Using persistent homology for hypothesis testing

Presentation by Elie Cueto

Articles by Willem Elbers & Rien Van de Weygaert

9/3-23

Journal Articles


MNRAS **486,** 1523–1538 (2019) doi:10.1093/mnras/stz908
Advance Access publication 2019 March 30

Persistent topology of the reionization bubble network – I. Formalism and phenomenology

Willem Elbers ★ and Rien van de Weygaert
Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700AV Groningen, the Netherlands

Journal Articles

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Royal Astronomical Society

MNRAS **520**, 2709–2726 (2023)
Advance Access publication 2023 January 13
https://doi.org/10.1093/mnras/stad120

Persistent topology of the reionization bubble network – II. Evolution and classification

Willem Elbers [1★] and Rien van de Weygaert [2]

[1] *Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham, DH1 3LE, UK*
[2] *Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700AV Groningen, the Netherlands*

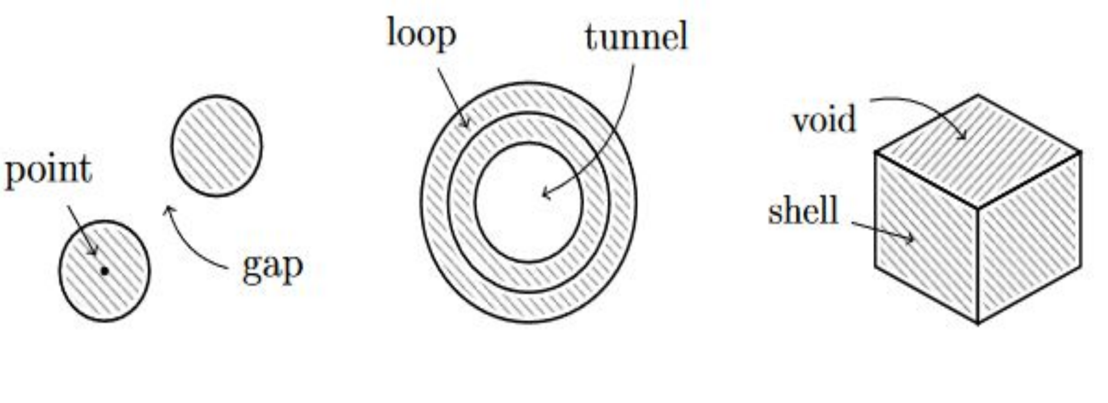
Disposition

- Introduction to persistent homology and Persistence diagrams
- Statistical treatment of Persistence diagrams
- Example of use in Astrophysics

Persistent homology: Counting Holes

Finding the number of n-D holes in a structure.

- Gaps between components
- Tunnels through structures
- Voids within shells
- Higher dimensional holes

