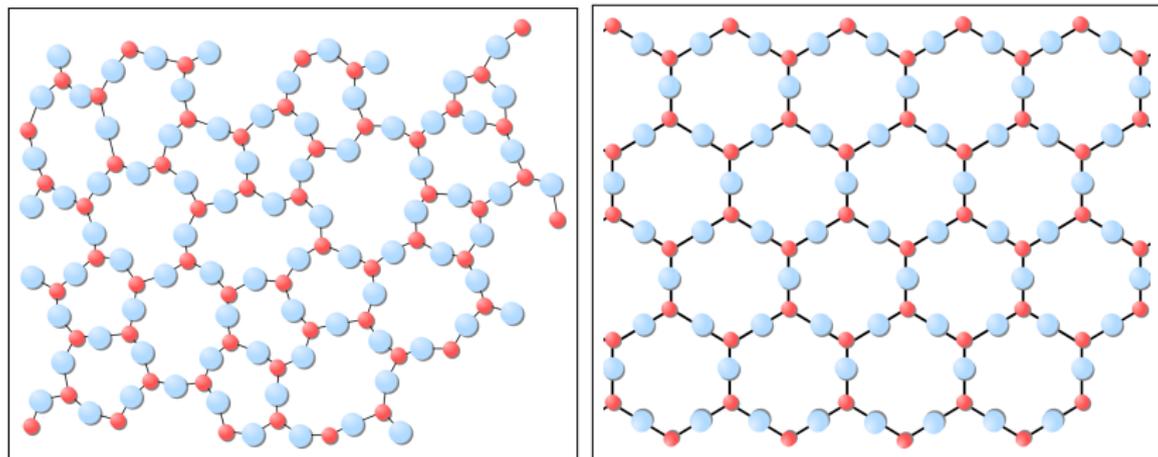
Oxide Glasses



Oxide glasses are present in our daily lives, but the relationship between the local structure and global physical properties of these materials is poorly understood. This is due in part to the lack of an appropriate language to describe that local structure.

B. Schweinhart, D. Rodney, and J.K. Mason, *Statistical Topology of Bond Networks with Applications to Silica* (2019).

Silica



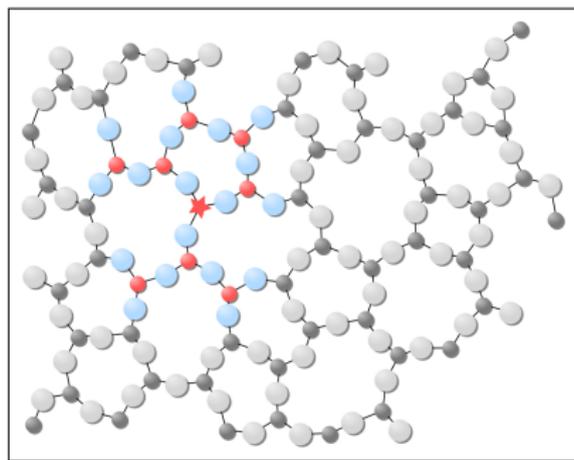
The bond network of silicon dioxide (silica) is a bipartite graph, where silicon atoms are (usually) adjacent to four oxygen atoms, and oxygen atoms are (usually) adjacent to two silicon atoms.

- Current goal: develop a rigorous methodology to classify local atomic environments appearing in oxide glass. It should differentiate glasses produced at different cooling rates, as well as different crystalline forms of SiO_2 .
- Future goal: relate global physical properties of glasses to local structure. One possible application: what local atomic environments are associated with crystal nucleation?
- The methodology is also applicable to other random graph models with relatively constrained degree distribution.

- Current goal: develop a rigorous methodology to classify local atomic environments appearing in oxide glass. It should differentiate glasses produced at different cooling rates, as well as different crystalline forms of SiO_2 .
- Future goal: relate global physical properties of glasses to local structure. One possible application: what local atomic environments are associated with crystal nucleation?
- The methodology is also applicable to other random graph models with relatively constrained degree distribution.

- Current goal: develop a rigorous methodology to classify local atomic environments appearing in oxide glass. It should differentiate glasses produced at different cooling rates, as well as different crystalline forms of SiO_2 .
- Future goal: relate global physical properties of glasses to local structure. One possible application: what local atomic environments are associated with crystal nucleation?
- The methodology is also applicable to other random graph models with relatively constrained degree distribution.

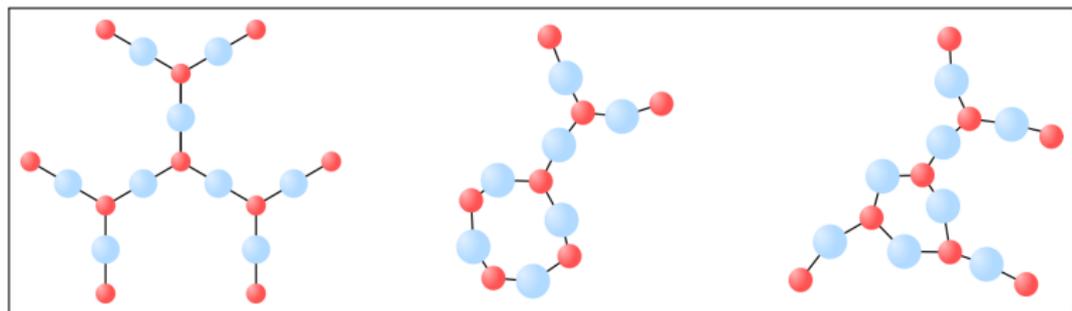
Swatches: Local Atomic Environments



If G is the bond network of an atomistic material, and v is an atom in G the **swatch** or **local atomic environment** of radius r centered at v is the ball of radius r in the graph distance.¹

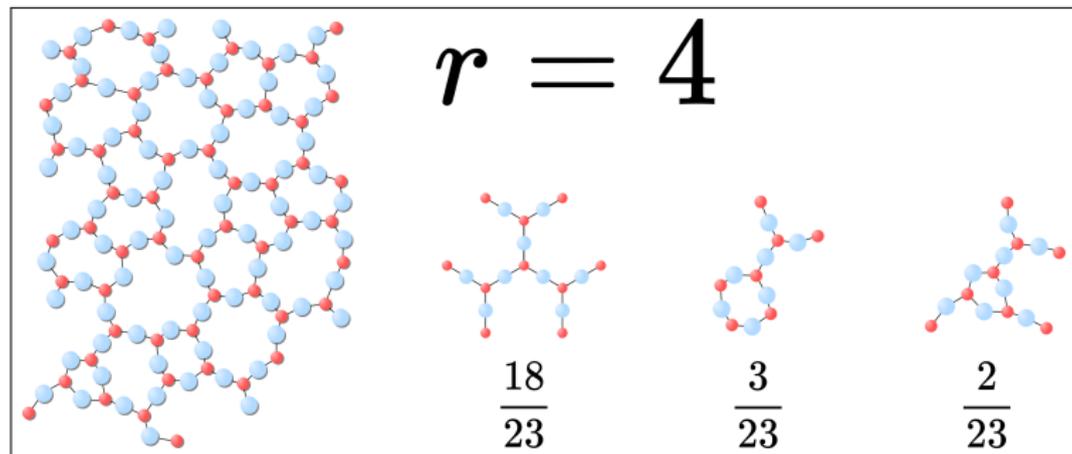
¹B. Schweinhart, J.K. Mason, and R.D. MacPherson, *Topological Similarity of Random Cell Complexes and Applications*, Physical Review E 93 (2016).

(In)equivalent Environments



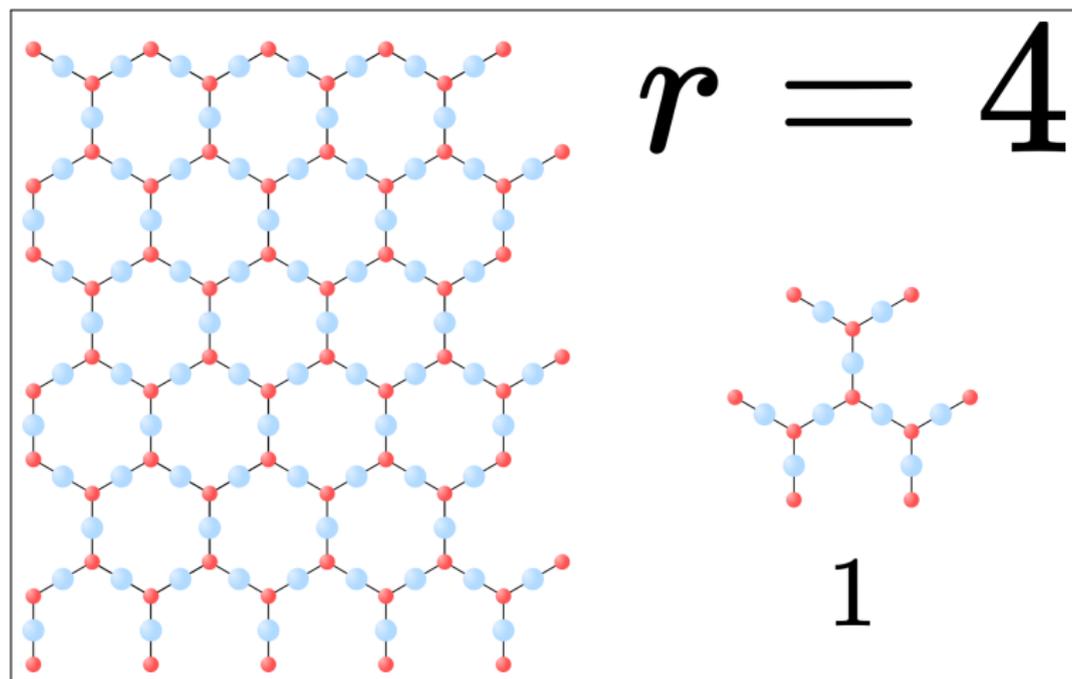
Two switches are equivalent if they are isomorphic as rooted, colored graphs.

The Cloth



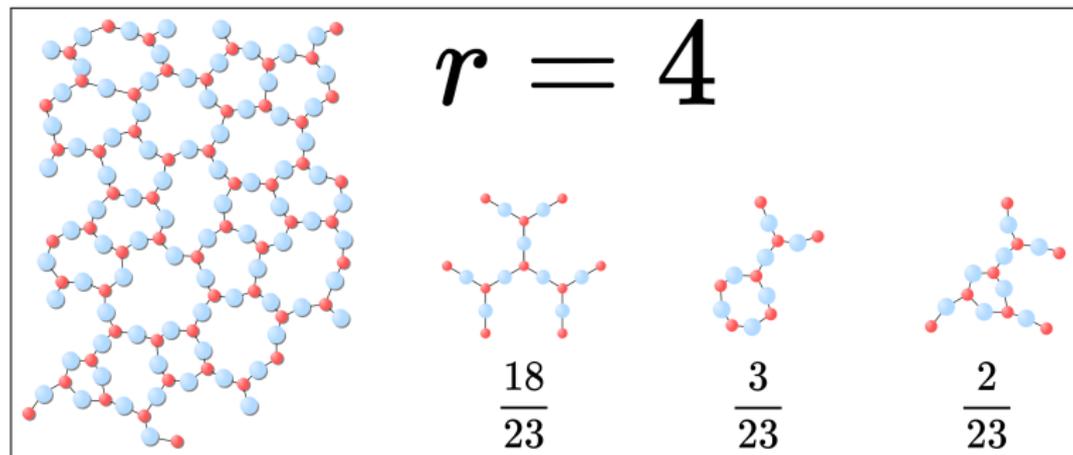
Given a graph G and a radius r , the counting measure on the vertices of G induces an empirical probability distribution of swatch types of radius r . This family of probability distributions is called the **cloth** of G . It characterizes the local topology of G .

Crystal Cloth



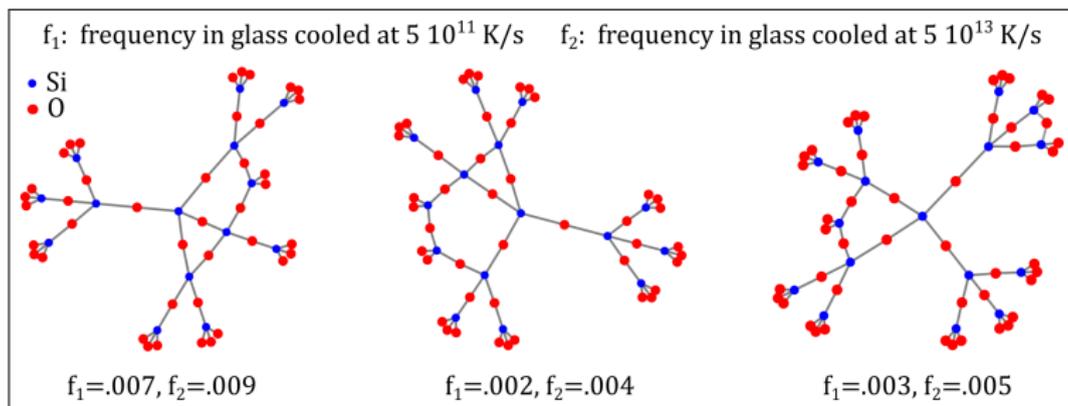
For many crystalline materials, the cloth is supported on one or two topological types.

