Statistics, Topology and Data Analysis TAMIDS @TEXAS A & M

Vasileios Maroulas

University of Tennessee

February 26, 2021

Joint with:

- **Materials Application**
  - David Keffer (UTK)
  - Peter Liaw (UTK)
  - Piotr Luszceck (UTK)
  - Kody Law (Manchester)
  - Cassie Micucci (Eastman)
  - Adam Spannaus (ORNL)
- **Bayesian and TDA**
  - Farzana Nasrin (UH)
  - Chris Oballe (Notre Dame)

Funded by:

ARO W911NF-17-1-0313; W911NF-21-1-0094

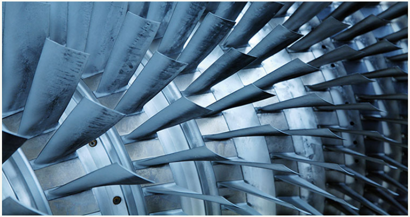
NSF MCB-1715794; DMS-1821241; DMS-2012609

ARL W911NF-17-2-0141; W911NF-19-2-0328

High Entropy Alloys

High Entropy Alloys (HEA) are a new, circa 2004, class of materials with unique properties

- Formed by mixing 5 or more elements
- Strength increased as temperature decreased to -321°F.
- Hardness increased as material was rolled to 0.07 mm, from an original thickness of 3mm
- Corrosion, oxidation resistance



Atom Probe Tomography (APT)

Local structure via APT to reconstruct a 3D atomic map.

- This process recovers approximately 10^8 data points, BUT
- Approximately 65% of the original data is not captured
- Recovered data is corrupted by noise
- **Uncover their true lattice structure from the APT dataset.**

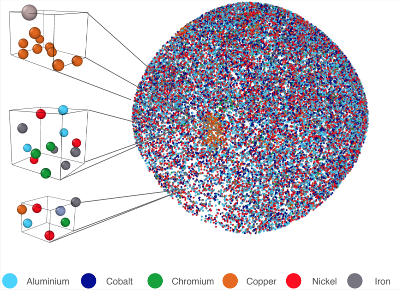


Figure: Image of the HEA $Al_{1.3}CoCrCuFeNi$ as seen via APT (Santodonato et al, 2015) with atomic neighborhoods shown in detail on the left. Certain patterns with distinct crystal structures exist, e.g., the orange region is copper-rich (left), but overall no pattern is identified.

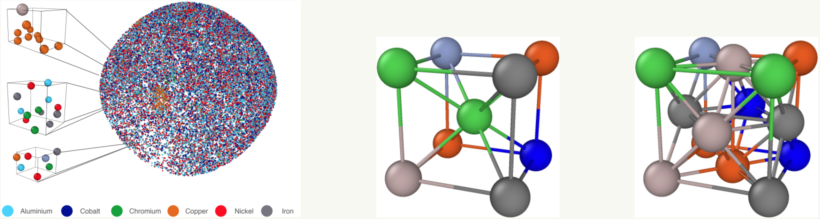
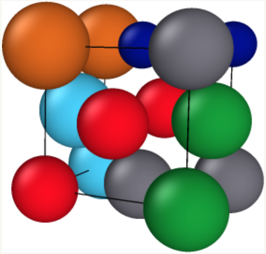
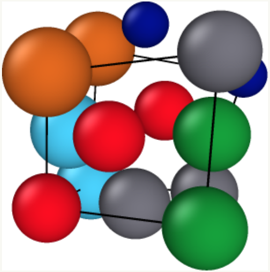


Figure: *Left:* Same image of HEA from APT data with atomic neighborhoods shown in detail on the left. Putting a single atomic cubic unit cell under a microscope, the true crystal structure of the material, which could be either *Center:* body-centered cubic (BCC) or *Right:* face-centered cubic (FCC) , is not revealed. This distinction is obscured due to further experimental noise. Notice there is an essential topological difference between the two structures: The BCC structure has one atom at its center, whereas the FCC is hollow in its center, but has one atom in the center of each of its faces.

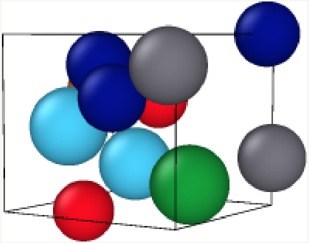
High Entropy Alloys



(a) Idealized FCC cell



(b) Distorted HEA FCC lattice

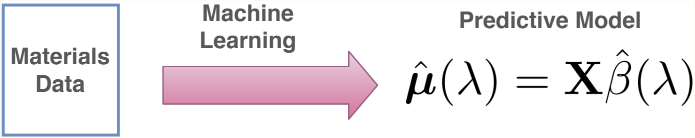


(c) FCC cell from APT experiment

Machine Learning for Materials Science

Applications of Machine Learning in Materials Science:

- Regression Modeling Steel Fatigue Prediction (Argawal et al., 2014)
- Materials Property Prediction (Zhou et al., 2018)
- Crystal Structure Classification (Zilletti et al., 2018)
- Microstructural Characterization of Neutron Scattering Data (deAlbuquerque et al., 2008)

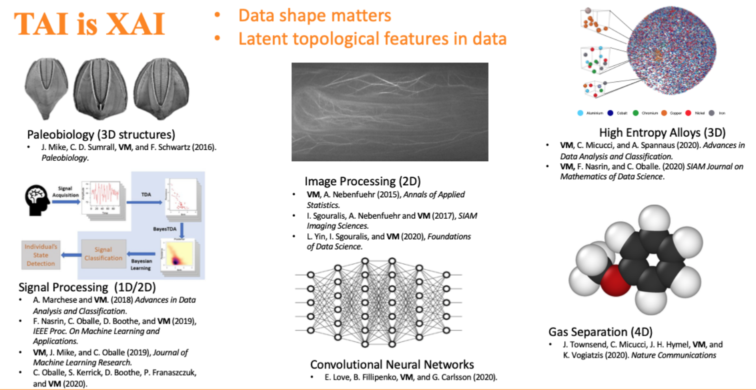


- Crystal Structure of HEAs is the dominant factor in determining the mechanical properties

Classification of crystal structures

- Two classes representing the crystal structure embedded in local neighborhoods of HEAs.
- Goal is to help material scientists to automatically classify into FCC vs BCC

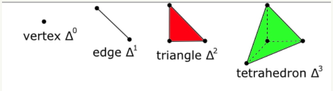
- Merge statistics and topology to understand the geometry of data and classify them.
- TDA/TAI has recently been introduced to several data problems.



- Moving into a quantum computing framework

Simplicial Complex

- Simplicial complexes are discretizations of real-life shapes
- Generalization of graphs with higher order relationships among the nodes.
- A simplicial complex is the union of simple pieces (simplices) i.e. vertices, edges, triangles etc.