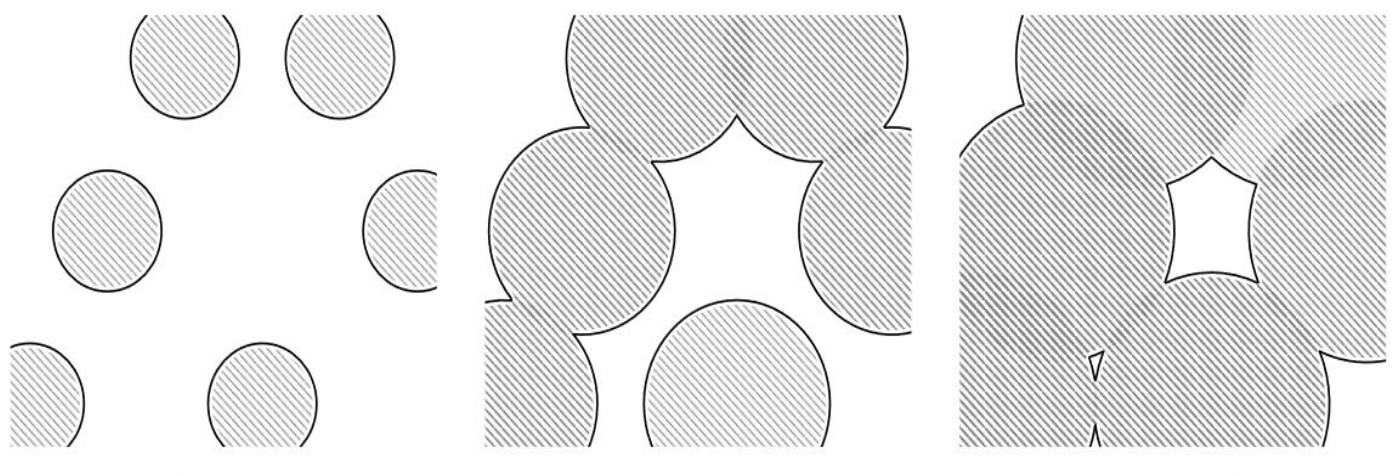


Plot: Elbers+19

Persistent homology: Counting Holes

Let spheres with a radius alpha surround each point.

