Cosmic Web Reconstruction through Density Ridges

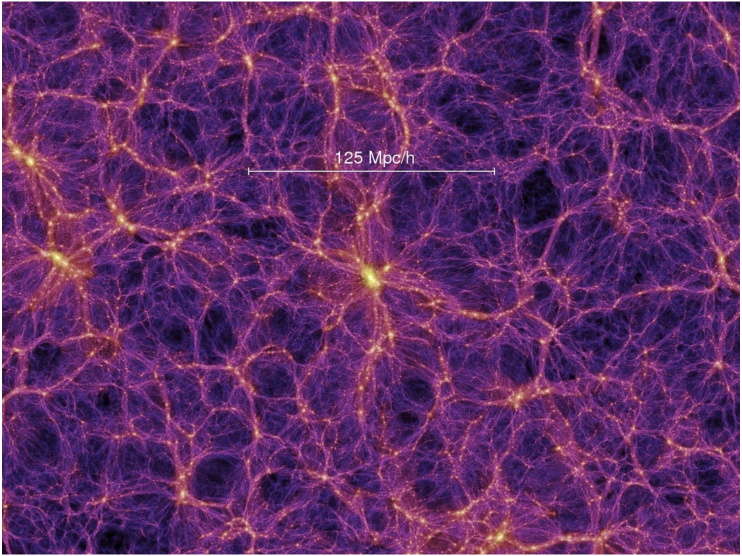
Yen-Chi Chen

Shirley Ho   Peter E. Freeman
Christopher R. Genovese   Larry Wasserman

Department of Statistics
McWilliams Center for Cosmology
Carnegie Mellon University

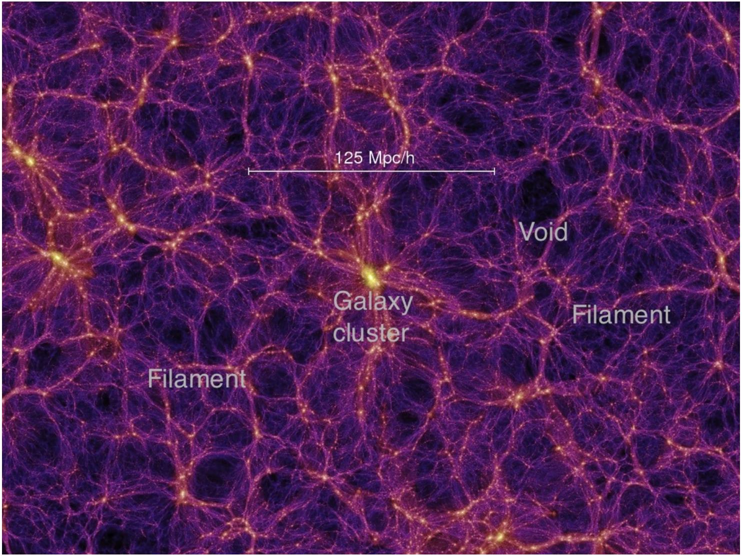
June 6, 2016

Cosmic Web: What Does Our Universe Look Like



Credit: Millennium Simulation

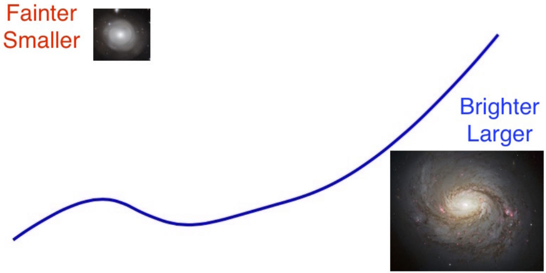
Cosmic Web: What Does Our Universe Look Like



Credit: Millennium Simulation

The Importance of Filaments

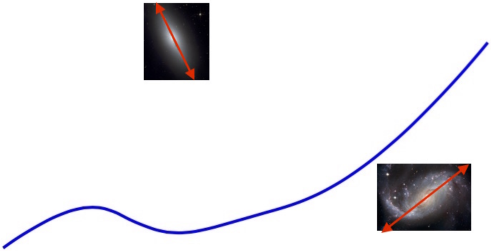
- A galaxy's brightness, size, and mass are associated with the distance to filaments.



→ **Chen** et al. 'Detecting Effects of Filaments on Galaxy Properties in Sloan Digital Sky Survey III' (2015)

The Importance of Filaments

- A galaxy's brightness, size, and mass are associated with the distance to filaments.
- A galaxy's alignment is associated with filaments.