Disordered Cloth



For disordered materials, the notion of a single “unit cell” is replaced by a probability distribution of local configurations.

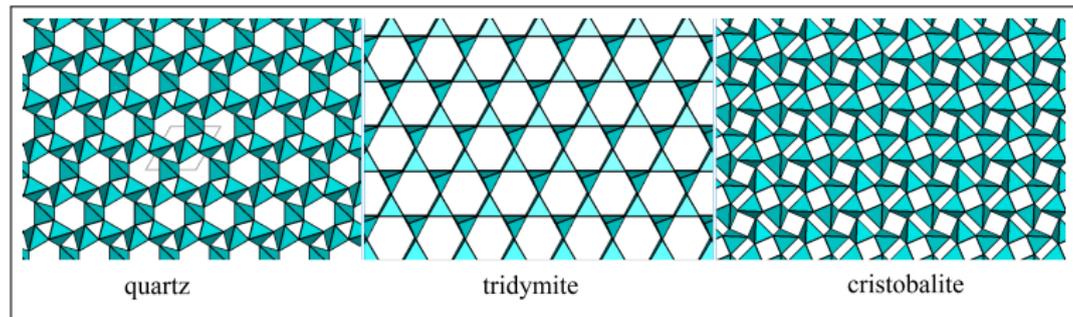
Application: Different Cooling Rates



This methodology distinguishes molecular dynamics simulations of glasses produced at different cooling rates. We can both see an overall pattern (faster cooling rate yields more “disorder”), and identify specific environments over-represented in the different preparations.²

²B. Schweinhart, D. Rodney, and J.K. Mason, *Statistical Topology of Bond Networks, with Applications to Silica*.

Crystalline SiO_2 and Radius 6



Proposed application: detect local atomic environments related to crystal nucleation.

Crystalline forms of SiO_2 such as quartz and cristobalite are indistinguishable below radius $r = 6$.

The Combinatorial Explosion

R	# of Classes per Atom
4	0.01
5	0.07
6	0.90
7	1.00

Number of graph isomorphism classes detected in a sample of 10^6 radius 6 atomic environments approaches $10^6!$ A different approach is needed.

In Statistical Topology of Bond Networks with Applications to Silica, we study notions of equivalence for local atomic environments that are coarser than graph isomorphism.

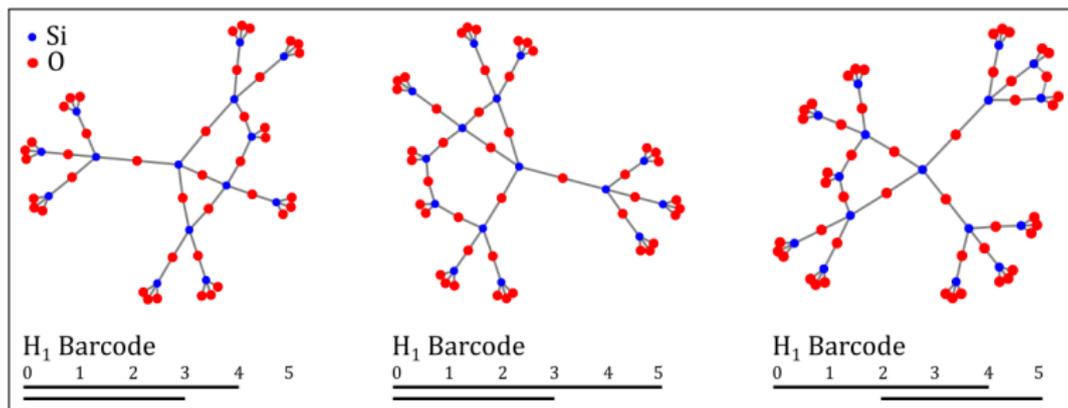
The Combinatorial Explosion

R	# of Classes per Atom
4	0.01
5	0.07
6	0.90
7	1.00

Number of graph isomorphism classes detected in a sample of 10^6 radius 6 atomic environments approaches $10^6!$ A different approach is needed.

In *Statistical Topology of Bond Networks with Applications to Silica*, we study notions of equivalence for local atomic environments that are coarser than graph isomorphism.

The H_1 Barcode

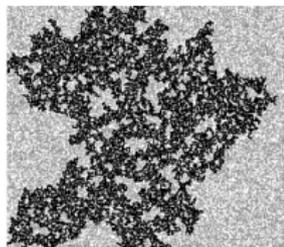


The H_1 barcode encodes information about algebraically independent rings in the local environment, and their distance from the root atom. At radius 6, it distinguishes glasses produced at different cooling rates as well as different crystalline forms of silica.

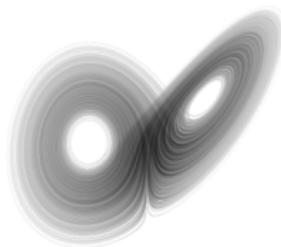
Presentation Outline

- 1 Brief Aside: The Structure of Glass
- 2 Background: Fractal Dimension and Persistent Homology
- 3 Previous Work and Definitions
- 4 Results for Minimum Spanning Trees (PH_0)
- 5 Computational Results
- 6 Proof Sketch

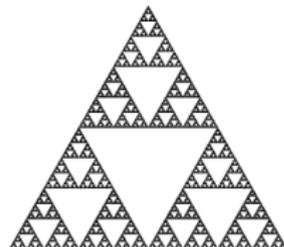
Fractal Dimension



Critical Percolation
 $d_H = 7/4$ (?)



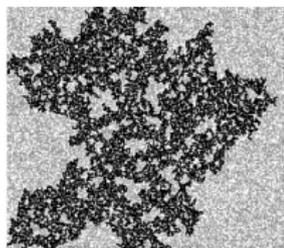
Lorenz Attractor
 $d_H \approx 2.06$



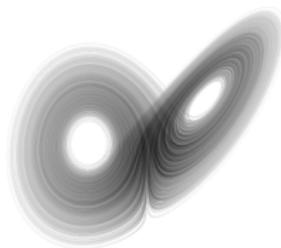
Sierpinski Triangle
 $d_H = \log_2(3)$

- Fractal dimension measures how the properties of a shape depend on scale.
- The first notion of a fractional dimension was proposed by Hausdorff in 1918. Since then, several other definitions have been proposed, including the box-counting, packing, and correlation dimensions.
- These dimensions agree on a wide class of “regular” sets.

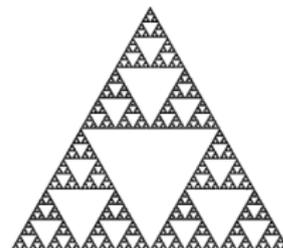
Fractal Dimension



Critical Percolation
 $d_H = 7/4$ (?)



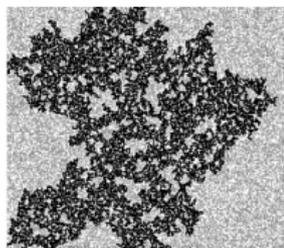
Lorenz Attractor
 $d_H \approx 2.06$



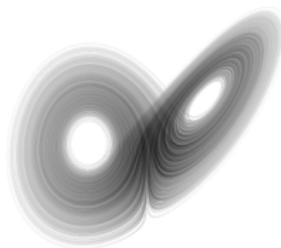
Sierpinski Triangle
 $d_H = \log_2(3)$

- Fractal dimension measures how the properties of a shape depend on scale.
- The first notion of a fractional dimension was proposed by Hausdorff in 1918. Since then, several other definitions have been proposed, including the box-counting, packing, and correlation dimensions.
- These dimensions agree on a wide class of “regular” sets.

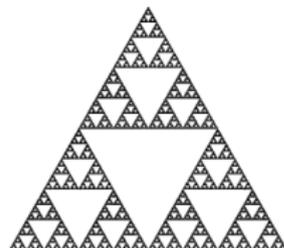
Fractal Dimension



Critical Percolation
 $d_H = 7/4$ (?)



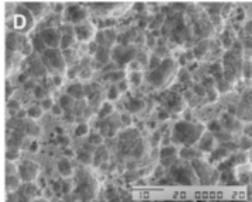
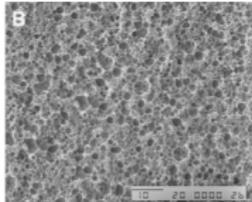
Lorenz Attractor
 $d_H \approx 2.06$



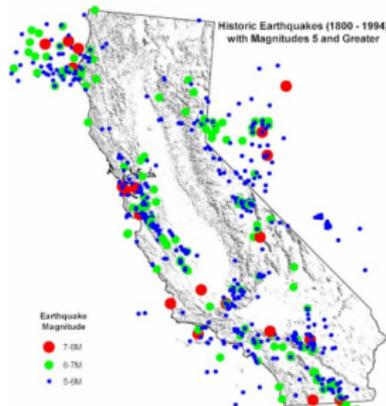
Sierpinski Triangle
 $d_H = \log_2(3)$

- Fractal dimension measures how the properties of a shape depend on scale.
- The first notion of a fractional dimension was proposed by Hausdorff in 1918. Since then, several other definitions have been proposed, including the box-counting, packing, and correlation dimensions.
- These dimensions agree on a wide class of “regular” sets.

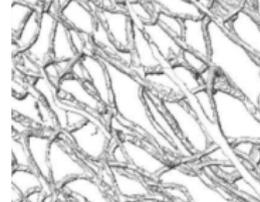
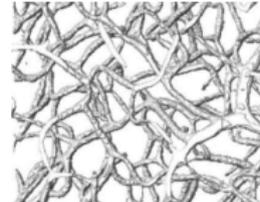
Fractal Applications



Aluminum Oxide¹



Earthquakes²



Vascular Networks³

Fractal dimension has applications in a wide variety of fields including medicine, ecology, materials science, and the analysis of large data sets. In some of these applications, it is important to estimate fractal dimension from random point samples.

¹Riscovič and Pavlovič, Scanning (2013), ²Kagan, Geophysical Journal International (2007), ³Lorhois and Cassot, Journal of Theoretical Biology (2010)

Filtrations



$$\epsilon = 0$$



$$\epsilon = \frac{1}{32\sqrt{3}}$$



$$\epsilon = \frac{1}{16\sqrt{3}}$$



$$\epsilon = \frac{1}{8\sqrt{3}}$$



$$\epsilon = \frac{1}{4\sqrt{3}}$$

A **filtration** of topological spaces is family $\{X_\alpha\}_{\alpha \in I}$ with ordered index set I , together with inclusions $i_{\alpha,\beta} : X_\alpha \hookrightarrow X_\beta$ for $\alpha < \beta$.

Example: if $S \subset \mathbb{R}^2$ we have the filtration of ϵ -neighborhoods

$\{S_\epsilon\}_{\epsilon \in \mathbb{R}^+}$.