Count the number of holes in each dimension as a function of alpha



(a) $\alpha = 0.7$ (b) $\alpha = 1.2$ (c) $\alpha = 1.5$

(a)	(b)	(c)
0D: 6	0D: 2	0D: 1
1D: 0	1D: 0	1D: 2

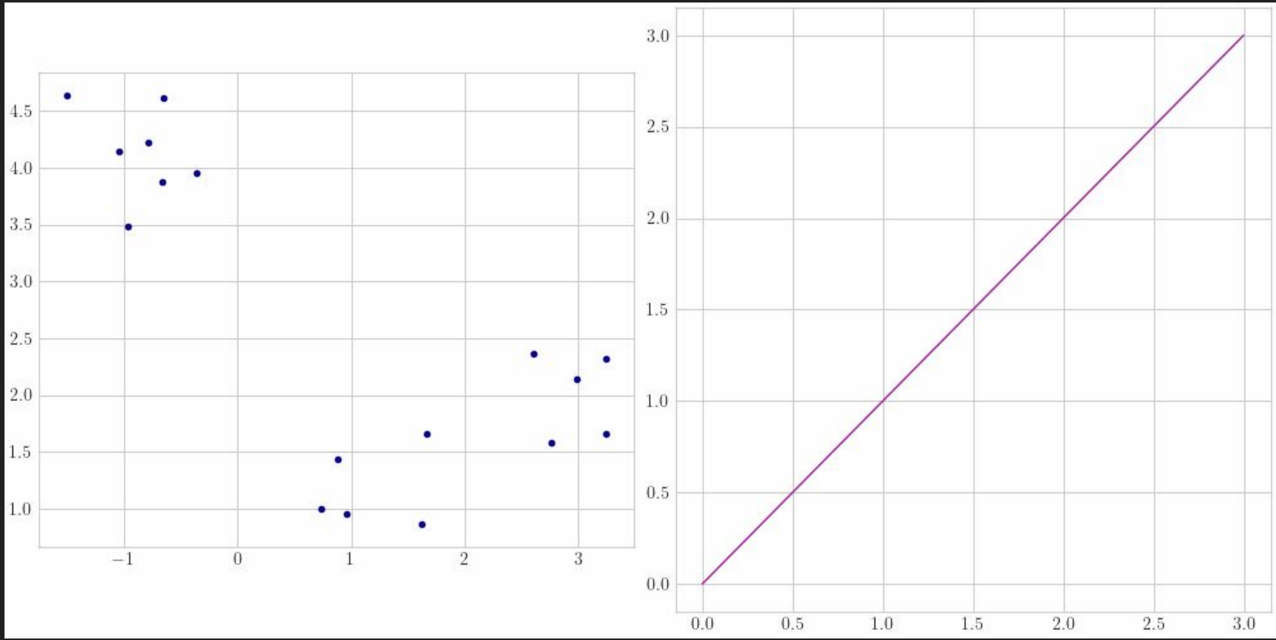
Plot: Elbers+19

Persistence diagrams

When is a feature born

When does a feature die

Visualising high dimensional data

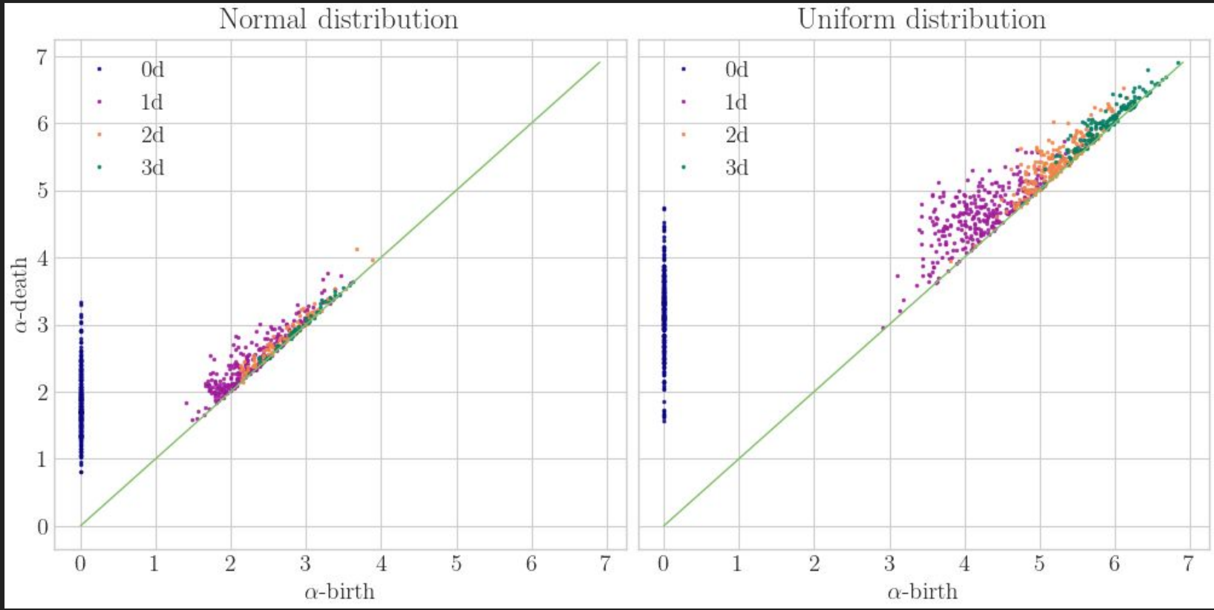


Persistence diagrams

When is a feature born

When does a feature die

Visualising high dimensional data



Finally: Statistics

We can find the "Distance" between two persistence diagrams:

$$d(X,Y) = \left[\inf_{\phi: X \to Y} \sum_{x \in X} ||x - \phi(x)||^2 \right]^{1/2}$$

With Monte Carlo simulations of our New and Improved model, as well as our null hypothesis model we can find a Fréchet Average:

$$F(Y) = \frac{1}{n} \sum_{i=1}^{n} d(Y, X_i)^2$$

Finally: Statistics

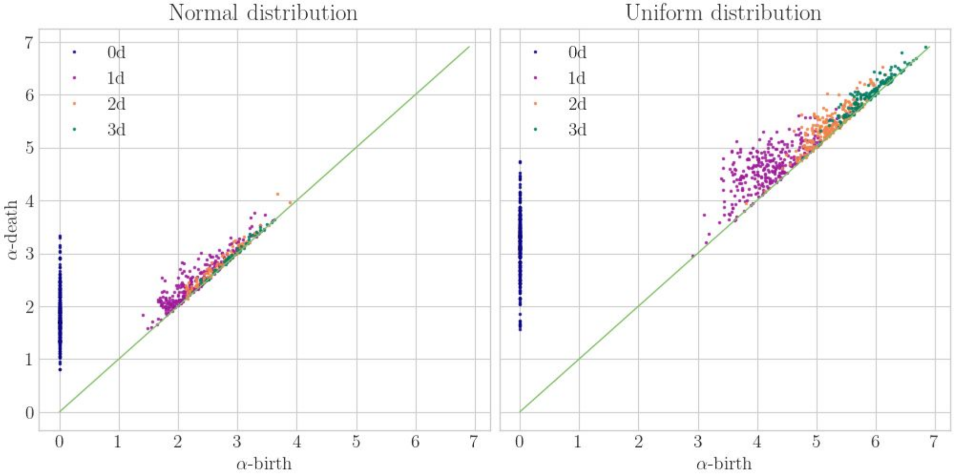
We can find the "Distance" between two persistence diagrams:

$$d(X,Y) = \left[\inf_{\phi: X \to Y} \sum_{x \in X} ||x - \phi(x)||^2 \right]^{1/2}$$

