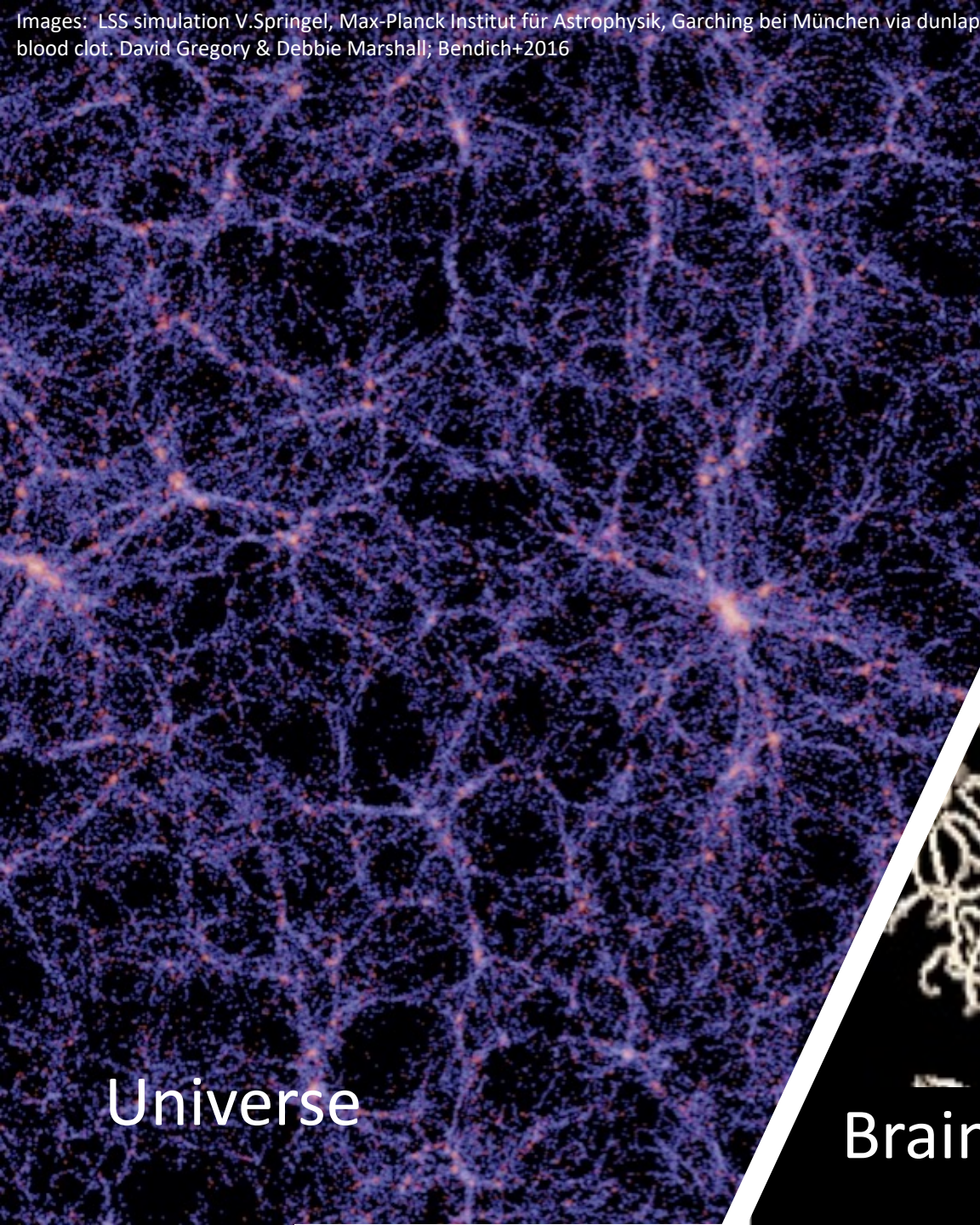


Getting something out of nothing: topological data analysis for cosmology

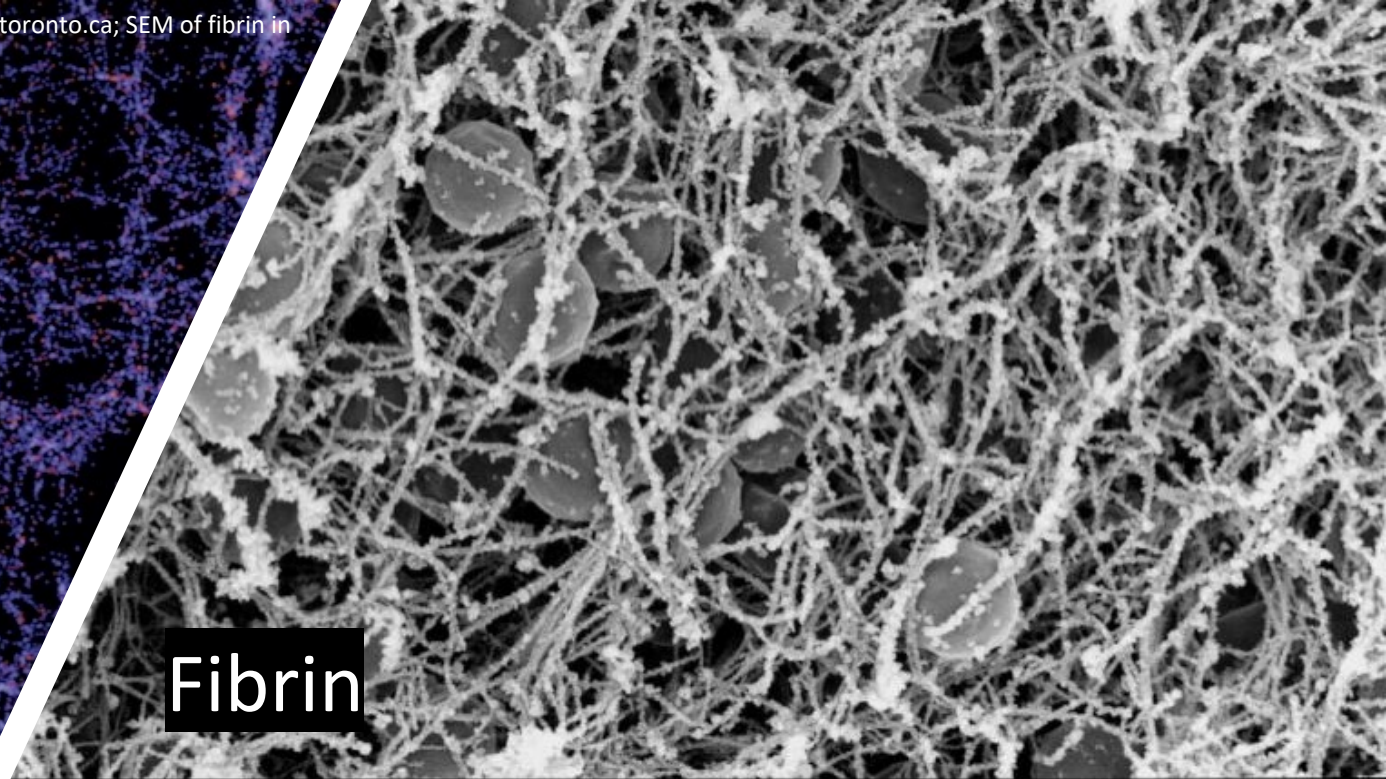
IAU-IAA Astrostatistics and Astroinformatics seminar
December 13, 2022

Jessi Cisewski-Kehe
Department of Statistics
University of Wisconsin-Madison

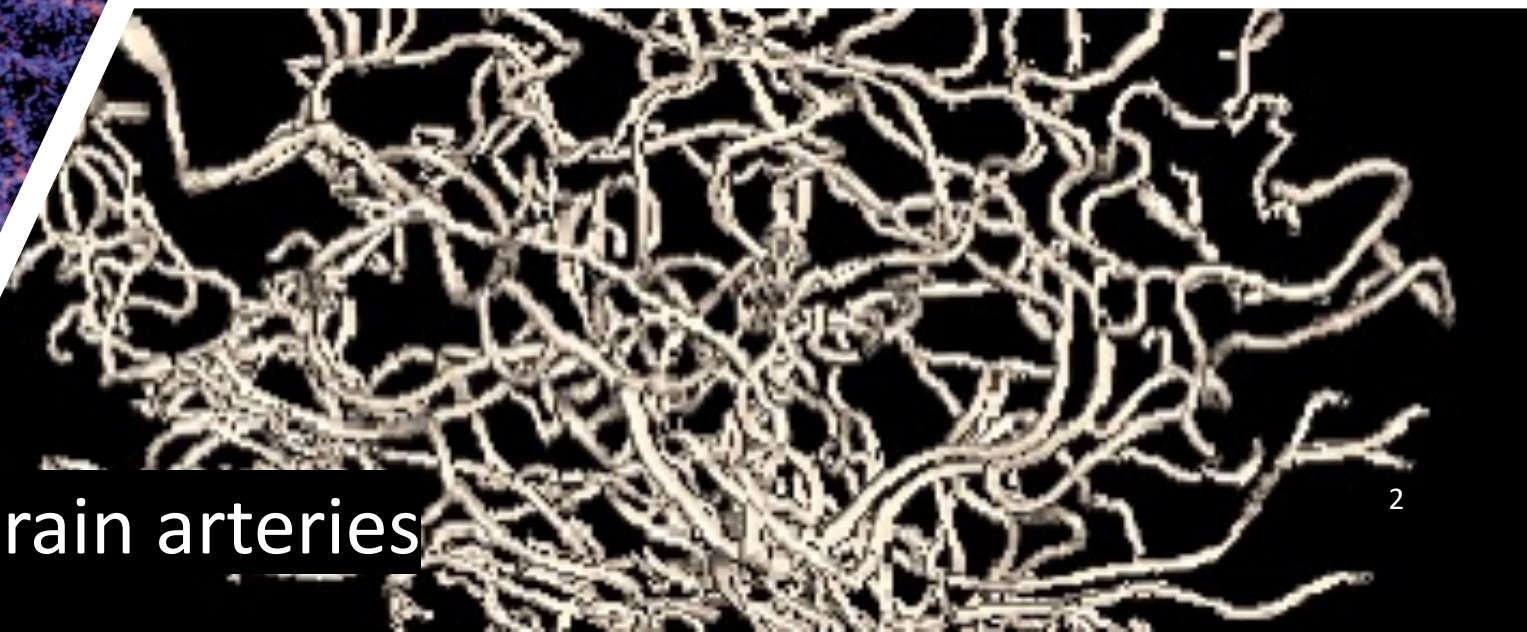
Support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation, and NSF-DMS-2038556.



Universe



Fibrin



Brain arteries

Outline

1. Background on topology for astronomy
2. Introduction to persistent homology
3. Applications of Persistent Homology
 - i. Hypothesis testing: Cold Dark Matter vs. Warm Dark Matter
 - ii. Finding holes in the Universe

Collaborators:

Brittany T. Fasy and Eric Berry (Montana State),

Wojciech Hellwing and Paweł Drozda (Polish Academy of Sciences),

Mark R. Lovell (U. of Iceland), Mike Wu (Stanford), Yen-Chi Chen (UW-Seattle), Xin Xu (Didi),

Sheridan Green (Susquehanna), Daisuke Nagai (Yale)

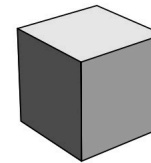
Euler characteristic (EC)

Leonhard Euler (1707-1783) made an observation about polyhedra:

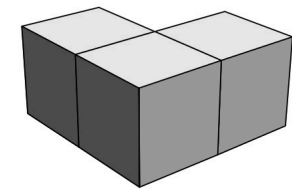
$$\text{\#Vertices} - \text{\#Edges} + \text{\#Faces} = 2$$

Name	Image	Vertices V	Edges E	Faces F	Euler characteristic: $V - E + F$
Tetrahedron		4	6	4	2
Hexahedron or cube		8	12	6	2
Octahedron		6	12	8	2
Dodecahedron		20	30	12	2
Icosahedron		12	30	20	2

If you glue together P polyhedral at a common face, this becomes $V - E + F - P = 1$.