- A face of $k-$simplex are all the $(k-1)-$simplex.
- Two simplices must intersect at a common face or not at all.

Construction of Simplicial complexes for data

Start with a point-cloud Π and create an abstract representation of vertices one for each point in your Π.

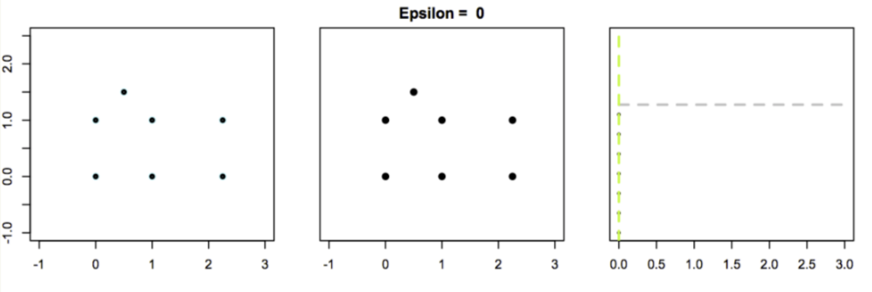


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

Construction of Simplicial complexes for data

Create circles of radius ϵ centered at each point.

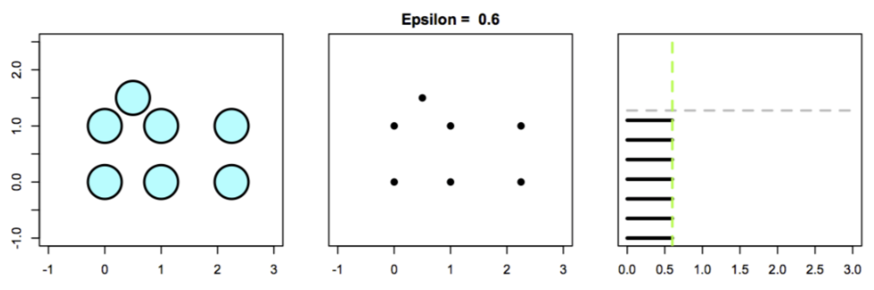


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

Construction of Simplicial complexes for data

Increase radius ϵ

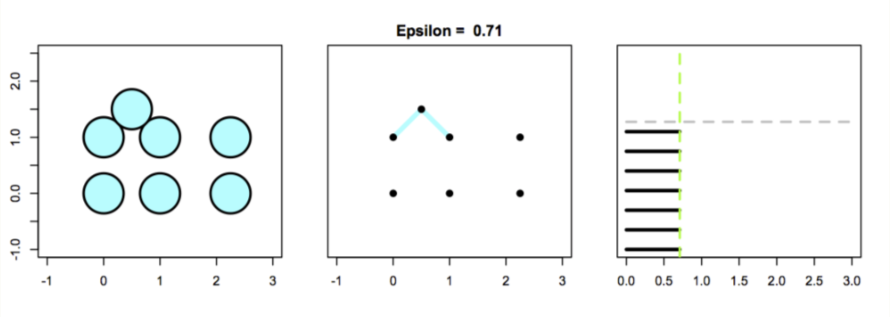


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

Construction of Simplicial complexes for data

Add edges between vertices v_i and v_j if the corresponding circles intersect.

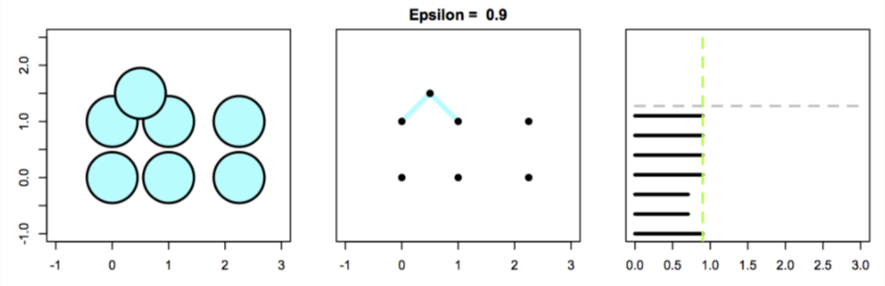


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

Construction of Simplicial complexes for data

- Add edges between vertices v_i and v_j if the corresponding circles intersect.
- Add triangles between vertices v_i, v_j and v_k if all three circles intersect, etc.

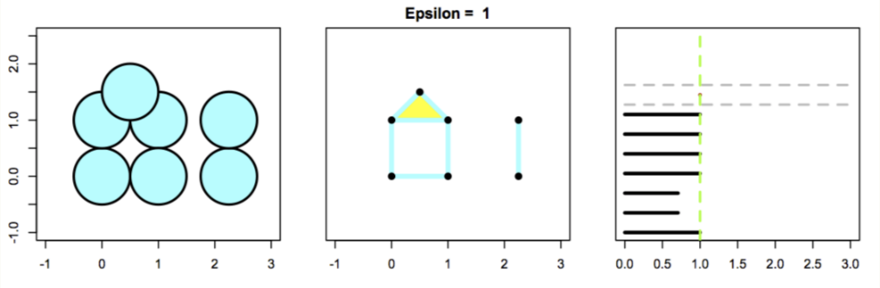


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

Construction of Simplicial complexes for data

- Add edges between vertices v_i and v_j if the corresponding circles intersect.
- Add triangles between vertices v_i, v_j and v_k if all three circles intersect, etc.

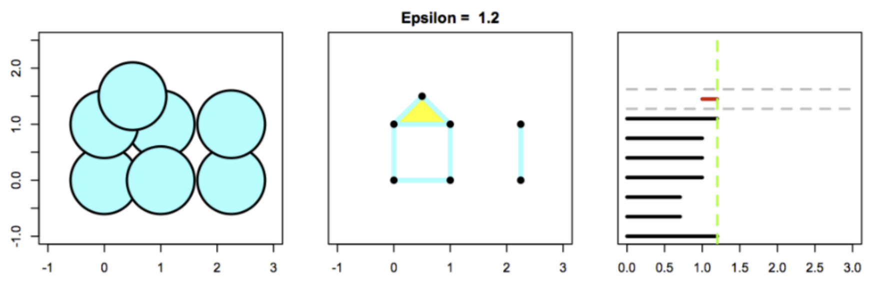


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

Construction of Simplicial complexes for data

- Add edges between vertices v_i and v_j if the corresponding circles intersect.
- Add triangles between vertices v_i, v_j and v_k if all three circles intersect, etc.