With Monte Carlo simulations of our New and Improved model, as well as our null hypothesis model we can find a Fréchet Average:

$$F(Y) = \frac{1}{n} \sum_{i=1}^{n} d(Y, X_i)^2$$

A model will tend to have a lower Fréchet average, if it is more similar to the data



These persistence diagrams a large distance, as they are constructed from data with different PDFs

Distances between diagrams

Mock data generated with

$$f(x, y) = (1 + ax + bx^2)(1 + ay + by^2)$$

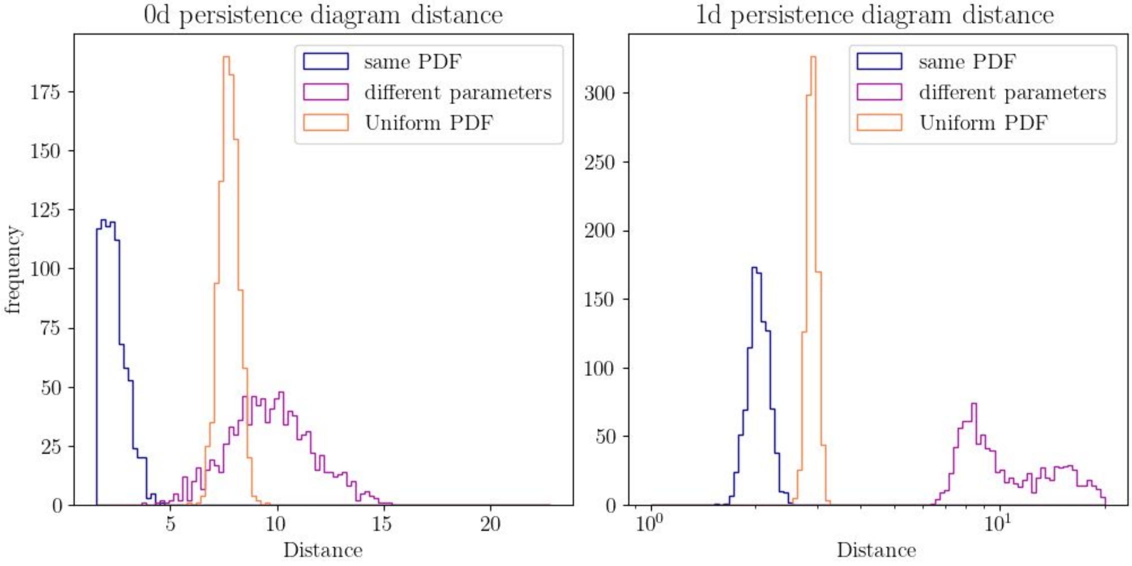
For a = 1, b = 2.4, -2<x<2, N = 1000

1000 simulations with same PDF,

Uniform PDF, and PDF with

a = -2, b = 2.4

The “Correct” PDF gives a much lower distance distribution



Persistence Fields

We can find a variance on each feature in our data:

