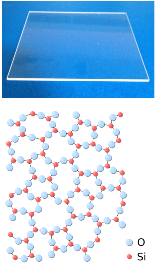
Topology and Geometry of Complex Systems

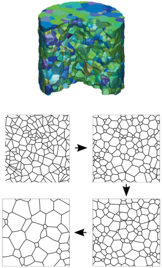
Benjamin Schweinhart

January 2020

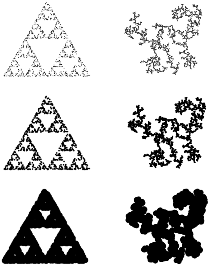
Topology and Geometry of Complex Systems



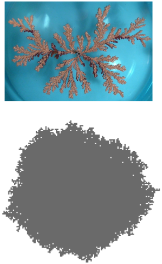
Local Structure of Bond Networks



Cellular Structures and Grain Growth



Fractal Dimension and Persistent Homology

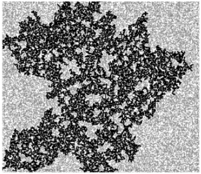


Stochastic Growth Models

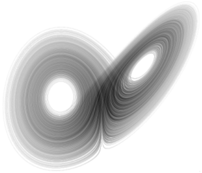
I study the topology and geometry of complex geometric objects. This falls under the purview of topological and geometric data analysis. I'm interested in both theory and applications, especially to materials science. Themes include local structure and randomness (stochastic topology).

Presentation Outline

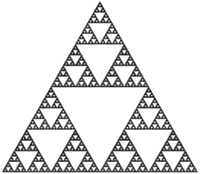
Fractal Dimension



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



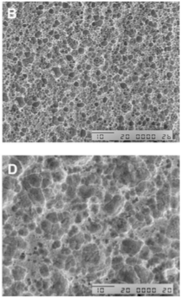
Lorenz Attractor
$d_H \approx 2.06$



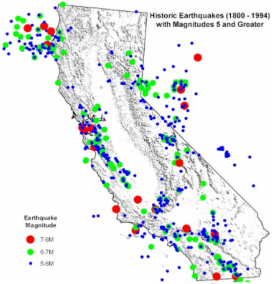
Sierpinski Triangle
$d_H = log_2(3)$

- 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.

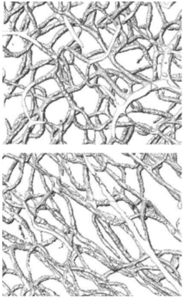
Fractal Applications



Aluminum Oxide[1]



Earthquakes[2]

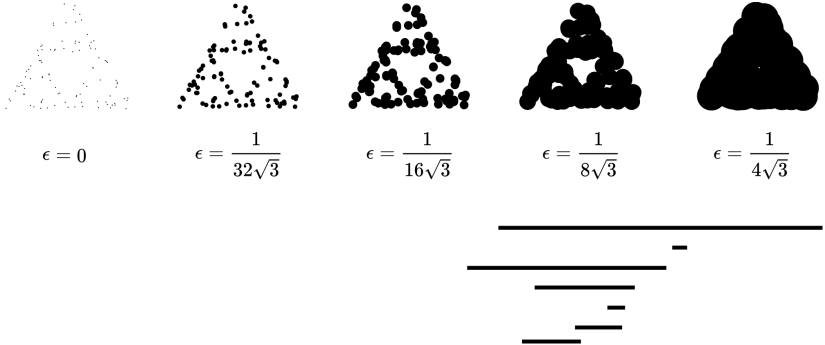


Vascular Networks[3]

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.

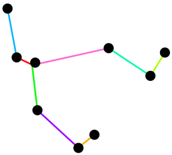
[1]Riscovič and Pavlovič, Scanning (2013), [2]Kagan, Geophysical Journal International (2007), [3]Lorthois and Cassot, Journal of Theoretical Biology (2010)

Persistent Homology



The Persistent Homology (*PH*) of a point cloud tracks the topological changes that occur as balls are grown around the points. Since its introduction by Edelsbrunner et al. in 2002, there has been a surge of interest in the theory and applications of *PH*.

Minimum Spanning Trees



$T(\mathbf{x})$

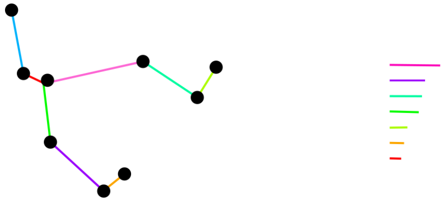
Definition (Minimum Spanning Tree)

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

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

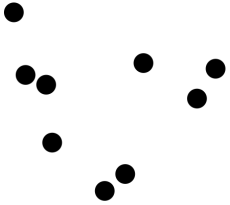
Minimum Spanning Trees and Persistent Homology (PH_0)



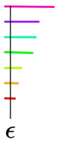
$T(\mathbf{x})$ $\quad$ $PH_0(\mathbf{x})$

If $\mathbf{x}$ is a finite point set in Euclidean space, $PH_0(\mathbf{x})$ is a set of intervals that track how components merge as balls are ground around $\mathbf{x}$. The PH_0 intervals are matched with the MST edges, with an interval corresponding to an edge of twice its length.