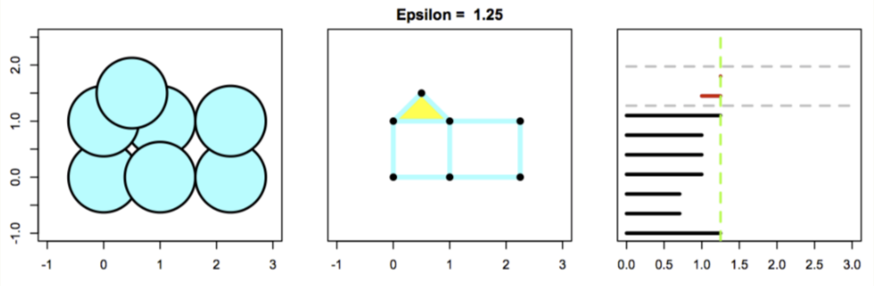


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

Construction of Simplicial complexes for data

- Add edges between vertices v_i and v_j if the corresponding circles intersect.
- Add triangles between vertices v_i, v_j and v_k if all three circles intersect, etc.

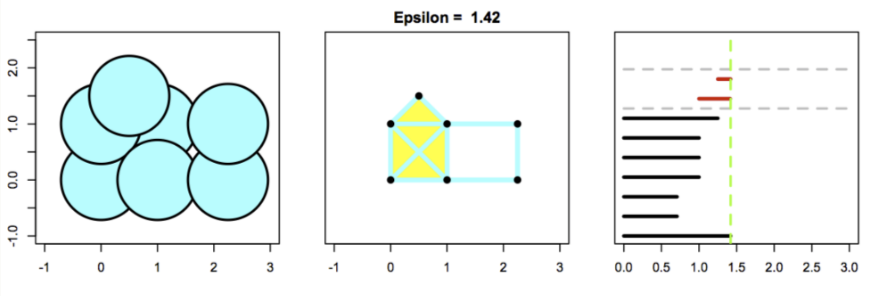


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

Construction of Simplicial complexes for data

Add triangles between vertices v_i, v_j and v_k if all three circles intersect, etc.

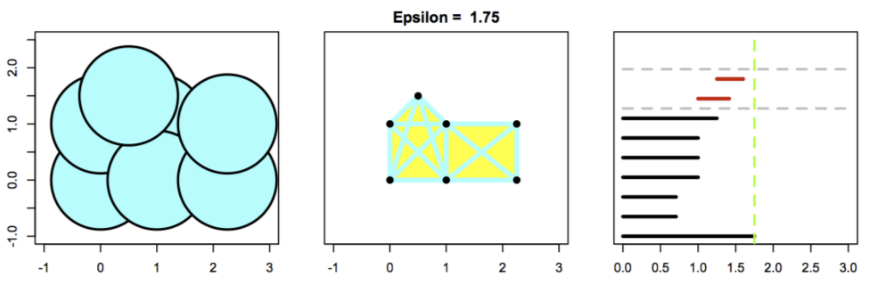
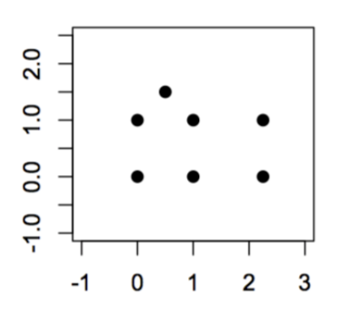
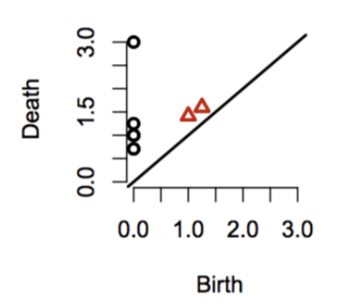


Figure: *Left:* Point Cloud; *Center:* Simplicial Complex; *Right:* Barcodes

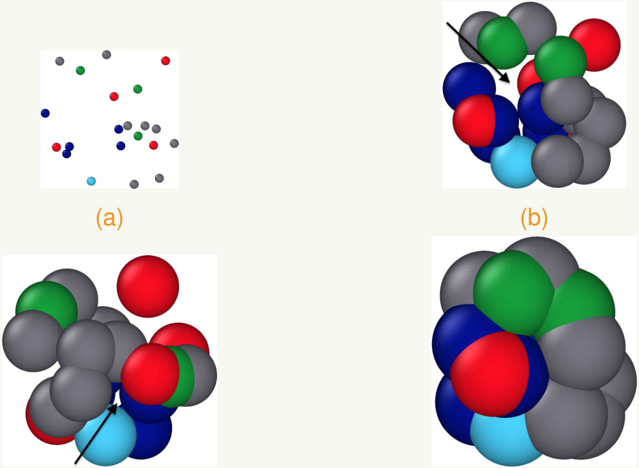
Persistence Diagrams

- Interested in is the *persistence* of the Betti numbers (number of connected components; number of holes).
- When do different connected components/holes form and how long do they last (with respect to ϵ)?
- The Betti numbers compactly encoded in a 2-dim plot which provides the birth time vs death time of these features



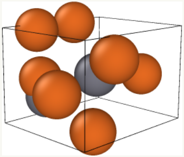


Vietoris-Rips Complex

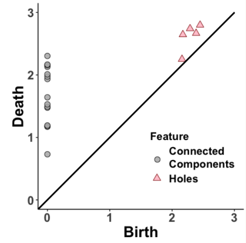


(a) (b) (c) (d)

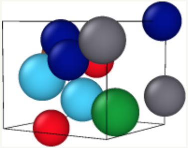
Persistence Diagrams for BCC and FCC Cells



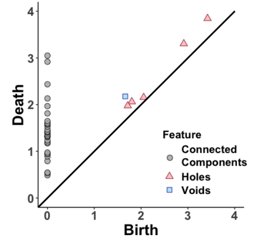
(a) BCC neighborhood, from APT experiment



(b) BCC Persistence Diagram



(c) FCC neighborhood, from APT experiment



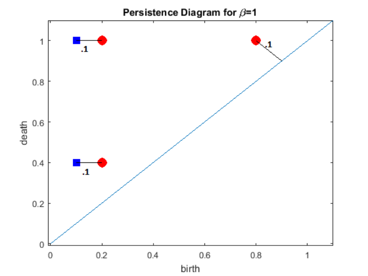
(d) FCC Persistence Diagram

Wasserstein Distance

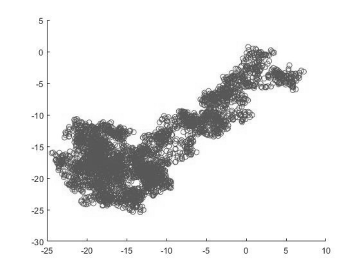
- Wasserstein Distance:

$$W_p(D_1, D_2) = (\inf_{\gamma} \sum_{x \in D_1} ||x - \gamma(x)||_{\infty}^{p})^{\frac{1}{p}}$$