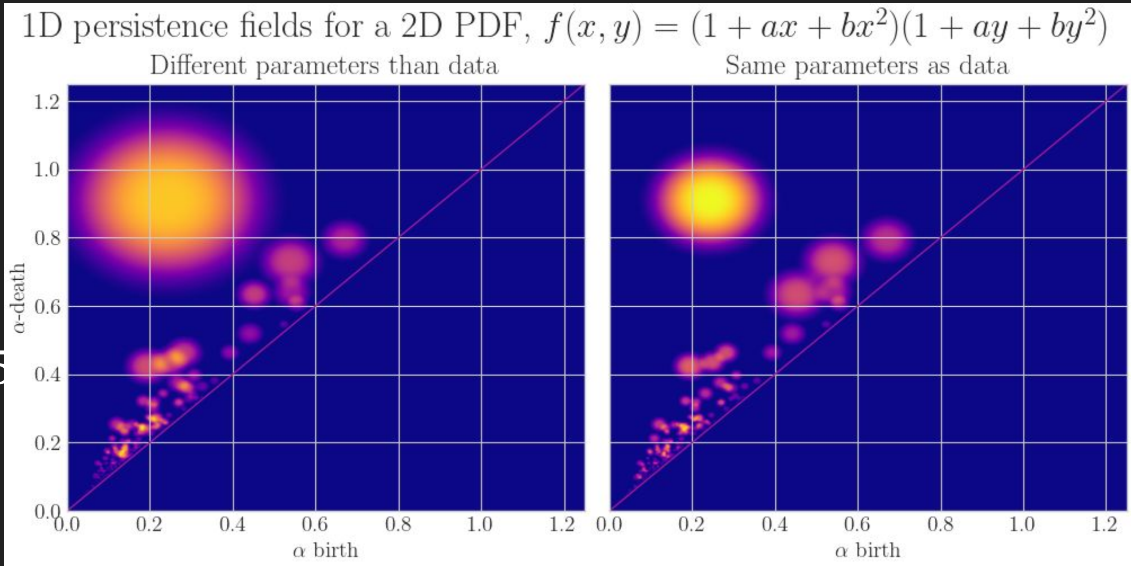
$$\sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n} ||y - \phi_i(y)||^2$$

With the variance we can construct Persistence Fields

For each feature here, r~σ and "brightness" ~ $(\alpha_{death} - \alpha_{birth})$^0.5

Highly persistent features tend to be more significant, but less commonly found.

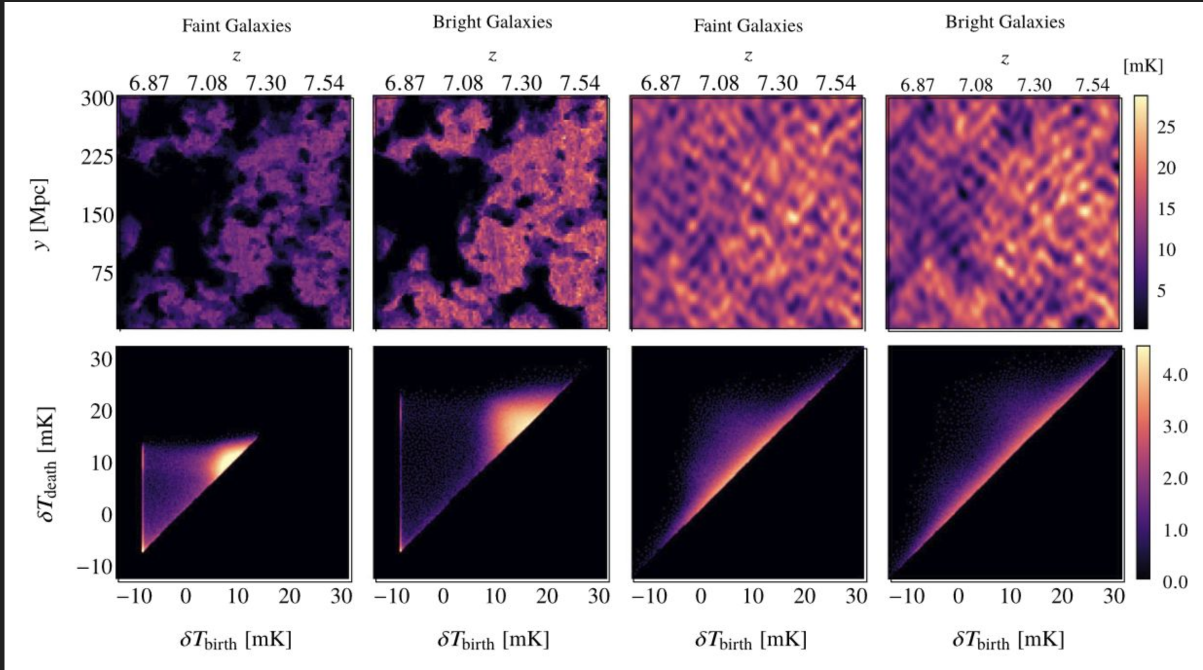
Larger radii here correspond the model being worse at recreating features of data