→ **Chen** et al. 'Investigating Galaxy-Filament Alignment in Hydrodynamic Simulations using Density Ridges' (2015)

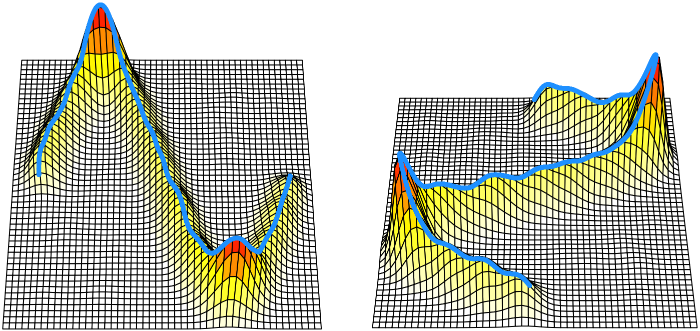
Density Ridges

We formalize the notion of filaments as *density ridges*.

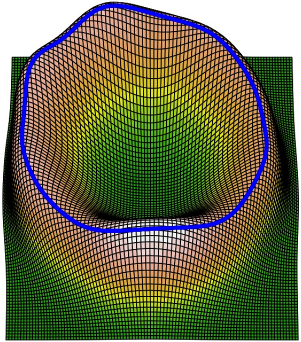
Example: Ridges in Mountains

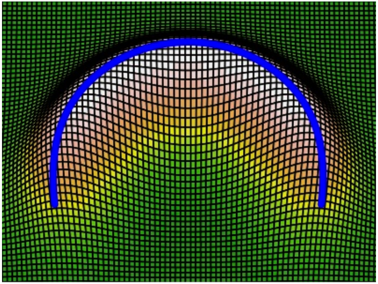
Credit: Google

Example: Ridges in Smooth Functions

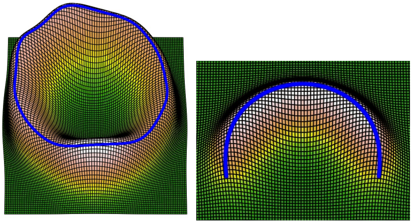


Example: Ridges in Smooth Functions

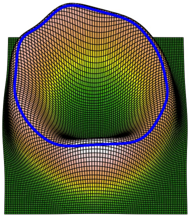
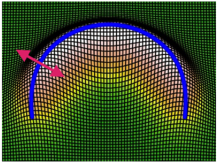




Ridges: Local Modes in Subspace

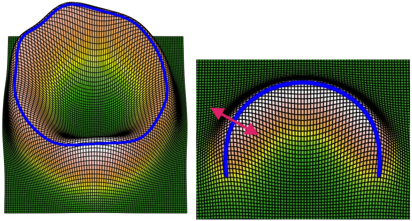


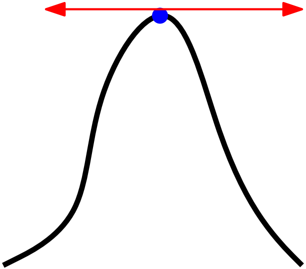
- A generalized local mode in a specific ‘subspace’.

- A generalized local mode in a specific ‘subspace’.

Ridges: Local Modes in Subspace





- A generalized local mode in a specific ‘subspace’.

Formal Definition of Density Ridges

- $p : \mathbb{R}^d \mapsto \mathbb{R}$, the density function.

Formal Definition of Density Ridges

- $p : \mathbb{R}^d \mapsto \mathbb{R}$, the density function.
- $(\lambda_j(x), v_j(x))$: jth eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.

Formal Definition of Density Ridges

- $p : \mathbb{R}^d \mapsto \mathbb{R}$, the density function.
- $(\lambda_j(x), v_j(x))$: jth eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of the 2nd eigenvector to the last eigenvector.

Formal Definition of Density Ridges

- $p : \mathbb{R}^d \mapsto \mathbb{R}$, the density function.
- $(\lambda_j(x), v_j(x))$: jth eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of the 2nd eigenvector to the last eigenvector.
- $V(x)V(x)^T$: a projection.