Minimum Spanning Trees and Persistent Homology (PH_0)



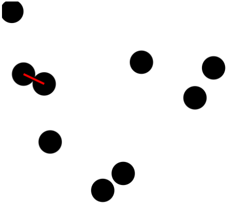
$T(\mathbf{x})$



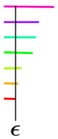
$PH_0(\mathbf{x})$

We can see the correspondence between $PH_0(\mathbf{x})$ and $T(\mathbf{x})$ by using Kruskal's algorithm.

Minimum Spanning Trees and PH_0



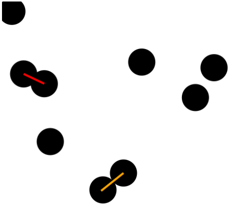
$T(\mathbf{x})$



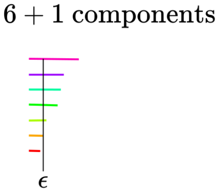
$PH_0(\mathbf{x})$

To find the first MST edge, grow balls around the point set until a component dies.

Minimum Spanning Trees and PH_0



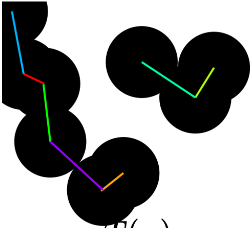
$T(\mathbf{x})$



$PH_0(\mathbf{x})$

Continue until an edge connects vertices in separate components to find the next MST edge.

Minimum Spanning Trees and PH_0



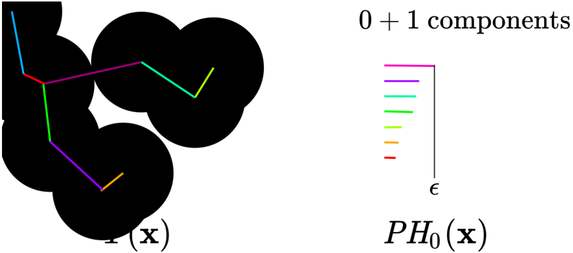
$T(\mathbf{x})$

$1 + 1$ components

$PH_0(\mathbf{x})$

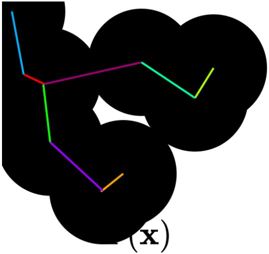
Let's skip to the end.

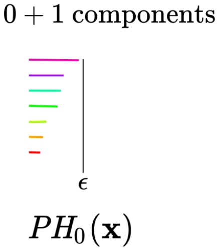
Minimum Spanning Trees and PH_0



When we are down to one component, the MST has been computed.

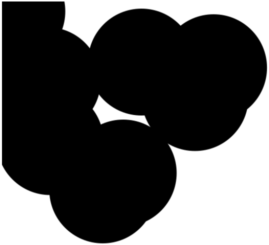
Minimum Spanning Trees and PH_0

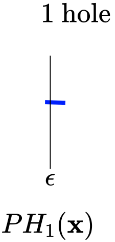




If we expand the balls a little more, a different topological change occurs.

One-dimensional Persistent Homology (PH_1)

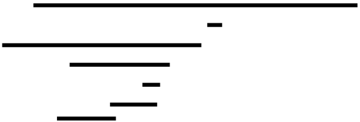




$PH_1(\mathbf{x})$ tracks the holes that form and disappear as balls are thickened around a set of points.

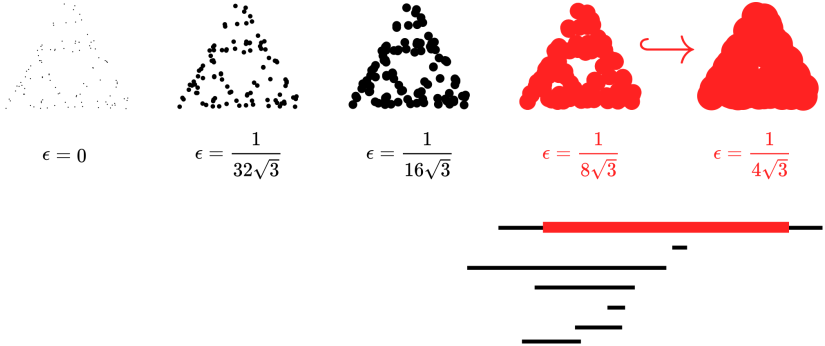
Persistent Homology

$\epsilon = 0$ $\quad \epsilon = \frac{1}{32\sqrt{3}}$ $\quad \epsilon = \frac{1}{16\sqrt{3}}$ $\quad \epsilon = \frac{1}{8\sqrt{3}}$ $\quad \epsilon = \frac{1}{4\sqrt{3}}$



More generally, i-dimensional persistent homology (PH_i) tracks how the i-dimensional homology changes as balls are thickened around a set of points. PH_i is a set of intervals corresponding to homology generators that are born and die in this process.

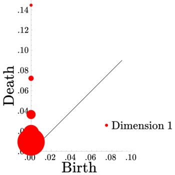
Persistent Homology



More generally, i-dimensional persistent homology (PH_i) tracks how the i-dimensional homology changes as balls are thickened around a set of points. PH_i is a set of intervals corresponding to homology generators that are born and die in this process.