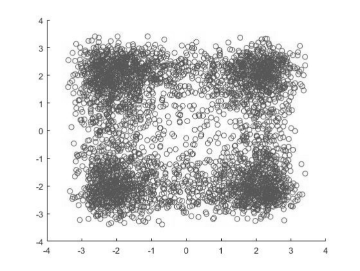
  where γ ranges over all bijections from D_1 to D_2.
- Penalty of unmatched points: distance to the diagonal. Matching to the diagonal is allowed in order to ensure bijections γ between D_1, D_2 exist.
- Assume ∞ many points along the diagonal of each persistence diagram with ∞ multiplicity
- No explicit penalty for cardinality differences between PDs



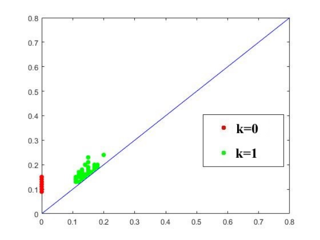
Example I



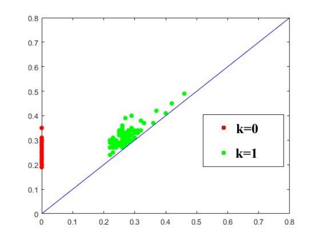
(a)



(b)

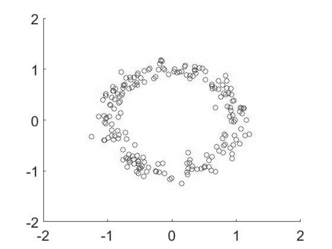


mean = 208.3; std=11.22

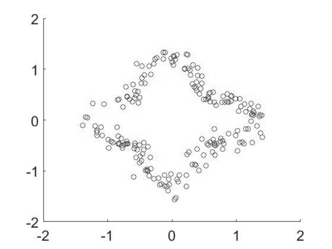


mean = 240; std=19.48

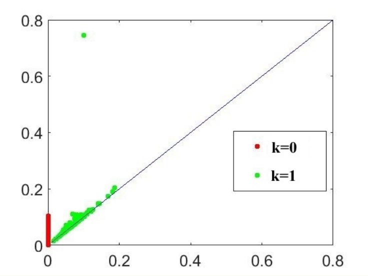
Example II



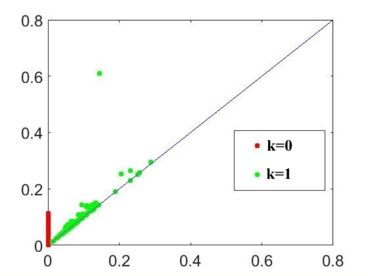
(a)



(b)



mean = 298.32; std=18.61



mean = 295.72; std=19.53

Need a distance

- Accounts for different cardinalities among persistence diagrams
- Penalizes outliers as well as the Wasserstein distance, but
- Bypasses the matching to the diagonal of persistence diagrams
- Differences in cardinality and geometry plays a role in the classification problem.
- The change in geometry between the two point cloud data is captured in the different behavior of the small persistence points.
- Other studies similarly arguing: Xia and Wei (2014); Robins and Turner (2016) ; Bubenik (2017)

Different Distance

Lemma 2.1 (A. Marchese and VM, 2018)

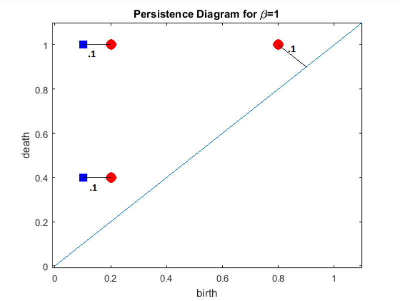
- *Given two persistence diagrams* $\mathbb{D}_1, \mathbb{D}_2 \in P_{W,k}$ *(space of PDs) s.t.* $|\mathbb{D}_1| = n \leq m = |\mathbb{D}_2|$
- $(x_1, ..., x_n) \in \mathbb{D}_1$, $(y_1, ..., y_m) \in \mathbb{D}_2$
- *Take* $c > 0$ *and* $1 < p < \infty$ *be fixed parameters and* Π_m *is the set of permutations of* $(1, ..., m)$.

$$d_p^c(\mathbb{D}_1, \mathbb{D}_2) = \left(\frac{1}{m}\left(\min_{\pi \in \Pi_m} \sum_{i=1}^{n} \min(c, ||x_i - y_{\pi(i)}||_\infty)^p + c^p(m-n)\right)\right)^{\frac{1}{p}}.$$

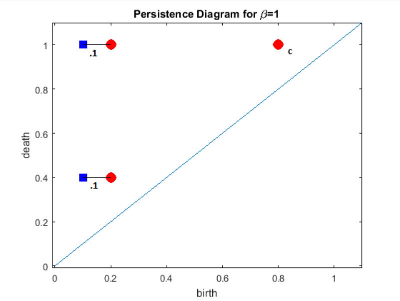
Then d_p^c *is a metric.*

- A. Marchese and **VM**. Signal classification with a point process distance on the space of persistence diagrams. *Advances in Data Analysis and Classification* 12 (3), pp 657-682, 2018.

Different Distance



(a) Wasserstein Distance



(b) d_p^c Distance

Different Distance

$$d_p^c(\mathbb{D}_1, \mathbb{D}_2) = (\frac{1}{m}(\min_{\pi \in \Pi_m} \sum_{i=1}^{n} \min(c, ||x_i - y_{\pi(i)}||_\infty)^p + c^p(m-n)))^{\frac{1}{p}}$$

- As p increases, the penalty for matching points is higher.
- As c increases, differences in cardinality penalized more.
  - *Smaller* c important for *small geometric differences*
  - *Larger* c vital for differentiating between *large geometric difference*

Proposition 2.1 (Stability of d_p^c, VM, C. Micucci, and A. Spannaus, ADAC, 2020)

Suppose A, A_i *finite nonempty point clouds in* $\mathbb{R}^n$, $d_p^c(A, A_i) \to 0$ *as* $i \to \infty$. *Then,*

$$d_p^c(\mathbb{D}, \mathbb{D}_i) \to 0 \text{ as } i \to \infty$$

where $\mathbb{D}, \mathbb{D}_i$ *persistence diagrams created from the Vietoris-Rips complex for* A *and* A_i*.*

VM, C. Micucci, and A. Spannaus, A Stable Cardinality Distance for Topological Classification, *Advances in Data Analysis and Classification,* 2020.

d_p^c Distance

Lemma 2.2 (A. Marchese and VM, 2018)

$(P_{W,k}, d_p^c)$ is Polish.

