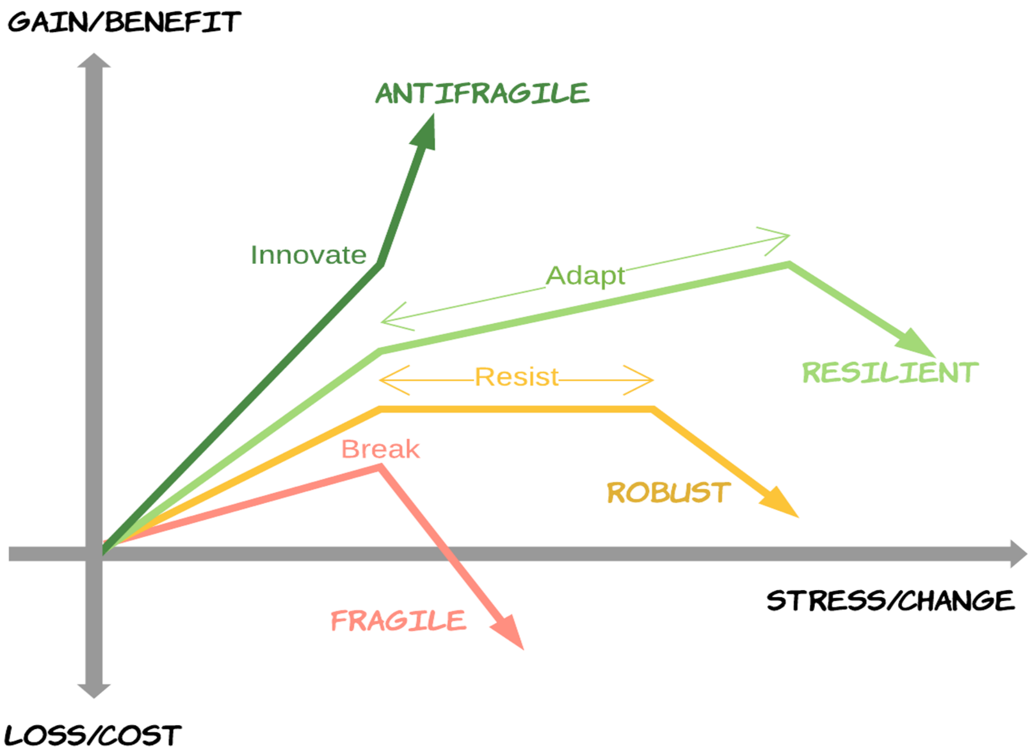
Antifragile Persistent Homology using Fisher Information

Alex Cole
UvA

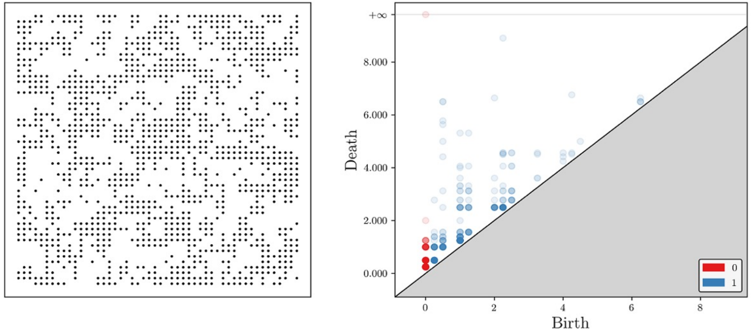
Jan Pieter vd Schaar
UvA

Karthik Viswanathan
UvA

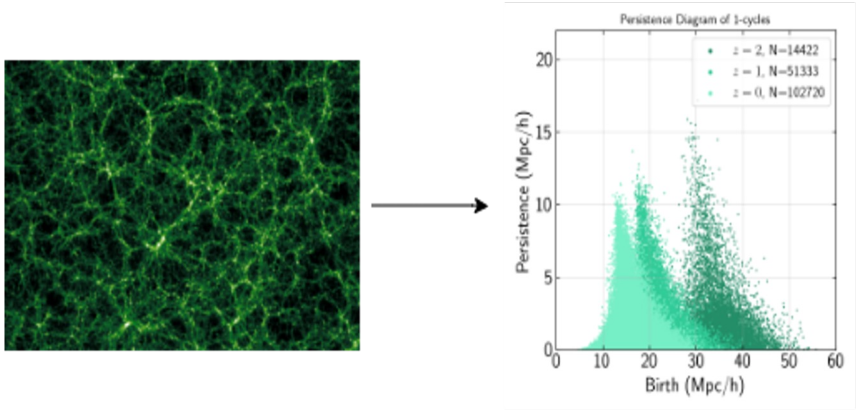


PC - Bilgin Ibryam

Motivation



PD for spins on a lattice. [1]

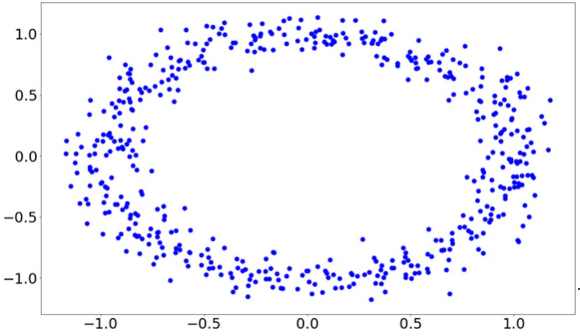


PD for LSS of universe. [2]

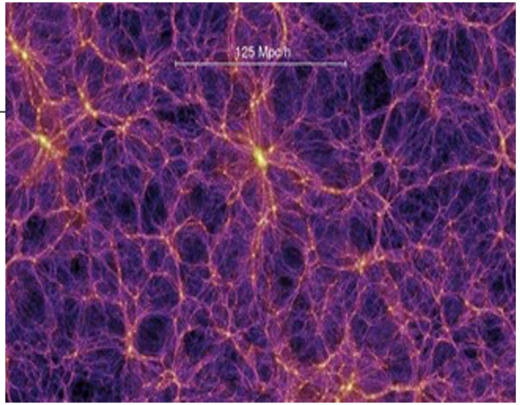
- Is there a **problem-adaptable** way of constructing filtrations and persistent summaries?

- How to *quantify* if the persistence diagram is a *good* summary of the input?

Problem Statement



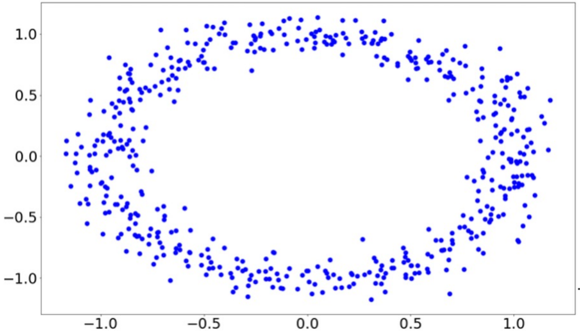
$$X \sim P(X|\theta)$$



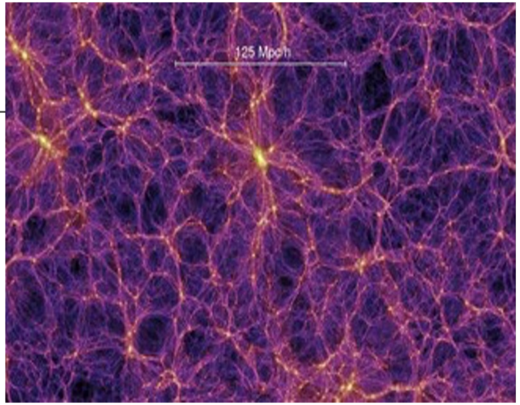
PC - Springel et al. 2015

- Given a point set drawn X from a distribution, in a scientific inference problem, find the filtration that is informative about θ.
- Eg. Points drawn from a circle (radius), LSS (f_{NL}), etc.