Formal Definition of Density Ridges

- $p : \mathbb{R}^d \mapsto \mathbb{R}$, the density function.
- $(\lambda_j(x), v_j(x))$: jth eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of the 2nd eigenvector to the last eigenvector.
- $V(x)V(x)^T$: a projection.
- Ridges:

$$R = \text{Ridge}(p) = \{x : V(x)V(x)^T \nabla p(x) = 0, \lambda_2(x) < 0\}.$$

Formal Definition of Density Ridges

- $p : \mathbb{R}^d \mapsto \mathbb{R}$, the density function.
- $(\lambda_j(x), v_j(x))$: jth eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of the 2nd eigenvector to the last eigenvector.
- $V(x)V(x)^T$: a projection.
- Ridges:

$$R = \text{Ridge}(p) = \{x : V(x)V(x)^T \nabla p(x) = 0, \lambda_2(x) < 0\}.$$

- Local modes:

$$\text{Mode}(p) = \{x : \nabla p(x) = 0, \lambda_1(x) < 0\}.$$

Dimension of Ridges

The dimension of a ridge is 1.

This is because ridges are points satisfying $V(x)V(x)^T\nabla p(x) = 0$.

$V(x)V(x)^T$ has rank $d-1$, so there are $d-1$ effective constraints.

By the Implicit Function Theorem, ridges have dimension 1.

We use the plug-in estimate:

$$\widehat{R}_n = \mathrm{Ridge}(\widehat{p}_n),$$

where $\widehat{p}_n = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)$ is the kernel density estimator (KDE).

[1] Ozertem, Umut, and Deniz Erdogmus. "Locally defined principal curves and surfaces." JMLR (2011).

Estimator and Algorithm

We use the plug-in estimate:

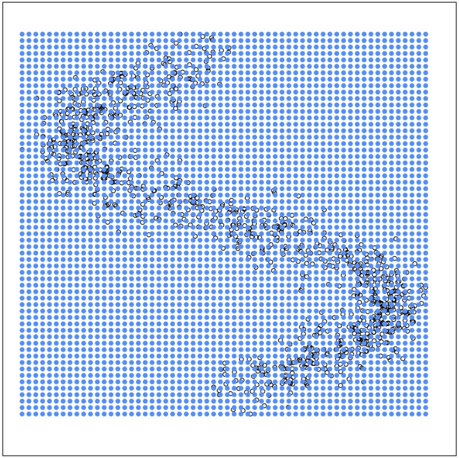
$$\widehat{R}_n = \text{Ridge}(\widehat{p}_n),$$

where $\widehat{p}_n = \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$ is the kernel density estimator (KDE).

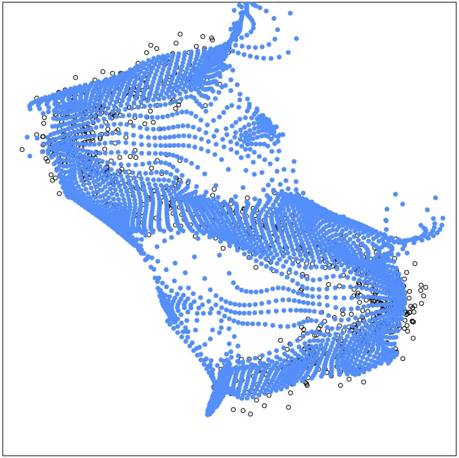
- In general, finding ridges from a given function is hard.
- The Subspace Constraint Mean Shift[1] (SCMS) algorithm allows us to find $\widehat{R}_n$, ridges of the KDE.

[1] Ozertem, Umut, and Deniz Erdogmus. "Locally defined principal curves and surfaces." JMLR (2011).