- Given a complete metric space, we are interested in the notion of the "mean" of a set of persistence diagrams.

d_p^c Distance

Lemma 2.2 (A. Marchese and VM, 2018)

$(P_{W,k}, d_p^c)$ *is Polish.*

- Given a complete metric space, we are interested in the notion of the "mean" of a set of persistence diagrams.
- Consider means and variances in the Fréchet sense.

d_p^c Distance

Lemma 2.2 (A. Marchese and VM, 2018)

$(P_{W,k}, d_p^c)$ *is Polish.*

- Given a complete metric space, we are interested in the notion of the “mean” of a set of persistence diagrams.
- Consider means and variances in the Fréchet sense.
- Consider a probability measure $\mathcal{D}$ on the space of $(P_{W,k}, \mathcal{B}(P_{W,k}))$ where $\mathcal{B}(P_{W,k})$ is the Borel $\sigma-$algebra on $P_{W,k}$ such that

$$F_{P_{W,k}}(\mathbb{D}_1) = \int_{P_{W,k}} d_p^c(\mathbb{D}_1, \mathbb{D}_2)^2 d\mathcal{D}(\mathbb{D}_2) < \infty \qquad \forall \mathbb{D}_1 \in P_{W,k}$$

Fréchet Means

Definition 2.3

Given a probability space $(P_{W,k}, \mathcal{B}(P_{W,k}), \mathcal{D})$, the Fréchet variance of $\mathcal{D}$ is

$$Var_{\mathcal{D}} = \inf_{\mathbb{D} \in P_{W,k}} [F_{P_{W,k}}(\mathbb{D}) = \int_{P_{W,k}} d_p^c(\mathbb{D}, \mathbb{D}_2)^2 \mathcal{D}(d\mathbb{D}_2)]$$

and the Fréchet expectation or Fréchet mean of $\mathcal{D}$ is

$$\mathbb{E}(\mathcal{D}) = \{\mathbb{D} | F_{P_{W,k}}(\mathbb{D}) = Var_{\mathcal{D}}\}$$

Fréchet Means

Definition 2.3

Given a probability space $(P_{W,k}, \mathcal{B}(P_{W,k}), \mathcal{D})$, the Fréchet variance of $\mathcal{D}$ is

$$Var_{\mathcal{D}} = \inf_{\mathbb{D} \in P_{W,k}} [F_{P_{W,k}}(\mathbb{D}) = \int_{P_{W,k}} d_p^c(\mathbb{D}, \mathbb{D}_2)^2 \mathcal{D}(d\mathbb{D}_2)]$$

and the Fréchet expectation or Fréchet mean of $\mathcal{D}$ is

$$\mathbb{E}(\mathcal{D}) = \{\mathbb{D} | F_{P_{W,k}}(\mathbb{D}) = Var_{\mathcal{D}}\}$$

- Fréchet means can be thought of as a generalization of centroids to metric spaces.

Fréchet Means

Definition 2.3

Given a probability space $(P_{W,k}, \mathcal{B}(P_{W,k}), \mathcal{D})$, the Fréchet variance of $\mathcal{D}$ is

$$Var_{\mathcal{D}} = \inf_{\mathbb{D} \in P_{W,k}} \left[F_{P_{W,k}}(\mathbb{D}) = \int_{P_{W,k}} d_p^c(\mathbb{D}, \mathbb{D}_2)^2 \mathcal{D}(d\mathbb{D}_2) \right]$$

and the Fréchet expectation or Fréchet mean of $\mathcal{D}$ is

$$\mathbb{E}(\mathcal{D}) = \{\mathbb{D} | F_{P_{W,k}}(\mathbb{D}) = Var_{\mathcal{D}}\}$$

- Fréchet means can be thought of as a generalization of centroids to metric spaces.

Theorem 2.4 (A. Marchese and **VM**, 2018)

Let $\mathcal{D}$ be a probability measure on $(P_{W,k}, \mathcal{B}(P_{W,k}))$. Then $\mathbb{E}(\mathcal{D}) \neq \emptyset$.

Classification Algorithm

- Fix β_l [β_0(connected components), β_1(holes), β_2(voids)]
- Take the PD training sets $T_{Y_1}^{\beta_l}, T_{Y_2}^{\beta_l}$ for each class.
- For new data x with corresponding β_l-persistence diagram $\mathbb{D}_x^{\beta_l}$, its distance from $\mathbb{D} \in T_{Y_k}^{\beta_l}$ is $d_p^c(\mathbb{D}_x^{\beta_l}, \mathbb{D})$.
- The average distance

$$d_{\beta_l}(x, Y_k) = \frac{1}{card(T_{Y_k}^{\beta_l})} \sum_{\mathbb{D} \in T_{Y_k}^{\beta_l}} d_p^c(\mathbb{D}_x^{\beta_l}, \mathbb{D})$$

- Assign the data x a label $\hat{Y}$ (one of $Y_1,\ Y_2$) defined by

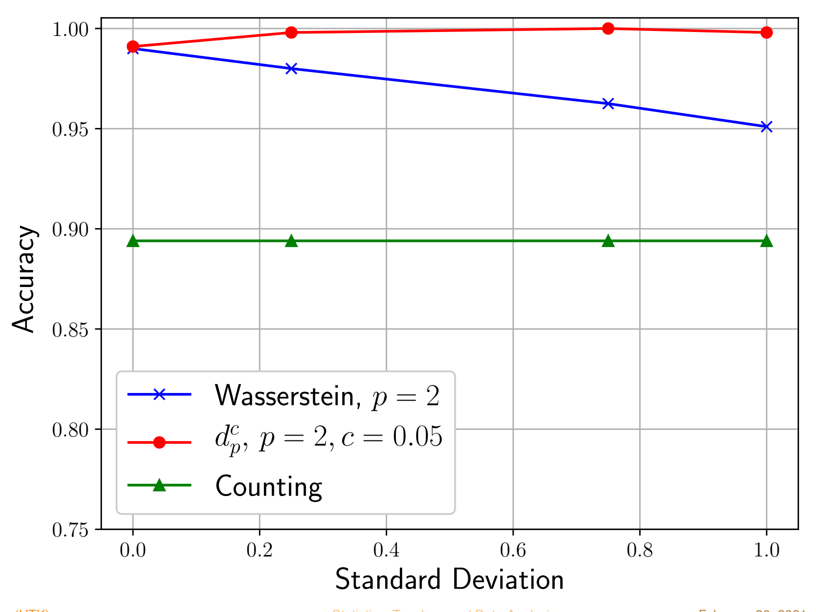
$$\hat{Y} = \text{argmin}_{1 \le k \le 2} \sum_{l=0}^{B_M} r_l d_{\beta_l}(x, Y_k)$$

where $\sum_{l=0}^{B_M} r_l = 1$ and r_l's are weights which determine how much each Betti number β_l is considered.