Problem Statement



$$X \sim P(X|\theta)$$



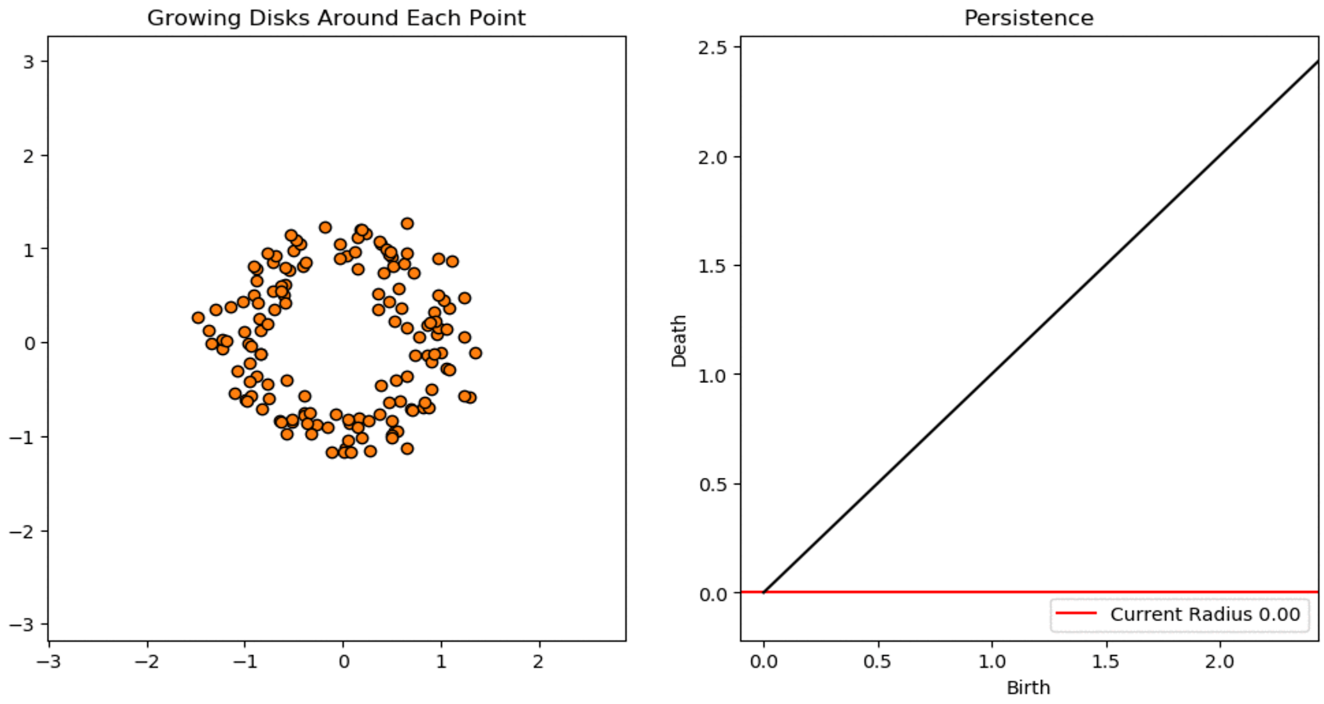
PC - Springel et al. 2015

- Given a point set drawn X from a distribution, in a scientific inference problem, find the filtration that is informative about θ.
- Eg. Points drawn from a circle (radius), LSS (f_{NL}), etc.
- Objective function - Fisher Information.

Outline

- Mathematical Preliminaries
  - Persistent Homology
  - Fisher Information
  - Vectorization of PD
- Finding the Optimal Filtration
  - The Pipeline
  - Examples
- Outlook

Persistent Homology - The Physics Way



- PD keeps track of the multiscale topology of point sets.
- It stores the scales at which topological features born (b) and get destroyed (d).

Gif Credit (GC) - Persistent Homology: A Non-Mathy Introduction with Examples.

Persistent Homology - The Math Way

- Filtrations - $f : K \to \mathbb{R}$

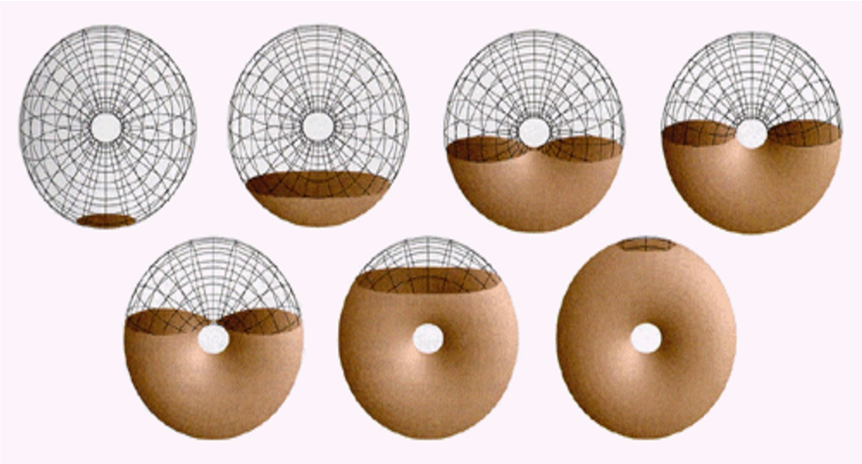
- Assigns a number to each simplex

  in K.

  $SubL(a) = f^{-1}[-\infty, a]$

- PH keeps track of changes in

  topology of SubL(a).

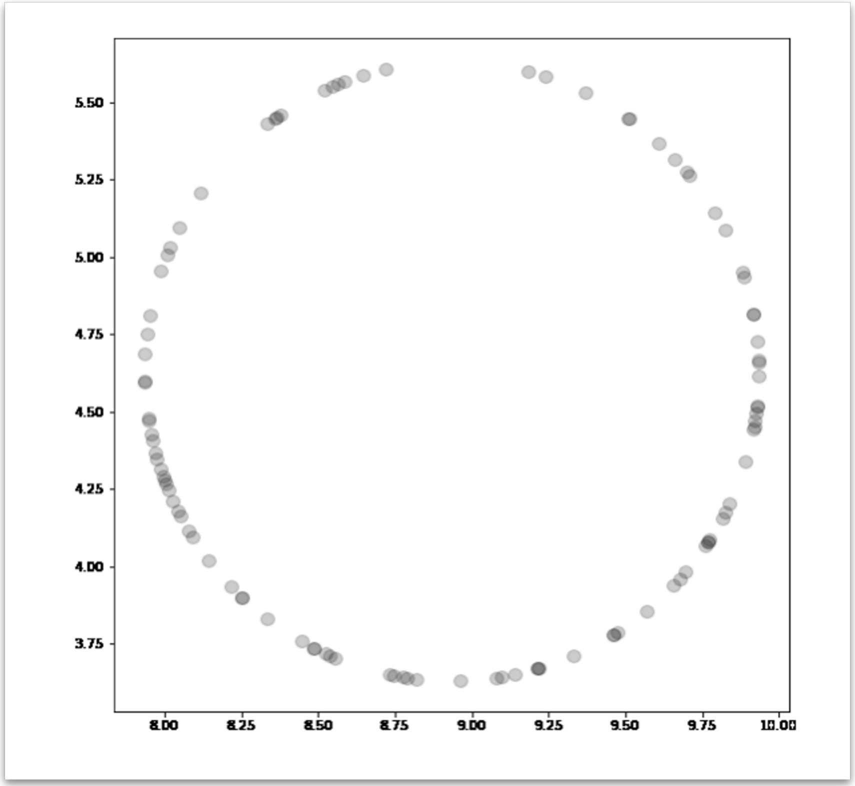


PC - Thomas Banchoff, *Slicing Doughnuts and Bagels*

Sublevel Sets - Example II

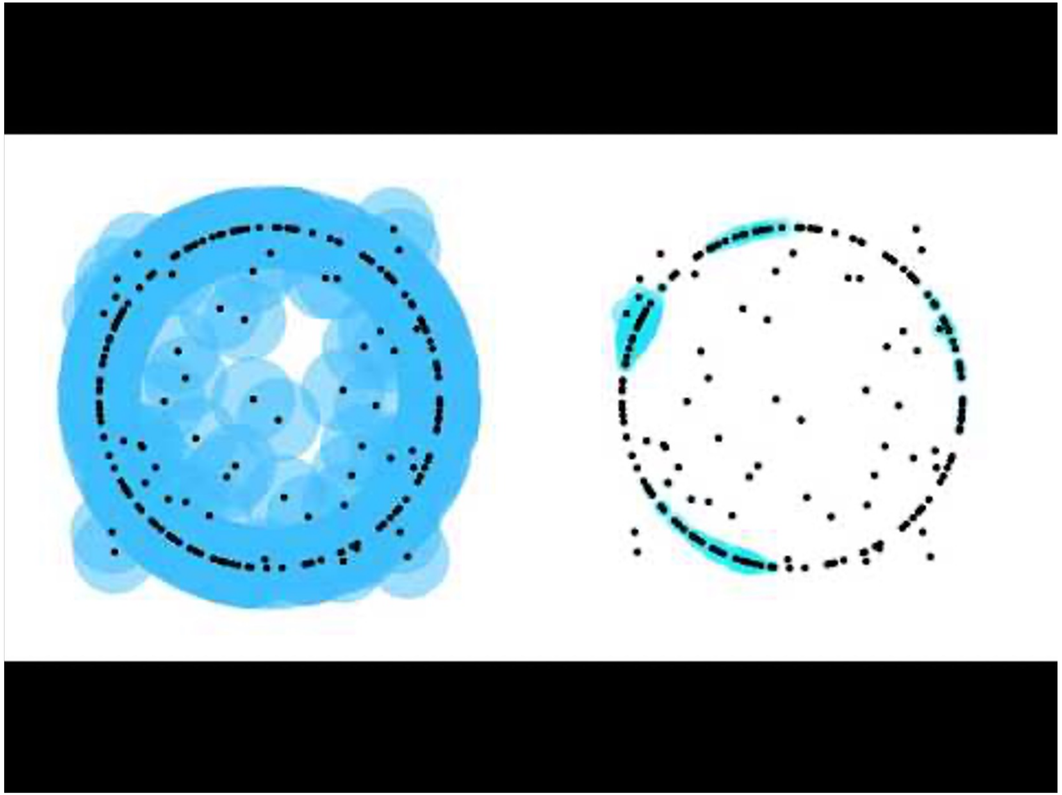
GC - Alex Cole

- Evolution of sublevel sets as the *level* “a” is varied.
- Shaded region - sublevel sets.
- A simplex (s) is added at *level* “a” if f(s) < a.