(a) $8 - 12 + 6 - 1 = 1$



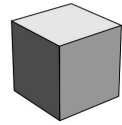
(b) $16 - 28 + 16 - 3 = 1$

Worsley, K.J., 1996. The geometry of random images. Chance, 9(1), pp.27-40

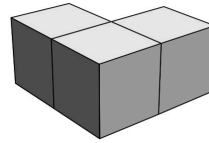
Table from https://en.wikipedia.org/wiki/Euler_characteristic

Holes

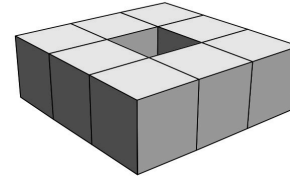
But these formulas do not hold in general...if a hole is present, it reduces $V - E + F - P$ by 1



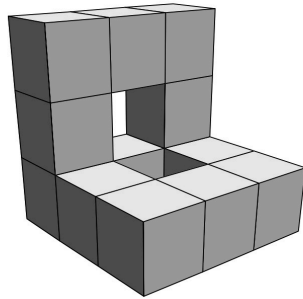
(a) $8 - 12 + 6 - 1 = 1$



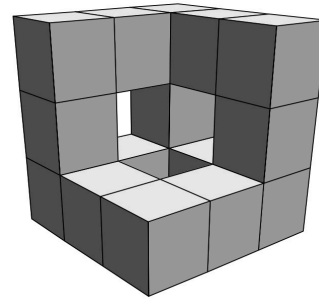
(b) $16 - 28 + 16 - 3 = 1$



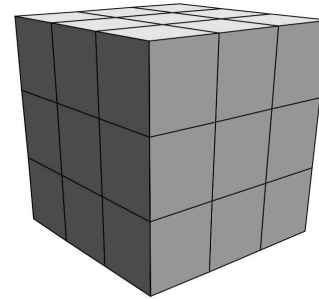
(c) $32 - 64 + 40 - 8 = 0$



(d) $48 - 100 + 64 - 13 = -1$



(e) $56 - 120 + 78 - 16 = -2$



(f) $64 - 144 + 108 - 26 = 2$

Hollow

Topology of the Universe

- 1980s-1990s: investigations of matter distribution
- Sponge-like topology of Large-scale Structure (LSS, Gott+ 1986):
 - “There has been a debate between **hierarchical clustering** models in which **clusters are high-density islands in a low-density sea**, and **cell structure** models in which **voids are isolated low-density islands in a high-density sea**.”
 - Third option: Sponge-like topology. “Sponges are characterized by many holes; chambers are connected by tunnels. Thus both the high- and low-density regions are multiply connected.”
 - Used genus (g) to describe galaxy distribution ($EC = 2 - 2g$ for closed surfaces)

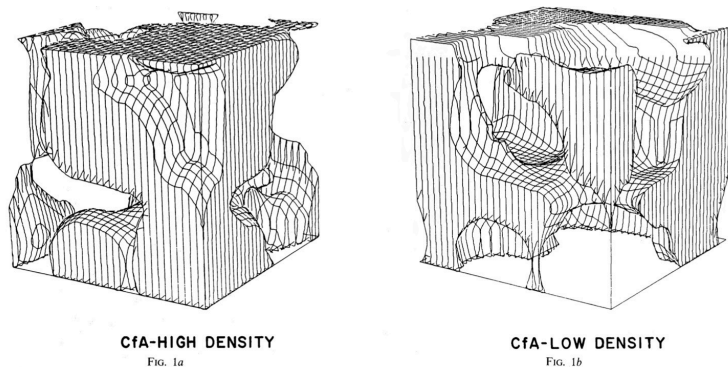


FIG. 1.—CfA volume limited sample. Earth is at the bottom front corner. Cube side length is $28.9h^{-1}$ Mpc. (a) High-density region; (b) low density region. Periodic boundary conditions assumed in Figs. 1–6, 9, 12, 13.

Image: Gott, Melott, and Dickinson 1986

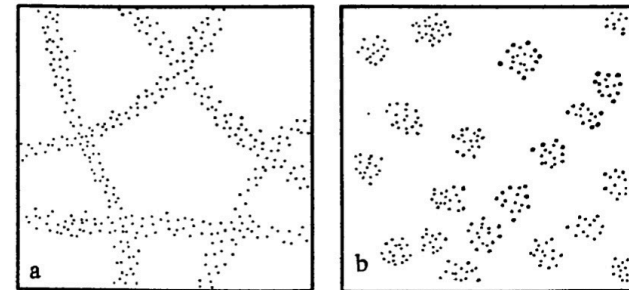


FIG. 2. Two different kinds of particle distribution (schematic): a) cell structure; b) isolated clusters.

Image: Shandarin 1983

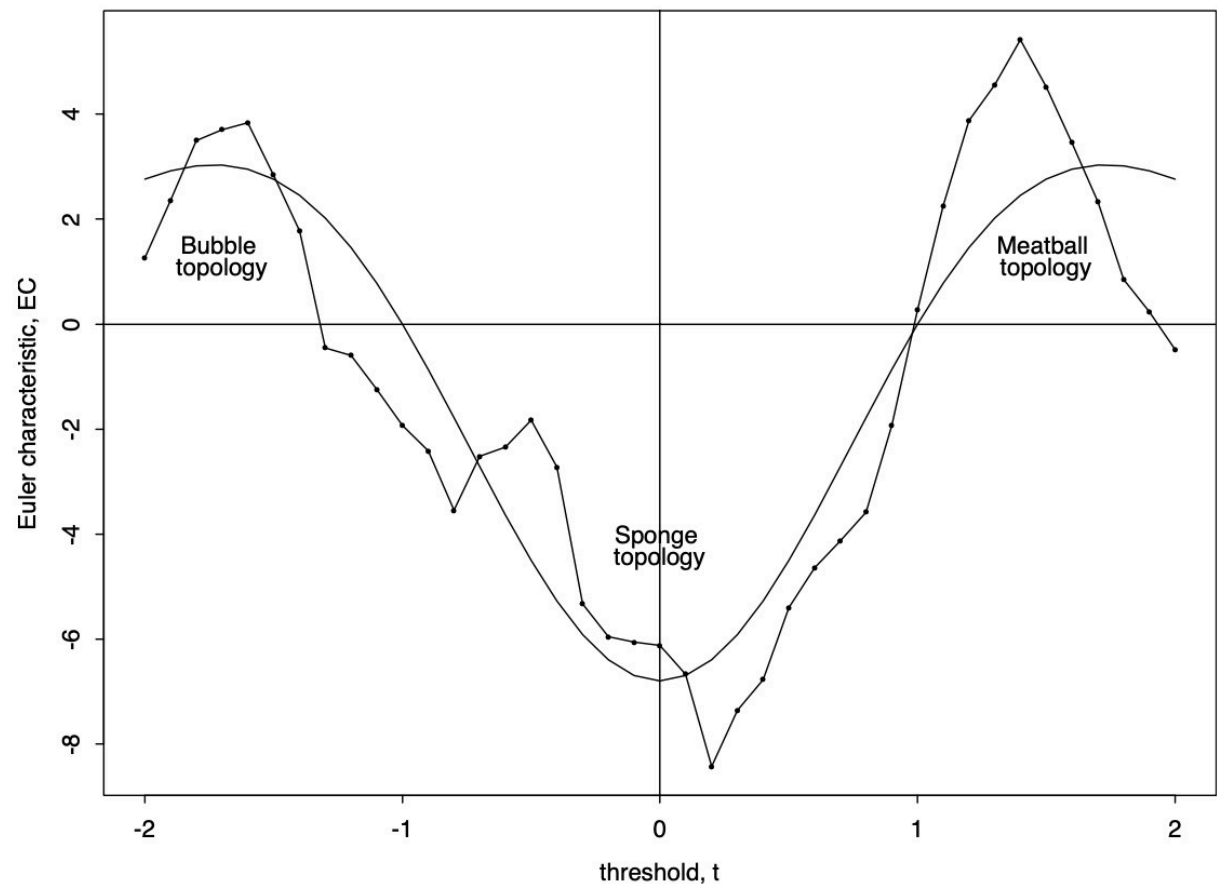
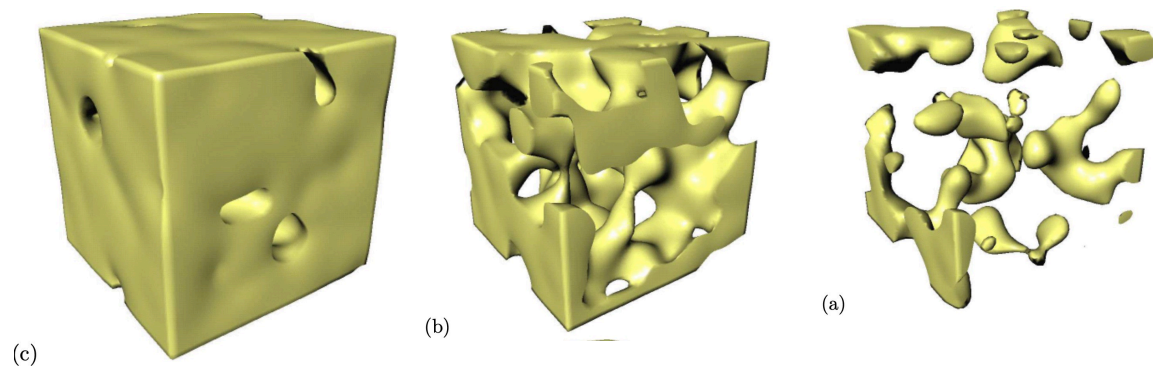


Figure 4 Plot of the observed EC of the set of high-density regions of the real galaxy data (points), and the expected EC from the formula (curve) for a Gaussian random field model, plotted against the density threshold. The observed and expected are in reasonable agreement, confirming the Gaussian random field model: high thresholds produce a “meatball” topology (Fig. 2a); medium thresholds produce a “sponge” topology (Fig. 2b), and low thresholds produce a “bubble” topology (Fig. 2c).

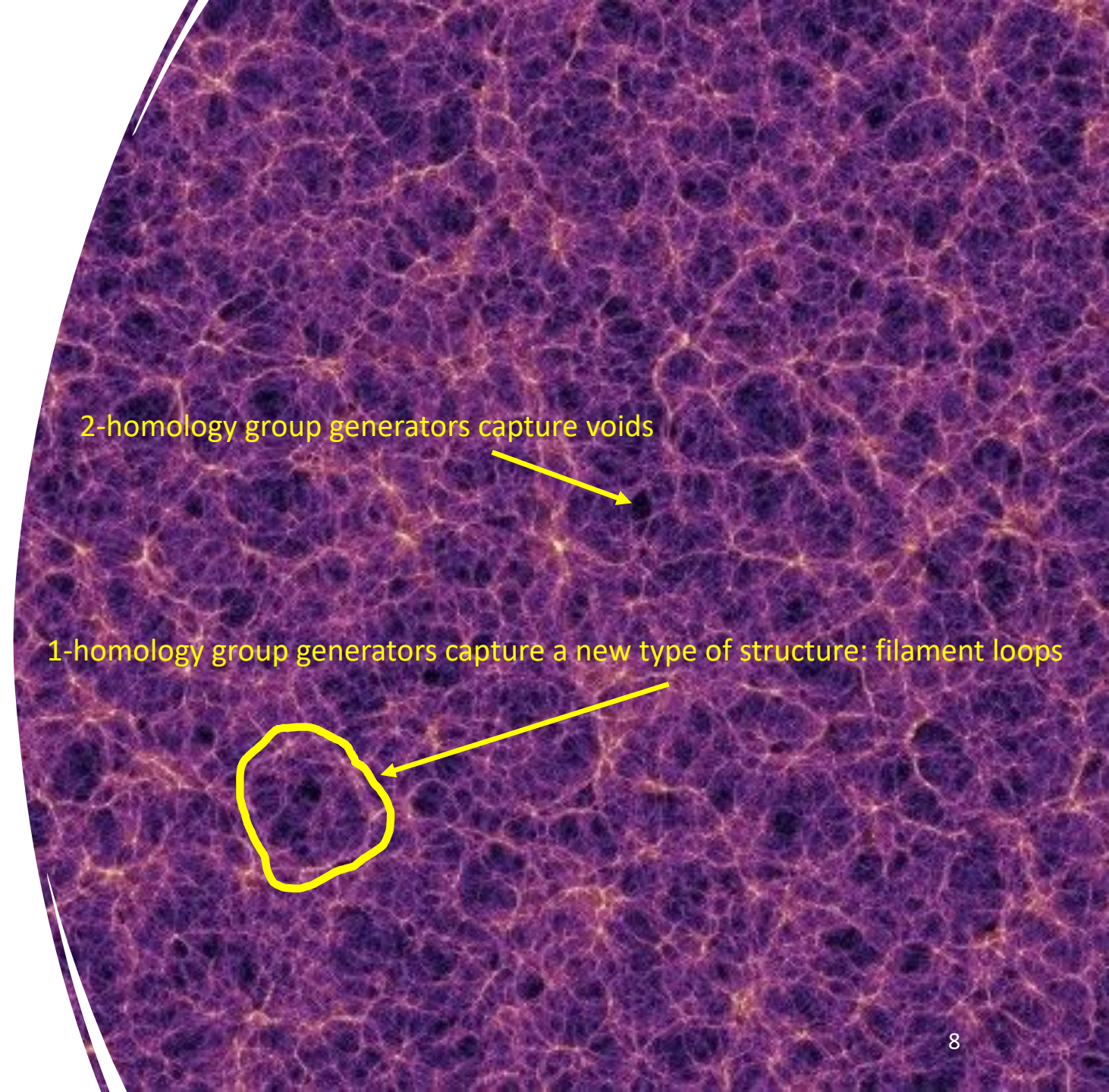


Holes in LSS

Millennium Simulation
(Springel+2005)

A cosmological simulation of the LSS of the
Universe (the *Cosmic Web*)