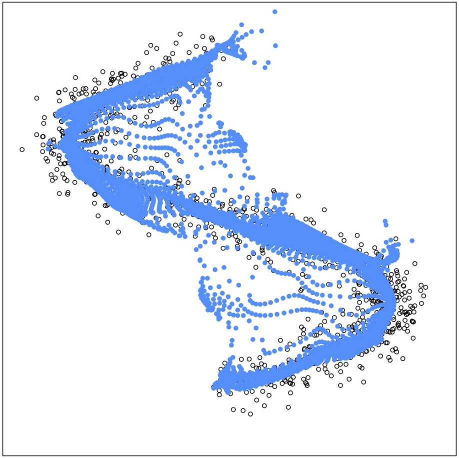
SCMS: Ridge Recovery Algorithm



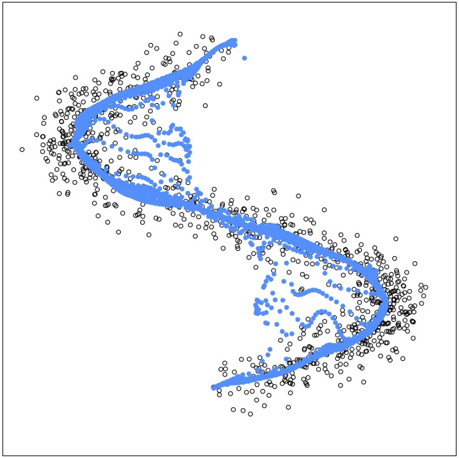
SCMS: Ridge Recovery Algorithm



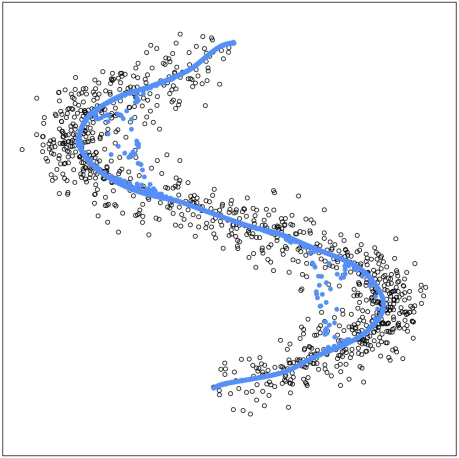
SCMS: Ridge Recovery Algorithm



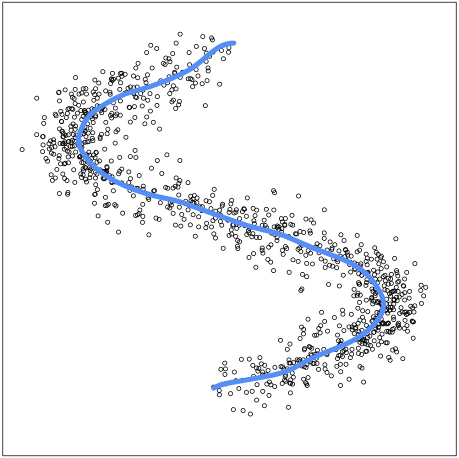
SCMS: Ridge Recovery Algorithm



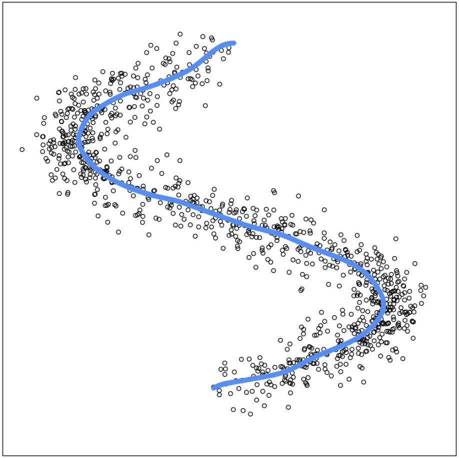
SCMS: Ridge Recovery Algorithm



SCMS: Ridge Recovery Algorithm

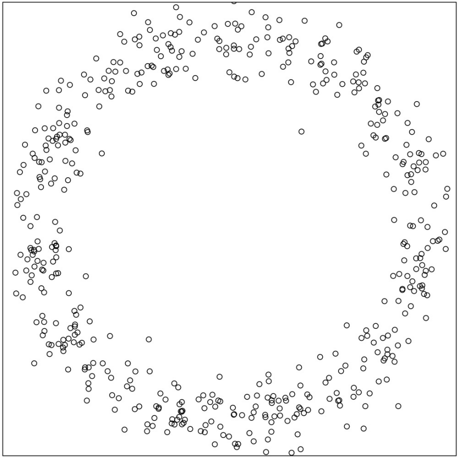


SCMS: Ridge Recovery Algorithm

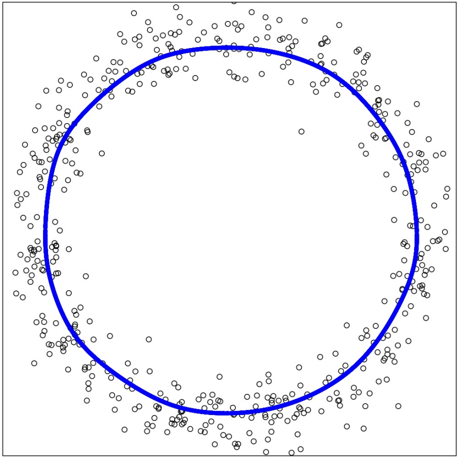


SCMS moves blue mesh points by gradient ascent and a projection.