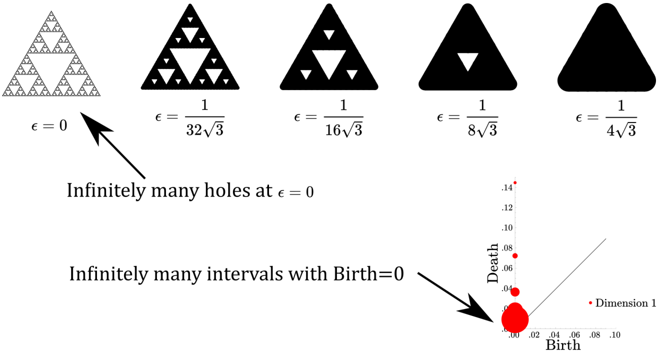
Persistent Homology

$\epsilon = 0$ $\quad \epsilon = \frac{1}{32\sqrt{3}}$ $\quad \epsilon = \frac{1}{16\sqrt{3}}$ $\quad \epsilon = \frac{1}{8\sqrt{3}}$ $\quad \epsilon = \frac{1}{4\sqrt{3}}$



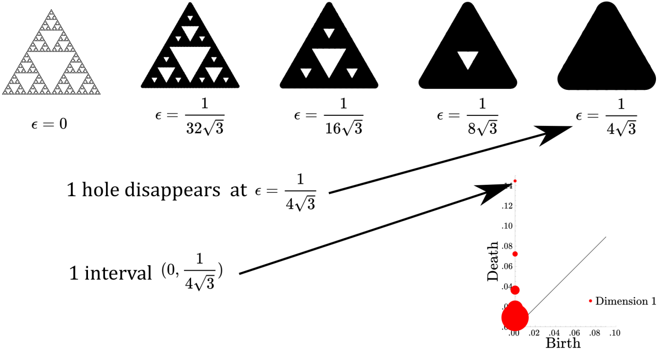
The information in *PH* is often summarized by a persistence diagram: a scatter plot of (birth, death) for each interval.

Persistent Homology



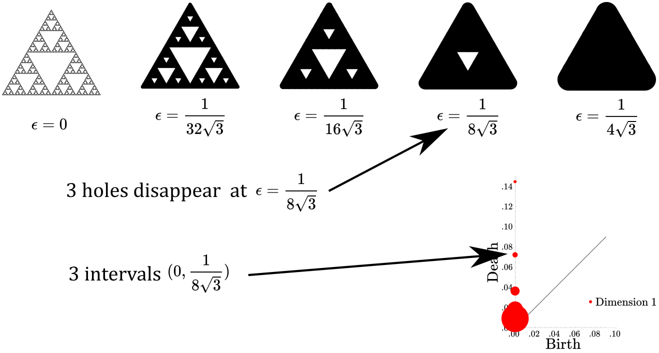
$PH_1(S_\epsilon)$ is a set of intervals, one for each component of the complement that disappears as ϵ increases (by Alexander duality).

Persistent Homology



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

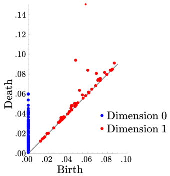
Persistent Homology



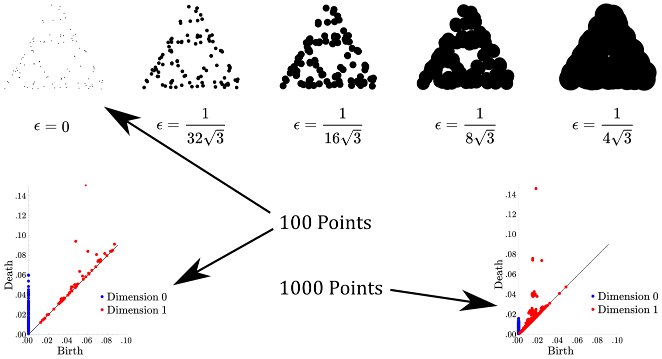
$PH_1(S_\epsilon)$ has one interval for each bounded component of the complement of S (by Alexander duality).

Persistent Homology of a Sample

$\epsilon = 0$ $\quad \epsilon = \frac{1}{32\sqrt{3}}$ $\quad \epsilon = \frac{1}{16\sqrt{3}}$ $\quad \epsilon = \frac{1}{8\sqrt{3}}$ $\quad \epsilon = \frac{1}{4\sqrt{3}}$

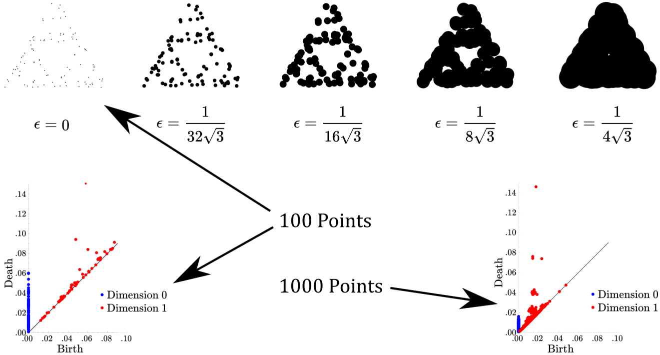


Persistent Homology of a Sample



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!

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," S. (2019).

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," S. and J. Jaquette (2019).

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* complexity 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 :

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

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 $0 < \alpha < m$,

$$\lim_{n\to\infty} n^{-\frac{m-\alpha}{m}} E^0_\alpha(x_1, \ldots, x_n) \to c(\alpha, m) \int_{\mathbb{R}^m} f(x)^{(m-\alpha)/m}\, dx$$

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 $E^i_\alpha(x_1, \dots, x_n)$ should scale as $n^{\frac{d-\alpha}{d}}$.