Filtrations



$$\epsilon = 0$$



$$\epsilon = \frac{1}{32\sqrt{3}}$$



$$\epsilon = \frac{1}{16\sqrt{3}}$$



$$\epsilon = \frac{1}{8\sqrt{3}}$$



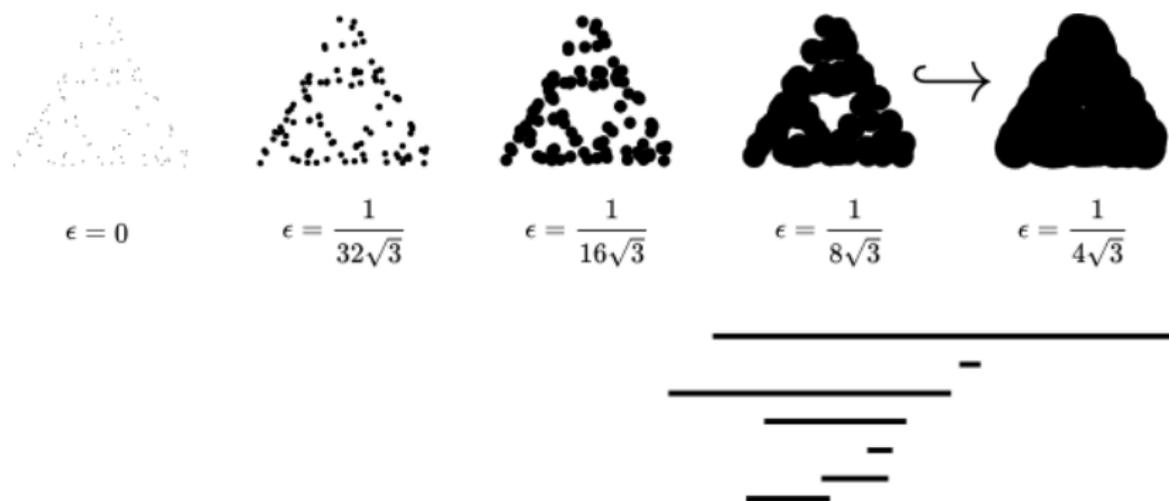
$$\epsilon = \frac{1}{4\sqrt{3}}$$

A **filtration** of topological spaces is family $\{X_\alpha\}_{\alpha \in I}$ with ordered index set I , together with inclusions $i_{\alpha,\beta} : X_\alpha \hookrightarrow X_\beta$ for $\alpha < \beta$.

Example: if $S \subset \mathbb{R}^2$ we have the filtration of ϵ -neighborhoods

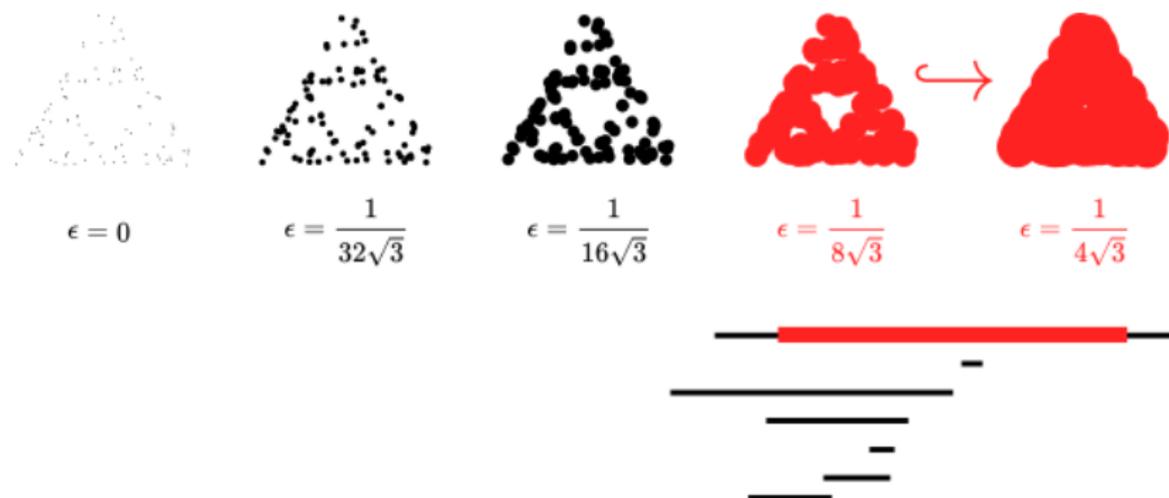
$\{S_\epsilon\}_{\epsilon \in \mathbb{R}^+}$.

Persistent Homology



Persistent Homology (*PH*) tracks how the homology changes through a filtration. PH_i is a set of intervals corresponding to homology generators that are born and die in this process.

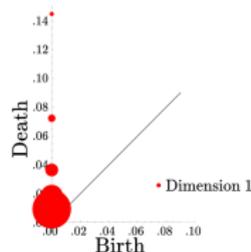
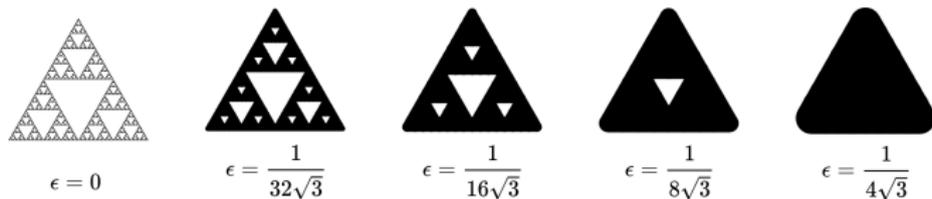
Definition of Persistent Homology



Given a filtration $\{X_\alpha\}$, $PH_i(X_\alpha)$ is the unique set of intervals so that

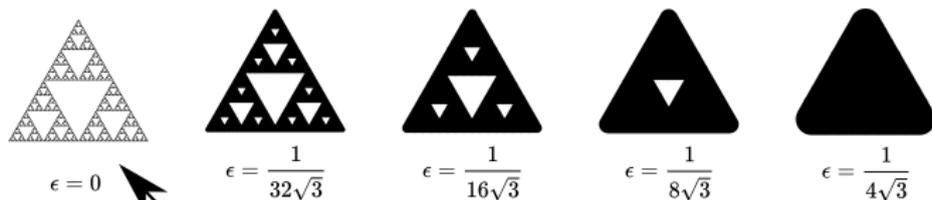
$$\text{rank}(i_{\alpha,\beta} : H_i(X_\alpha) \rightarrow H_i(X_\beta)) = \# \{I \in PH_i(X_\alpha) : [\alpha, \beta] \subseteq I\} .$$

Persistent Homology



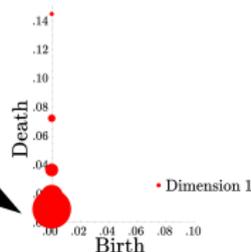
The information in PH is often summarized by a persistence diagram: a scatter plot of (birth, death) for each interval.

Persistent Homology



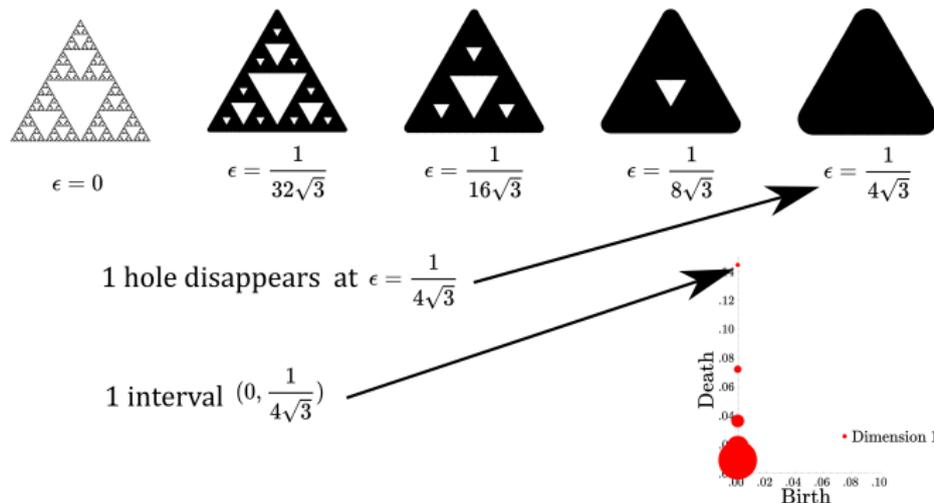
Infinitely many holes at $\epsilon = 0$

Infinitely many intervals with Birth=0



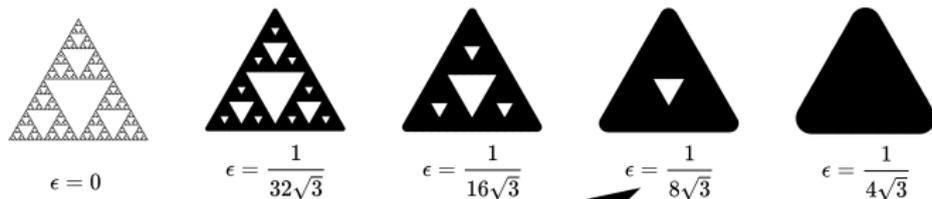
$PH_1(S_\epsilon)$ is a set of intervals, one for each component of the complement that disappears as ϵ increases (by Alexander duality).

Persistent Homology



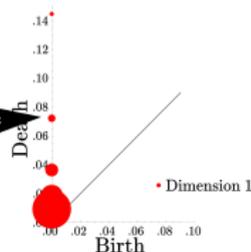
$PH_1(S_\epsilon)$ is a set of intervals, one for each component of the complement that disappears as ϵ increases (by Alexander duality).

Persistent Homology



3 holes disappear at $\epsilon = \frac{1}{8\sqrt{3}}$

3 intervals $(0, \frac{1}{8\sqrt{3}})$



$P\mathcal{H}_1(S_\epsilon)$ has one interval for each bounded component of the complement of S (by Alexander duality).

Persistent Homology of a Sample



$$\epsilon = 0$$



$$\epsilon = \frac{1}{32\sqrt{3}}$$



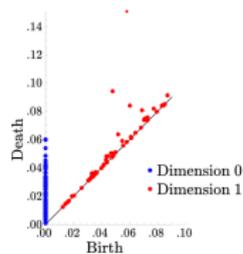
$$\epsilon = \frac{1}{16\sqrt{3}}$$



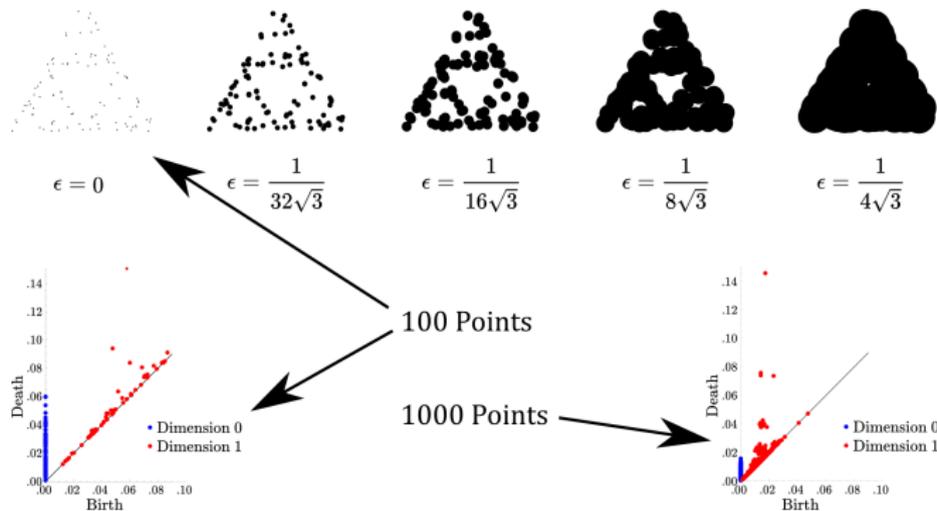
$$\epsilon = \frac{1}{8\sqrt{3}}$$



$$\epsilon = \frac{1}{4\sqrt{3}}$$

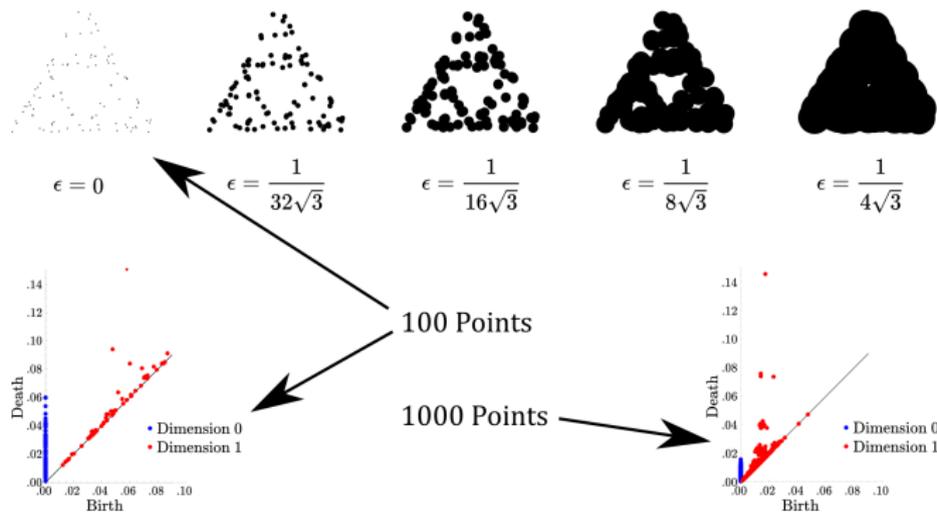


Persistent Homology of a Sample



If we take the persistent homology of larger and larger samples, the diagram begins to approach that of the support.