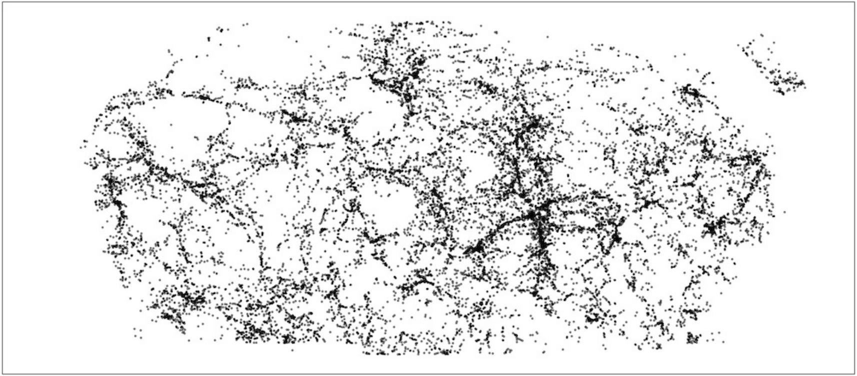
Example for Estimated Density Ridges



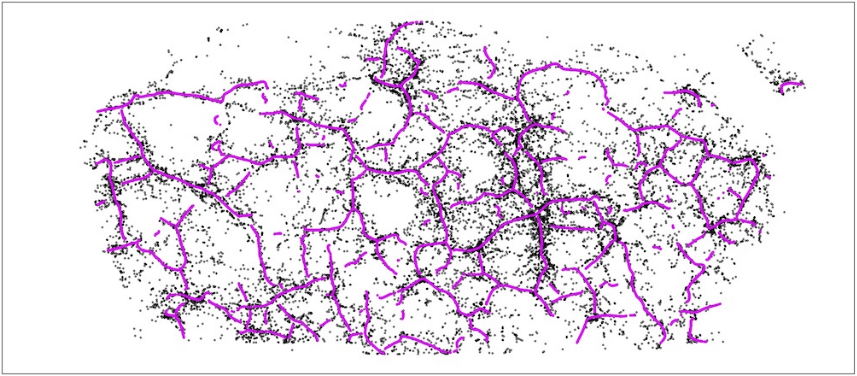
Example for Estimated Density Ridges



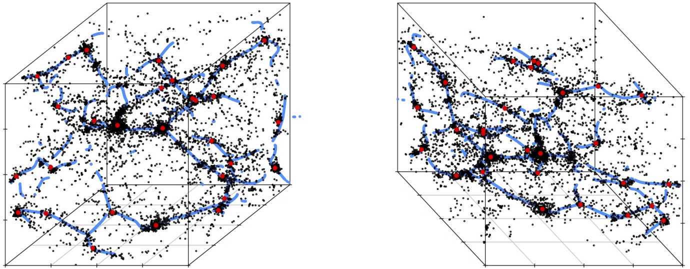
Example for Estimated Density Ridges



Example for Estimated Density Ridges



3D Example for Estimated Ridges



Blue curves: density ridges.
Red points: density local modes.

Statistical Properties

Our ridge estimator $\widehat{R}_n$ is a consistent estimator under various uncertainty measurements.

- (Pointwise error)
  $\epsilon_n(x) = d(x, \widehat{R}_n) \xrightarrow{P} 0$, for each $x \in R$.
- ($\mathcal{L}_2$ error)
  $L_{2,n} = \int_R \epsilon_n(x)^2 dx \xrightarrow{P} 0$.
- ($\mathcal{L}_\infty$ error)
  $L_{\infty,n} = \sup_{x \in R} \epsilon_n(x) \xrightarrow{P} 0$.

Statistical Properties

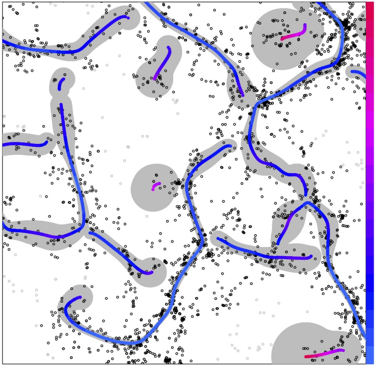
Our ridge estimator $\widehat{R}_n$ is a consistent estimator under various uncertainty measurements.

- (Pointwise error)
  $\epsilon_n(x) = d(x, \widehat{R}_n) \xrightarrow{P} 0$, for each $x \in R$.
- ($\mathcal{L}_2$ error)
  $L_{2,n} = \int_R \epsilon_n(x)^2 dx \xrightarrow{P} 0$.
- ($\mathcal{L}_\infty$ error)
  $L_{\infty,n} = \sup_{x \in R} \epsilon_n(x) \xrightarrow{P} 0$.