Variational Filtrations

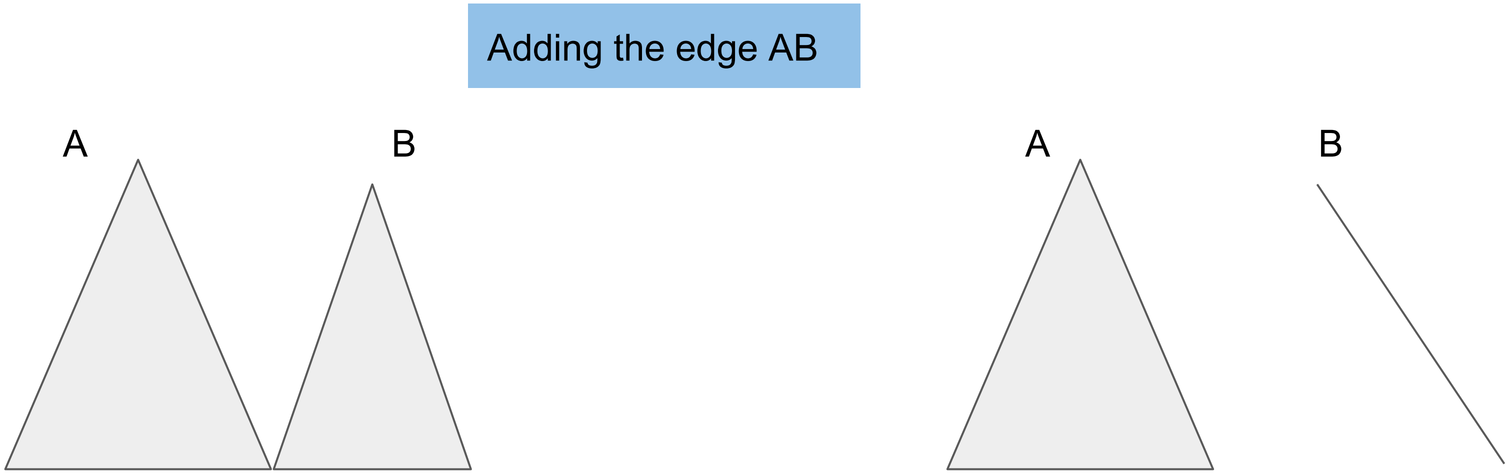
- Uniform growth of balls → Parametrized.
- The filtration f is parametrized by a variational family ϕ.
- Right - DTM filtrations.
    - Density controlled.
    - Less dense implies late start of growth of balls.



GC - Raphaël Tinarrage

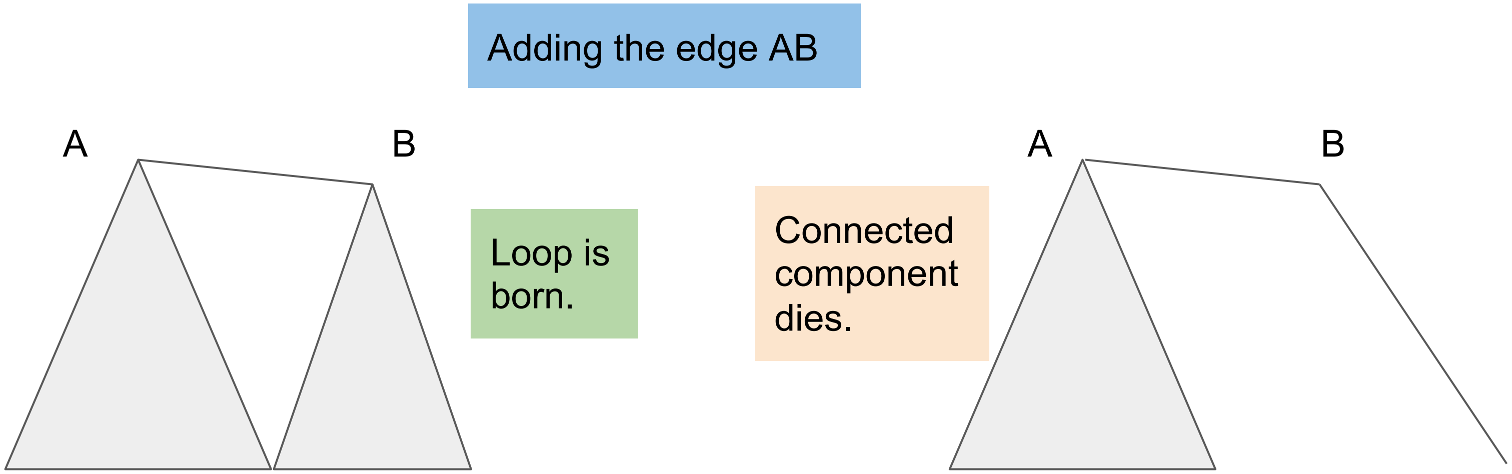
Persistence Pairs

- Adding a p-simplex to K *always* either creates a p cycle, or destroys a (p - 1)-cycle.
  - Eg. Adding an edge (1-simplex) can create a loop (1 cycle) or destroy a component (0 cycle).



Persistence Pairs

- Adding a p-simplex to K *always* either creates a p cycle, or destroys a (p - 1)-cycle.
  - Eg. Adding an edge (1-simplex) can create a loop (1 cycle) or destroy a component (0 cycle).

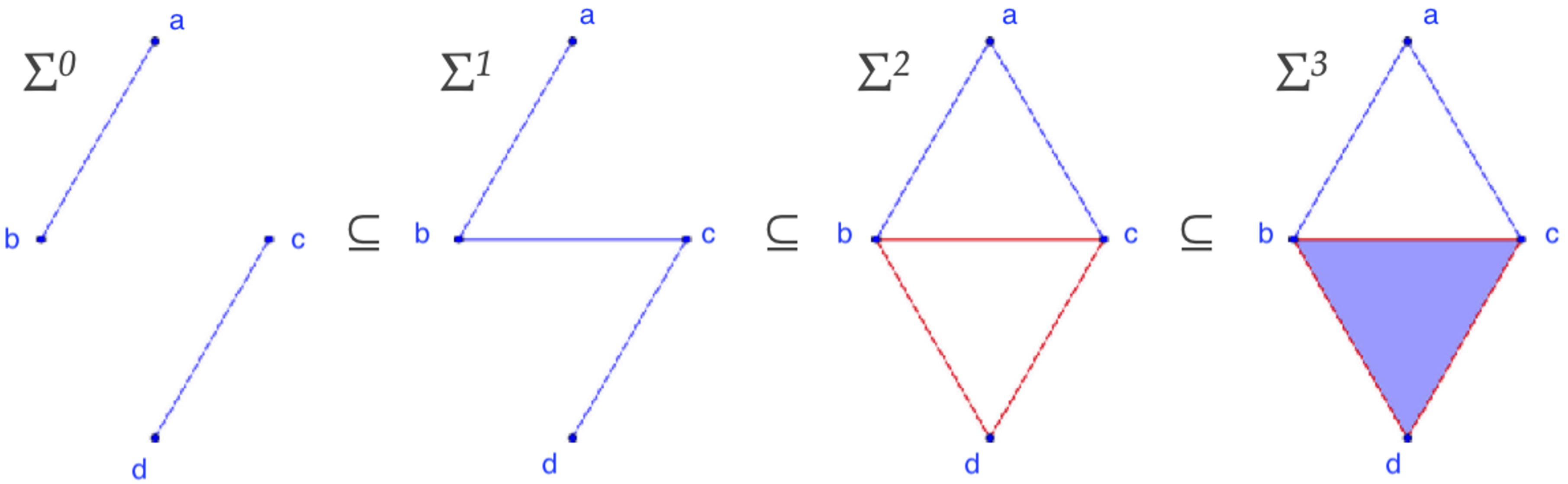


Persistence Pairs

- Adding a p-simplex to K *always* either creates a p cycle, or destroys a (p - 1)-cycle.
    - Eg. Adding an edge (1-simplex) can create a loop (1 cycle) or destroy a component (0 cycle).
- Persistence pairs are simplices that create (σ_b) and destroy (σ_d) topological features.
- The birth and death of a topological feature can be written in terms of the filtration “f” and the persistence pairs.

$$b = f(\sigma_b), d = f(\sigma_d)$$

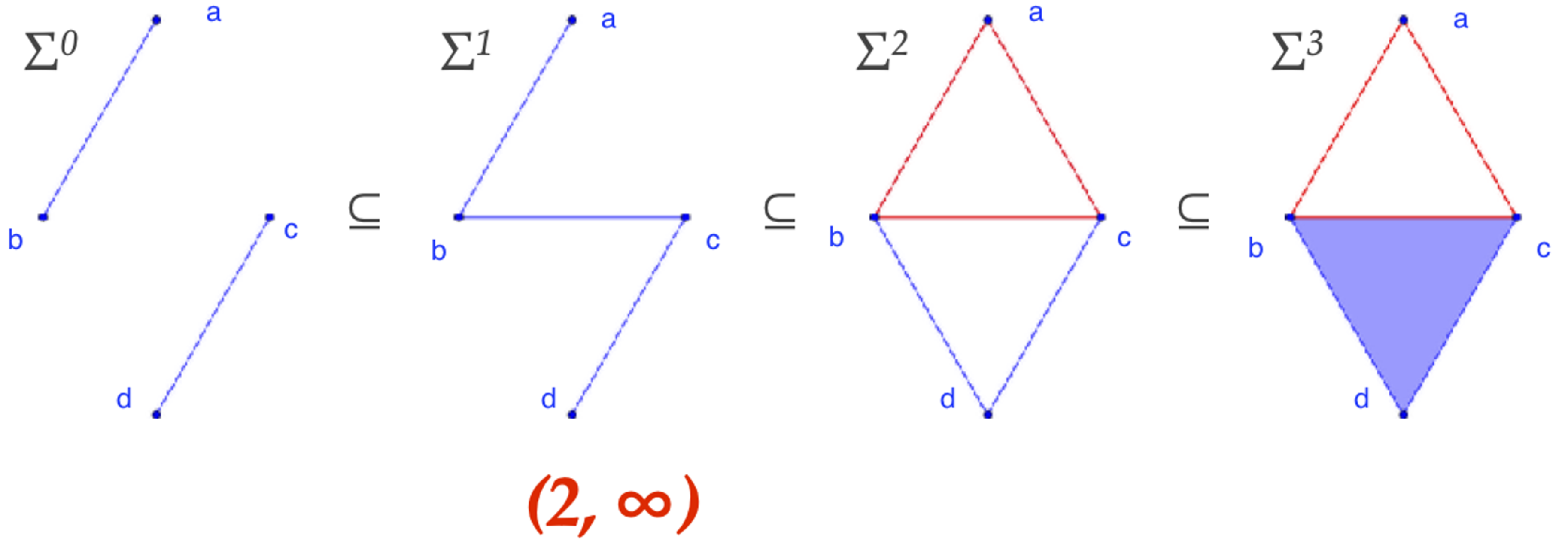
Let f(s) be the step at which ‘s’ was added. f(bc = 1), f(ac) = 2.