Definition

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

where

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

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 $E_\alpha^0(x_1, \ldots, x_n) = o(n^{(m-\alpha)/m})$. 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).

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

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

r 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 boundaries 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)

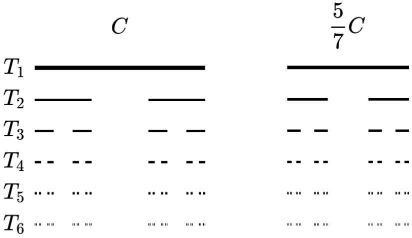
Let μ be a d-Ahlfors regular measure on a metric space, and let $\{x_i\}_{i\in\mathbb{N}}$ be i.i.d. samples from μ. If $0 < \alpha < d$,

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

with high probability as $n \to \infty$, 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



Theorem (S., 2019)

There is a d-Ahlfors regular measure so that $n^{-\frac{d-\alpha}{d}} E_\alpha^0(x_1, \ldots, x_n)$ 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" in DCG (S., 2018). Our cleanest result is for $\mathbb{R}^2$:

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

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

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

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

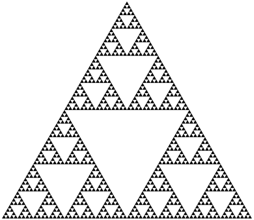
Computational Results

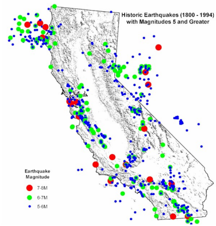
Question

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 (CNSNS, 2019).

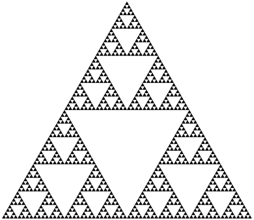
Computational Results

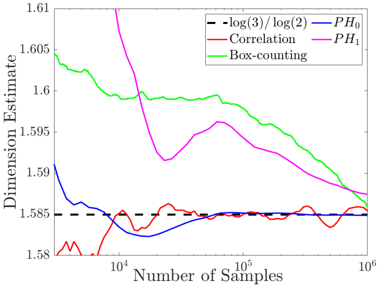




We compare the performance of algorithms to estimate the PH_i, box-counting, and correlation dimensions, for three classes of examples: self-similar fractals, chaotic attractors, and empirical earthquake data.

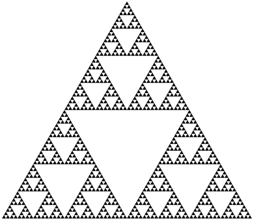
Sierpinski Triangle

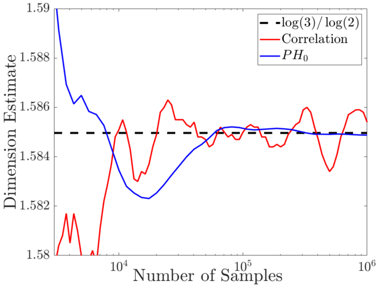




In general, the PH_0 and correlation dimensions perform comparably well. In cases where the true dimension is known, they approach it at about the same rate. In most cases, the box-counting and higher PH_i dimensions perform worse.

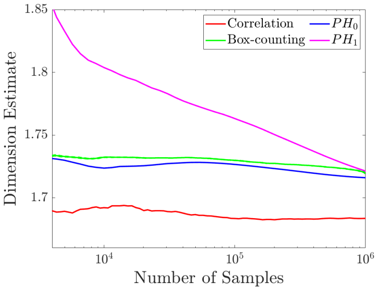
Sierpinski Triangle





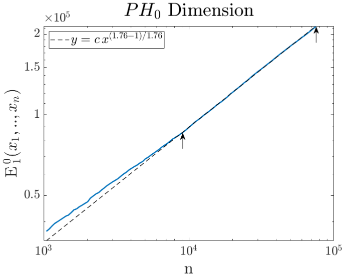
We found one simple rule for fitting a power law to estimate the PH_0 which worked well for all examples, in contrast to the correlation dimension and (especially) the box-counting dimension.

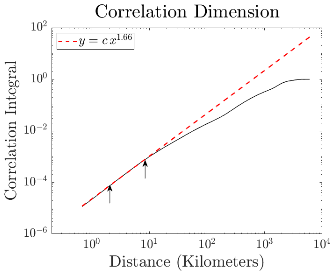
Ikeda attractor



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

Earthquake Data





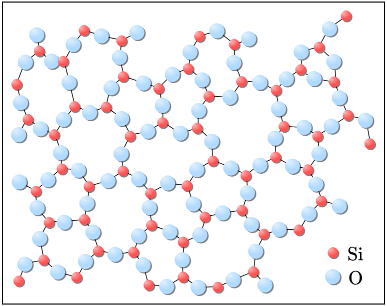
We applied the dimension estimation algorithms to the Hauksson–Shearer Southern California earthquake catalog, and found a PH_0 dimension estimate of 1.76 and a correlation dimension estimate of 1.66. This is in line with previous studies.

Future Directions

- If μ has connected support, do sharper asymptotics hold for $E_\alpha^i(x_1, \ldots, x_n)$, as in Steele's theorem (a Beardwood-Halton-Hammersley type result)?
- Sharper results for $i > 0$?
- Is there a relationship between the difficulty of dimension estimation, and the complexity of the persistent homology of the support?