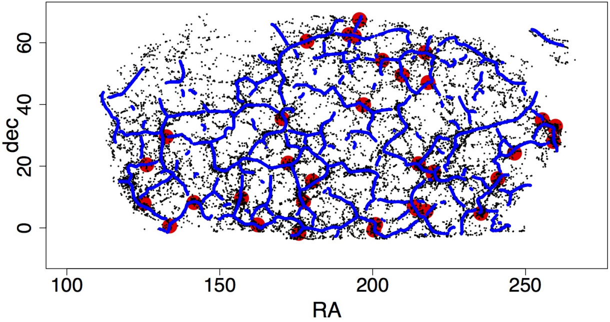
Moreover, we may use the bootstrap to estimate the uncertainty (error) of ridges.

Basic idea: comparing original ridges and bootstrap ridges.

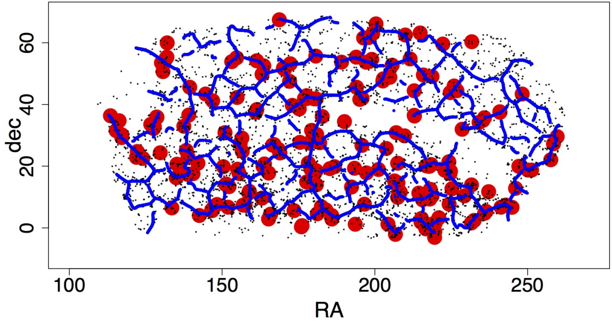
Uncertainty of Ridges



- Blue: filaments. Red: galaxy clusters (redMaPPer).

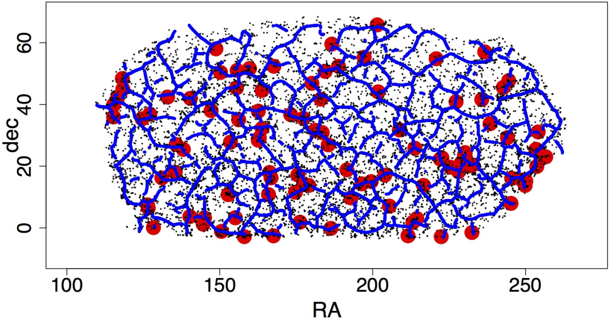


- Blue: filaments. Red: galaxy clusters (redMaPPer).



SDSS: Comparing to Clusters

- Blue: filaments. Red: galaxy clusters (redMaPPer).

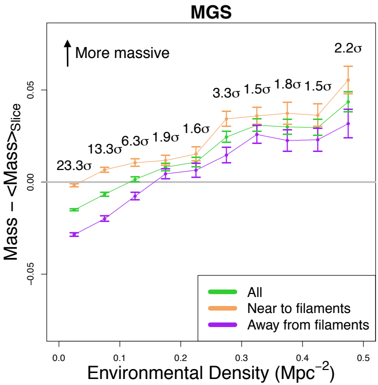


SDSS: Filament Effects VS Environments

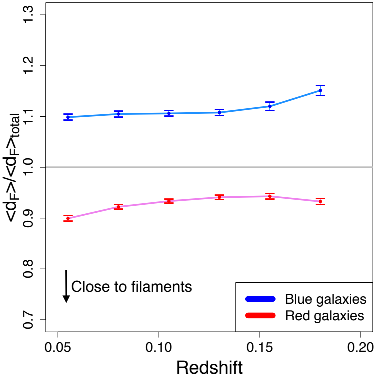
Do filaments have an extra effect other than environments?

SDSS: Filament Effects VS Environments

Do filaments have an extra effect other than environments?
$\longrightarrow$Yes!

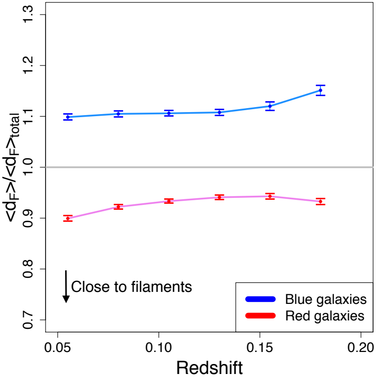


SDSS: Color



$<d_F>/<d_F>_{total}$
1.3
1.2
1.1
1.0
0.9
0.8
0.7
Close to filaments
Blue galaxies
Red galaxies
0.05
0.10
0.15
0.20
Redshift

SDSS: Color



Similar pattern also appears for other galaxy properties such as brightness, size, and age.