(2, 3)

PC - Ulderio Fugacci

$$\sigma_b = bd, \sigma_d = bcd$$

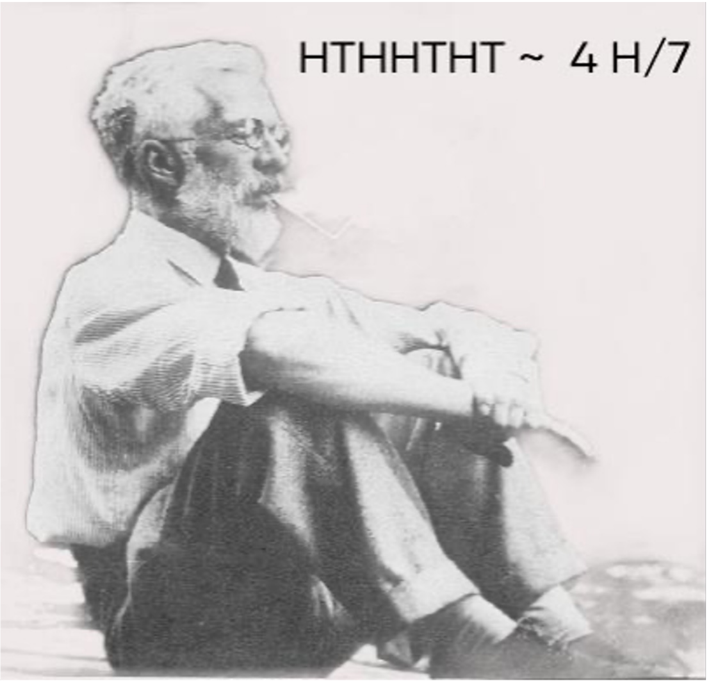


PC - Ulderio Fugacci

$$\sigma_b = ac$$

Fisher Information

Fisher Information - Motivation



PC - Google

- FI is used to quantify the information stored in a summary statistic.
- Is the summary sensitive to the model parameters (θ) as much as the raw data?
- Given a biased coin with P(H) = θ, a summary statistic is the number of head.

Mathematical Formulation of FI

$$p_\theta(X) = P(X|\theta)$$

- I(θ) measures the change in the functional form of the log-likelihood.

$$I_X(\theta) = E_X\left(\left(\frac{\partial}{\partial\theta}\log p_\theta(X)\right)^2\right) = E_X\left(-\partial_\theta^2 \log p_\theta(X)\right) = \int dX\ p_\theta(X)(-\partial_\theta^2 \log p_\theta(X))$$

Mathematical Formulation of FI

$$\boxed{p_\theta(X) = P(X|\theta)}$$

- I(θ) measures the change in the functional form of the log-likelihood.

$$I_X(\theta) = E_X\left(\left(\frac{\partial}{\partial\theta}\log p_\theta(X)\right)^2\right) = E_X\left(-\partial_\theta^2 \log p_\theta(X)\right) = \int dX\ p_\theta(X)(-\partial_\theta^2 \log p_\theta(X))$$

- Example - Coin Toss with P(H) = θ, I(θ) = $((1 - \theta)\theta)^{-1}$. High information near θ = 0, 1.

- The Fisher Information signifies how well we can **contrast** data drawn from P(X|θ) and P(X|$\theta + \varepsilon$).

FI for Gaussian Distributions

- When P(X|θ) = Gaussian ($\mu(\theta)$, C),

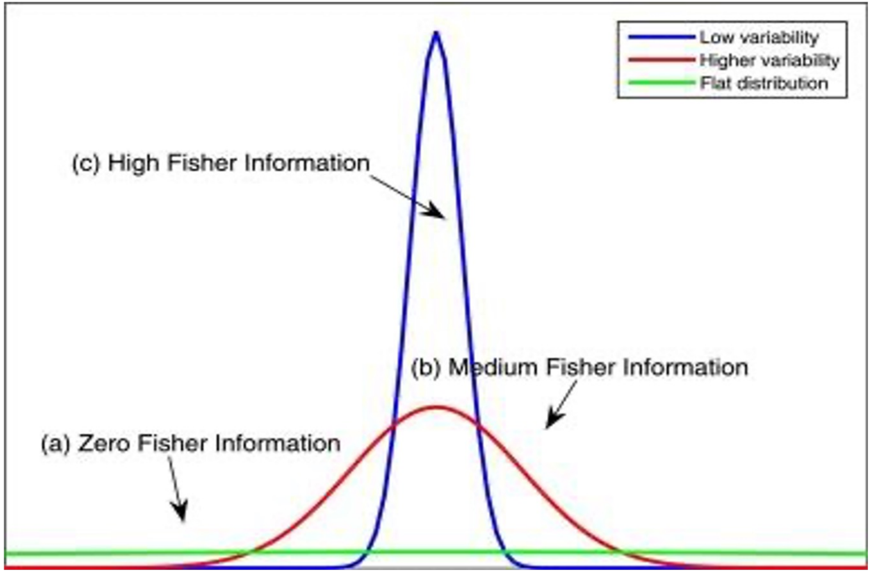
$$I(\theta) = \frac{d\mu}{d\theta}^T C^{-1} \frac{d\mu}{d\theta}$$

FI for Gaussian Distributions

- When P(X|θ) = Gaussian ($\mu(\theta)$, C),

$$I(\theta) = \frac{d\mu}{d\theta}^T C^{-1} \frac{d\mu}{d\theta}$$

- Two elements -
  - The summary is expected to be precise - low variance.
  - The summary is sensitive to θ.



PC - Avan Al-Saffar, Eun-jin Kim

FI for Gaussian Distributions

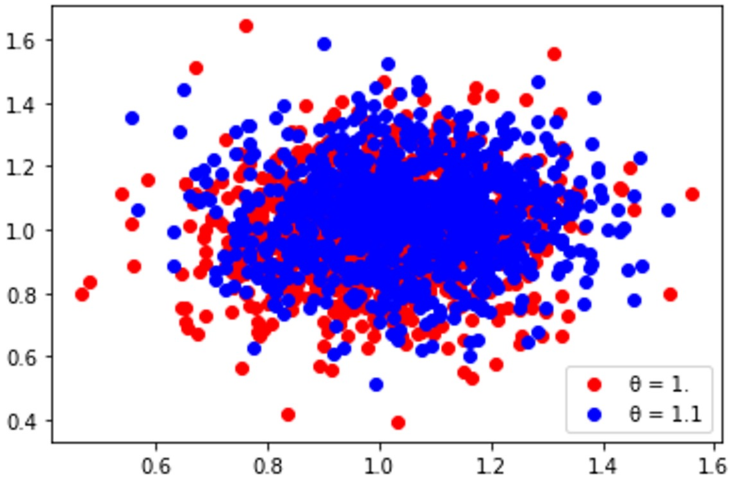
$$I(\theta) = \frac{d\mu}{d\theta}^T C^{-1} \frac{d\mu}{d\theta}$$

- We would like the summary of the data to "move" as much as possible when we vary θ, implying high $\frac{d\mu}{d\theta}$.

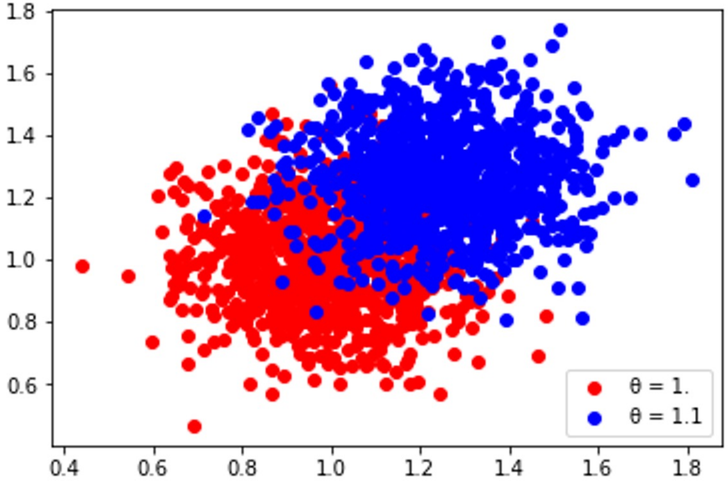
FI for Gaussian Distributions

$$I(\theta) = \frac{d\mu}{d\theta}^T C^{-1} \frac{d\mu}{d\theta}$$

- We would like the summary of the data to "move" as much as possible when we vary θ, implying high $\frac{d\mu}{d\theta}$.