Presentation Outline

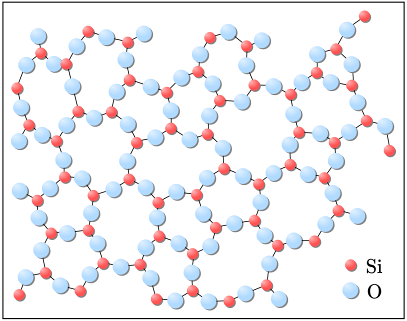
Oxide Glasses

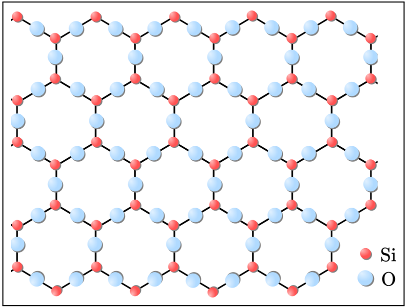


Oxide glasses are present in our daily lives, but the relationship between the local structure and global physical properties of these materials is poorly understood. This is due in part to the lack of an appropriate language to describe that local structure.

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

Silica



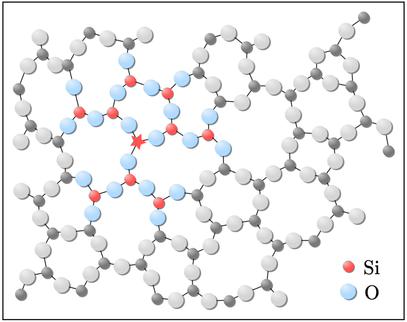


The bond network of silicon dioxide (silica) is a bipartite graph, where silicon atoms are (usually) adjacent to four oxygen atoms, and oxygen atoms are (usually) adjacent to two silicon atoms.

Idea

- Current goal: develop a rigorous methodology to describe the local structure of oxide glasses. It should differentiate various silica glasses, and 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 environments are associated with crystal nucleation?
- Our methodology is also applicable to other systems that can be represented by a sparse graph. Examples: zeolites, metallic glasses/sphere packings, quasicrystals, circulatory networks, ...
- Software package on GitHub (link on my website).

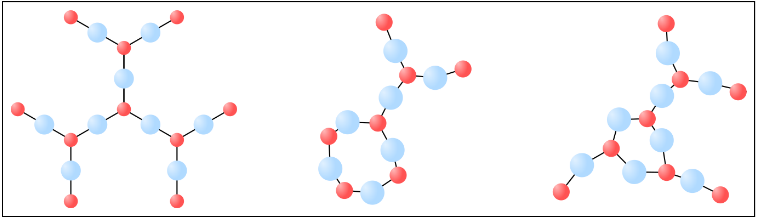
Local Environments



If G is a graph, and v is a vertex of G the **local environment** of radius r centered at v is the ball of radius r in the graph distance.[1]

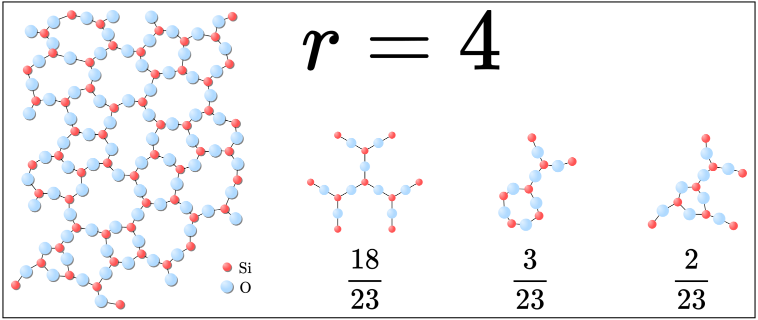
[1] B. Schweinhart, J.K. Mason, and R.D. MacPherson, *Topological Similarity of Random Cell Complexes and Applications*, Physical Review E 93 (2016).

(In)equivalent Environments



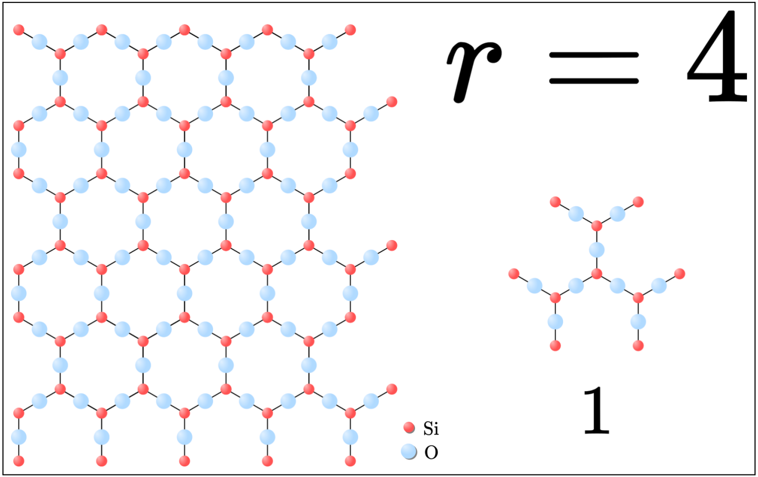
Two environments are equivalent if they are isomorphic as rooted, colored graphs.