Lighter = high density regions
Darker = low density regions



Motivation: TDA for Cosmology

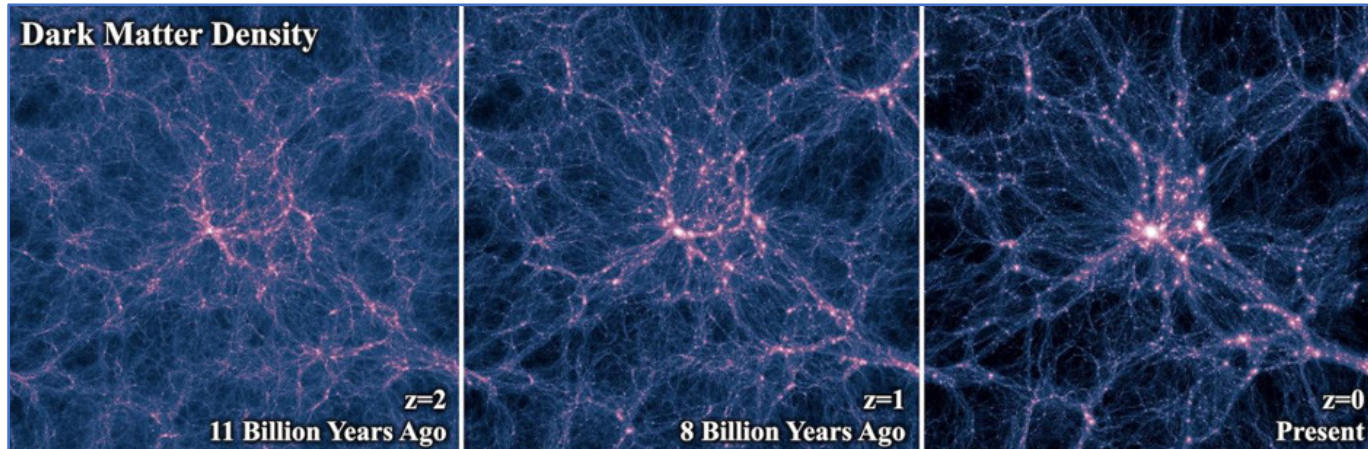


Image credit: Nelson / Illustris Collaboration (from <http://spaceref.com>)

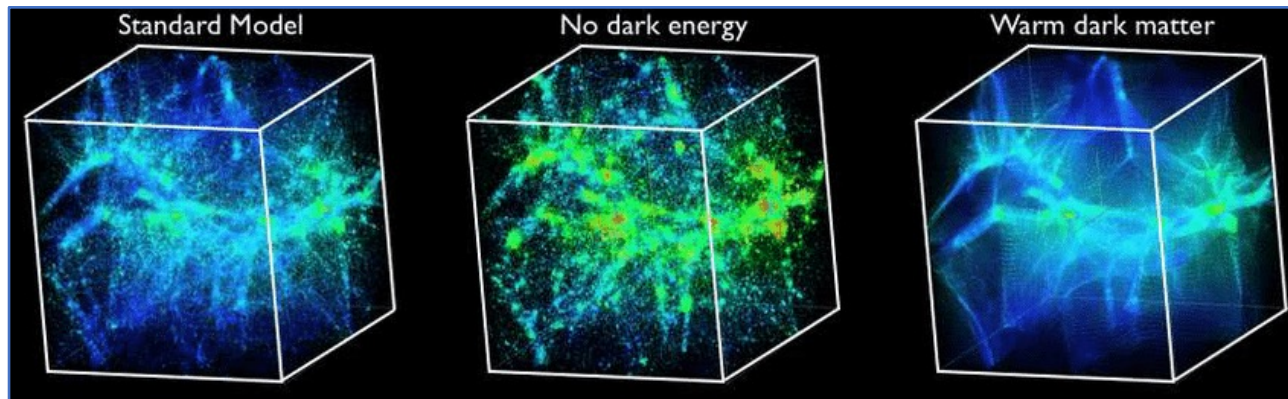


Image credit: Figure 1.10 from Vennin, V., 2014. Cosmological inflation: theoretical aspects and observational constraints (Doctoral dissertation, Université Pierre et Marie Curie).

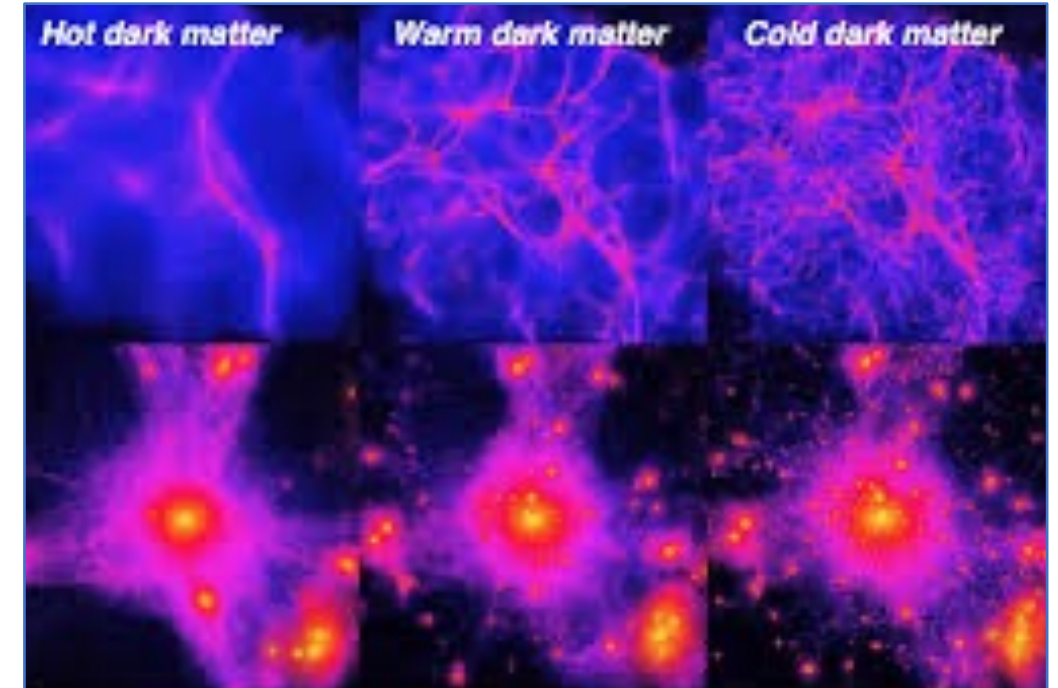


Image credit: Ben Moore (University of Zurich)

Can persistent homology discriminate LSS generated under different models?

Persistent homology

Homology

Count holes



Persistent homology

Count holes in data

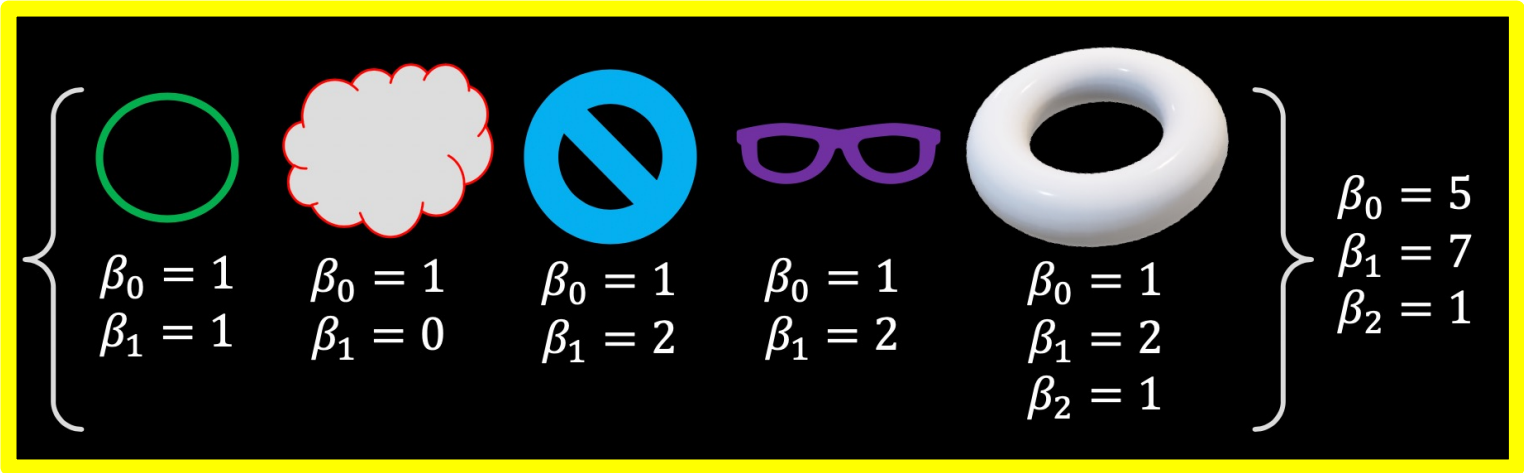
Betti Numbers

β_0 = Connected Components

β_1 = Loops

β_2 = Voids

...

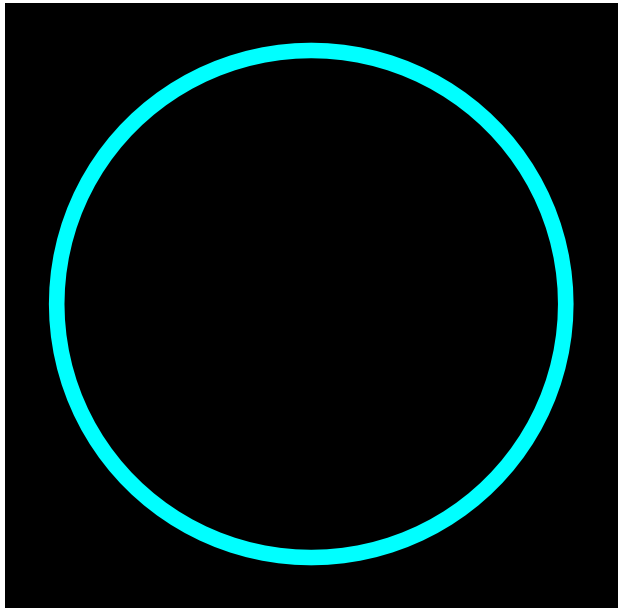


Euler characteristic = $\beta_0 - \beta_1 + \beta_2 - \dots$

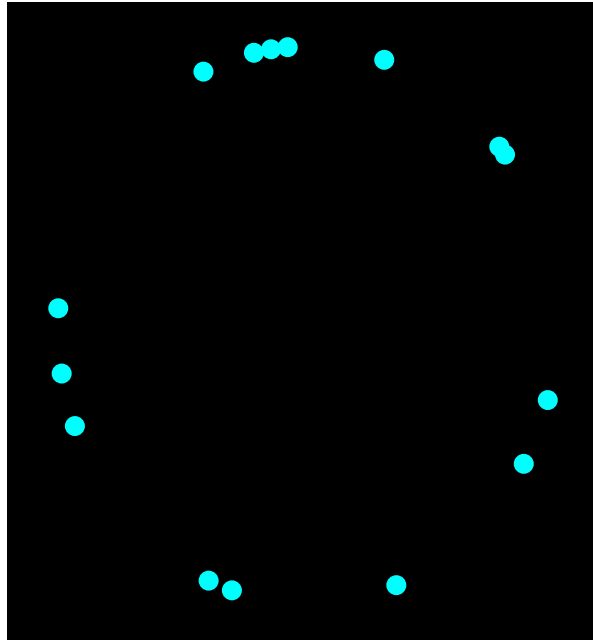
Homology: considering data

β_0 = Connected Components
 β_1 = Loops

$$\beta_0 = 1, \beta_1 = 1$$



$$\beta_0 = 15, \beta_1 = 0$$



?

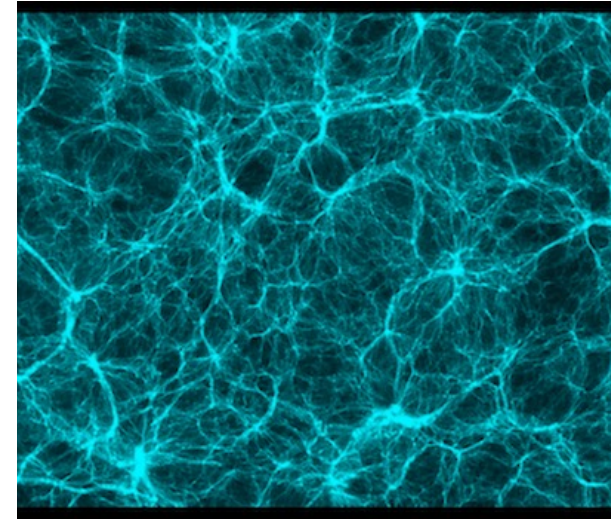
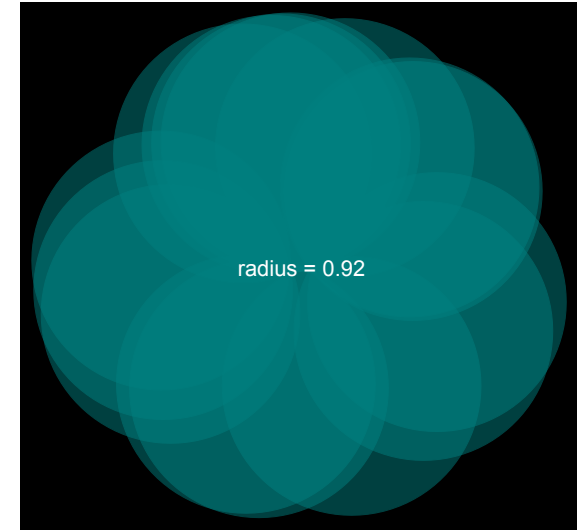
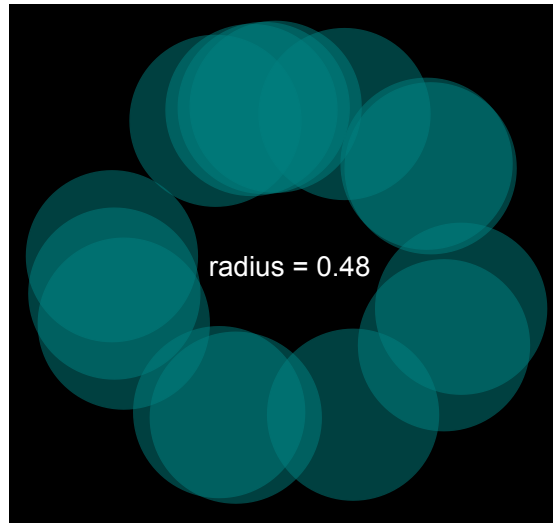
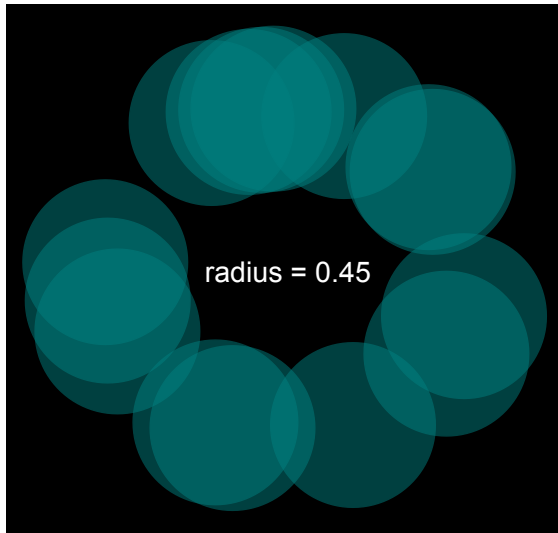