<



FI for Gaussian Distributions

- "Movement + spread" is measured covariantly using FI.

- I($\boldsymbol{\theta}$) generalizes to the Fisher Determinant when $\boldsymbol{\theta}$ is multidimensional.

$$I(\theta) = \frac{d\mu}{d\theta}^T C^{-1} \frac{d\mu}{d\theta} \rightarrow \det \frac{d\mu}{d\theta_\alpha}^T C^{-1} \frac{d\mu}{d\theta_\beta}$$

FI for Gaussian Distributions

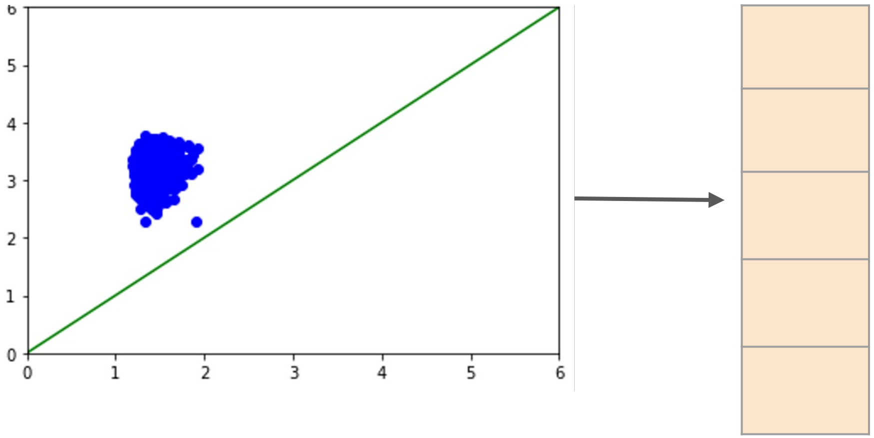
- "Movement + spread" is measured covariantly using FI.
- I($\boldsymbol{\theta}$) generalizes to the Fisher Determinant when $\boldsymbol{\theta}$ is multidimensional.

$$I(\theta) = \frac{d\mu}{d\theta}^T C^{-1} \frac{d\mu}{d\theta} \rightarrow \det \frac{d\mu}{d\theta_\alpha}^T C^{-1} \frac{d\mu}{d\theta_\beta}$$

- We use this expression
  - To measure the information in the persistence summaries, making a Gaussian approximation.
  - As the **loss function** to learn the optimal filtration.

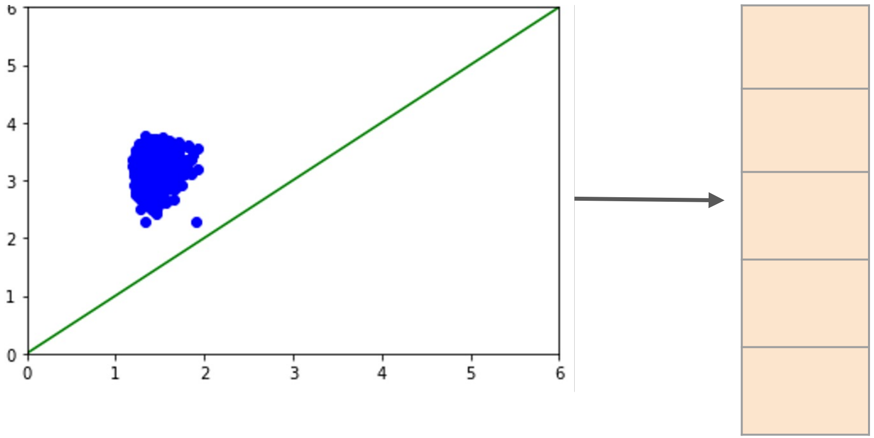
Vectorizing the Persistence Diagram

$$Vec : \{PD\} \rightarrow \mathbb{R}^k$$



Vectorizing the Persistence Diagram

$$Vec : \{PD\} \to \mathbb{R}^k$$



- For statistical analysis, it is easier to work with vectors than sets.
- *Vec* can be parametrized by a variational family and an *optimal vectorization* can be learnt [3].

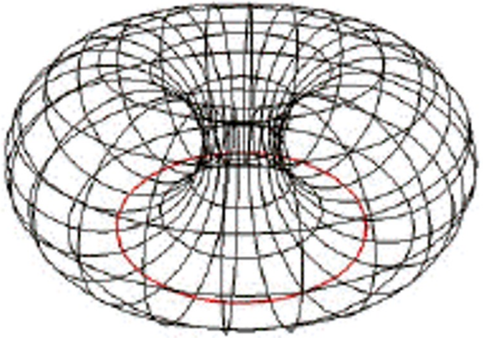
Finding the *Optimal* Filtration

Slicing Bagels < Dunkin' Donuts

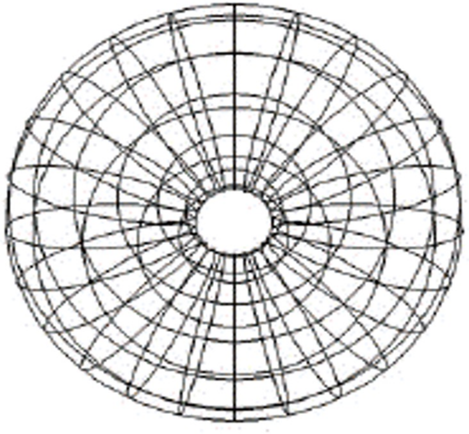
- Estimating the diameter of the *center hole* in a torus.

Slicing Bagels < Dunkin' Donuts

- Estimating the diameter of the *center hole* in a torus.