Persistent Homology of a Sample



We also have a cluster of small intervals that are usually written off as “noise.” We can use this noise to estimate fractal dimension!

Minimum Spanning Trees



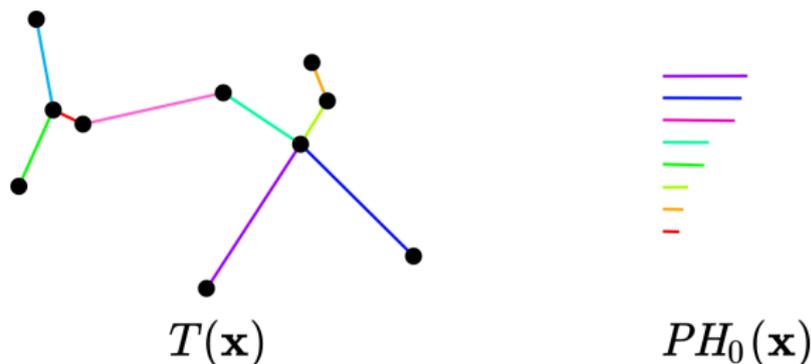
Definition (Minimum Spanning Tree)

Let \mathbf{x} be a finite metric space. The **minimum spanning tree** on \mathbf{x} , denoted $T(\mathbf{x})$ is the connected graph with vertex set \mathbf{x} that minimizes the sum of the length of the edges.

In fact, for any $\alpha > 0$, $T(\mathbf{x})$ minimizes the weighted sum

$$\sum_{e \in T(\mathbf{x})} |e|^\alpha .$$

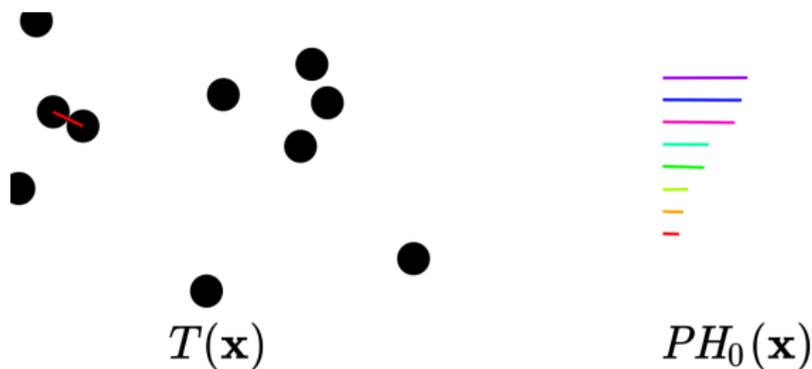
Minimum Spanning Trees and PH_0



Observation: If \mathbf{x} is a finite metric space, then there is a bijection between the edges of $T(\mathbf{x})$ and the intervals of $PH_0(\mathbf{x})$. An edge corresponds to an interval of half the length.³

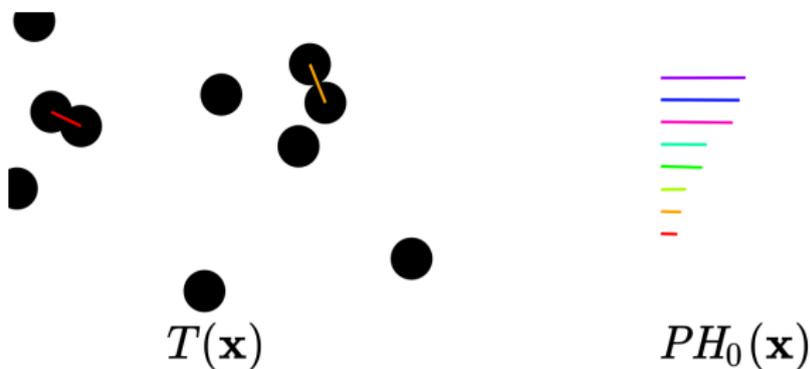
³Note: this depends on whether persistent homology is taken with respect to the Rips or Čech complex.

Minimum Spanning Trees and PH_0



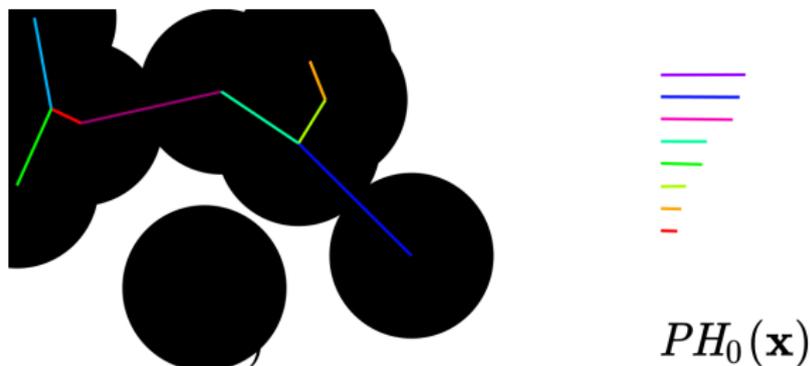
The proof follows Kruskal's algorithm to compute the MST: expand balls until two from different components overlap.

Minimum Spanning Trees and PH_0



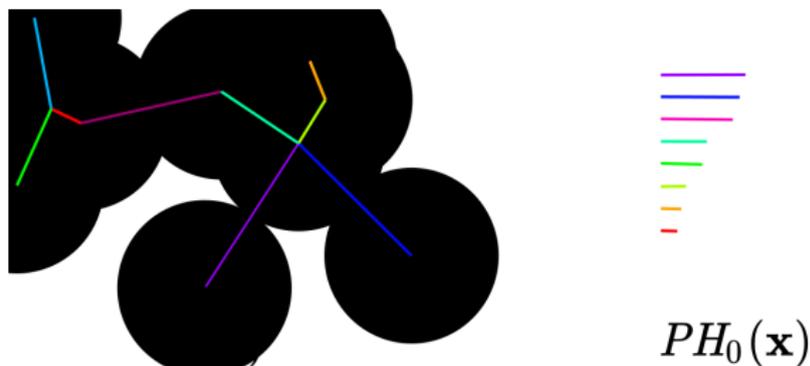
The proof follows Kruskal's algorithm to compute the MST: expand balls until two from different components overlap.

Minimum Spanning Trees and PH_0



The proof follows Kruskal's algorithm to compute the MST:
expand balls until two from different components overlap.

Minimum Spanning Trees and PH_0



The proof follows Kruskal's algorithm to compute the MST: expand balls until two from different components overlap.

Main Questions

Question

Can the fractal dimension of a metric measure space be estimated from the persistent homology of random point samples?

“Fractal Dimension and the Persistent Homology of Random Geometric Complexes,” (2018).

Question

How does the practical performance of the PH_i -dimension compare to classical methods such as box-counting or the correlation algorithm?

“Fractal Dimension Estimation with Persistent Homology: A Comparative Study,” with J. Jaquette (2019).

Main Questions

Question

Can the fractal dimension of a metric measure space be estimated from the persistent homology of random point samples?

“Fractal Dimension and the Persistent Homology of Random Geometric Complexes,” (2018).

Question

How does the practical performance of the PH_i -dimension compare to classical methods such as box-counting or the correlation algorithm?

“Fractal Dimension Estimation with Persistent Homology: A Comparative Study,” with J. Jaquette (2019).

Presentation Outline

- 1 Brief Aside: The Structure of Glass
- 2 Background: Fractal Dimension and Persistent Homology
- 3 Previous Work and Definitions**
- 4 Results for Minimum Spanning Trees (PH_0)
- 5 Computational Results
- 6 Proof Sketch

Previous Work on Fractal Dimension and PH

Several authors have defined fractal dimensions based on PH , and compared computational estimates with known dimensions:

- Robins (PhD thesis, 2000): Persistent Betti numbers of fractals, proved results for H_0 of disconnected sets.
- MacPherson and S. (“Measuring Shape with Topology,” 2012): PH of shapes, studied probability distributions of polymers.
- Adams et. al. (“A Fractal Dimension for Measures via PH ,” 2019): PH of random point samples; definition very similar to the one here. Computational experiments that motivated the current work.
- S. (“ PH and the Upper Box Dimension,” 2018): PH of extremal point sets. First rigorous results relating PH to a classically defined fractal dimension.

Previous Work on Fractal Dimension and PH

Several authors have defined fractal dimensions based on PH , and compared computational estimates with known dimensions:

- Robins (PhD thesis, 2000): Persistent Betti numbers of fractals, proved results for H_0 of disconnected sets.
- MacPherson and S. ("Measuring Shape with Topology," 2012): PH of shapes, studied probability distributions of polymers.
- Adams et. al. ("A Fractal Dimension for Measures via PH ," 2019): PH of random point samples; definition very similar to the one here. Computational experiments that motivated the current work.
- S. ("PH and the Upper Box Dimension," 2018): PH of extremal point sets. First rigorous results relating PH to a classically defined fractal dimension.

Previous Work on Fractal Dimension and PH

Several authors have defined fractal dimensions based on PH , and compared computational estimates with known dimensions:

- Robins (PhD thesis, 2000): Persistent Betti numbers of fractals, proved results for H_0 of disconnected sets.
- MacPherson and S. (“Measuring Shape with Topology,” 2012): PH of shapes, studied probability distributions of polymers.
- Adams et. al. (“A Fractal Dimension for Measures via PH ,” 2019): PH of random point samples; definition very similar to the one here. Computational experiments that motivated the current work.
- S. (“ PH and the Upper Box Dimension,” 2018): PH of extremal point sets. First rigorous results relating PH to a classically defined fractal dimension.

Previous Work on Fractal Dimension and PH

Several authors have defined fractal dimensions based on PH , and compared computational estimates with known dimensions:

- Robins (PhD thesis, 2000): Persistent Betti numbers of fractals, proved results for H_0 of disconnected sets.
- MacPherson and S. ("Measuring Shape with Topology," 2012): PH of shapes, studied probability distributions of polymers.
- Adams et. al. ("A Fractal Dimension for Measures via PH ," 2019): PH of random point samples; definition very similar to the one here. Computational experiments that motivated the current work.
- S. (" PH and the Upper Box Dimension," 2018): PH of extremal point sets. First rigorous results relating PH to a classically defined fractal dimension.

Previous Work on Fractal Dimension and PH

Several authors have defined fractal dimensions based on PH , and compared computational estimates with known dimensions:

- Robins (PhD thesis, 2000): Persistent Betti numbers of fractals, proved results for H_0 of disconnected sets.
- MacPherson and S. (“Measuring Shape with Topology,” 2012): PH of shapes, studied probability distributions of polymers.
- Adams et. al. (“A Fractal Dimension for Measures via PH ,” 2019): PH of random point samples; definition very similar to the one here. Computational experiments that motivated the current work.
- S. (“ PH and the Upper Box Dimension,” 2018): PH of extremal point sets. First rigorous results relating PH to a classically defined fractal dimension.

Weighted Lifetime Sums

Definition (α -Weighted Lifetime Sum)

If X is a bounded metric space, define

$$E_{\alpha}^i(X) = \sum_{(b,d) \in PH_i(X)} (d - b)^{\alpha}.$$

When $i = 0$ and X is finite the sum can be taken over the edges of the minimum spanning tree on X :

$$E_{\alpha}^0(X) = \frac{1}{2^{\alpha}} \sum_{e \in T(X)} |e|^{\alpha}.$$

Weighted Lifetime Sums

Definition (α -Weighted Lifetime Sum)

If X is a bounded metric space, define

$$E_{\alpha}^i(X) = \sum_{(b,d) \in PH_i(X)} (d - b)^{\alpha}.$$

When $i = 0$ and X is finite the sum can be taken over the edges of the minimum spanning tree on X :

$$E_{\alpha}^0(X) = \frac{1}{2^{\alpha}} \sum_{e \in T(X)} |e|^{\alpha}.$$

Steele's Theorem

Theorem (Steele, 1988)

Let μ be a compactly supported probability measure on \mathbb{R}^m , $m \geq 2$, and let $\{x_n\}_{n \in \mathbb{N}}$ be i.i.d. samples from μ . If $0 < \alpha < m$,

$$\lim_{n \rightarrow \infty} n^{-\frac{m-\alpha}{m}} E_{\alpha}^0(x_1, \dots, x_n) \rightarrow c(\alpha, m) \int_{\mathbb{R}^m} f(x)^{(m-\alpha)/m} dx$$

with probability one, where $f(x)$ is the probability density of the absolutely continuous part of μ , and $c(\alpha, m)$ is a positive constant that depends only on α and m .