Image of LSS: <http://astro.berkeley.edu>

Persistent homology is a multi-scale version of homology

(e.g., Edelsbrunner+2002; Edelsbrunner and Harer 2008; Carlsson 2009)

Persistent homology: Vietoris Rips filtration



Birth of loop: diameter = $2 \times 0.48 = 0.96$

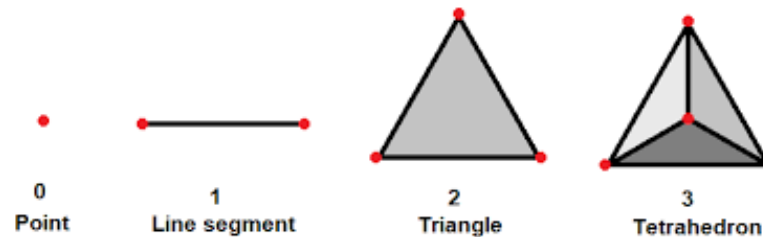
Death of loop: diameter = $2 \times 0.92 = 1.84$

Persistence (or lifetime) of loop: $1.84 - 0.96 = 0.88$

- Define $S_\epsilon = \bigcup_{i=1}^n B(Y_i, \epsilon)$ (union of balls with radius ϵ centered at observations $Y_1, \dots, Y_n$)
- Persistent homology tracks the changing homology of S_ϵ across a range of ϵ 's
- The birth and death times of different holes (homology group generators) are plotted on a persistence diagram.

Simplex

- A **p-simplex** σ is the convex hull of $p+1$ affinely independent points $x_0, x_1, \dots, x_p \in \mathbb{R}^d$. Denote $\sigma = \text{conv}\{x_0, \dots, x_p\}$, and the dimension of σ is p .
 - A 0-simplex is a vertex, 1-simplex an edge, 2-simplex a triangle, and 3-simplex a tetrahedron

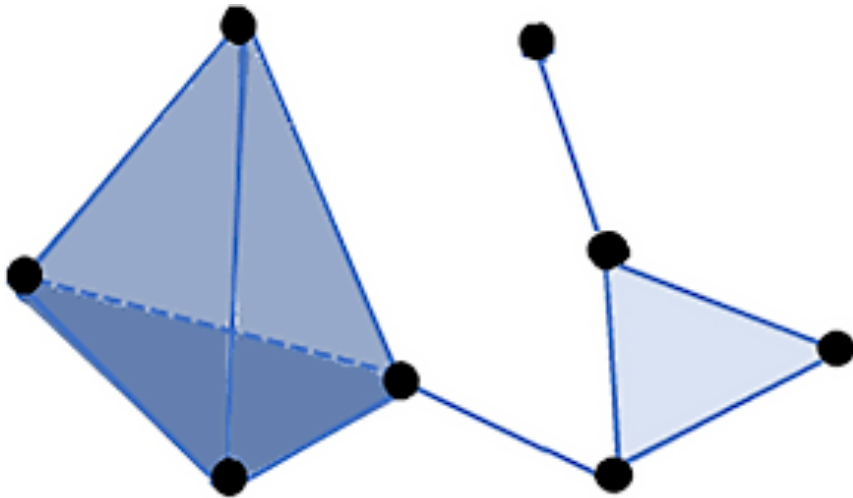


- A **face** of σ is $\text{conv}S$ where $S \subset \{x_0, \dots, x_p\}$ is a subset of the $p+1$ vertices.

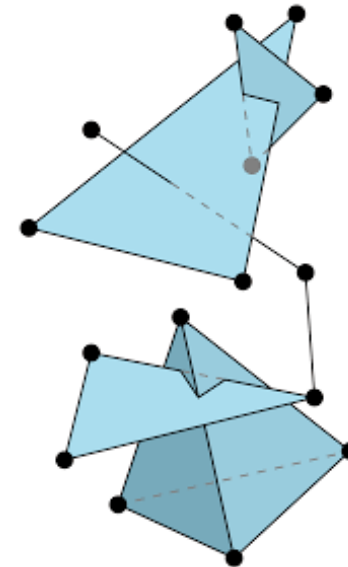
Simplicial complex

A **simplicial complex** K is a finite collection of simplices such that

- 1) $\sigma \in K$ and τ being a face of σ implies $\tau \in K$, and
- 2) $\sigma, \sigma' \in K$ implies $\sigma \cap \sigma'$ is either empty or a face of both σ and σ' .



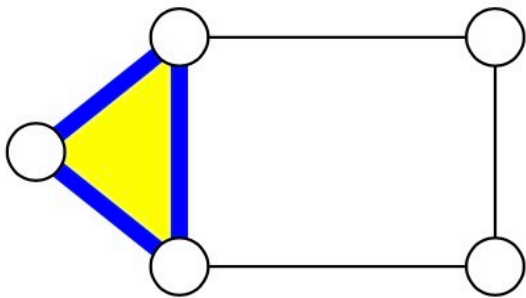
Simplicial complex



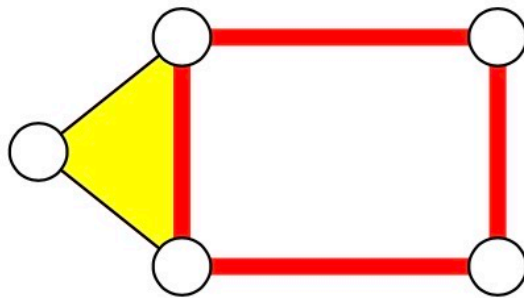
Not a simplicial complex

Finding a hole (basic illustration)

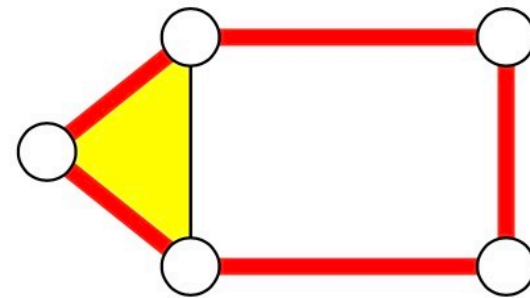
- Loops (1-cycles) for simplicial complex below: $Z_1 = \{0, c_1, c_2, c_3\}$
- Uninteresting loops: $B_1 = \{0, c_1\}$ (“boundary cycles”)
- Interesting 1-cycles are $c_2 + B_1 = c_3 + B_1 = \{c_2, c_3\}$



C_1

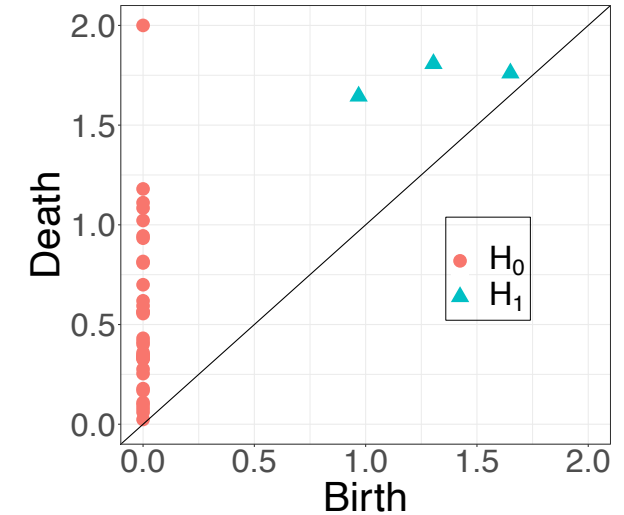
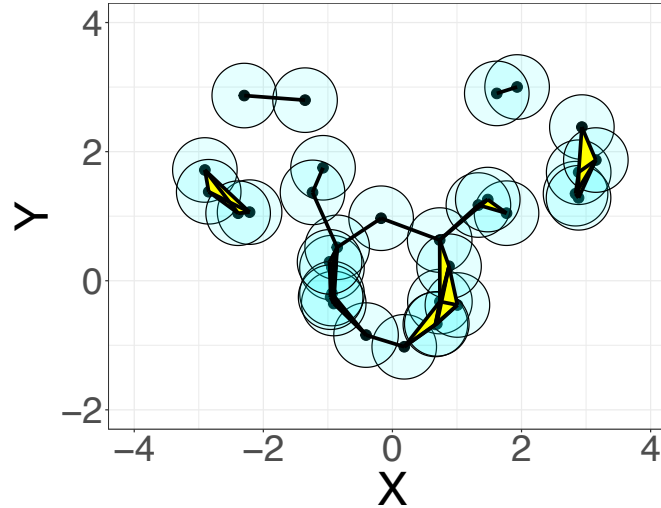
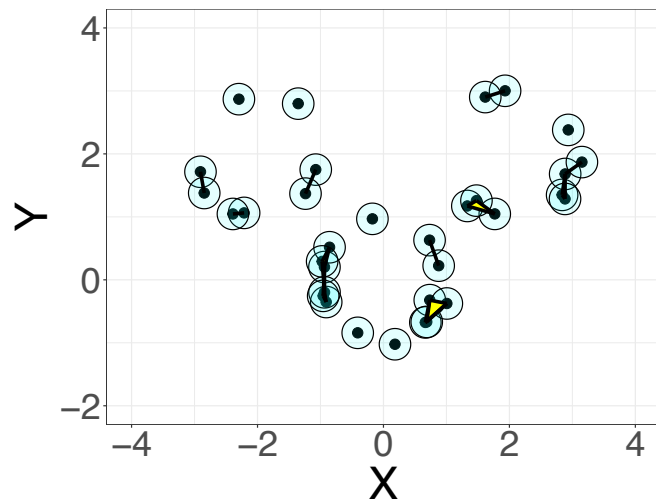


C_2



C_3

Persistent homology summaries



A persistence diagram D is a collection of birth (b_j) and death (d_j) times of homology group generators of a particular dimension (r_j):

$$D = \{(r_j, b_j, d_j) \mid j = 1, \dots, n_l\}$$

where n_l is the number of homology group generators off the diagonal.

- Rather than defining the filtration using a Rips Complex over the data points, a function can be used for persistent homology
- Kernel density estimates or Distance-to-Measure (DTM) functions (Chazal+2011) are popular approaches in TDA for turning a point-cloud of data into a function

Some examples from astronomy



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

Willem Elbers¹★ and Rien van de Weygaert

Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700AV Groningen, the Netherlands



The topology of the cosmic web in terms of persistent Betti numbers

Pratyush Pranav,^{1,2}★ Herbert Edelsbrunner,³ Rien van de Weygaert,¹ Gert Vegter,⁴ Michael Kerber,⁵ Bernard J. T. Jones¹ and Mathijs Wintraecken^{4,6}

A&A 648, A74 (2021)
<https://doi.org/10.1051/0004-6361/202039048>
© ESO 2021

**Astronomy
&
Astrophysics**

Persistent homology in cosmic shear: Constraining parameters with topological data analysis

Sven Heydenreich¹, Benjamin Brück², and Joachim Harnois-Déraps^{3,4}

Alpha, Betti and the Megaparsec Universe: On the Topology of the Cosmic Web

Rien van de Weygaert¹, Gert Vegter², Herbert Edelsbrunner³, Bernard J.T. Jones¹, Pratyush Pranav¹, Changbom Park⁴, Wojciech A. Hellwing⁵, Bob Eldering², Nico Kruithof², E.G.P. (Patrick) Bos¹, Johan Hidding¹, Job Feldbrugge¹, Eline ten Have⁶, Matti van Engelen², Manuel Caroli⁷, and Monique Teillaud⁷



Persistent homology of the cosmic web – I. Hierarchical topology in Λ CDM cosmologies

Georg Wilding^{1,2,3}★, Keimpe Nevenzeel,¹ Rien van de Weygaert,^{1,3} Gert Vegter,^{2,3} Pratyush Pranav^{1,4,5}, Bernard J. T. Jones¹, Konstantinos Efstathiou^{2,3,6} and Job Feldbrugge^{7,8}

Hypothesis testing

Cold Dark Matter vs. Warm Dark Matter

References:

C-K, Fasy, Hellwing, Lovell, Drozda, and Wu (2022)

Berry, Chen, C-K, and Fasy (2020)