10-fold cross validation

- Generated 1000 unit neighborhoods (500 of each type)
- Data is partitioned into 10 different sets
- 9 of the partitions are used for training purposes
- 1 partition is used for testing
- Done 10 times so that every partition acts as the testing data exactly once
- The accuracy is averaged among all partitions

Results on Synthetic APT data



Statistics and Persistence Diagrams

- Summary statistics such as center and variance (Bobrowski et al., 2014; Mileyko et al., 2011; Turner et al., 2014; Marchese and **VM**, 2017)
- Birth and death estimates (Emmett et al., 2014)
- Confidence sets (Fasy et al., 2014)
- Need a framework to understand the above summary statistics through a single viewpoint

Bayesian framework for Persistence Diagrams

- First Bayesian discussion in TDA context: Y. Mileyko, S. Mukherjee, and J. Harer (2011)
- A conditional probability setting on PDs where the likelihood for the observed point cloud has been substituted by the likelihood for its associated topological summary

	Bayesian for RVs	**Bayesian for Random PDs**
Prior	Modeled by a density f	???
Likelihood	Depends on observed data	???
Posterior	Compute the posterior density	???

Recall: $f(x| \text{ data }) \propto \ell(\text{ data } |x) f(x)$

Prior Distribution

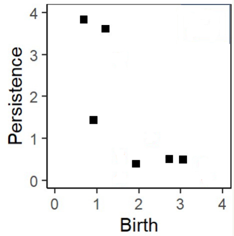


Figure: Sample PD from the prior

- Consider PDs as samples from a point process
- Poisson point process
- Need the intensity density $\lambda(\cdot)$ to characterize it
- Cardinality distribution: $c(n) = e^{-\mu} \frac{\mu^n}{n!}$ where $\mu := \int_{\mathbb{X}} \lambda(x)dx$
- Spatial distribution: $p(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{\lambda(x_i)}{\mu}$

- Another approach is to consider random set theory and establish kernels on the space of persistence diagrams
- **VM**, J. Mike, C. Oballe, Nonparametric Estimation of Probability Density Functions of Random Persistence Diagrams. *Journal of Machine Learning Research*, 20 (151), pp.1-49, 2019.

Likelihood

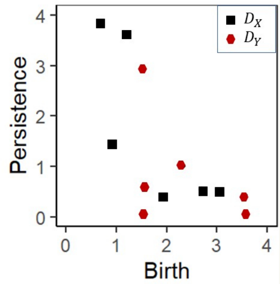


Figure: A sample D_X from prior Poisson PP $\mathcal{D}_X$ and an observed persistence diagram D_Y

- Marked point process
- Point process Ψ_M consists of points $(x_i, m(x_i)) \in \mathbb{X} \times \mathbb{M}$, where $m(x_i)$ are called marks.
- Ψ is a Poisson PP.
- Marks are drawn independently from a kernel $\ell : \mathbb{X} \times \mathbb{M} \to \mathbb{R}_{\geq 0}$.

Likelihood

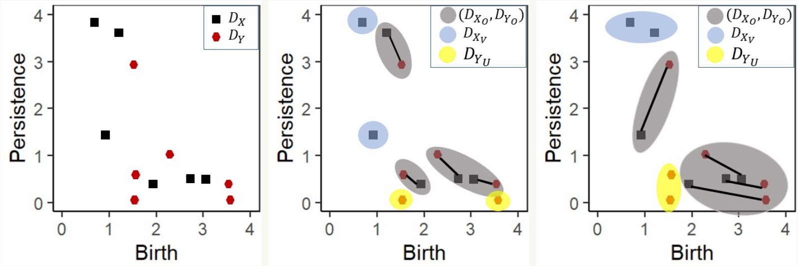


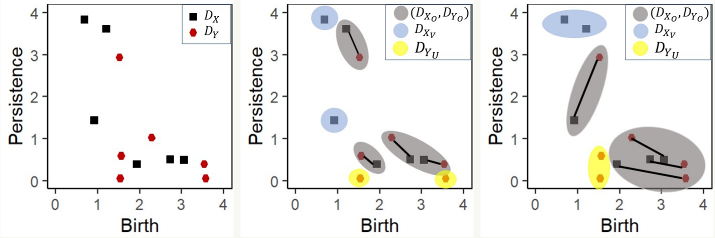
Figure: *(a) A sample D_X from prior Poisson PP $\mathcal{D}_X$ and an observed persistence diagram D_Y. (b) and (c) are the decomposition of D_X into D_{X_O} & D_{X_V} and D_Y into D_{Y_O} & D_{Y_U}.*

Bayes Theorem for Persistent Homology

Theorem 3.1 (VM, F. Nasrin, C. Oballe, SIMODS, 2020)

Let λ_{D_X} be the prior intensity, and ℓ the likelihood which is associated with the stochastic kernel of the marked point process. The posterior intensity is given by

$$\lambda_{\mathbb{D}_X|D_{Y_{1:m}}}(x) = \left(1-\alpha(x)\right)\lambda_{\mathcal{D}_X}(x) + \frac{\alpha(x)}{m}\sum_{i=1}^{m}\sum_{y\in D_{Y^i}} \frac{\ell(y|x)\lambda_{\mathcal{D}_X}(x)}{\lambda_{\mathcal{D}_{Y_U}}(y) + \int_{\mathcal{W}} \ell(y|u)\alpha(u)\lambda_{\mathcal{D}_X}(u)du}. \tag{1}$$



VM, F. Nasrin, and C. Oballe. A Bayesian Framework for Peristent Homology. *SIAM Journal on Mathematics of Data Science*, 2(1), pp. 48-74, 2020.

Conjugate family of priors

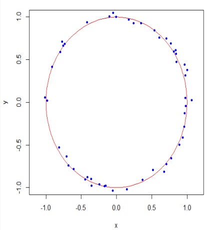
Corollary 3.2 (VM, F. Nasrin, C. Oballe, SIMODS, 2020)

Let the prior intensity λ_{D_X} be a Gaussian mixture, the likelihood ℓ associated with the stochastic kernel of the marked point process is a Gaussian density, then the posterior intensity,

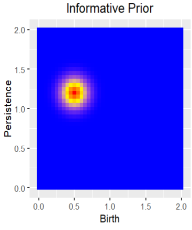
$$\lambda_{\mathcal{D}_X | D_{Y_{1:m}}}(x) = (1-\alpha)\lambda_{\mathcal{D}_X}(x) + \frac{\alpha}{m}\sum_{i=1}^{m}\sum_{y \in D_{Y_i}}\sum_{j=1}^{N} c_j^{x|y}\, \mathcal{N}^*(x; \mu_j^{x|y}, \sigma_j^{x|y} I);$$

VM, F. Nasrin, and C. Oballe. A Bayesian Framework for Peristent Homology. *SIAM Journal on Mathematics of Data Science*, 2(1), pp. 48-74, 2020.