A Filament Catalogue for the SDSS

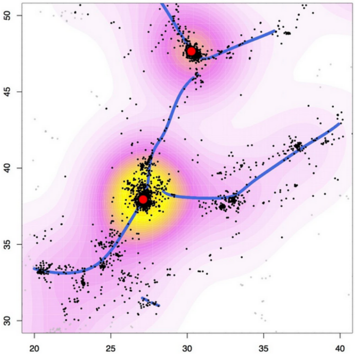
Cosmic Web Reconstruction

Home | Introduction | Density Ridges | Catalogue | People | Contact

Search this site

A filament is a one-dimensional, smooth, connected structure embedded in a multi-dimensional space that characterizes the high density regions. Matter in the Universe tend to aggregate around these filaments which weave our Universe into an intricate network structure. We use ridges of density function (density ridges) as tracers for the filaments and apply our method to the SDSS DR12 to construct a filament catalogue.

This is a joint project of department of Statistics, department of Physics, and the McWilliams center for Cosmology at Carnegie Mellon University. This is also a project of the astrostatistics group at Carnegie Mellon University.



Data: the MBII simulation, we thank Ananth Tenneti for proving this data.

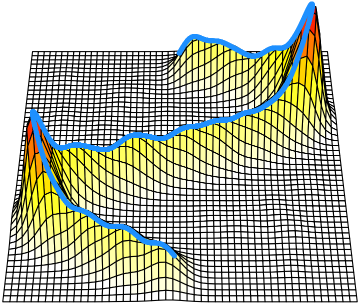
References:
Cosmology--

1. Chen, Yen-Chi, Shirley Ho, Peter E. Freeman, Christopher R. Genovese, and Larry Wasserman. ``*Cosmic Web Reconstruction through Density Ridges: Method and Algorithm*." arXiv: 1501.05303
2. Chen, Yen-Chi, Shirley Ho, Peter E. Freeman, Christopher R. Genovese, and Larry Wasserman. ``*Cosmic Web Reconstruction through Density Ridges: Catalogue*." (In progress)
3. Chen, Yen-Chi, Shirley Ho, Ananth Tenneti, Rachel Mandelbaum, Rupert Croft, Tiziana DiMatteo, Christopher R. Genovese, and Larry Wasserman. ``*Investigating Galaxy-Filament Alignment in Hydrodynamic Simulations using Density Ridges*." arXiv: 1508.04149
4. Chen, Yen-Chi, Shirley Ho, Rachel Mandelbaum, Peter E. Freeman, Christopher R. Genovese, and Larry Wasserman. ``*Detecting Effects of Filaments on Galaxy Properties in Sloan Digital Sky Survey III*." (In progress)