Copernicus Complexio (COCO) simulations

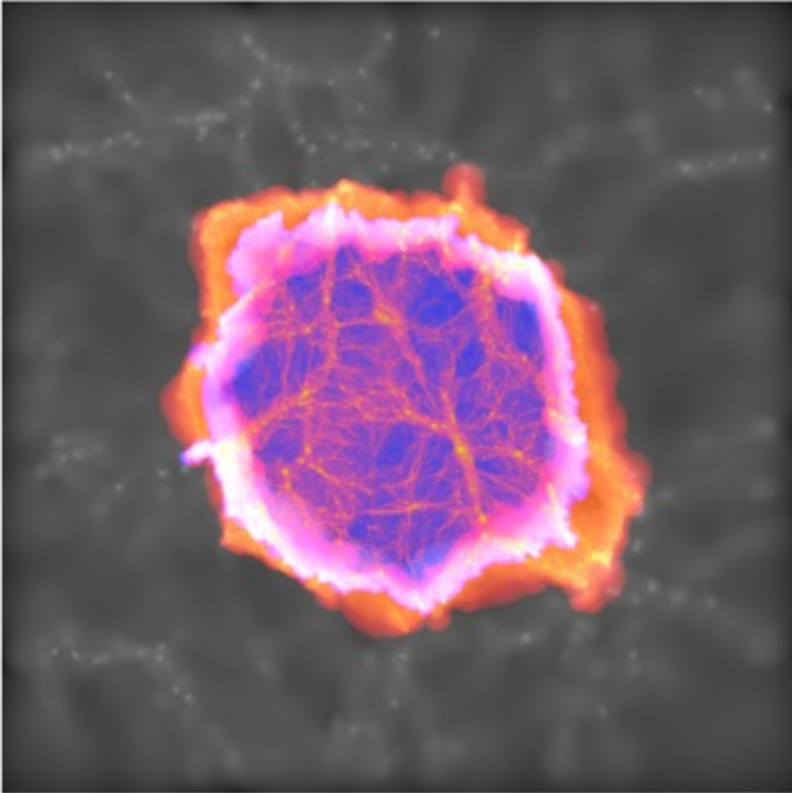


Figure 1. Projected density along the z -axis of a $70.4 \times 70.4 \times 1.5 h^{-1}$ Mpc slice centred on the middle of the COCO simulation at redshift $z = 0$. The various colours show the density at different resolution levels: lowest resolution (grey), medium resolution (orange and purple) and high resolution (blue to yellow). Note the amazing level of the cosmic web details seen inside the high-resolution region.

Figure credit: Hellwing+2016

- COCO cosmological simulation data
 - Cold Dark Matter (CDM, Hellwing+2016)
 - Warm Dark Matter (WDM, Bose+2016)
- N-body simulations of structure formation
- Simulation box: 70.4 Mpc/h per side
 - High-resolution central sphere with radius 17.4 Mpc/h
- Objects in data are dark matter **halos** with a particular mass

Note:

1 pc = 1 parsec $\approx$ 3.26 light years

h = encodes uncertainty about the expansion rate of the Universe

Two-sample hypothesis test

Given two sets of persistence diagram,

$$D_1^{(1)}, \dots, D_{n_1}^{(1)} \sim \mathcal{D}^{(1)}$$

$$D_1^{(2)}, \dots, D_{n_2}^{(2)} \sim \mathcal{D}^{(2)}$$

where $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(2)}$ are the true underlying distributions of persistence diagrams for group 1 and 2, respectively (existence of distributions established in Mileyko+2011).

Then consider the two-sample hypothesis testing framework with

$$H_0 : \mathcal{D}^{(1)} = \mathcal{D}^{(2)} \text{ vs. } H_1 : \mathcal{D}^{(1)} \neq \mathcal{D}^{(2)}$$

Test statistic: Persistence diagrams are difficult objects to work with (e.g., computationally expensive to compute distances between two diagrams)

→ consider functional summaries of persistence diagrams

Functional summaries of persistence diagrams

Many functional summaries have been proposed

(e.g., Chazal+2014; Adams+2017; Bubenik 2015; Chen+2015; Biscio and Møller 2019)

Functional summary

$\mathbb{F}: \mathcal{D} \rightarrow \mathcal{F}$ where $\mathcal{D}$ is the space of persistence diagrams and $\mathcal{F}$ is a collection of functions.

Random diagrams, $D_1, \dots, D_n \in \mathcal{D}$ become random functions $F_1 = \mathbb{F}(D_1), \dots, F_n = \mathbb{F}(D_n) \in \mathcal{F}$

Test statistics

Let $F_{s,i} = \mathbb{F}(D_i^{(s)})$, for diagram i of sample $s = 1, 2$, and $\hat{F}_s = n_s^{-1} \sum_{i=1}^{n_s} F_{s,i}(t)$
 $\hat{F}_s$ is a consistent estimator of the population mean functional summary, $\mathbb{E}(F_s(t))$

→ Use $T = d(\hat{F}_1, \hat{F}_2)$ as test statistic for some metric $d(\cdot, \cdot)$

$$d(\hat{F}_1, \hat{F}_2) = \int_{\mathbb{T}} |\hat{F}_1(t) - \hat{F}_2(t)| dt$$

Reference: Berry, Chen, C-K, and Fasy (2020)

Landscape and Silhouette functions

Landscape functions (Bubenik 2015, Bubenik and Dlotko 2017) are the collection of functions

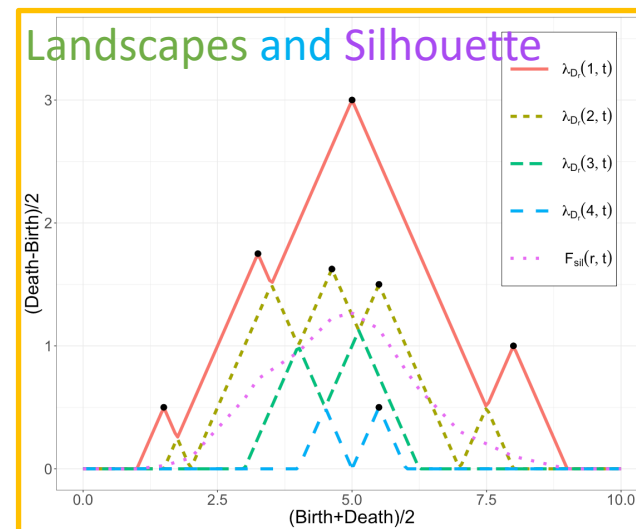
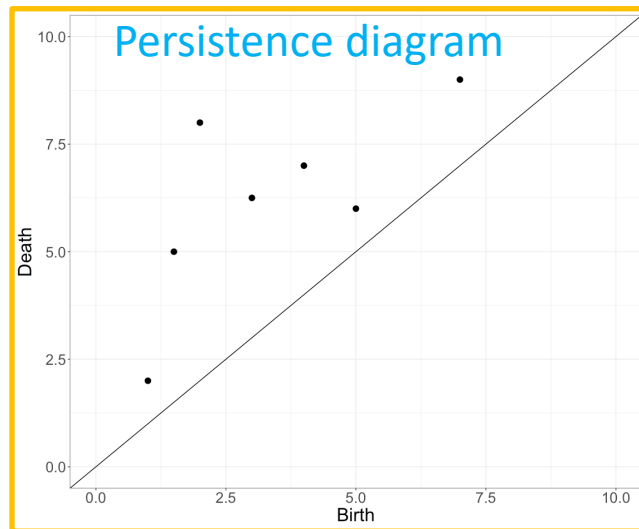
$$\lambda_D(k, t) = \text{kmax}_{p_j \in D} \Lambda_{p_j}(t), \text{ where } \Lambda_{p_j}(t) = \begin{cases} t - b_j & t \in [b_j, \frac{d_j+b_j}{2}] \\ d_j - t & t \in [\frac{d_j+b_j}{2}, d_j] \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{F}_{\text{land}}(\mathcal{I}, r, t) = \bigoplus_{k \in \mathcal{I}} \lambda_{D_r}(k, t)$$

where kmax selects the kth largest value, $k \in \mathbb{N}$, $p_j = \left(\frac{b_j+d_j}{2}, \frac{d_j-b_j}{2}\right)$, $j = 1, \dots, n_l$, r is the homology group dimension and $\mathcal{I}$ be an index set of landscape layers.

Weighted Silhouette Functions are weighted combinations of landscape function, defined as

$$\mathcal{F}_{\text{sil}}(r, t \mid p) = \frac{\sum_{j=1}^{n_r} |d_{r,j} - b_{r,j}|^p \Lambda_{p_{r,j}}(t)}{\sum_{j=1}^{n_r} |d_{r,j} - b_{r,j}|^p}$$



Euler Characteristic and Betti Functions

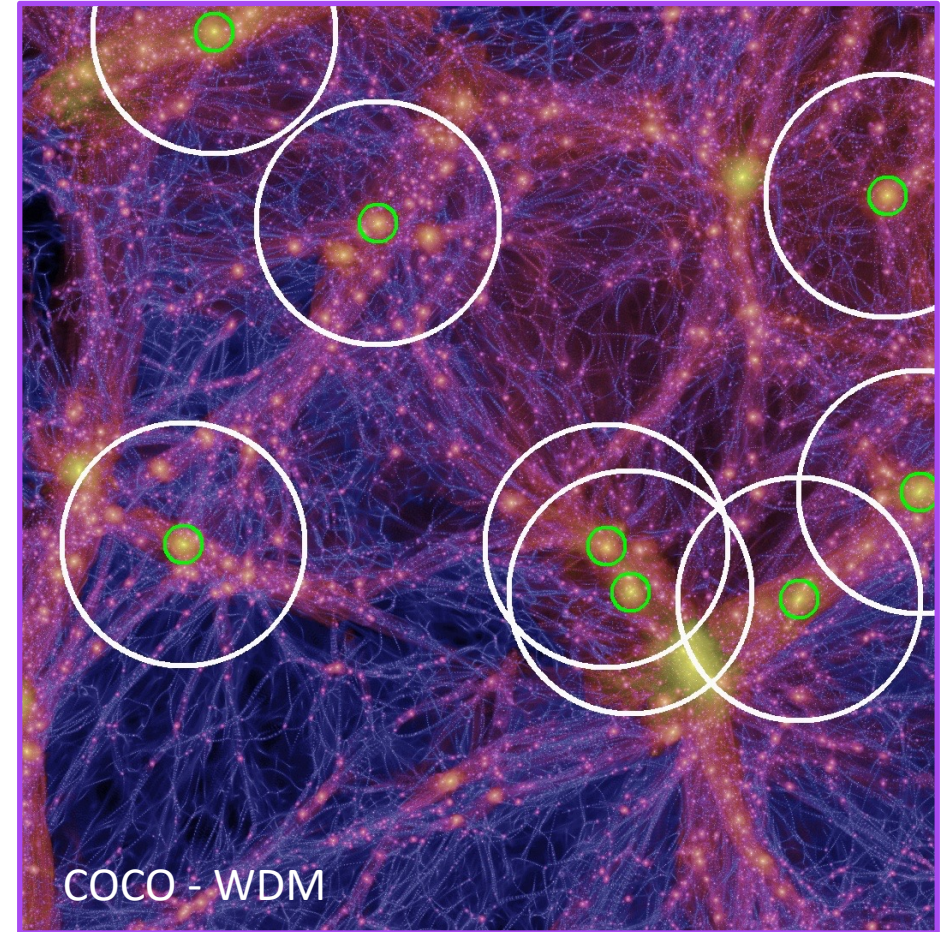
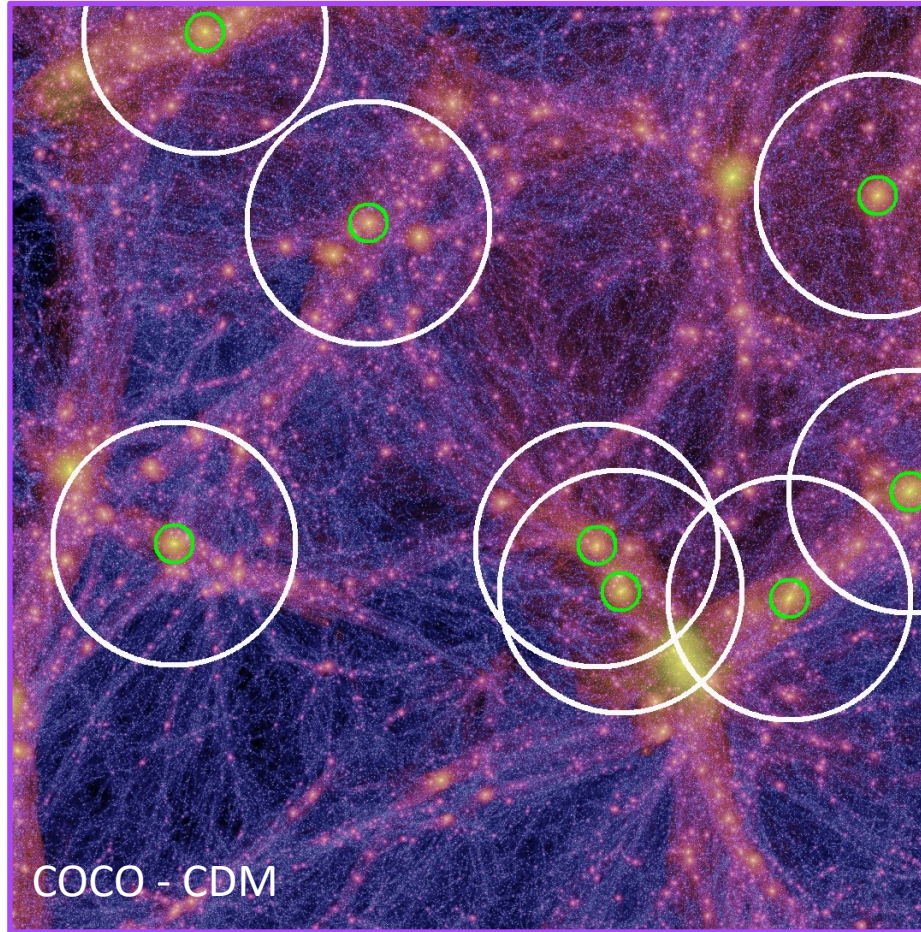
The **Betti functions** track the number of homology group generators (i.e., the homology group rank) that persist across the filtration parameter t for each dimension r , defined as

$$F_{\text{betti}}(r, t) = |\{(r, b_j, d_j) : b_j \leq t, d_j > t\}|$$

The **Euler Characteristic function** is the alternating sum of the Betti functions, defined as

$$\mathcal{F}_{\text{ec}}(t) = \sum_{r=0}^2 (-1)^r \mathcal{F}_{\text{betti}}(r, t)$$

COCO Milky-Way-analog Halo Samples



- 23x23x10 Mpc slide of COCO-CDM (left) and COCO-WDM (right).
- Eight of the 77 volumes used in this study are displayed: the MW-analog halos are enclosed by a green circle with a radius 3 Mpc indicated by a white circle.
- Selected haloes have masses between $0.5 \times 10^{12} M_{\text{Sun}}$ and $2 \times 10^{12} M_{\text{Sun}}$ with no haloes with masses $> 0.5 \times 10^{12} M_{\text{Sun}}$ within 0.7 Mpc.