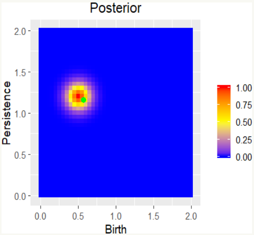
Example 1



1.0
0.5
0.0
-0.5
-1.0
y
-1.0
-0.5
0.0
0.5
1.0
x

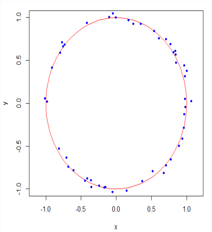


Informative Prior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0

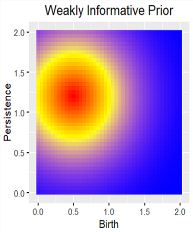


Posterior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0
1.00
0.75
0.50
0.25
0.00

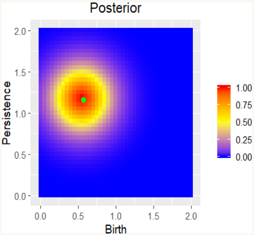
Example 1



y
x
-1.0
-0.5
0.0
0.5
1.0

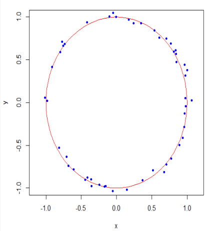


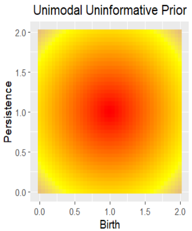
Weakly Informative Prior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0

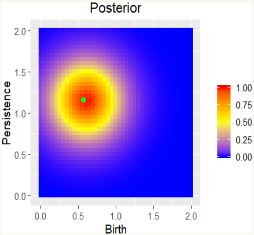


Posterior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0
1.00
0.75
0.50
0.25
0.00

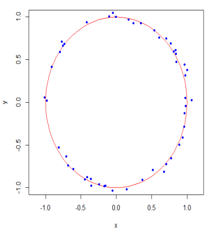
Example 1

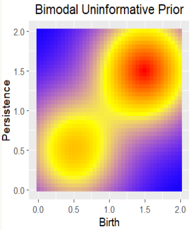

y
x
-1.0
-0.5
0.0
0.5
1.0

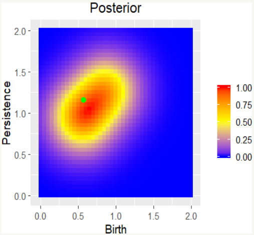

Unimodal Uninformative Prior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0


Posterior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0
1.00
0.75
0.50
0.25
0.00

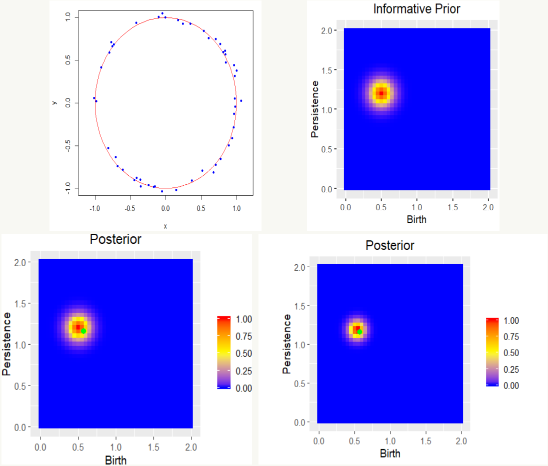
Example 1


1.0
0.5
0.0
-0.5
-1.0
y
-1.0
-0.5
0.0
0.5
1.0
x

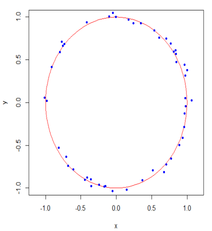

Bimodal Uninformative Prior
2.0
1.5
1.0
0.5
0.0
Persistence
0.0
0.5
1.0
1.5
2.0
Birth

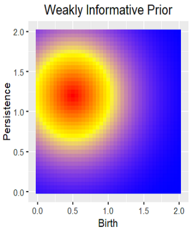

Posterior
2.0
1.5
1.0
0.5
0.0
Persistence
0.0
0.5
1.0
1.5
2.0
Birth
1.00
0.75
0.50
0.25
0.00

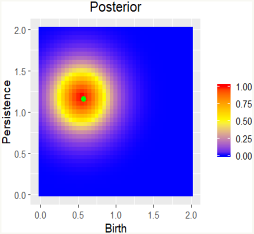
Example 2 (Trusting Data Less vs More)

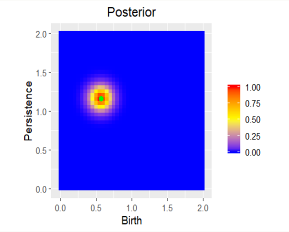


Example 2 (Trusting Data Less vs More)

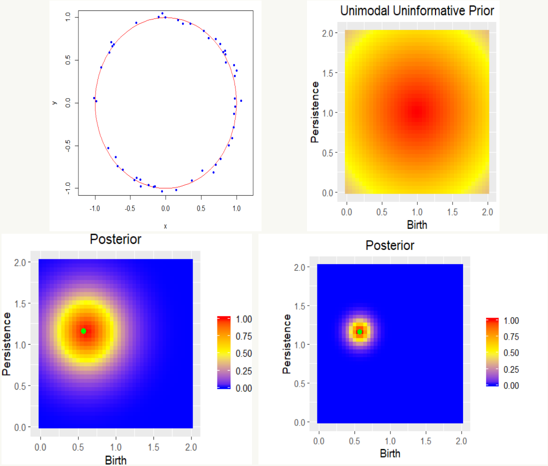




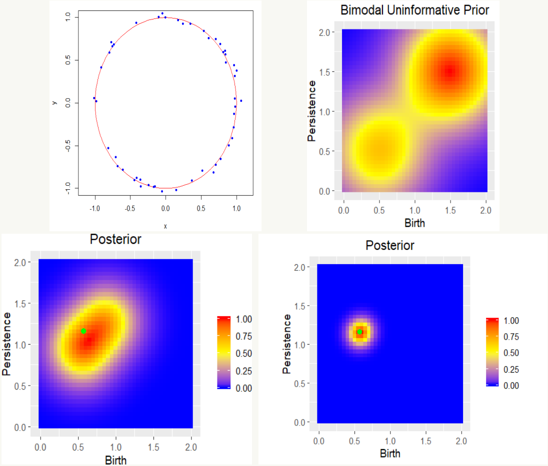




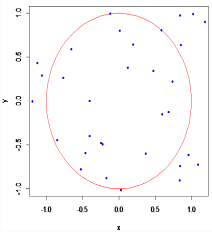
Example 2 (Trusting Data Less vs More)