The Empirical Distribution



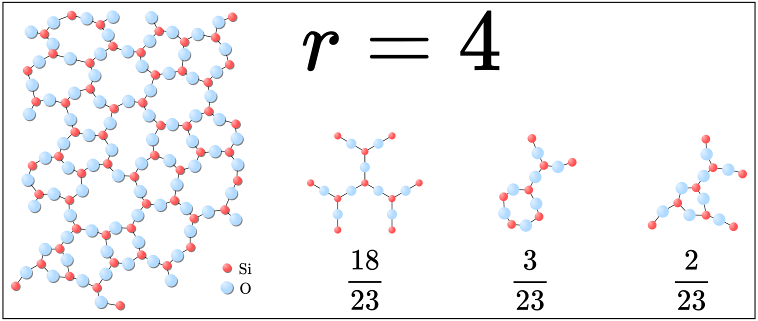
Given a graph G and a radius r, the counting measure on the vertices of G induces an empirical probability distribution of graph isomorphism classes of radius r. This family of probability distributions is called the **empirical distribution** of G. It characterizes the local topology of G.

Crystal Distribution



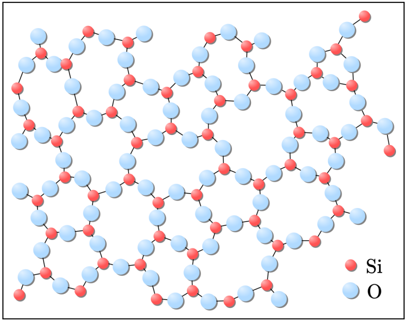
For many crystalline materials, the empirical distribution is supported on one or two topological types.

Disordered Distribution



For disordered materials, the notion of a single "unit cell" is replaced by a probability distribution of local environments.

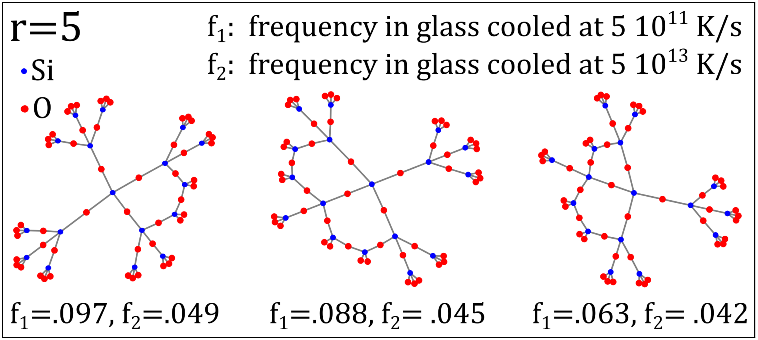
Application: Different Cooling Rates



Silica glasses produced at different cooling rates exhibit subtle differences in their physical properties. We apply our methodology to find differences in the local structure.

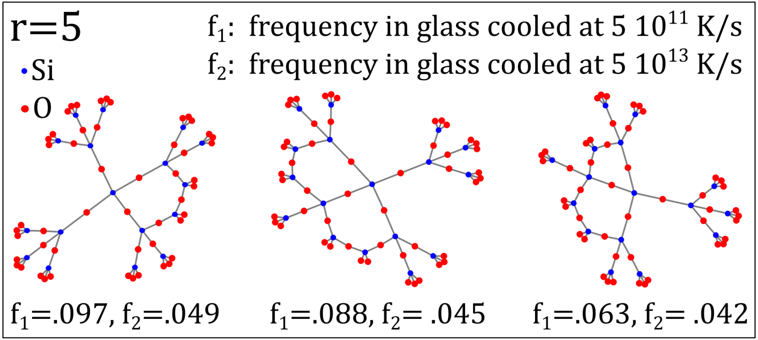
Data set: molecular dynamics simulations of silica glasses produced at three different cooling rates, with 10^5 silicon atoms for each cooling rate.

Application: Different Cooling Rates



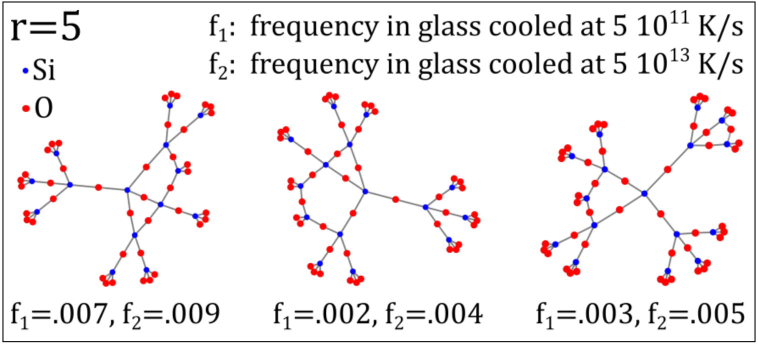
Our methodology distinguishes molecular dynamics simulations of glasses produced at different cooling rates.

Application: Different Cooling Rates



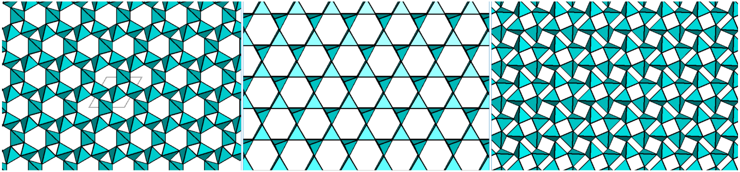
The most common types are the same in all preparations, but are less common for faster cooling rates and there is more probability mass in the tail. Interpretation: glasses produced at faster cooling rates are more disordered.

Application: Different Cooling Rates



Topological types that are over-represented in glasses produced at a faster cooling rate tend to have multiple relatively short rings.