Comparison Methods

Persistence Diagram Test (PDT), Robinson & Turner 2017) compares persistence diagrams directly using the following test statistic

$$\mathcal{T}_{\text{PDT}}(D_{1,1:n_1|r}, D_{2,1:n_2|r} \mid p, q) = \sum_{l=1}^2 \frac{1}{2n_l(n_l-1)} \sum_{i=1}^{n_l} \sum_{j=1}^{n_l} W_p(D_{l,i|r}, D_{l,j|r})^q$$

$D_{l,1:n_l|r}$ = set of n_l persistence diagrams for homology dimension r from population $l = 1, 2$

$W_p(\cdot, \cdot)$ is the **p -Wasserstein distance**, with $1 \leq p \leq \infty$ and $1 \leq q < \infty$

$$W_p(D_1, D_2) = \left(\inf_{\eta: D_1 \rightarrow D_2} \sum_{x \in D_1} \|x - \eta(x)\|_\infty^p \right)^{\frac{1}{p}}$$

We use the **Bottleneck distance** with $p = \infty$ and $q = 1$

$$W_\infty(D_1, D_2) = \inf_{\eta: D_1 \rightarrow D_2} \sup_{x \in D_1} \|x - \eta(x)\|_\infty$$

Spatial Point Process Functions:

G-function (distribution function of the nearest-neighbor distance)

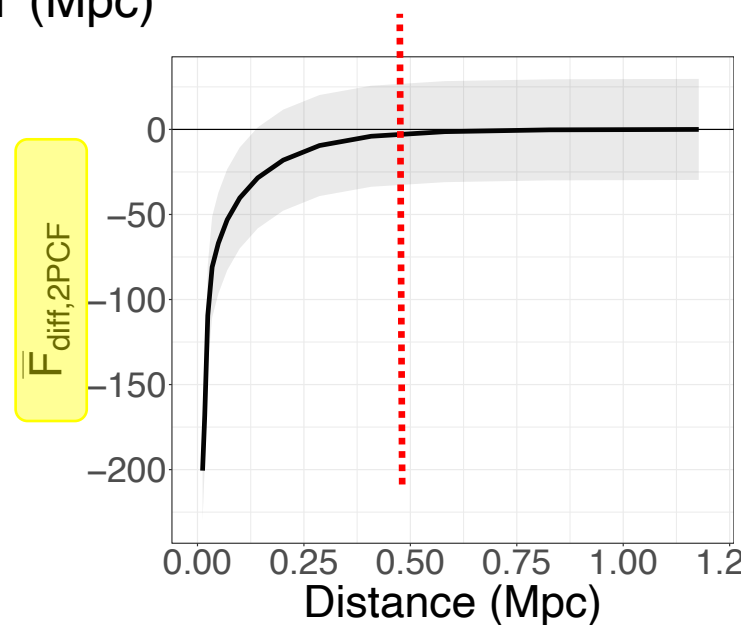
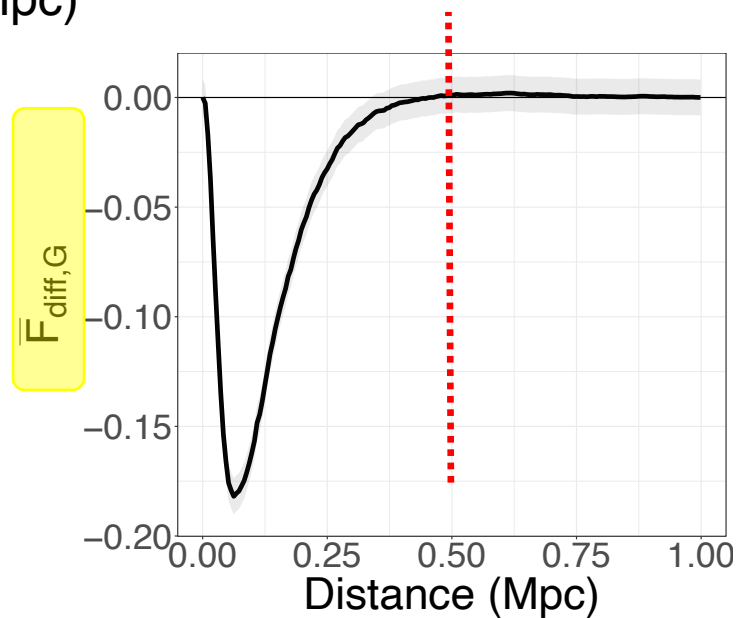
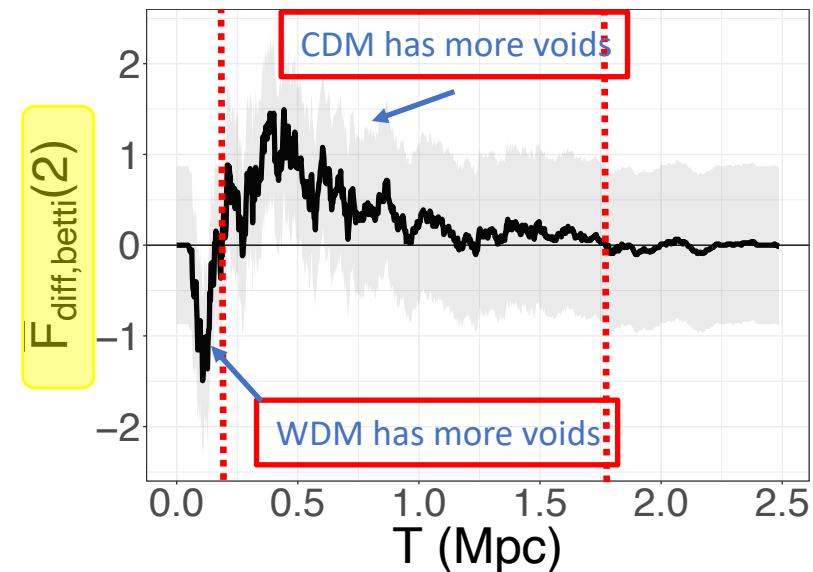
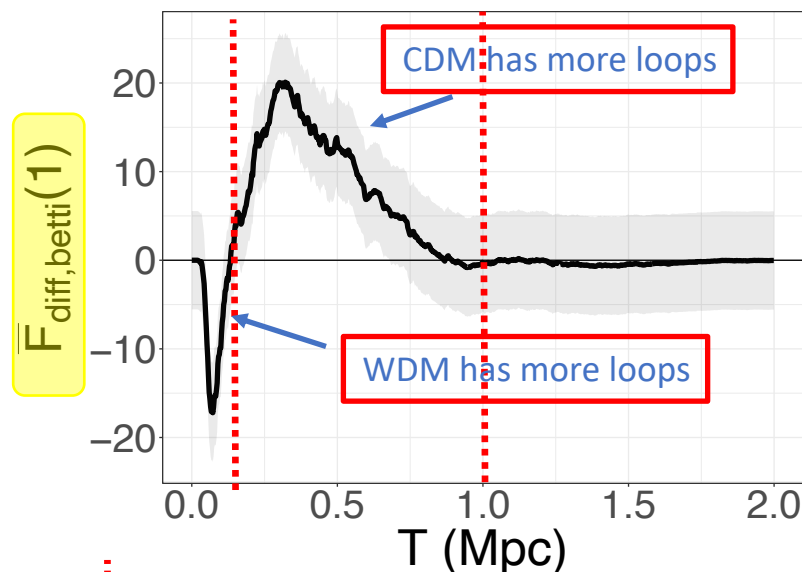
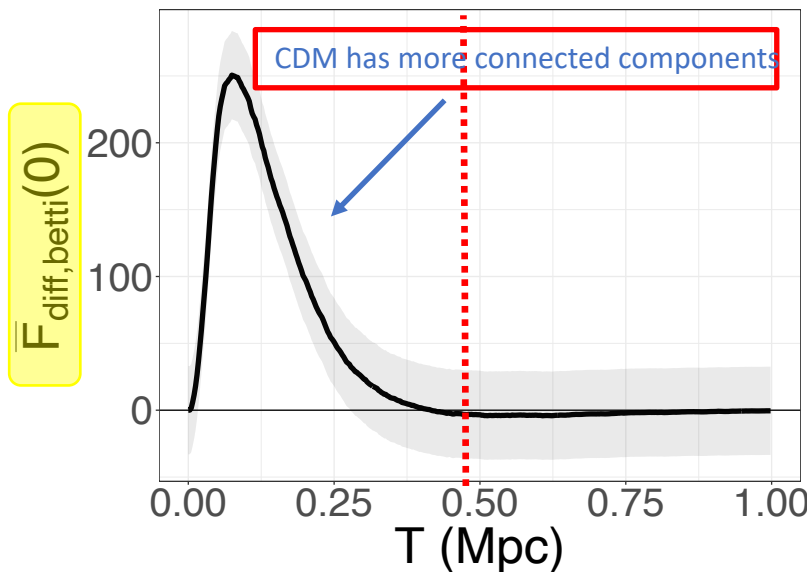
Pair-correlation function (aka two-point correlation function, 2PCF)

Results (p-values, 20,000 permutations)

Test statistic	Notation	Homology dimension	p_{perm}	p_{matched}
Landscape	$\mathcal{F}_{\text{land}}(1 : 10, 0, t)$	0	0	0
Silhouette	$\mathcal{F}_{\text{sil}}(0, t \mid p = 0.5)$	0	0.009	0
Silhouette	$\mathcal{F}_{\text{sil}}(0, t \mid p = 1)$	0	0.001	0
Silhouette	$\mathcal{F}_{\text{sil}}(0, t \mid p = 2)$	0	0	0
Betti	$\mathcal{F}_{\text{betti}}(0, t)$	0	0.001	0
PDT	$\mathcal{T}_{\text{PDT}}(D_{1,\cdot 0}, D_{2,\cdot 0} \mid \infty, 1)$	0	0.003	0
Landscape	$\mathcal{F}_{\text{land}}(1 : 10, 1, t)$	1	0	0
Silhouette	$\mathcal{F}_{\text{sil}}(1, t \mid p = 0.5)$	1	0.009	0
Silhouette	$\mathcal{F}_{\text{sil}}(1, t \mid p = 1)$	1	0.007	0
Silhouette	$\mathcal{F}_{\text{sil}}(1, t \mid p = 2)$	1	0.014	0
Betti	$\mathcal{F}_{\text{betti}}(1, t)$	1	0	0
PDT	$\mathcal{T}_{\text{PDT}}(D_{1,\cdot 1}, D_{2,\cdot 1} \mid \infty, 1)$	1	0.255	0.036
Landscape	$\mathcal{F}_{\text{land}}(1 : 10, 2, t)$	2	0	0
Silhouette	$\mathcal{F}_{\text{sil}}(2, t \mid p = 0.5)$	2	0.123	0.003
Silhouette	$\mathcal{F}_{\text{sil}}(2, t \mid p = 1)$	2	0.158	0.009
Silhouette	$\mathcal{F}_{\text{sil}}(2, t \mid p = 2)$	2	0.135	0.008
Betti	$\mathcal{F}_{\text{betti}}(2, t)$	2	0.001	0
PDT	$\mathcal{T}_{\text{PDT}}(D_{1,\cdot 2}, D_{2,\cdot 2} \mid \infty, 1)$	2	0.009	0
Euler characteristic	$\mathcal{F}_{\text{ec}}(t)$	0-2	0.001	0
G-function	$\mathcal{F}_{\text{G}}(t)$	N/A	0	0
2PCF	$\mathcal{F}_{2\text{PCF}}(t)$	N/A	0	0

Interpretation

$$\bar{F}_{\text{diff}}(t) = n_s^{-1} \sum_{i=1}^{n_s} (F_{c,i}(t) - F_{w,i}(t))$$