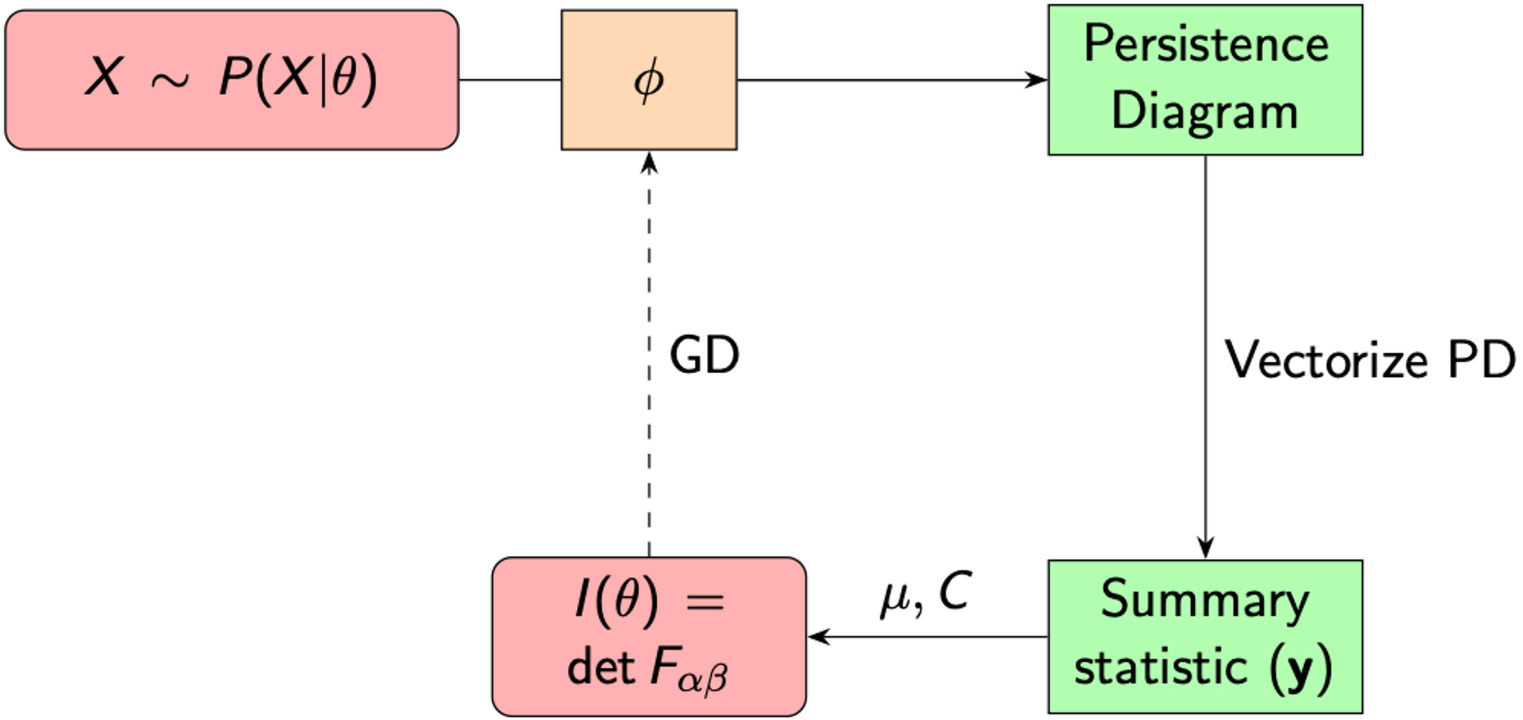
<



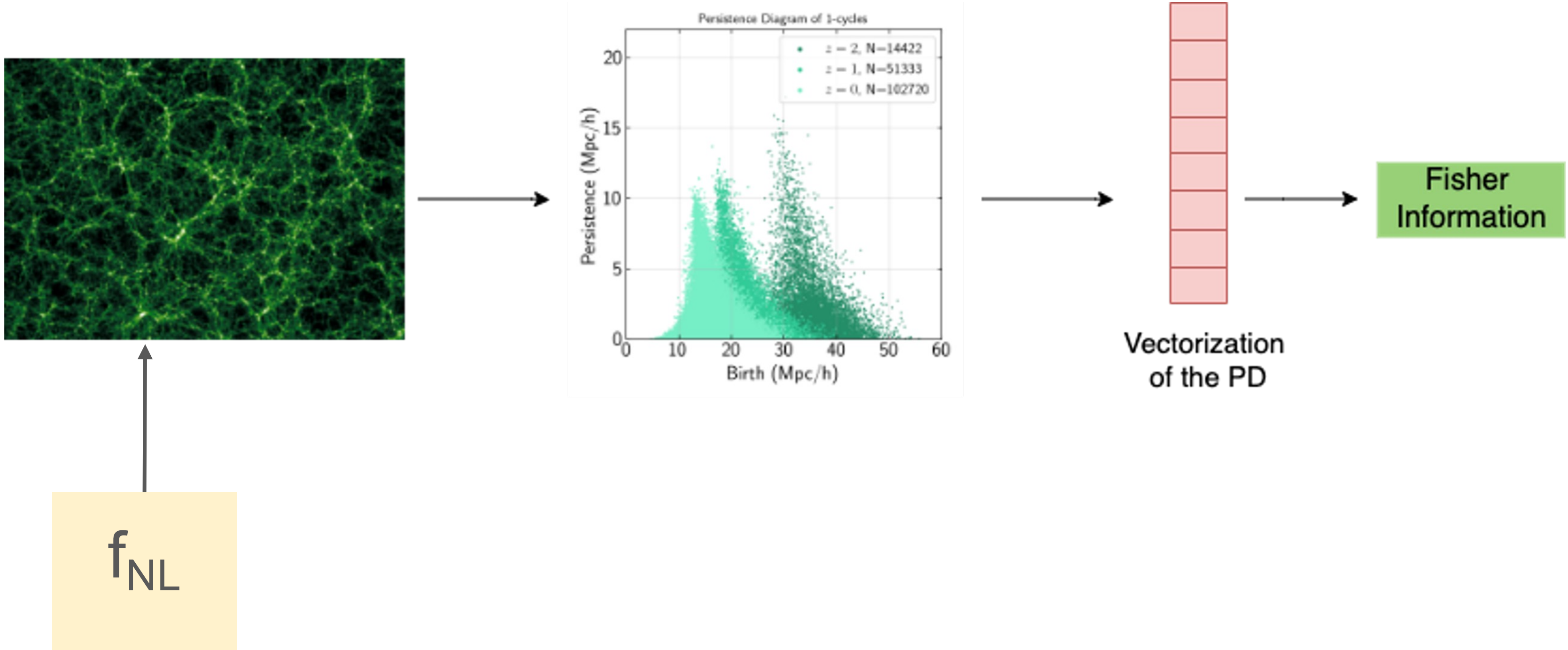
d = ?

PC - Thomas Banchoff, *Slicing Doughnuts and Bagels*

The Pipeline



Pipeline Example



Why use Fisher Information to choose the filtration?

- Cramér–Rao Bound
  - $$\mathrm{var}(\hat{\theta}) \geq 1/I(\theta)$$
  - Better estimators require more informative summary statistics.
- FI measures the relevance of the filtration for a specific *task,* facilitating customization.

Implementation Details

- Update equation for $$\phi : \phi_{\text{new}} = \phi_{\text{old}} + \alpha \, \nabla_{\phi} \, I(\theta)$$

- Chain rule to calculate the intermediate gradients.

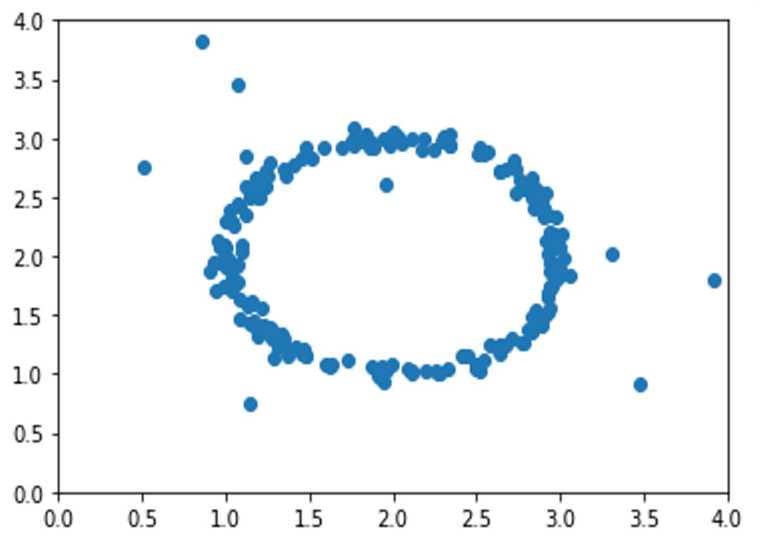
- Differentiating the persistence diagram - Identify the persistence pairs, and their dependence on ϕ.

$$b = f_{\phi}(\sigma_b), d = f_{\phi}(\sigma_d)$$

- Implemented using GUDHI, PyTorch.

Input Data

- Input points chosen with radius ~ Gaussian (1, 0.2).
- Angles chosen uniformly randomly.
- Some noisy background points.



Variational Family

- ϕ = (k, p, denWeight)

$$f(v) = \left(\frac{\sum_{j \in neighbors(v)} dist(v,j)^k}{N} \right)^{1/k} + denWeight * density(v)$$

$$f(e = (v, w)) = \left(\frac{\ell(e)^p + \max(f(v), f(w))^p}{2} \right)^{1/p}$$

Variational Family

- ϕ = (k, p, denWeight)

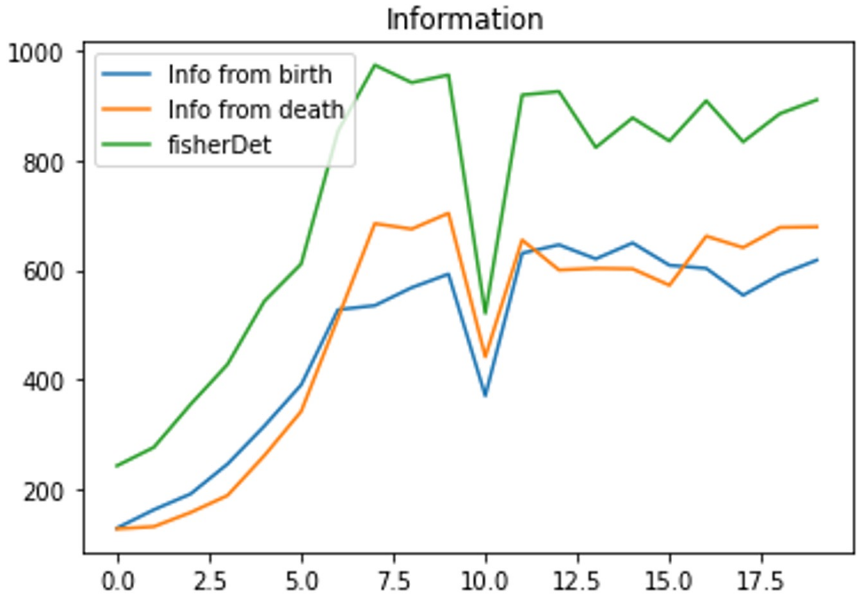
$$f(v) = \left(\frac{\sum_{j \in neighbors(v)} dist(v,j)^k}{N} \right)^{1/k} + denWeight * density(v)$$

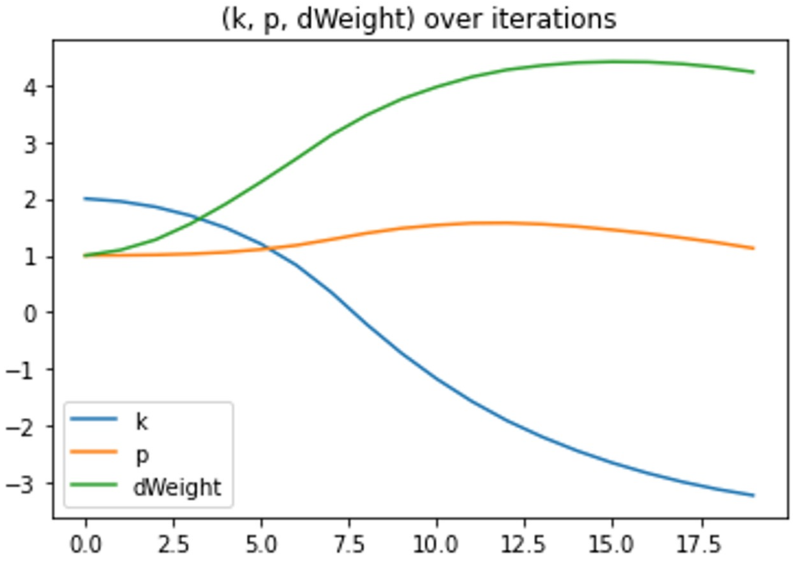
$$f(e = (v,w)) = \left(\frac{\ell(e)^p + \max(f(v), f(w))^p}{2} \right)^{1/p}$$

- y = (b, d) of the most persistent pair. Need not be hand crafted generally.