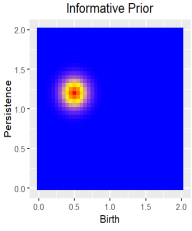
Example 2 (Trusting Data Less vs More)



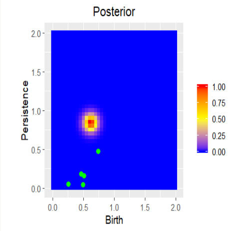
Example 3



1.0
0.5
0.0
-0.5
-1.0
y
-1.0
-0.5
0.0
0.5
1.0
x

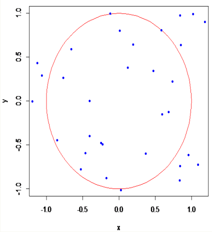


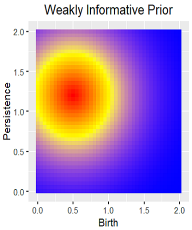
Informative Prior
2.0
1.5
1.0
0.5
0.0
Persistence
0.0
0.5
1.0
1.5
2.0
Birth

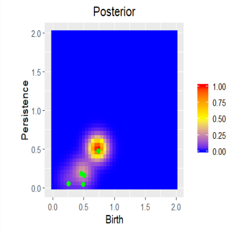


Posterior
2.0
1.5
1.0
0.5
0.0
Persistence
0.0
0.5
1.0
1.5
2.0
Birth
1.00
0.75
0.50
0.25
0.00

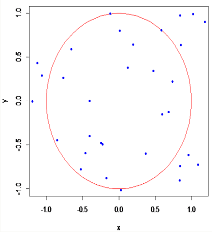
Example 3

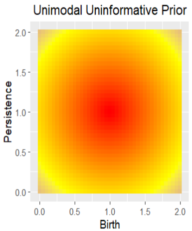

y
x
-1.0
-0.5
0.0
0.5
1.0

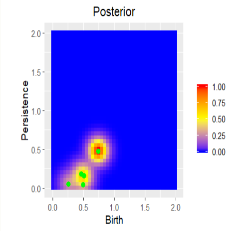

Weakly Informative Prior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0


Posterior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0
1.00
0.75
0.50
0.25
0.00

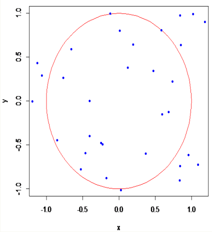
Example 3

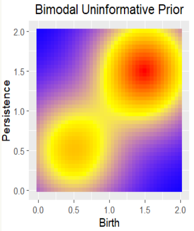

1.0
0.5
0.0
-0.5
-1.0
y
-1.0
-0.5
0.0
0.5
1.0
x

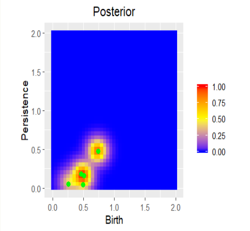

Unimodal Uninformative Prior
Persistence
Birth
2.0
1.5
1.0
0.5
0.0


Posterior
Persistence
Birth
2.0
1.5
1.0
0.5
0.0
1.00
0.75
0.50
0.25
0.00

Example 3


1.0
0.5
0.0
-0.5
-1.0
y
-1.0
-0.5
0.0
0.5
1.0
x


Bimodal Uninformative Prior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0


Posterior
Persistence
Birth
0.0
0.5
1.0
1.5
2.0
1.00
0.75
0.50
0.25
0.00

HEAs Classification

- Considered 100,000 of each crystal structure (synthesized at Liaw's research group and ORNL)

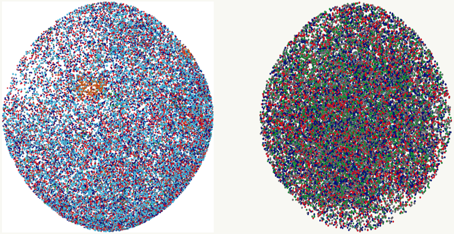


Figure: *Left:* BCC: $Al_{1.3}CoCrCuFeNi$ vs *Right:* FCC: $Al_{0.3}CoCrFeNi$. Note that the copper-rich FCC regions have been removed from the $Al_{1.3}CoCrCuFeNi$ as a preprocessing step

HEAs Classification

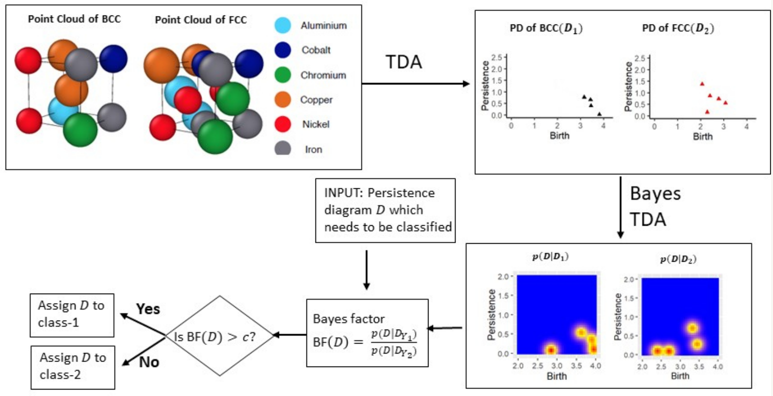


Figure: Flowchart for Classification Scheme

- Used 50% of data as training sets and 10-fold cross–validation
- Accuracy: 94%

Conclusion

- Classification of crystal structure of HEAs using statistical learning and topology
- Use d_p^c distance, or
- Use a Generalized Bayesian perspective allowing the flexibility to use historical data/or purely data driven approach via a uniform prior
- Computing ratios of posterior distributions of PDs.

TABLE: *The parallels between the Bayesian for RVs and for random PDs.*

	Bayesian for RVs	**Bayesian for Random PDs**
Prior	Modeled by a density f	Modeled by a PPP with intensity λ
Likelihood	Depends on observed data	ℓ that depends on observed PDs
Posterior	Compute the posterior density	A PPP with posterior intensity

- Install from Github using maroulaslab/BayesTDA.