Steele's Theorem

Theorem (Steele, 1988)

Let μ be a compactly supported probability measure on \mathbb{R}^m , $m \geq 2$, and let $\{x_n\}_{n \in \mathbb{N}}$ be i.i.d. samples from μ . If $0 < \alpha < m$,

$$\lim_{n \rightarrow \infty} n^{-\frac{m-\alpha}{m}} E_{\alpha}^0(x_1, \dots, x_n) \rightarrow c(\alpha, m) \int_{\mathbb{R}^m} f(x)^{(m-\alpha)/m} dx$$

with probability one, where $f(x)$ is the probability density of the absolutely continuous part of μ , and $c(\alpha, m)$ is a positive constant that depends only on α and m .

Persistent Homology Dimension

Let μ be a probability measure on a metric space, $\{x_i\}_{i \in \mathbb{N}}$ be i.i.d. samples from μ , and $\alpha \in \mathbb{R}^+$. Idea: if the support of μ is d -dimensional, then $E_\alpha^i(x_1, \dots, x_n)$ should scale as $n^{\frac{d-\alpha}{d}}$.

Definition

$$\dim_{PH_i^\alpha}(\mu) = \frac{\alpha}{1 - \beta}$$

where

$$\beta = \limsup_{n \rightarrow \infty} \frac{\log(\mathbb{E}(E_\alpha^i(x_1, \dots, x_n)))}{\log(n)}.$$

Persistent Homology Dimension

Let μ be a probability measure on a metric space, $\{x_i\}_{i \in \mathbb{N}}$ be i.i.d. samples from μ , and $\alpha \in \mathbb{R}^+$. Idea: if the support of μ is d -dimensional, then $E_\alpha^i(x_1, \dots, x_n)$ should scale as $n^{\frac{d-\alpha}{d}}$.

Definition

$$\dim_{PH_i^\alpha}(\mu) = \frac{\alpha}{1 - \beta}$$

where

$$\beta = \limsup_{n \rightarrow \infty} \frac{\log(\mathbb{E}(E_\alpha^i(x_1, \dots, x_n)))}{\log(n)}.$$

Steele's Theorem

Corollary (Steele, 1988)

If μ is a nonsingular, compactly supported probability distribution on \mathbb{R}^m and $0 < \alpha < m$

$$\dim_{PH_0^\alpha}(\mu) = m.$$

Steele's Theorem and Fractal Dimension

Quote (Steele, 1988)

*One feature of the previous theorem that should be noted is that if μ has bounded support and μ is singular with respect to Lebesgue measure, then we have with probability one that $E_{\alpha}^0(x_1, \dots, x_n) = o(n^{(m-\alpha)/m})$. Part of the appeal of this observation is the indication that the length of the minimum spanning tree is a measure of the **dimension** of the support of the distribution. This suggests that the asymptotic behavior of the minimum spanning tree might be a useful adjunct to the concept of dimension in the modeling applications and analysis of fractals.*

Minimum Spanning Trees on Fractals

- However, despite many subsequent stronger results for absolutely continuous measures, very little was known about random minimum spanning trees from singular measures.
- Only previous rigorous result: Kozma, Lotker, and Stupp (2011) on the length of the longest edge of a random minimum spanning tree drawn from a Ahlfors regular measure with connected support.
- Computational experiments: Weygaert, Jones, and Martinez (1992).

Minimum Spanning Trees on Fractals

- However, despite many subsequent stronger results for absolutely continuous measures, very little was known about random minimum spanning trees from singular measures.
- Only previous rigorous result: Kozma, Lotker, and Stupp (2011) on the length of the longest edge of a random minimum spanning tree drawn from a Ahlfors regular measure with connected support.
- Computational experiments: Weygaert, Jones, and Martinez (1992).

Minimum Spanning Trees on Fractals

- However, despite many subsequent stronger results for absolutely continuous measures, very little was known about random minimum spanning trees from singular measures.
- Only previous rigorous result: Kozma, Lotker, and Stupp (2011) on the length of the longest edge of a random minimum spanning tree drawn from a Ahlfors regular measure with connected support.
- Computational experiments: Weygaert, Jones, and Martinez (1992).

Presentation Outline

- 1 Brief Aside: The Structure of Glass
- 2 Background: Fractal Dimension and Persistent Homology
- 3 Previous Work and Definitions
- 4 Results for Minimum Spanning Trees (PH_0)**
- 5 Computational Results
- 6 Proof Sketch

Ahlfors Regularity

Definition (Ahlfors Regularity)

A probability measure μ supported on a metric space X is d -Ahlfors regular if there exist positive real numbers c and r_0 so that

$$\frac{1}{c} r^d \leq \mu(B_r(x)) \leq c r^d$$

for all $x \in X$ and $r < r_0$.

Ahlfors regularity is a standard hypothesis that implies that the fractal dimension of a measure is well defined. That is, the various classical notions of dimension coincide and equal d .

Ahlfors Regular Examples

- The natural measures on the Cantor set, Sierpinski triangle, as well as any self-similar fractal defined by an iterated function system whose correct-dimensional Hausdorff measure is positive (this is weaker than the usual open-set condition).
- A natural measure on the boundary of a certain hyperbolic groups, such as the fundamental group of a compact, negatively curved manifold.
- Bounded probability densities on a compact Riemannian manifolds.

Main Theorem for Minimum Spanning Trees

Theorem (S., 2018)

Let μ be a d -Ahlfors regular measure on a metric space, and let $\{x_i\}_{i \in \mathbb{N}}$ be i.i.d. samples from μ . If $0 < \alpha < d$,

$$C_0 \leq n^{-\frac{d-\alpha}{d}} E_\alpha^0(x_1, \dots, x_n) \leq C_1$$

with high probability as $n \rightarrow \infty$, where C_0 and C_1 are positive constants that do not depend on n . In particular,

$$\dim_{PH_0^\alpha}(\mu) = d.$$

Main Theorem for Minimum Spanning Trees

Theorem (S., 2018)

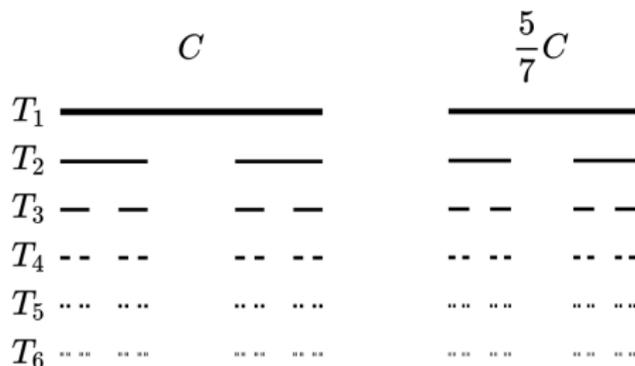
Let μ be a d -Ahlfors regular measure on a metric space, and let $\{x_i\}_{i \in \mathbb{N}}$ be i.i.d. samples from μ . If $0 < \alpha < d$,

$$C_0 \leq n^{-\frac{d-\alpha}{d}} E_\alpha^0(x_1, \dots, x_n) \leq C_1$$

with high probability as $n \rightarrow \infty$, where C_0 and C_1 are positive constants that do not depend on n . In particular,

$$\dim_{PH_0^\alpha}(\mu) = d.$$

Sharpness of Our Result



There is a d -Ahlfors regular measure so that $n^{-\frac{d-\alpha}{d}} E_{\alpha}^0(x_1, \dots, x_n)$ oscillates between two constants with high probability. The idea is to “interleave” the Cantor set with the Cantor set rescaled by $5/7$.

Higher Dimensional Persistent Homology

Higher dimensional results are more difficult because of extremal questions about the number of PH_i intervals of a set of n points. See “Persistent Homology and the Upper Box Dimension” (S., 2018). Our cleanest result is for \mathbb{R}^2 (for the Čech complex):

Theorem (S., 2018)

Let μ be a d -Ahlfors regular measure on \mathbb{R}^2 with $d > 1.5$, and let $\{x_i\}_{i \in \mathbb{N}}$ be i.i.d. samples from μ . If $0 < \alpha < d$, there are constants $0 < C_0 \leq C_1$ so that

$$C_0 \leq n^{-\frac{d-\alpha}{d}} E_\alpha^1(x_1, \dots, x_n) \leq C_1$$

with high probability as $n \rightarrow \infty$. In particular,

$$\dim_{PH_1^\alpha}(\mu) = d.$$

Higher Dimensional Persistent Homology

Higher dimensional results are more difficult because of extremal questions about the number of PH_i intervals of a set of n points. See “Persistent Homology and the Upper Box Dimension” (S., 2018). Our cleanest result is for \mathbb{R}^2 (for the Čech complex):

Theorem (S., 2018)

Let μ be a d -Ahlfors regular measure on \mathbb{R}^2 with $d > 1.5$, and let $\{x_i\}_{i \in \mathbb{N}}$ be i.i.d. samples from μ . If $0 < \alpha < d$, there are constants $0 < C_0 \leq C_1$ so that

$$C_0 \leq n^{-\frac{d-\alpha}{d}} E_\alpha^1(x_1, \dots, x_n) \leq C_1$$

with high probability as $n \rightarrow \infty$. In particular,

$$\dim_{PH_1^\alpha}(\mu) = d.$$

Higher Dimensional Persistent Homology

Higher dimensional results are more difficult because of extremal questions about the number of PH_i intervals of a set of n points. See “Persistent Homology and the Upper Box Dimension” (S., 2018). Our cleanest result is for \mathbb{R}^2 (for the Čech complex):

Theorem (S., 2018)

Let μ be a d -Ahlfors regular measure on \mathbb{R}^2 with $d > 1.5$, and let $\{x_i\}_{i \in \mathbb{N}}$ be i.i.d. samples from μ . If $0 < \alpha < d$, there are constants $0 < C_0 \leq C_1$ so that

$$C_0 \leq n^{-\frac{d-\alpha}{d}} E_\alpha^1(x_1, \dots, x_n) \leq C_1$$

with high probability as $n \rightarrow \infty$. In particular,

$$\dim_{PH_1^\alpha}(\mu) = d.$$

Presentation Outline

- 1 Brief Aside: The Structure of Glass
- 2 Background: Fractal Dimension and Persistent Homology
- 3 Previous Work and Definitions
- 4 Results for Minimum Spanning Trees (PH_0)
- 5 Computational Results**
- 6 Proof Sketch

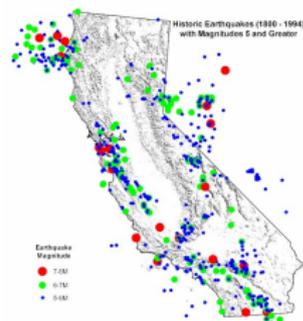
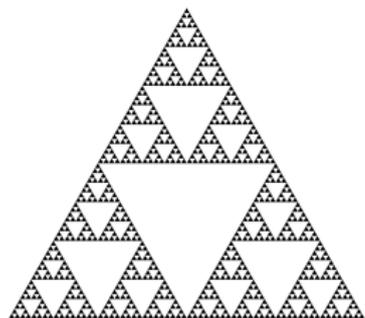
Computational Results

Question

How does the practical performance of the PH_i -dimension compare to classical methods such as box-counting or the correlation algorithm?

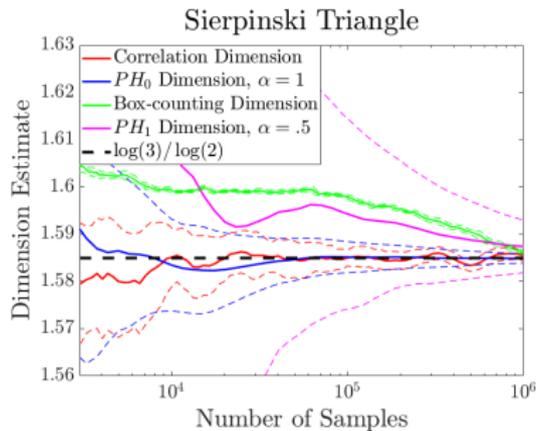
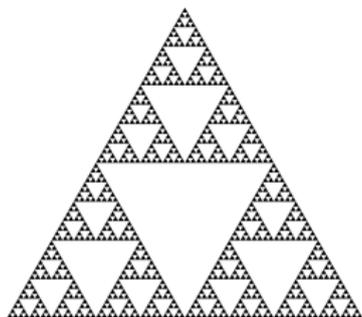
“Fractal Dimension Estimation with Persistent Homology: A Comparative Study,” with J. Jaquette (2019).