Crystalline SiO_2 and Radius 6



Proposed application: detect local environments related to crystal nucleation.

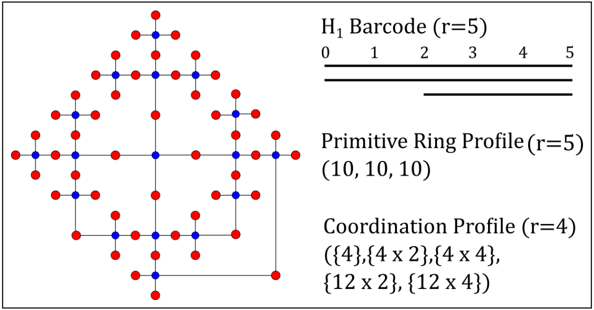
Crystalline forms of SiO_2 such as quartz and cristobalite are indistinguishable below radius $r = 6$.

The Combinatorial Explosion

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

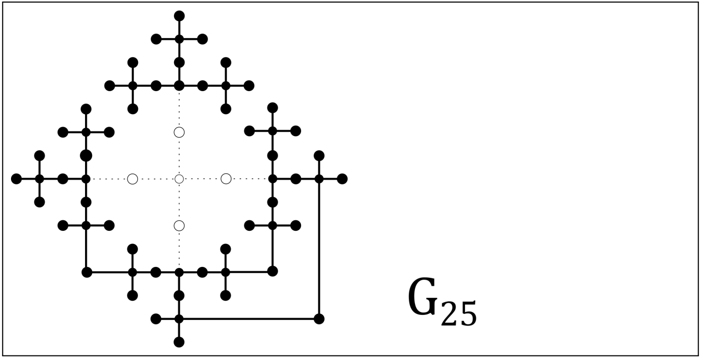
The number of graph isomorphism classes detected in a sample of 10^5 radius 6 environments approaches 10^5! A different approach is needed.

Coarser Equivalence Classes



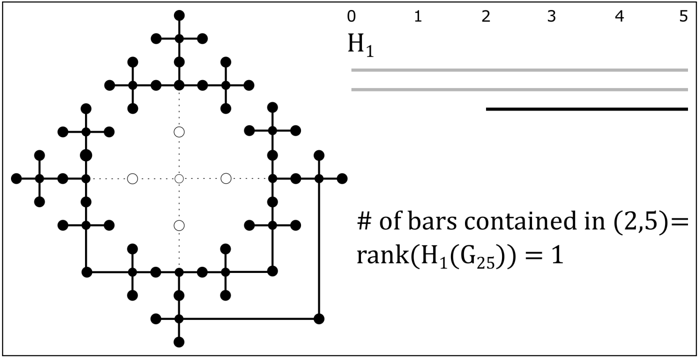
In *Statistical Topology of Bond Networks with Applications to Silica*, we study several notions of equivalence for local environments that are coarser than graph isomorphism. Of these, H_1 barcode equivalence is perhaps the most promising.

New Approach: H_1 Barcode



The shell annulus $G_{i,j}$ is induced subgraph on vertices at distances between i and j from the root.

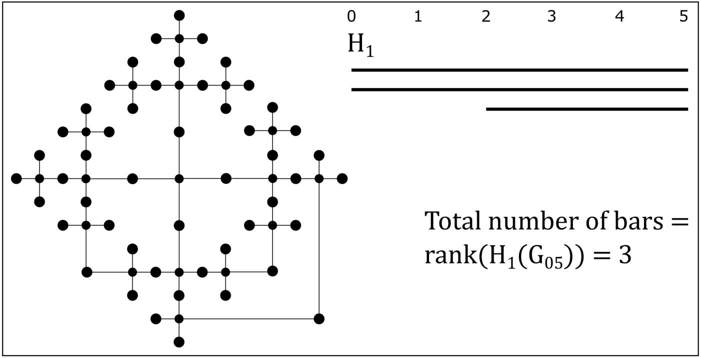
New Approach: H_1 Barcode



The H_1 barcode is a set of intervals so that

$$\text{rank } H_1(G_{i,j}) = \# \text{ of intervals contained in } (i,j).$$

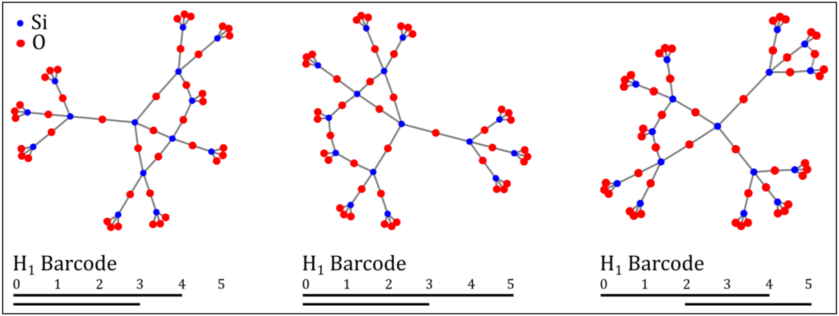
New Approach: H_1 Barcode



The H_1 barcode is a set of intervals so that

$$\text{rank } H_1(G_{i,j}) = \# \text{ of intervals contained in } (i, j).$$

Application: Different Cooling Rates

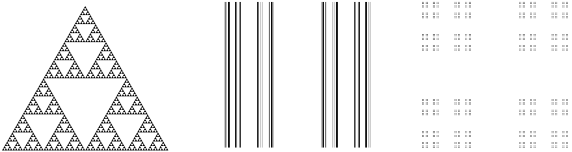


At radius 6, the H_1 barcode differentiates glasses produced at differing cooling rates as well as various crystalline forms of silica. In *Statistical Topology of Bond Networks with Applications to Silica*, we compare its performance to other coarser notions of equivalence using tools from information theory.