Performance Graphs

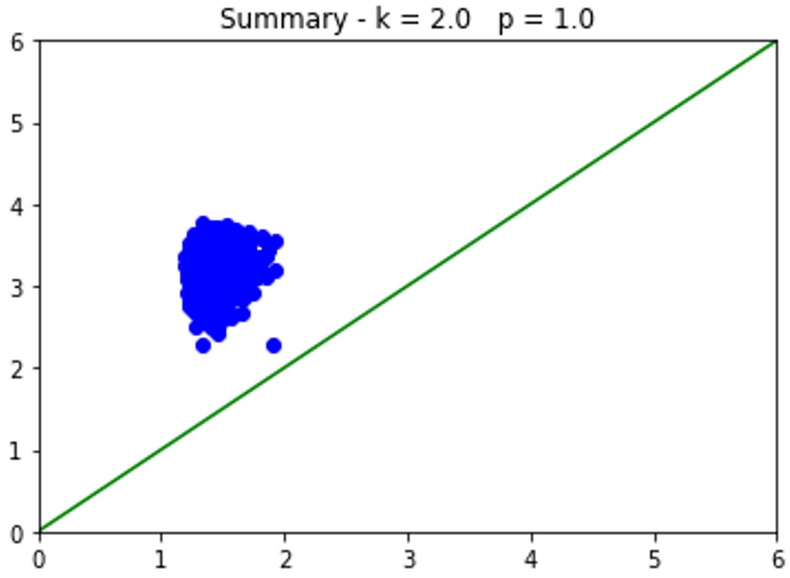
dWeight is *learned* to increase over iterations.

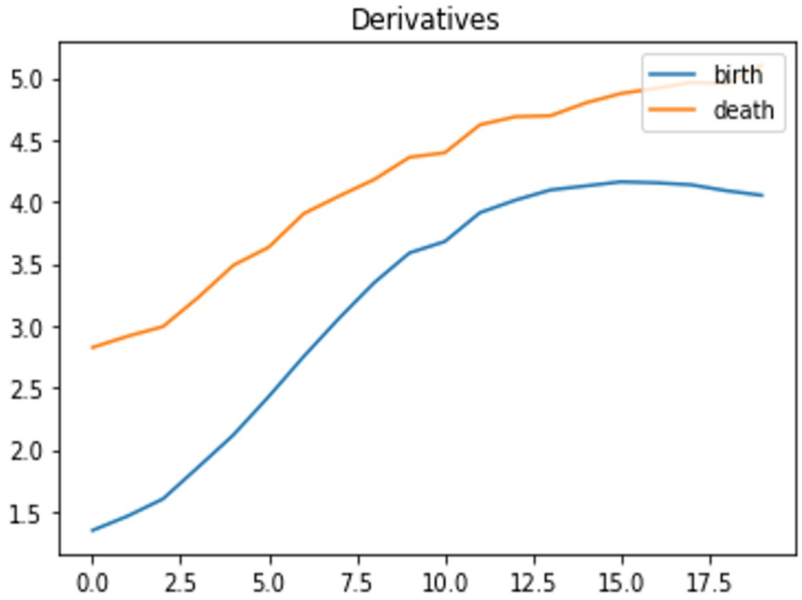




The dip in FI is due to standard gradient descent issues.