Computational Results



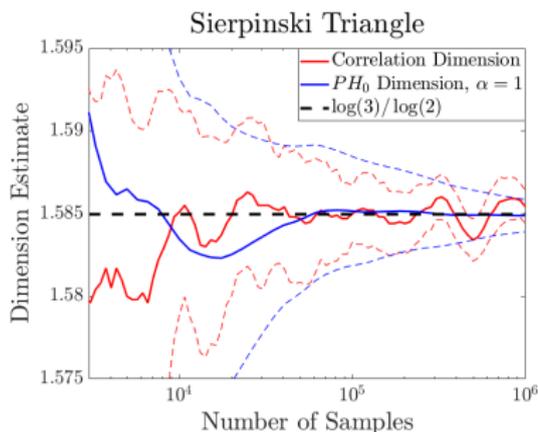
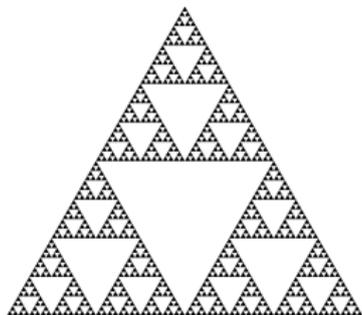
We compare the performance of algorithms to estimate the PH_i , box-counting, and correlation dimensions, for three classes of examples: self-similar fractals, chaotic attractors, and empirical earthquake data.

Sierpinski Triangle



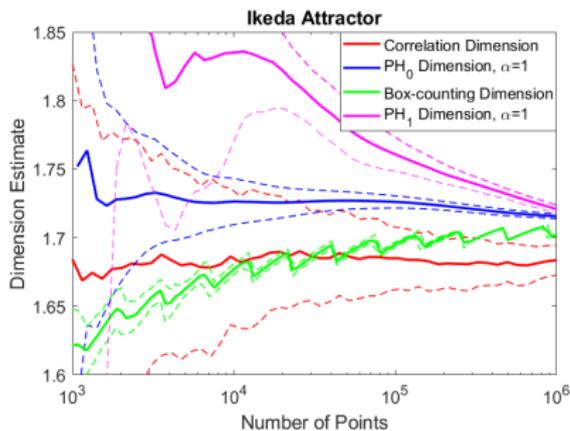
In general, the PH_0 and correlation dimensions perform comparably well. In cases where the true dimension is known, they approach it at about the same rate. In most cases, the box-counting and higher PH_i dimensions perform worse.

Sierpinski Triangle



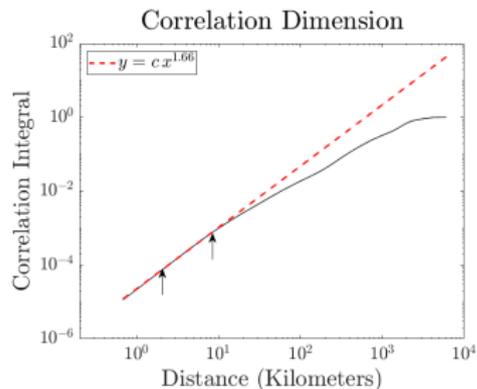
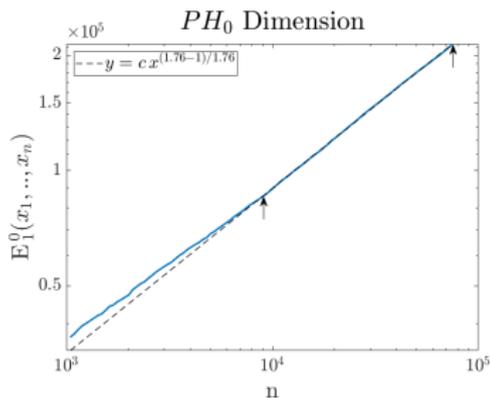
We found one simple rule for fitting a power law to estimate the PH_0 which worked well for all examples, in contrast to the correlation dimension and (especially) the box-counting dimension.

Ikeda attractor



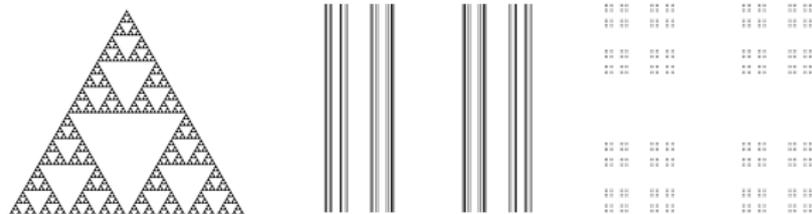
Different notions of dimension may disagree for non-regular sets.

Earthquake Data



We applied the dimension estimation algorithms to the Hauksson–Shearer Southern California earthquake catalog, and found a PH_0 dimension estimate of 1.76 and a correlation dimension estimate of 1.66. This is in line with previous studies.

PH complexity

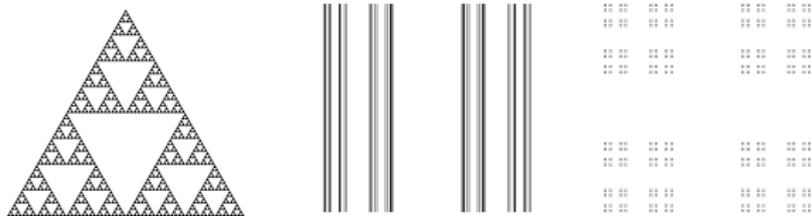


Definition (MacPherson–S.,2012)

$$\text{comp}_{PH}^i(X) = \inf \left\{ \alpha : E_\alpha^i(X) < \infty \right\} .$$

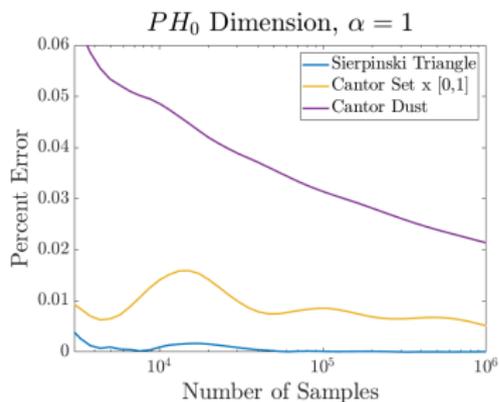
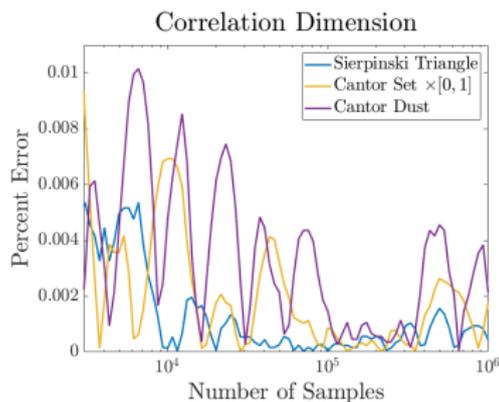
Measures the complexity of a shape, rather than the dimension.

PH complexity



Example	True Dim.	$\text{comp}_{PH_0}(X)$	$\text{comp}_{PH_1}(X)$
S	$\frac{\log(3)}{\log(2)}$	0	$\frac{\log(3)}{\log(2)}$
$C \times I$	$1 + \frac{\log(2)}{\log(3)}$	$\frac{\log(2)}{\log(3)}$	0
$C \times C$	$\frac{2 \log(2)}{\log(3)}$	$\frac{2 \log(2)}{\log(3)}$	$\frac{2 \log(2)}{\log(3)}$

An Indicator of Difficulty?



$$\text{comp}_{PH}^0(S) < \text{comp}_{PH}^0(C \times I) < \text{comp}_{PH}^0(C \times C)$$

Future Directions

- If μ has connected support, do sharper asymptotics hold for $E_{\alpha}^i(x_1, \dots, x_n)$, as in Steele's theorem (a Beardwood-Halton-Hammersley type result)?
- Sharper results for $i > 0$?
- Can we formalize the observed relationship between comp_{PH}^0 , and the difficulty of dimension estimation?
- Thank you for your attention!

Future Directions

- If μ has connected support, do sharper asymptotics hold for $E_{\alpha}^i(x_1, \dots, x_n)$, as in Steele's theorem (a Beardwood-Halton-Hammersley type result)?
- Sharper results for $i > 0$?
- Can we formalize the observed relationship between comp_{PH}^0 , and the difficulty of dimension estimation?
- Thank you for your attention!

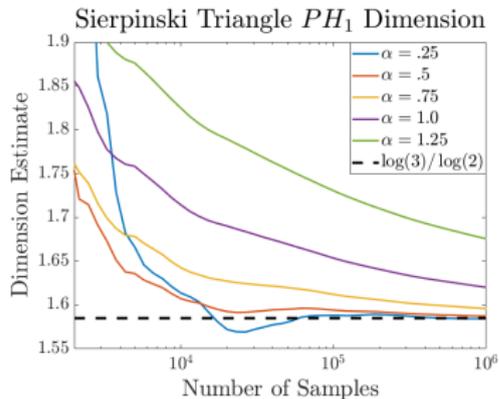
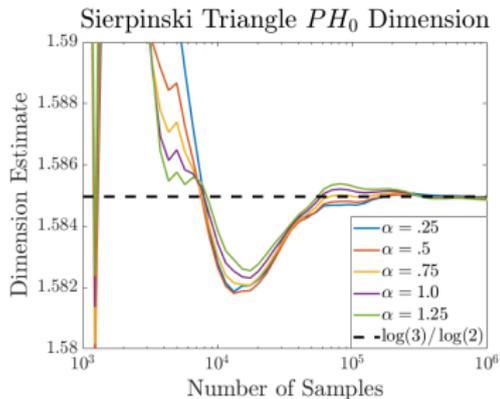
Future Directions

- If μ has connected support, do sharper asymptotics hold for $E_{\alpha}^i(x_1, \dots, x_n)$, as in Steele's theorem (a Beardwood-Halton-Hammersley type result)?
- Sharper results for $i > 0$?
- Can we formalize the observed relationship between comp_{PH}^0 , and the difficulty of dimension estimation?
- Thank you for your attention!

Future Directions

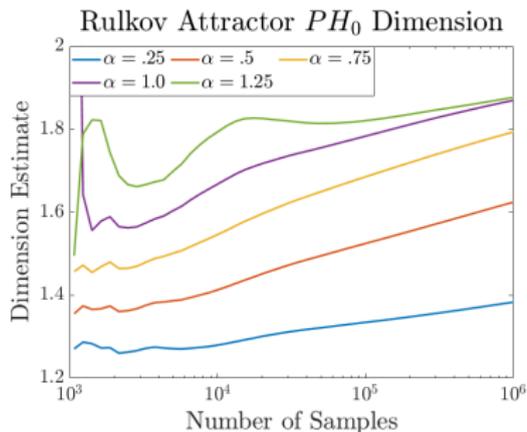
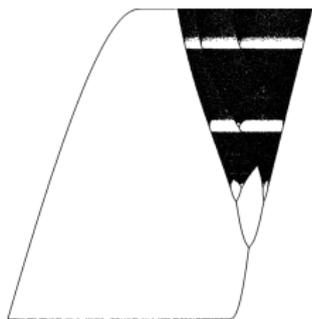
- If μ has connected support, do sharper asymptotics hold for $E_{\alpha}^i(x_1, \dots, x_n)$, as in Steele's theorem (a Beardwood-Halton-Hammersley type result)?
- Sharper results for $i > 0$?
- Can we formalize the observed relationship between comp_{PH}^0 , and the difficulty of dimension estimation?
- Thank you for your attention!

Dependence on α



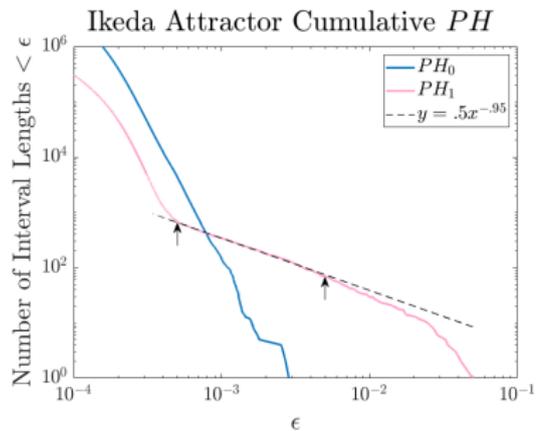
For self-similar fractals, lower values of α produced better convergence for \dim_{PH}^i when $\text{comp}_{PH}^i \neq 0$, but there wasn't much difference otherwise.