Persistent homology put to use: Epoch of Reionisation

Distinguishing different models of reionisation.

Even When introducing noise levels consistent with how observations will look, different models can consistently be differentiated

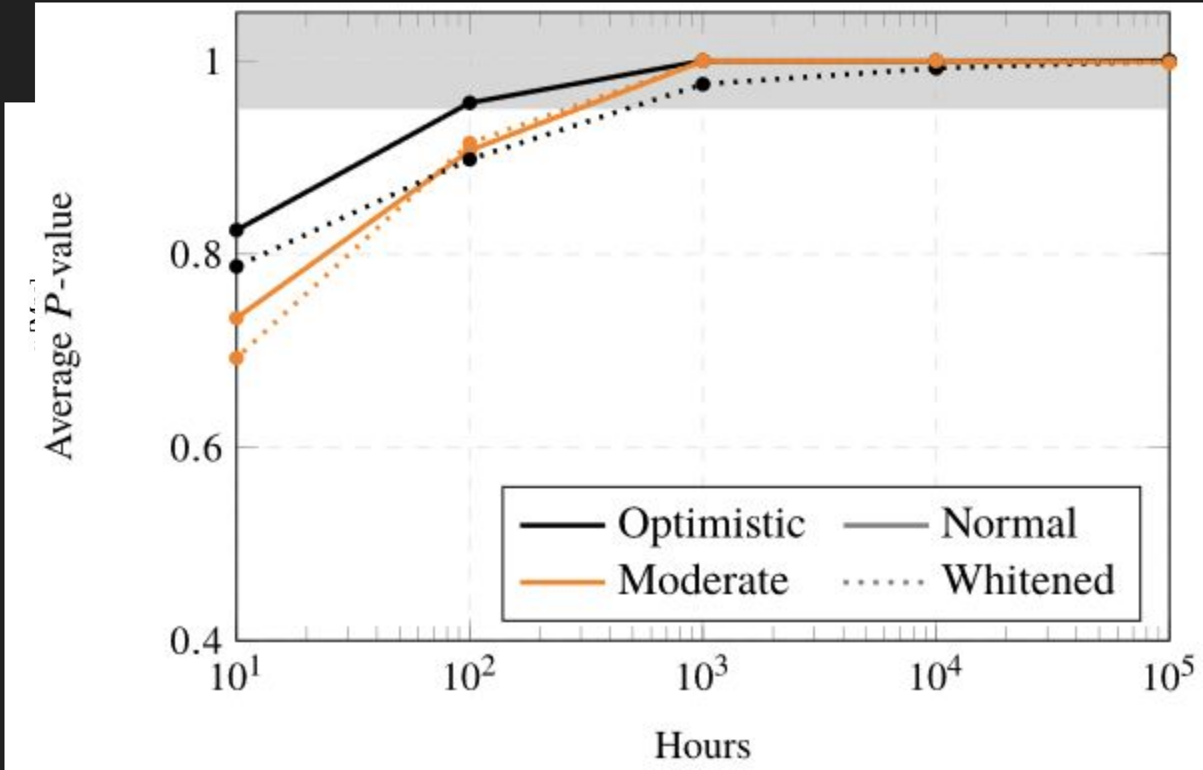


Plot: Elbers+23

Persistent homology put to use: Epoch of Reionisation

Distinguishing different models of reionisation.

Even When introducing noise levels consistent with how observations will look, different models can consistently be differentiated



Plot: Elbers+23

Conclusions

Visualising High dimensional data

Determining how well different models reproduce data

Very robust to noisy data