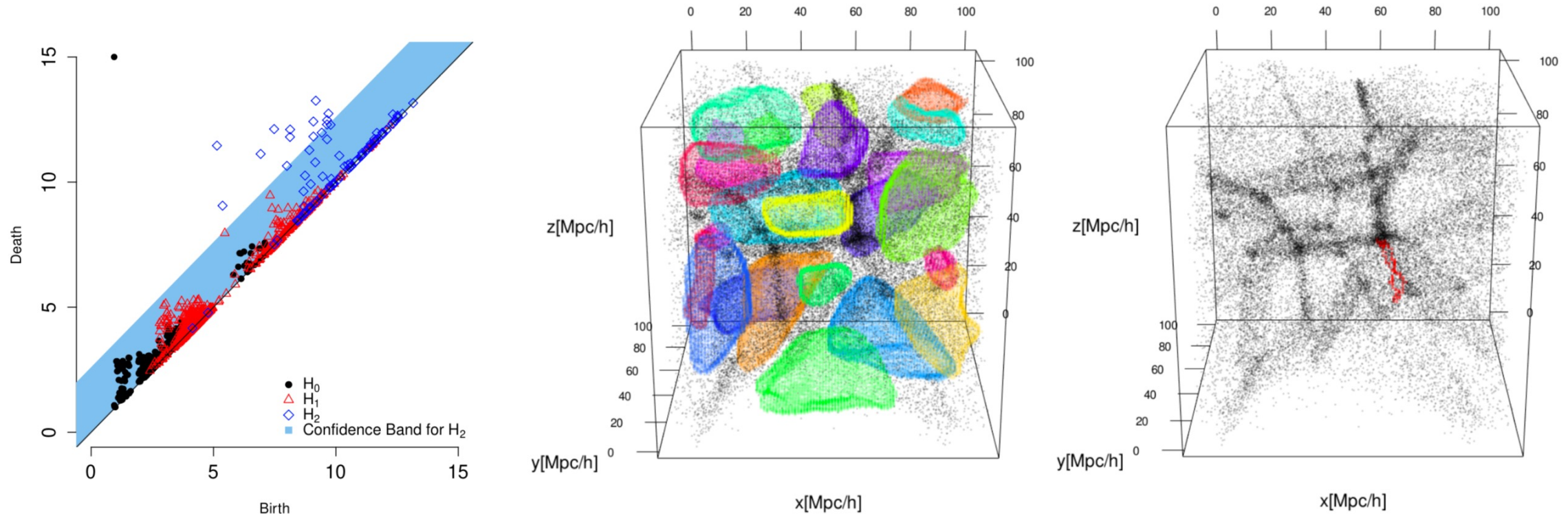
Finding holes in the Universe

Representations of homology group generators

Reference: Xu, C-K, Green, and Nagai (2019)

Visualization of Holes

Significant Cosmic Holes in Universe (SCHU) identifies representations of cosmic voids and loops of filaments in cosmological datasets and assigns statistical significance



(left) Persistence diagram with 90% confidence band for H_2 ; (middle) 23 voids identified with SCHU; (right) example of a filament loop with a p-value = 0.011

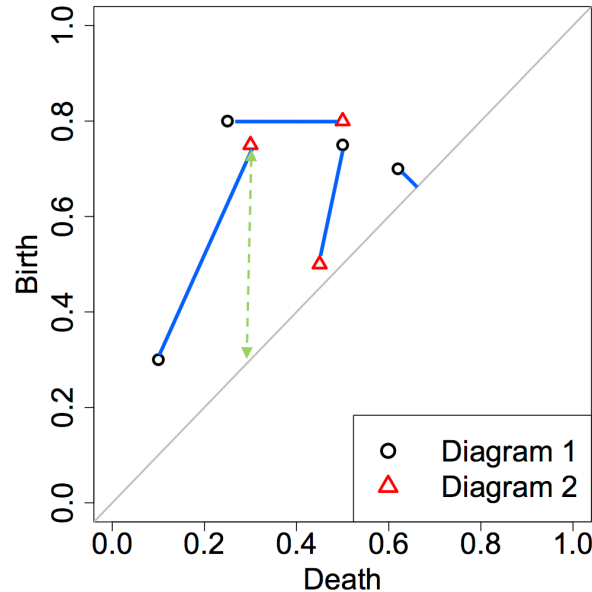
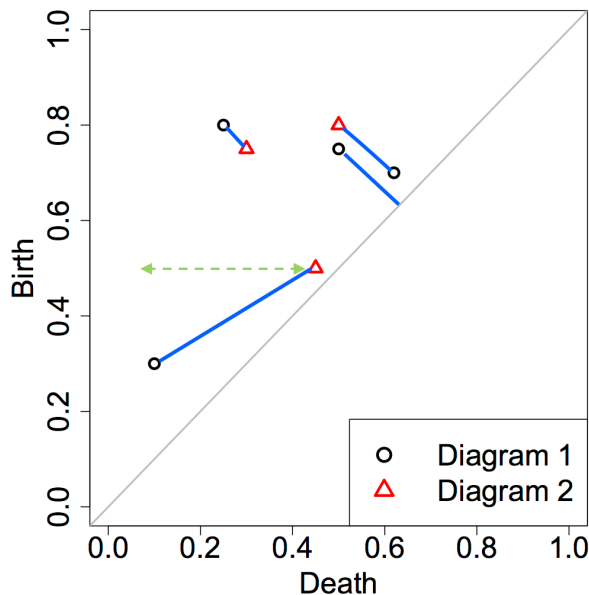
Data based on simulation motivated by Voronoi Foam (Icke and van de Weygaert1987; van de Weygaert+1989; van de Weygaert1994)

Confidence sets on persistence diagrams

- Inference on persistence diagrams requires a notion of distance.
Popular choice: bottleneck distance
- Bottleneck distance between two diagrams, D_1 and D_2 and bijection η :

$$W_\infty(D_1, D_2) = \inf_{\eta: D_1 \rightarrow D_2} \sup_{x \in D_1} \|x - \eta(x)\|_\infty$$

- Examples of two matchings for Diagrams 1 and 2:

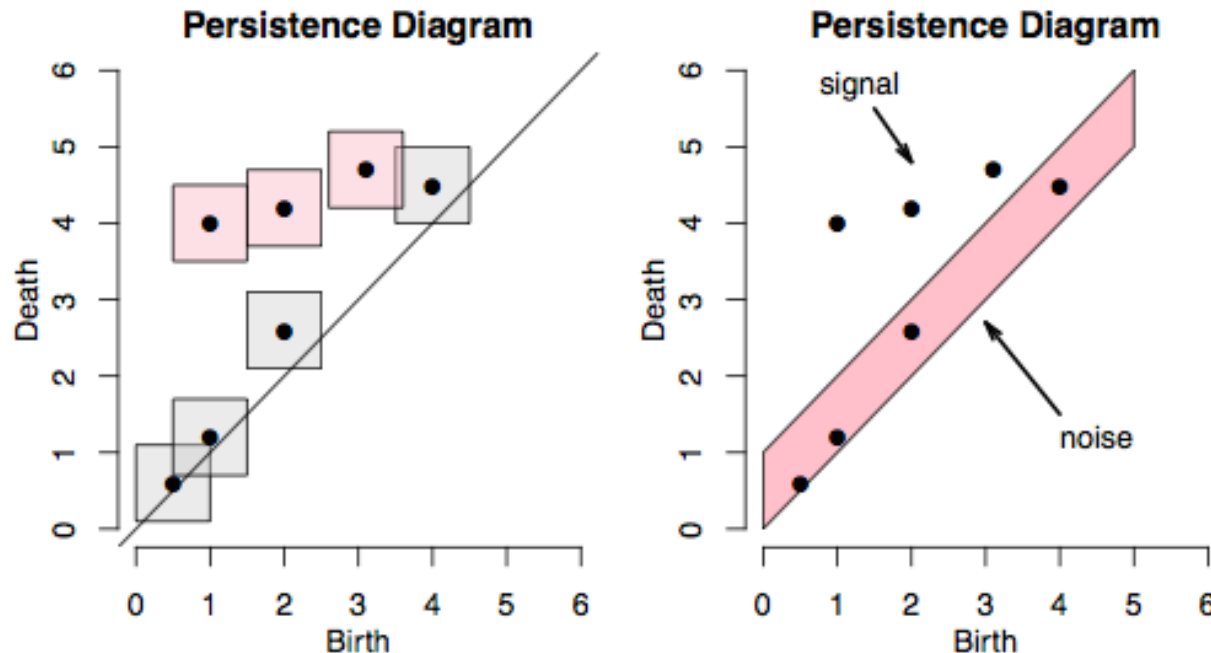


Fasy+2014 define a $(1 - \alpha)$ confidence set, C_n , for a persistence diagram D as an interval $[0, c_n]$, with $c_n \equiv c_n(Y_1, \dots, Y_n)$, such that

$$\limsup_{n \rightarrow \infty} P \left(W_\infty(\hat{D}_n, D) > c_n \right) \leq \alpha$$

where $\hat{D}_n$ is the observed diagram. Then

$$C_n \equiv \{ \tilde{D} : W_\infty(\hat{D}_n, \tilde{D}) \leq c_n \}$$

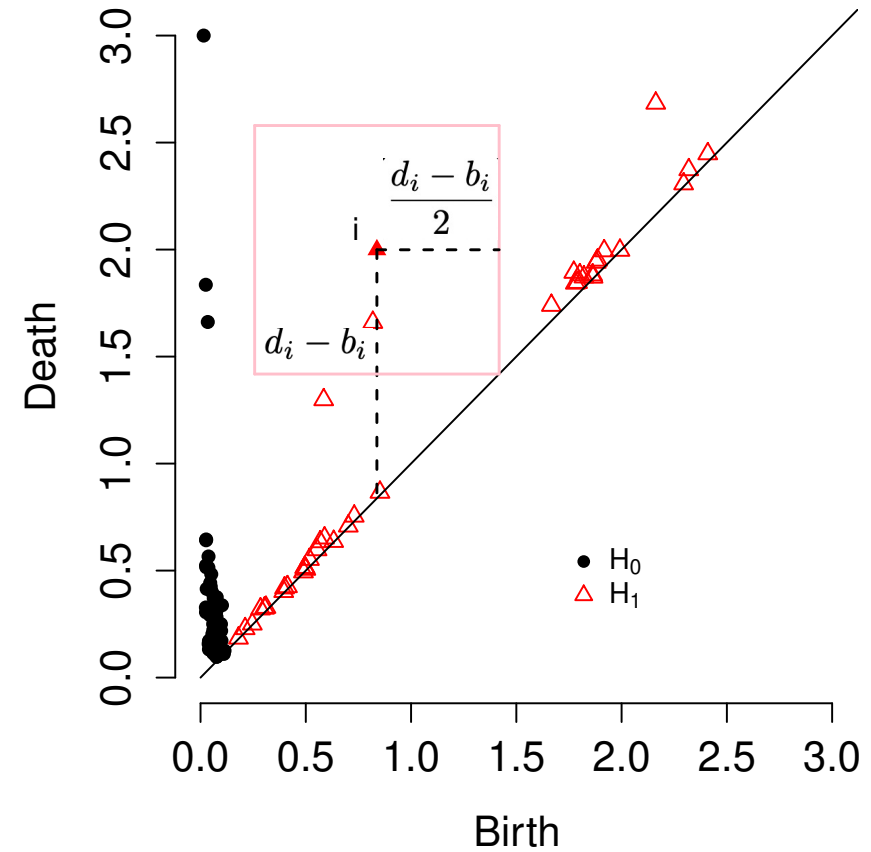


Relies on the Stability Theorem of Cohen-Steiner et al. (2007): $W_\infty(D(f), D(g)) \leq \|f - g\|_\infty$

(the L_∞ -norm between two functions, f and g , bounds the bottleneck distance between their corresponding persistence diagrams)

- Approximate the distribution of $W_\infty(\hat{D}_n, D)$ with a bootstrap sample of size N , $w^{(1)}, \dots, w^{(N)}$
 $\rightarrow$ denote the resulting empirical distribution as $\hat{F}_n(w)$
- Instead of confidence bands, SCHU assigns p-values to each homology group generator i as

$$\text{p-value}(i) = 1 - \hat{F}_n \left(\frac{d_i - b_i}{2} \right)$$



Reference: Xu, C-K, Green, and Nagai (2019), <https://github.com/xinxuyale/SCHU>
 R: TDA package

Simulation Study

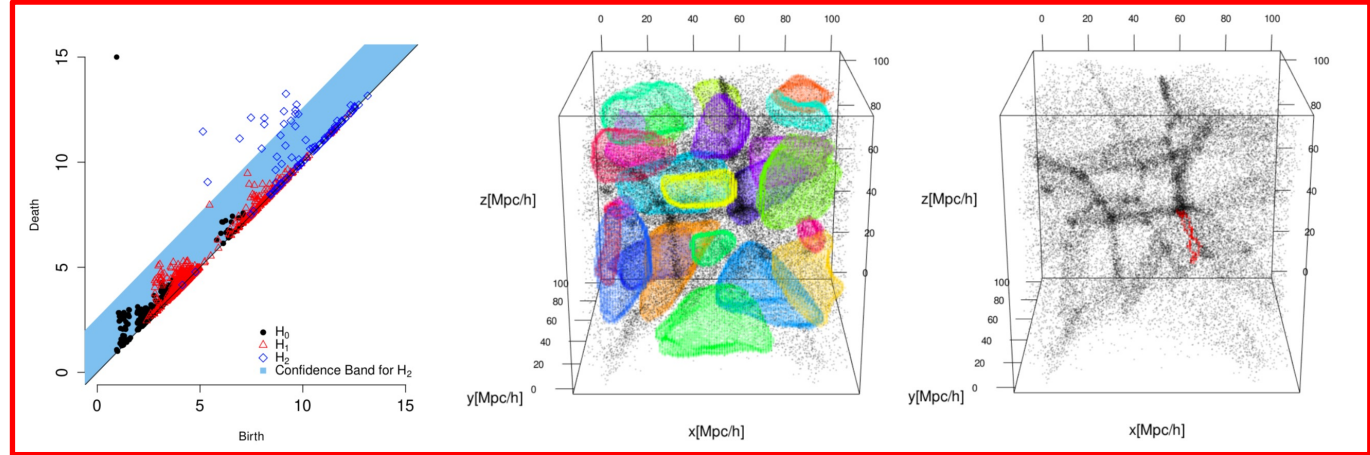
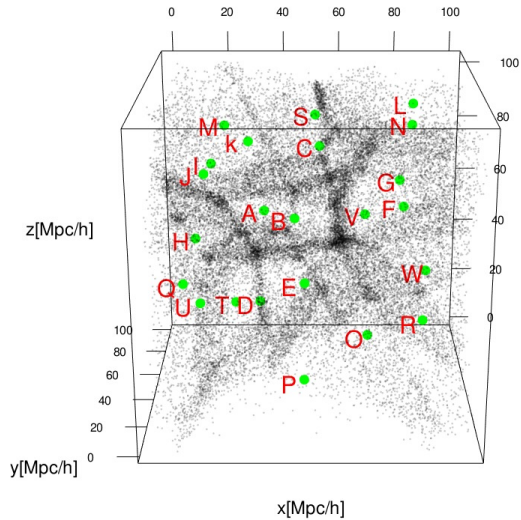


Table 1

p -values, birth and death times of all the 23 void generators found by SCHU, as shown in Fig. 10(c).