Dependence on α



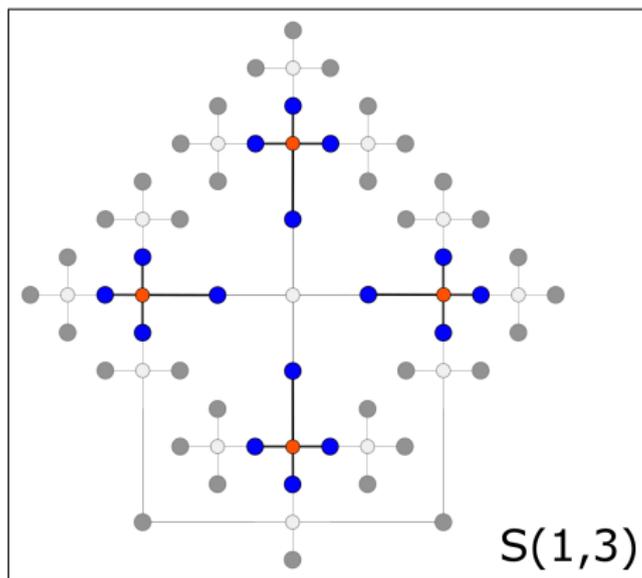
For non-regular sets, different values of α may give different values for \dim_{PH}^i .

Ikeda attractor

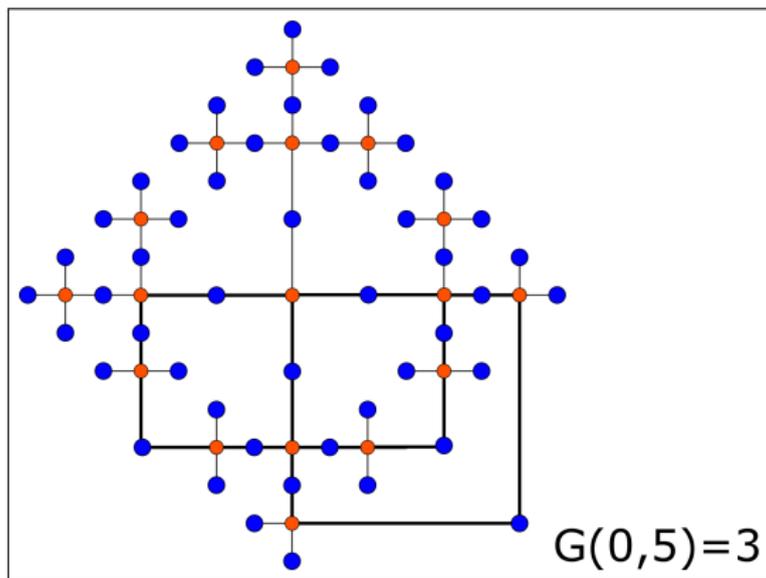


$$\text{comp}_{PH}^1 \approx .95?$$

Shell Annuli

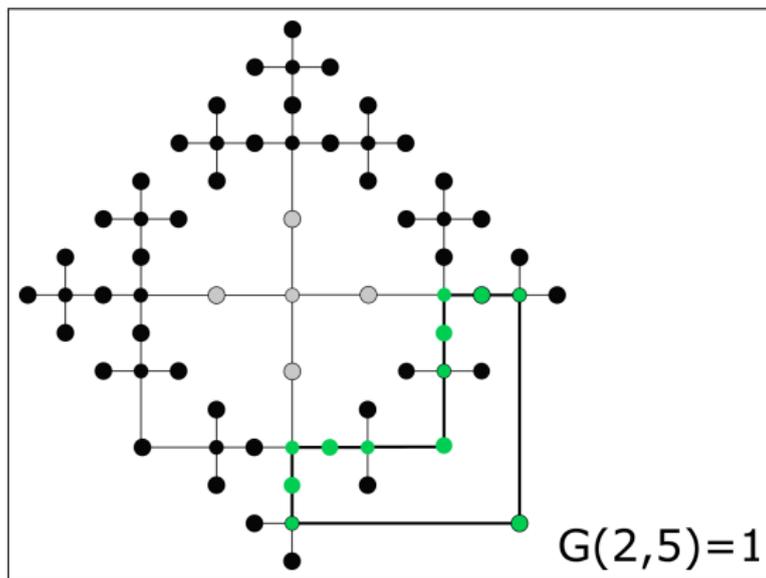


If S is a swatch of radius R and $0 \leq i \leq j \leq r$, let $S(i,j)$ be the subgraph consisting of atoms between shells i and j and the bonds between them.

H_1 

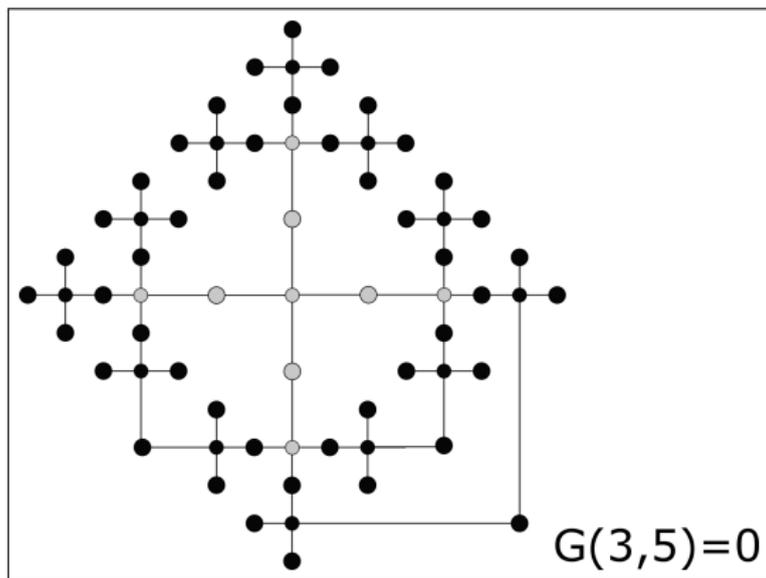
$$G(i,j) = \text{rank}(H_1(S(i,j)))$$

That is, $G(i,j)$ is the # of “independent rings” of $S(i,j)$.

H_1 

$$G(i,j) = \text{rank}(H_1(S(i,j)))$$

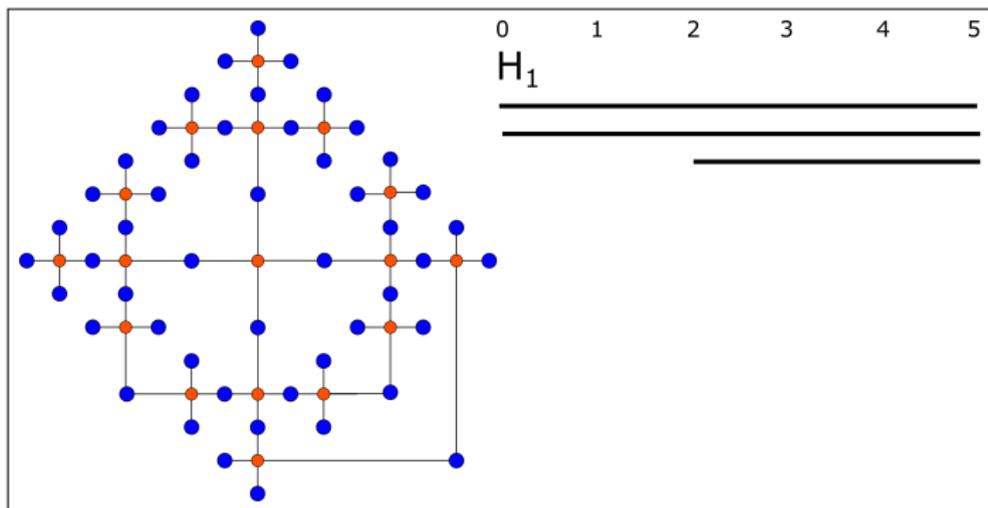
That is, $G(i,j)$ is the # of “independent rings” of $S(i,j)$.

H_1 

$$G(i,j) = \text{rank}(H_1(S(i,j)))$$

That is, $G(i,j)$ is the # of “independent rings” of $S(i,j)$.

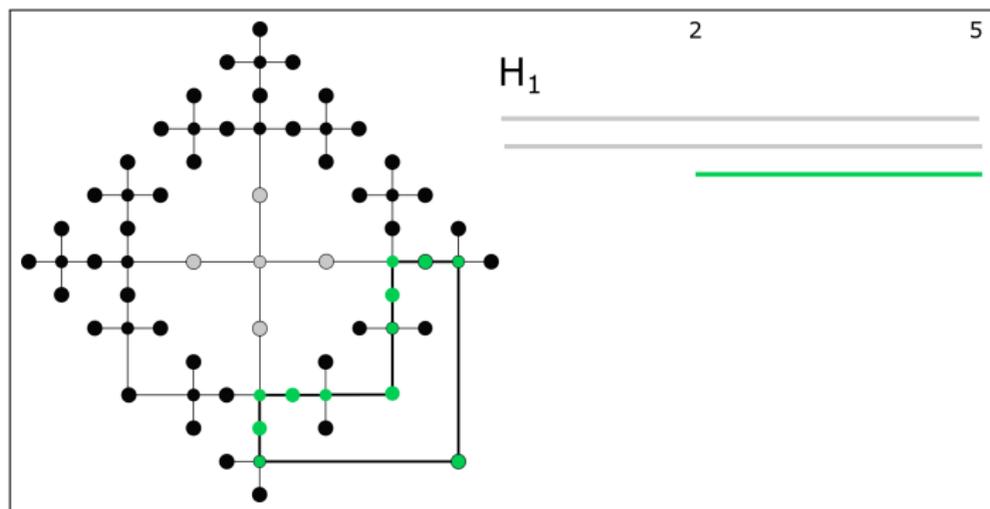
H_1 Barcode



Our local environments are one dimensional, so $G(i, j) \leq G(k, l)$ if $(i, j) \subseteq (k, l)$. As such, there is a unique set of intervals so that

$$G(i, j) = \# \text{ of intervals contained in } (i, j).$$

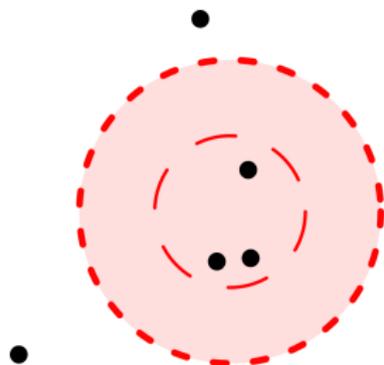
H_1 Barcode



Our local environments are one dimensional, so $G(i, j) \leq G(k, l)$ if $(i, j) \subseteq (k, l)$. As such, there is a unique set of intervals so that

$$G(i, j) = \# \text{ of intervals contained in } (i, j).$$

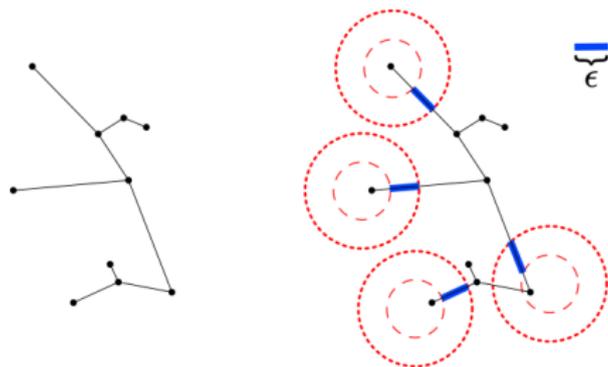
Lower Bound Proof Sketch



For a ball $B_{2\epsilon}(y)$ and a finite set $\mathbf{x} \subset X$ define the occupancy event

$$\omega(B_{2\epsilon}(y), \mathbf{x}) = \begin{cases} 1 & \mathbf{x} \cap B_{\epsilon}(y) \neq \emptyset \text{ and } \mathbf{x} \cap (B_{2\epsilon}(y) \setminus B_{\epsilon}(y)) = \emptyset \\ 0 & \text{otherwise} \end{cases} .$$

Lower Bound Proof Sketch



Lemma (Disjoint Ball Lemma)

Let \mathcal{B} be a set of disjoint balls of radius 2ϵ centered at points of X , and let \mathbf{x} be a finite subset of X . Then

$$F(\mathbf{x}, \epsilon) \geq \sum_{B \in \mathcal{B}} \omega(B, \mathbf{x}) - 1.$$

Lower Bound Proof Sketch

Let $n \in \mathbb{N}$ and $\epsilon = n^{-1/d}$. Let $B_1^n, \dots, B_{s_n}^n$ be a maximal collection of disjoint balls of radius 2ϵ centered at points of X .