Future Directions

- More applications to glasses: finding correlations between frequencies of local environments and physical properties, phase transitions (glass transition, crystallization), optical properties.
- Applications to other systems: zeolites, metallic glasses/sphere packings, quasicrystals, circulatory networks, ...
- Physical interpretation of the H_1 barcode? (Degrees of freedom per atom.)
- Generalization of the H_1 barcode to higher dimensional homology.
- Thank you for your attention!

PH complexity

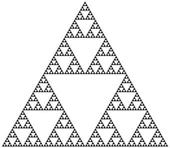
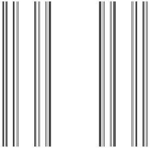
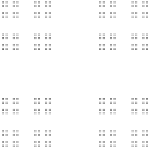


Definition (MacPherson–S.,2012)

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

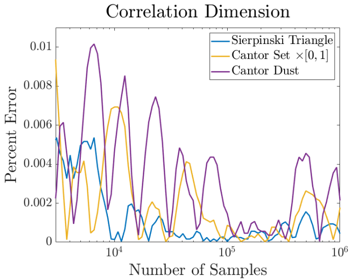
Measures the complexity of a shape, rather then the dimension.

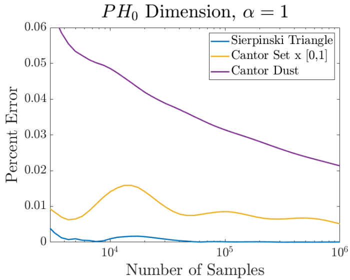
PH complexity

Example	True Dim.	$\mathrm{comp}_{PH_0}(X)$	$\mathrm{comp}_{PH_1}(X)$
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
$C \times C$	$\frac{2\log(2)}{\log(3)}$	$\frac{2\log(2)}{\log(3)}$	$\frac{2\log(2)}{\log(3)}$

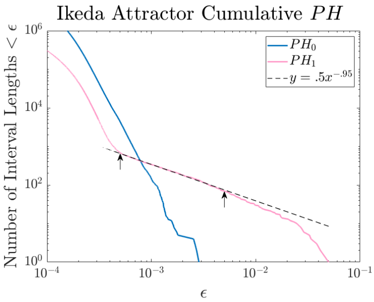
An Indicator of Difficulty?





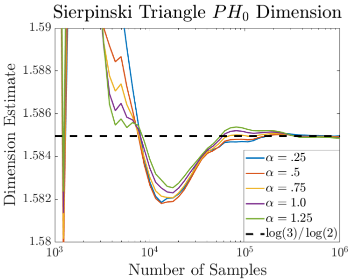
$$\text{comp}^0_{PH}(S) < \text{comp}^0_{PH}(C \times I) < \text{comp}^0_{PH}(C \times C)$$

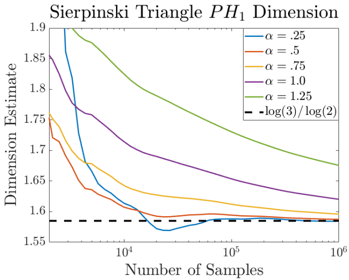
Ikeda attractor



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

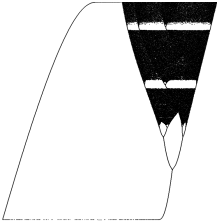
Dependence on α

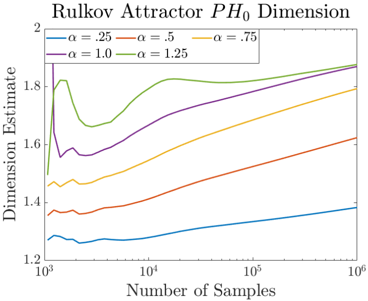




For self-similar fractals, lower values of α produced better convergence for $\dim^i_{PH}$ when $\text{comp}^i_{PH} \neq 0$, but there wasn't much difference otherwise.

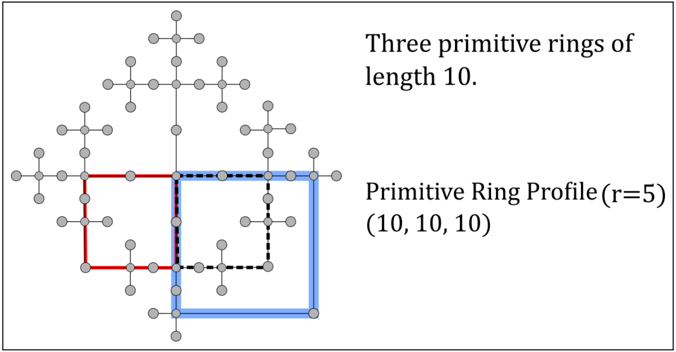
Dependence on α





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

Classical Approach: Ring Statistics



Ring statistics are perhaps the most commonly used structural descriptor in the oxide glass literature (i.e. how many rings of a given length?). There is considerable disagreement over which rings are important. Here, we count the number of "primitive rings" containing the root and call this the primitive ring profile.