Generator	A	B	C	D	E	F	G	H	I	J	K	L
Birth	9.16	5.37	5.14	7.48	8.12	9.64	9.70	9.64	8.74	8.92	9.46	9.64
Death	10.79	9.06	11.46	12.12	11.80	11.71	12.74	12.48	10.26	11.28	10.22	12.30
p -value	0.866	0.001	<0.001	<0.001	0.001	0.406	0.034	0.058	0.931	0.192	>0.999	0.084
Generator	M	N	O	P	Q	R	S	T	U	V	W	
Birth	8.13	9.42	9.08	8.97	9.08	8.68	9.18	10.14	6.93	9.79	8.00	
Death	12.10	11.97	11.82	9.92	12.42	9.63	13.25	11.05	11.12	12.29	10.65	
p -value	0.001	0.114	0.071	>0.999	0.009	>0.999	0.001	>0.999	0.001	0.133	0.088	

- Mapped the volume center of each of the 23 most persistent H_2 generators to its nearest void seed point
- Each of the 23 void seeds were uniquely matched with a corresponding generator suggesting that SCHU was able to accurately locate the cosmic voids

Data based on simulation motivated by Voronoi Foam (Icke and van de Weygaert1987; van de Weygaert+1989; van de Weygaert1994)

Simulation box: $100 \text{ h}^{-1} \text{Mpc}$ per dimension and a grid spacing of $1 \text{ h}^{-1} \text{Mpc}$.

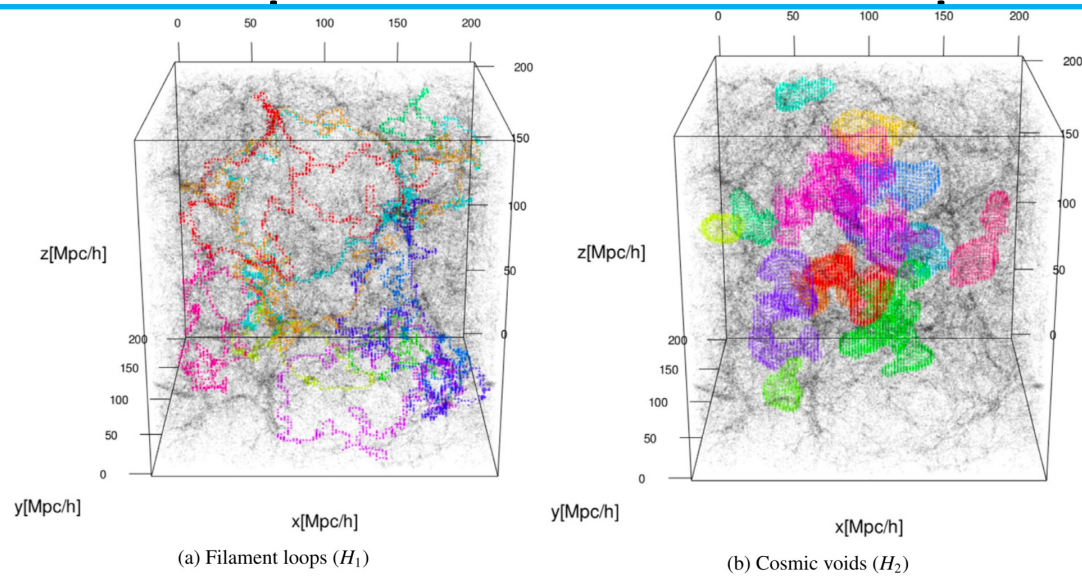
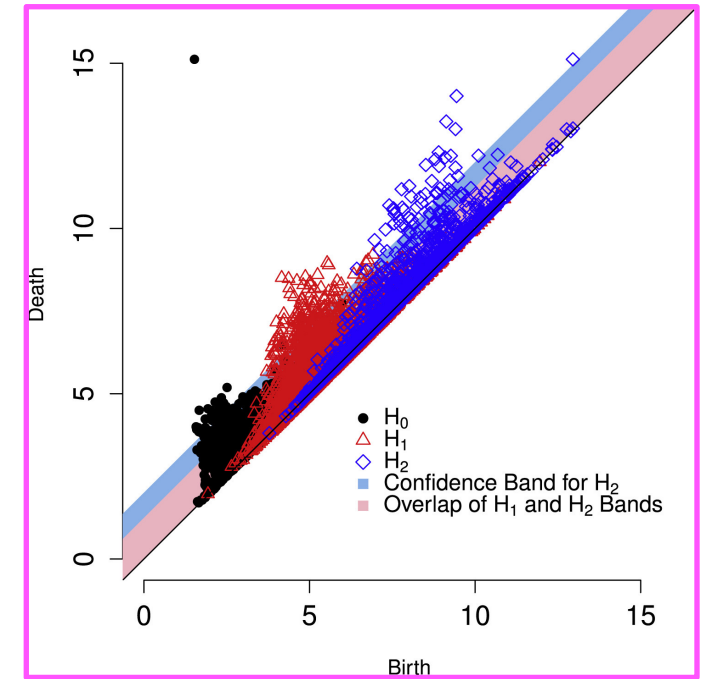
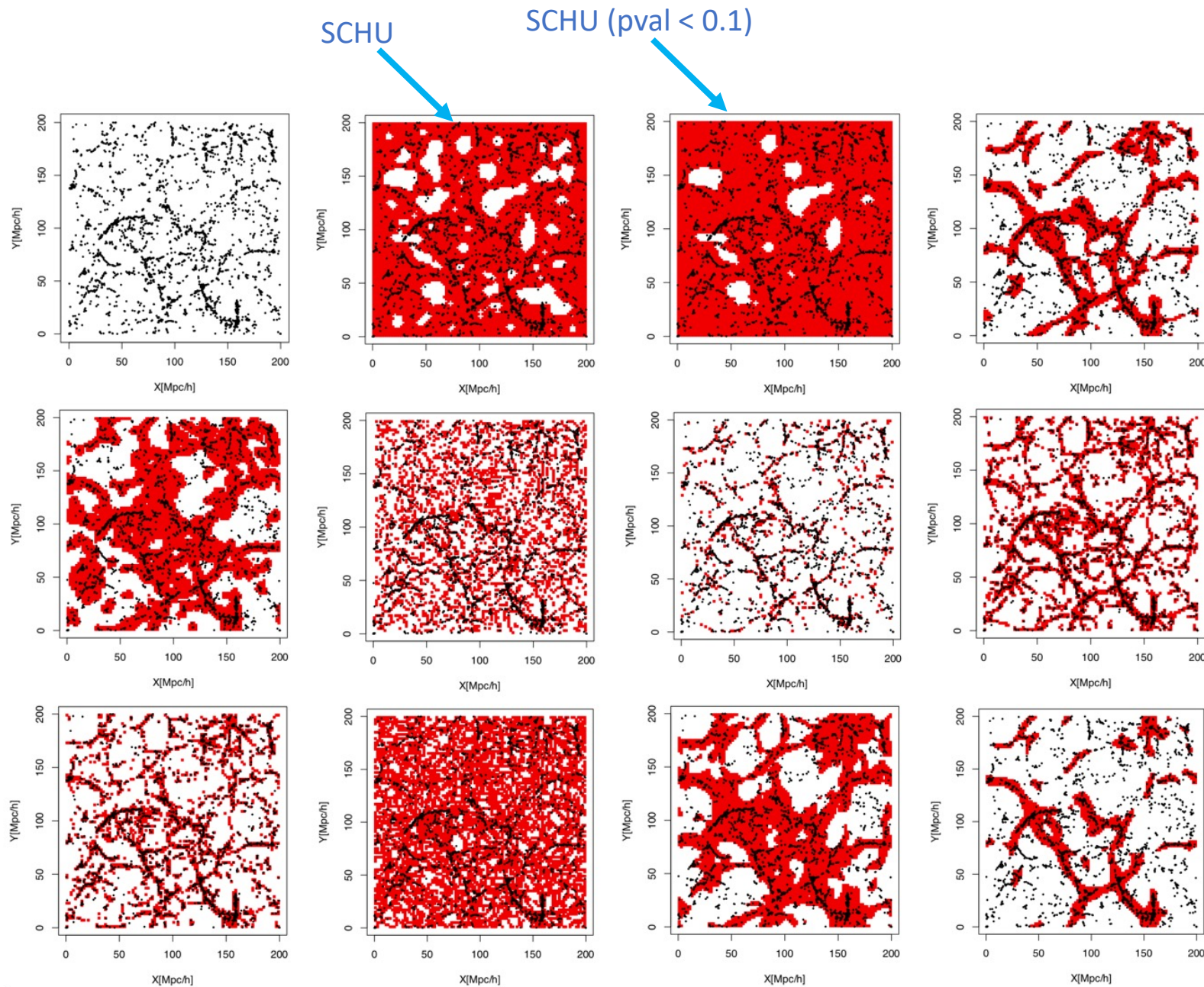


Fig. 15. Filament loops (a) and voids (b) identified in the Libeskind et al. (2018) dataset using SCHU. The DTM function was constructed with $m_0 = 0.0002$. The most significant 10 filament loops (a) and the most significant 15 cosmic voids generators (b) are shown in different colors. All of their p -values are less than 0.001. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



- Libeskind+2018 carried out comparison of methods for finding different cosmological structures
 - Used GADGET-2 dark-matter only N-body simulation code with 512^3 particles
 - The simulation cube is $200 \text{ h}^{-1} \text{ Mpc}$ per side
- Libeskind+2018 used this dataset to compare several cosmic environment classification methods
- Compare cosmic voids because filament loops are not defined/detected by other methods
- Persistence diagram (right): displays 90% confidence band for H_1 (pink) and H_2 (blue), used DTM filtration with $m_0=0.002$ (corresponds to using 58 nearest neighbors)



2h⁻¹Mpc slice of Libeskind et al. (2018)
data: white regions denote voids.

(row 1) GADGET data, SCHU, SCHU
 (p<0.1), Classic;

(row 2) DisPerSE, MMF2, MSWA,
 NEXUS;

(row 3) ORIGAMI, SpineWeb, TWeb,
 VWeb

Concluding remarks

Spatially complex data are prevalent in science (e.g., Cosmic Web, fibrin)

→ However, analyzing these data is not always straightforward

Persistent homology can extract summaries from data characterized by holes

→ Functional summaries of persistence diagrams tend to be easier to work with than the diagrams for inference (e.g. hypothesis testing) or prediction

Reference: Berry, E., Chen, Y.C., Cisewski-Kehe, J. and Fasy, B.T., 2020. Functional summaries of persistence diagrams. *Journal of Applied and Computational Topology*, 4(2), pp.211-262.

Code: https://github.com/JessiCisewskiKehe/generalized_landscapes

TDA-based hypothesis tests detect differences between the WDM and CDM COCO data of MW-analog halo neighborhoods

→ Statistically significant differences detected on different scales from spatial point process functions

Reference: Cisewski-Kehe, J., Fasy, B.T., Hellwing, W., Lovell, M.R., Drazda, P., Wu, M., 2022. Differentiating small-scale subhalo distributions in CDM and WDM models using persistent homology, *Physical Review D*, 106(2), p.023521.

Code: <https://github.com/JessiCisewskiKehe/DarkMatterTDA>

Representations of homology group generators may be used for visualization

→ Significant Cosmic Holes in Universe (SCHU) can identify cosmic voids and filaments loops and assign statistical significance

References: Xu, X., Cisewski-Kehe, J., Green, S.B. and Nagai, D., 2019. Finding cosmic voids and filament loops using topological data analysis. *Astronomy and Computing*, 27, pp.34-52.

Code: <https://github.com/xinxuyale/SCHU>

Many thanks!