Lemma

There is a positive real number $\gamma > 0$ so that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{s_n} \omega(B_j^n, \mathbf{x}_n) \geq \gamma$$

in probability as $n \rightarrow \infty$.

Idea: use Ahlfors regularity to control s_n , apply standard probabilistic arguments.

Lower Bound Proof Sketch

Let $n \in \mathbb{N}$ and $\epsilon = n^{-1/d}$. Let $B_1^n, \dots, B_{s_n}^n$ be a maximal collection of disjoint balls of radius 2ϵ centered at points of X .

Lemma

There is a positive real number $\gamma > 0$ so that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{s_n} \omega(B_j^n, \mathbf{x}_n) \geq \gamma$$

in probability as $n \rightarrow \infty$.

Idea: use Ahlfors regularity to control s_n , apply standard probabilistic arguments.

Lower Bound Proof Sketch

Let $n \in \mathbb{N}$ and $\epsilon = n^{-1/d}$. Let $B_1^n, \dots, B_{s_n}^n$ be a maximal collection of disjoint balls of radius 2ϵ centered at points of X .

Lemma

There is a positive real number $\gamma > 0$ so that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{s_n} \omega(B_j^n, \mathbf{x}_n) \geq \gamma$$

in probability as $n \rightarrow \infty$.

Idea: use Ahlfors regularity to control s_n , apply standard probabilistic arguments.

Lower Bound Proof Sketch

$$\lim_{n \rightarrow \infty} n^{-\frac{d-\alpha}{d}} E_{\alpha}^0(\mathbf{x}_n) \geq \lim_{n \rightarrow \infty} n^{-\frac{d-\alpha}{d}} n^{-\alpha/d} F(\mathbf{x}_n, n^{-1/d})$$

$$\geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{s_n} \omega(B_j^n, \mathbf{x}_n)$$

ball occupancy

$$\geq \gamma$$

previous lemma

in probability as $n \rightarrow \infty$.

Lower Bound Proof Sketch

$$\begin{aligned}\lim_{n \rightarrow \infty} n^{-\frac{d-\alpha}{d}} E_{\alpha}^0(\mathbf{x}_n) &\geq \lim_{n \rightarrow \infty} n^{-\frac{d-\alpha}{d}} n^{-\alpha/d} F(\mathbf{x}_n, n^{-1/d}) \\ &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{s_n} \omega(B_j^n, \mathbf{x}_n) \\ &\geq \gamma\end{aligned}$$

ball occupancy

previous lemma

in probability as $n \rightarrow \infty$.

Lower Bound Proof Sketch

$$\begin{aligned}\lim_{n \rightarrow \infty} n^{-\frac{d-\alpha}{d}} E_{\alpha}^0(\mathbf{x}_n) &\geq \lim_{n \rightarrow \infty} n^{-\frac{d-\alpha}{d}} n^{-\alpha/d} F(\mathbf{x}_n, n^{-1/d}) \\ &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^{s_n} \omega(B_j^n, \mathbf{x}_n) \\ &\geq \gamma\end{aligned}$$

ball occupancy

previous lemma

in probability as $n \rightarrow \infty$.

Upper Bound Proof Sketch

$\mathbf{x}_n = \{x_1, \dots, x_n\} \subset X$. We prove an extremal upper bound on $E_\alpha^0(\mathbf{x}_n)$ that is stronger than what was stated in the previous slide.

Idea: control the number of edges of length greater than ϵ . Let

$$F(\mathbf{x}_n, \epsilon) = \left| \{e \in T(\mathbf{x}_n) : |e| > \epsilon\} \right|.$$

Upper Bound Proof Sketch

$\mathbf{x}_n = \{x_1, \dots, x_n\} \subset X$. We prove an extremal upper bound on $E_\alpha^0(\mathbf{x}_n)$ that is stronger than what was stated in the previous slide.

Idea: control the number of edges of length greater than ϵ . Let

$$F(\mathbf{x}_n, \epsilon) = \left| \{e \in T(\mathbf{x}_n) : |e| > \epsilon\} \right|.$$

Upper Bound Proof Sketch

Lemma (ϵ -approximation Lemma)

If X supports a d -Ahlfors regular μ , there exists a $C > 0$ so that for any $Y \subseteq X$ and any $\epsilon > 0$ there is a finite point set $Y^\epsilon \subset Y$ so that

- $d_H(Y, Y^\epsilon) < \epsilon$
- $|Y^\epsilon| < C\epsilon^{-d}$.

$$\begin{aligned} F(\mathbf{x}_n, \epsilon) &\leq F(\mathbf{x}_n^{\epsilon/2}, 0) && \text{by bottleneck stability} \\ &= \left| \mathbf{x}_n^{\epsilon/2} \right| - 1 \\ &\leq 2^d C \epsilon^{-d} \end{aligned}$$

Upper Bound Proof Sketch

Lemma (ϵ -approximation Lemma)

If X supports a d -Ahlfors regular μ , there exists a $C > 0$ so that for any $Y \subseteq X$ and any $\epsilon > 0$ there is a finite point set $Y^\epsilon \subset Y$ so that

- $d_H(Y, Y^\epsilon) < \epsilon$
- $|Y^\epsilon| < C\epsilon^{-d}$.

$$\begin{aligned} F(\mathbf{x}_n, \epsilon) &\leq F(\mathbf{x}_n^{\epsilon/2}, 0) && \text{by bottleneck stability} \\ &= \left| \mathbf{x}_n^{\epsilon/2} \right| - 1 \\ &\leq 2^d C \epsilon^{-d} \end{aligned}$$

Upper Bound Proof Sketch

Lemma (ϵ -approximation Lemma)

If X supports a d -Ahlfors regular μ , there exists a $C > 0$ so that for any $Y \subseteq X$ and any $\epsilon > 0$ there is a finite point set $Y^\epsilon \subset Y$ so that

- $d_H(Y, Y^\epsilon) < \epsilon$
- $|Y^\epsilon| < C\epsilon^{-d}$.

$$\begin{aligned} F(\mathbf{x}_n, \epsilon) &\leq F(\mathbf{x}_n^{\epsilon/2}, 0) && \text{by bottleneck stability} \\ &= \left| \mathbf{x}_n^{\epsilon/2} \right| - 1 \\ &\leq 2^d C \epsilon^{-d} \end{aligned}$$

Upper Bound Proof Sketch

Lemma (ϵ -approximation Lemma)

If X supports a d -Ahlfors regular μ , there exists a $C > 0$ so that for any $Y \subseteq X$ and any $\epsilon > 0$ there is a finite point set $Y^\epsilon \subset Y$ so that

- $d_H(Y, Y^\epsilon) < \epsilon$
- $|Y^\epsilon| < C\epsilon^{-d}$.

$$\begin{aligned} F(\mathbf{x}_n, \epsilon) &\leq F(\mathbf{x}_n^{\epsilon/2}, 0) && \text{by bottleneck stability} \\ &= \left| \mathbf{x}_n^{\epsilon/2} \right| - 1 \\ &\leq 2^d C \epsilon^{-d} \end{aligned}$$

Upper Bound Proof Sketch

Lemma (ϵ -approximation Lemma)

If X supports a d -Ahlfors regular μ , there exists a $C > 0$ so that for any $Y \subseteq X$ and any $\epsilon > 0$ there is a finite point set $Y^\epsilon \subset Y$ so that

- $d_H(Y, Y^\epsilon) < \epsilon$
- $|Y^\epsilon| < C\epsilon^{-d}$.

$$\begin{aligned} F(\mathbf{x}_n, \epsilon) &\leq F(\mathbf{x}_n^{\epsilon/2}, 0) && \text{by bottleneck stability} \\ &= \left| \mathbf{x}_n^{\epsilon/2} \right| - 1 \\ &\leq 2^d C \epsilon^{-d} \end{aligned}$$

Upper Bound Proof Sketch

$T(\mathbf{x}_n)$ has $n - 1$ edges so

$$\begin{aligned} F(\mathbf{x}_n, \epsilon) &\leq f(\epsilon) := \min(n - 1, 2^{-d} C \epsilon^{-d}) \\ &= \begin{cases} n - 1 & \epsilon \leq \kappa \\ 2^{-d} C \epsilon^{-d} & \epsilon \geq \kappa \end{cases}, \end{aligned}$$

where

$$\kappa = \frac{1}{2} \left(\frac{D}{n - 1} \right)^{1/d}.$$

Upper Bound Proof Sketch

$T(\mathbf{x}_n)$ has $n - 1$ edges so

$$\begin{aligned} F(\mathbf{x}_n, \epsilon) &\leq f(\epsilon) := \min(n - 1, 2^{-d} C \epsilon^{-d}) \\ &= \begin{cases} n - 1 & \epsilon \leq \kappa \\ 2^{-d} C \epsilon^{-d} & \epsilon \geq \kappa \end{cases}, \end{aligned}$$

where

$$\kappa = \frac{1}{2} \left(\frac{D}{n - 1} \right)^{1/d}.$$

Upper Bound Proof Sketch

$$\begin{aligned} E_{\alpha}^0(\mathbf{x}_n) &= \sum_{e \in T(\mathbf{x}_n)} |e|^{\alpha} \\ &= \int_{\epsilon=0}^{\infty} -\frac{\partial |F(\mathbf{x}_n, \epsilon)|}{\partial \epsilon} \epsilon^{\alpha} d\epsilon \\ &\leq \dots && \text{integration by parts} \\ &\leq \alpha \int_{\epsilon=0}^{\text{diam } X} f(\epsilon) \epsilon^{\alpha-1} d\epsilon \\ &\leq \dots \\ &\leq D_{\alpha} n^{\frac{d-\alpha}{d}}, \end{aligned}$$

where

$$D_{\alpha} = 2^{\alpha} D^{\frac{\alpha}{d}} \left(1 + D \frac{\alpha}{d-\alpha}\right).$$

Upper Bound Proof Sketch

$$\begin{aligned} E_\alpha^0(\mathbf{x}_n) &= \sum_{e \in T(\mathbf{x}_n)} |e|^\alpha \\ &= \int_{\epsilon=0}^{\infty} -\frac{\partial |F(\mathbf{x}_n, \epsilon)|}{\partial \epsilon} \epsilon^\alpha d\epsilon \\ &\leq \dots && \text{integration by parts} \\ &\leq \alpha \int_{\epsilon=0}^{\text{diam } X} f(\epsilon) \epsilon^{\alpha-1} d\epsilon \\ &\leq \dots \\ &\leq D_\alpha n^{\frac{d-\alpha}{d}}, \end{aligned}$$

where

$$D_\alpha = 2^\alpha D^{\frac{\alpha}{d}} \left(1 + D \frac{\alpha}{d-\alpha}\right).$$

Upper Bound Proof Sketch

$$\begin{aligned} E_\alpha^0(\mathbf{x}_n) &= \sum_{e \in T(\mathbf{x}_n)} |e|^\alpha \\ &= \int_{\epsilon=0}^{\infty} -\frac{\partial |F(\mathbf{x}_n, \epsilon)|}{\partial \epsilon} \epsilon^\alpha d\epsilon \\ &\leq \dots && \text{integration by parts} \\ &\leq \alpha \int_{\epsilon=0}^{\text{diam } X} f(\epsilon) \epsilon^{\alpha-1} d\epsilon \\ &\leq \dots \\ &\leq D_\alpha n^{\frac{d-\alpha}{d}}, \end{aligned}$$

where

$$D_\alpha = 2^\alpha D^{\frac{\alpha}{d}} \left(1 + D \frac{\alpha}{d-\alpha}\right).$$

Upper Bound Proof Sketch

$$\begin{aligned} E_\alpha^0(\mathbf{x}_n) &= \sum_{e \in T(\mathbf{x}_n)} |e|^\alpha \\ &= \int_{\epsilon=0}^{\infty} -\frac{\partial |F(\mathbf{x}_n, \epsilon)|}{\partial \epsilon} \epsilon^\alpha d\epsilon \\ &\leq \dots && \text{integration by parts} \\ &\leq \alpha \int_{\epsilon=0}^{\text{diam } X} f(\epsilon) \epsilon^{\alpha-1} d\epsilon \\ &\leq \dots \\ &\leq D_\alpha n^{\frac{d-\alpha}{d}}, \end{aligned}$$

where

$$D_\alpha = 2^\alpha D^{\frac{\alpha}{d}} \left(1 + D \frac{\alpha}{d-\alpha}\right).$$