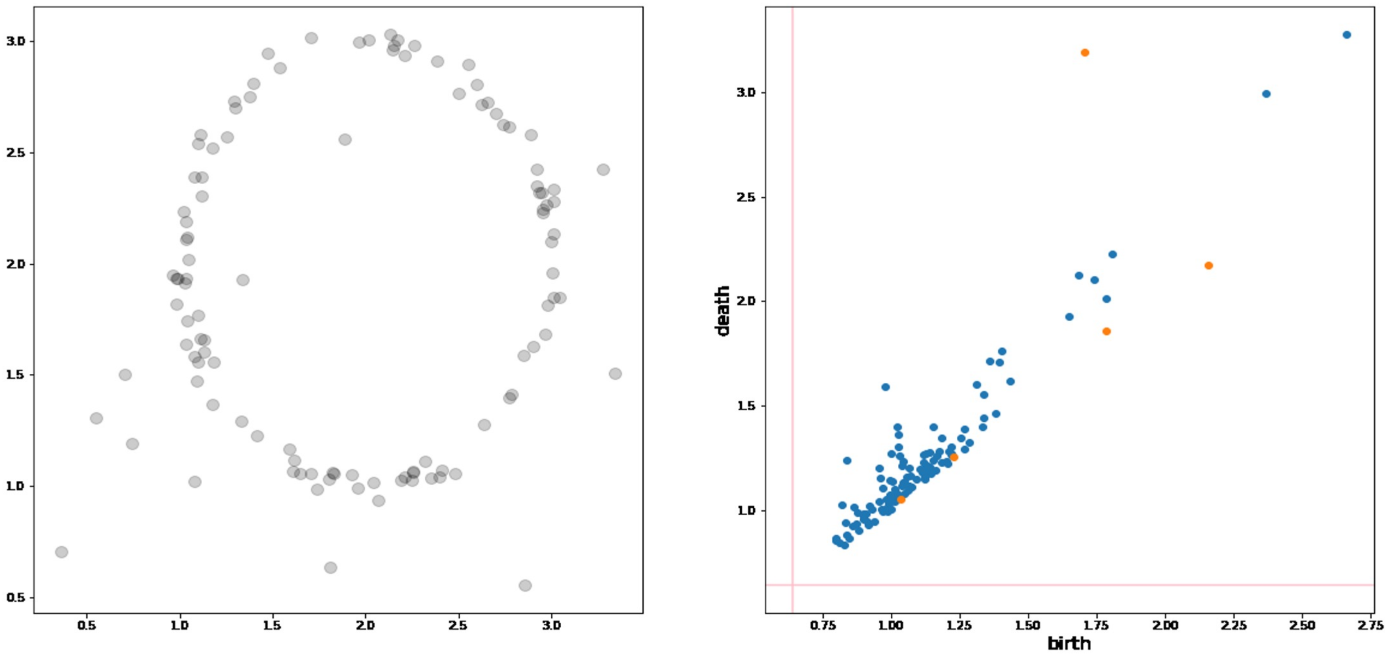
Performance Graphs





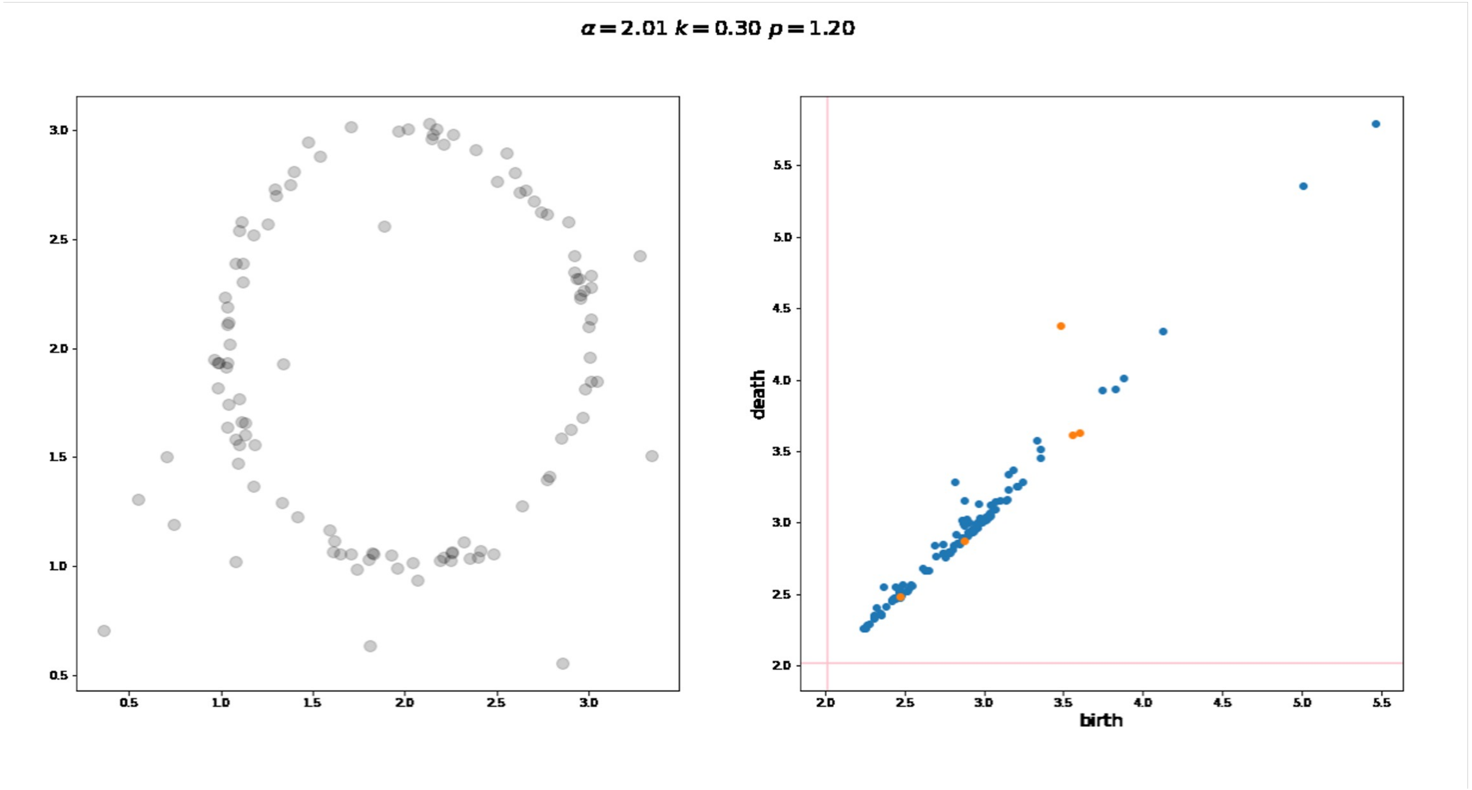
Initial Filtration

$\alpha = 0.64$ $k = 1.00$ $p = 1.00$

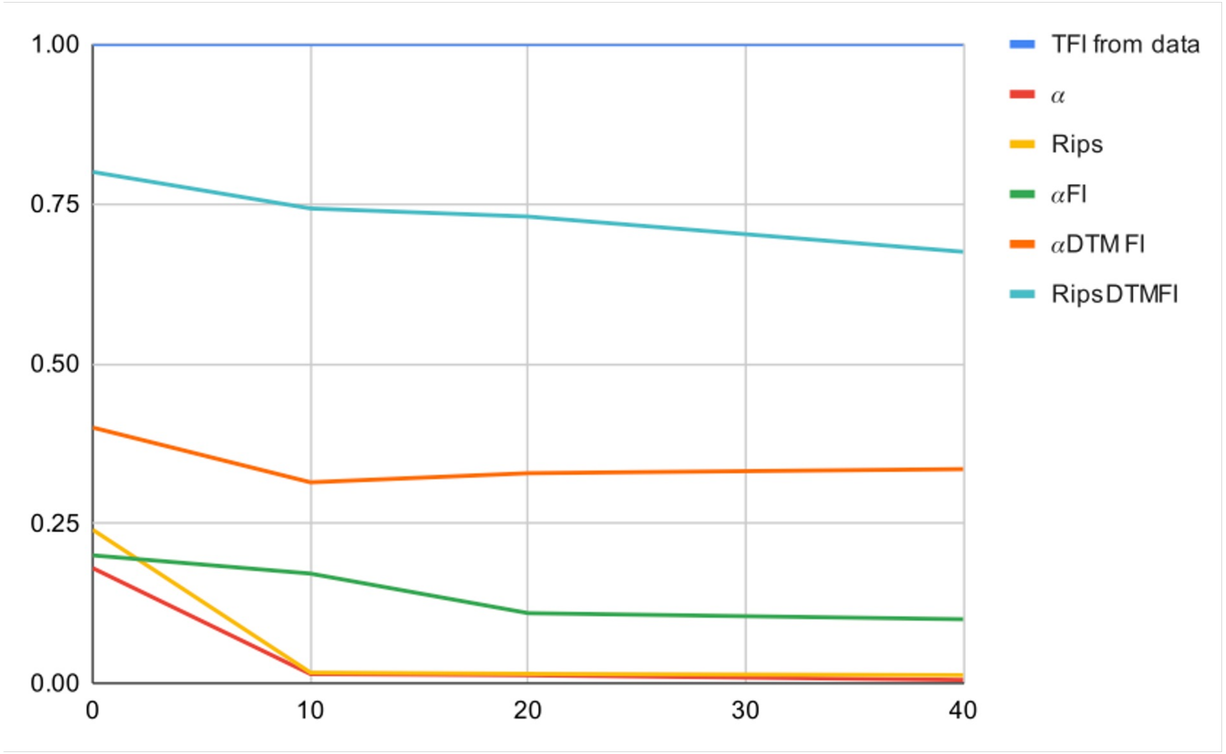


Learnt Filtration

Learns to filter the noisy background points.



Results



X axis - #bg noisy points.
Y axis - FI(filtration)/TFI

Outlook

Summary

Point set sampled from P(X|θ).

Antifragile PH

- Filtration that maximizes I(θ) using gradient descent.
- An informative vectorized persistence summary.

Upshots

- Customized PH for a given problem.
  - Adaptive filtration and vectorization.
- Can detect less persistent features that are informative.

- From the optimal filtration, we can
  - Learn about the sensitivity of the topological features to θ.
  - Interpret the higher order statistics in data.

Future Work

- Forecast Δf_{NL} using LSS data [2]. Study the effect of NG on geometry of LSS.

- Improvements on the pipeline -
  - More generalized variational family.
  - Improved loss function (FI→ *swyft* [4]).
  - Better vectorizations.
  - Automatic differentiation, better runtime. (Pytorch→ JAX)

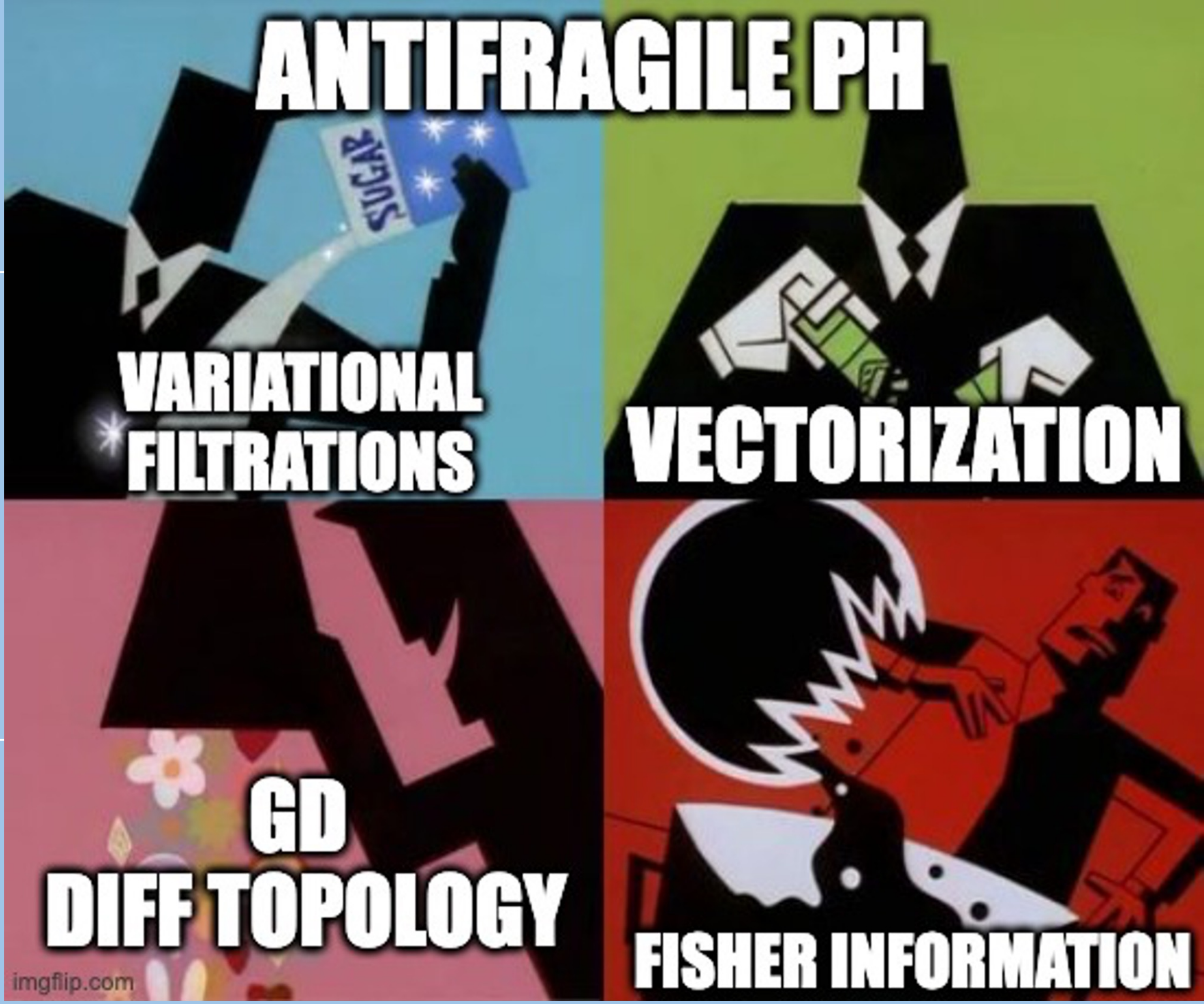
Future Work

- Forecast Δf_{NL} using LSS data [2]. Study the effect of NG on geometry of LSS.

- Improvements on the pipeline -
  - More generalized variational family.
  - Improved loss function (FI$\rightarrow$ *swyft* [4]).
  - Better vectorizations.
  - Automatic differentiation, better runtime. (Pytorch$\rightarrow$ JAX)

- Can we reproduce RG transformations using PH by minimizing information loss in coarse graining?

THANK YOU!


ANTIFRAGILE PH
SUGAR
VARIATIONAL FILTRATIONS
VECTORIZATION
GD DIFF TOPOLOGY
FISHER INFORMATION
imgflip.com

References

1. Alex Cole, Gregory J. Loges, and Gary Shiu - *Quantitative and interpretable order parameters for phase transitions from persistent homology.*

2. Matteo Biagetti, Juan Calles, Lina Castiblanco, Alex Cole and Jorge Noreña - *Fisher Forecasts for Primordial non-Gaussianity from Persistent Homology.*

3. Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, Yuhei Umeda - *PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures.*

4. Benjamin Kurt Miller, Alex Cole, Gilles Louppe, Christoph Weniger - *Simulation-efficient marginal posterior estimation with swyft: stop wasting your precious time*