https://sites.google.com/site/yenchicr/

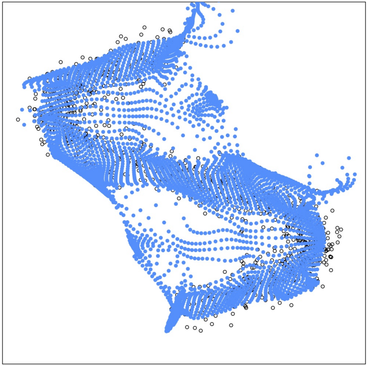
Summary

1. Model: density ridges.



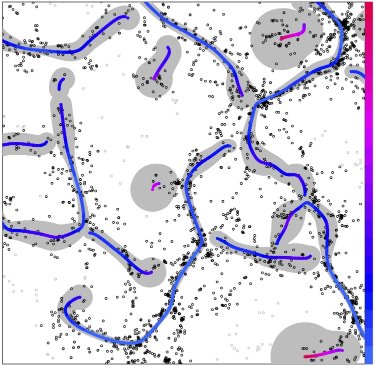
Summary

1. Model: density ridges.
2. Algorithm: SCMS.



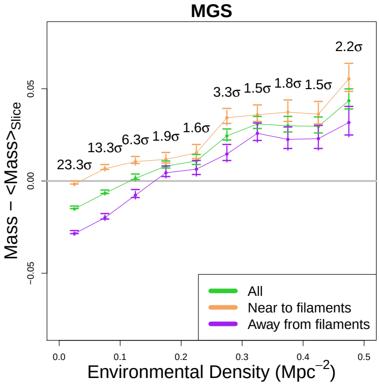
Summary

1. Model: density ridges.
2. Algorithm: SCMS.
3. Uncertainty measures.



Summary

1. Model: density ridges.
2. Algorithm: SCMS.
3. Uncertainty measures.
4. Applications in various problems.

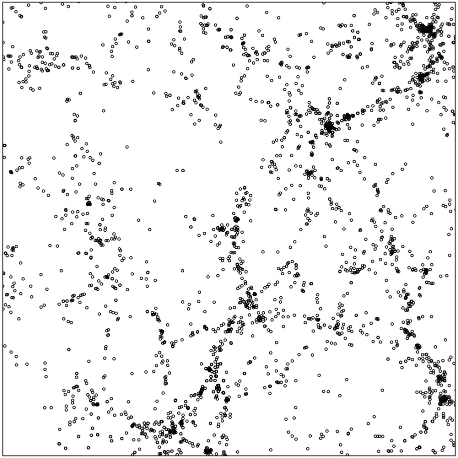


Thank you!

References

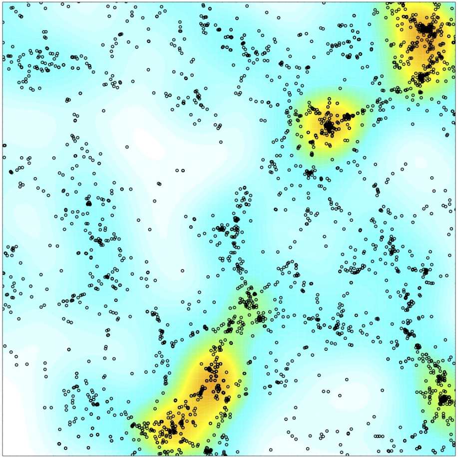
1. Chen, Yen-Chi, Shirley Ho, Peter E. Freeman, Christopher R. Genovese, and Larry Wasserman. "Cosmic Web Reconstruction through Density Ridges: Method and Algorithm." To appear in Monthly Notices of the Royal Astronomical Society..
2. Chen, Yen-Chi, et al. "Investigating Galaxy-Filament Alignments in Hydrodynamic Simulations using Density Ridges." arXiv preprint arXiv:1508.04149 (2015).
3. Chen, Yen-Chi, Christopher R. Genovese, and Larry Wasserman. "Asymptotic theory for density ridges." The Annals of Statistics 43.5 (2015): 1896-1928.
4. Conroy, Charlie, James E. Gunn, and Martin White. "The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies." The Astrophysical Journal 699.1 (2009): 486.
5. Eberly, David. Ridges in image and data analysis. Vol. 7. Springer Science & Business Media, 1996.
6. Genovese, Christopher R., et al. "Nonparametric ridge estimation." The Annals of Statistics 42.4 (2014): 1511-1545.
7. Ozertem, Umut, and Deniz Erdogmus. "Locally defined principal curves and surfaces." The Journal of Machine Learning Research 12 (2011): 1249-1286.

1. Rawdata

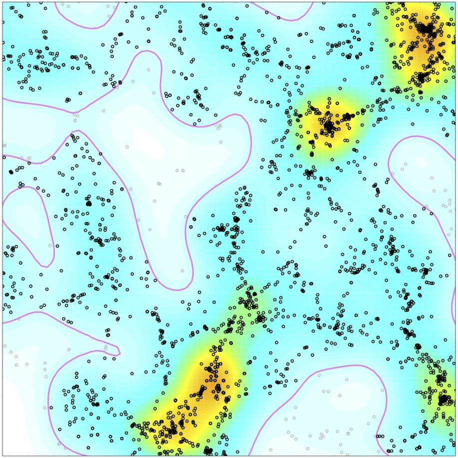


Algorithm

1. Rawdata
2. Density Reconstruction

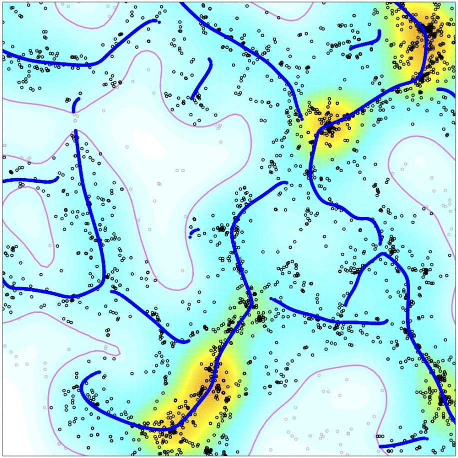


1. Rawdata
2. Density Reconstruction
3. Thresholding

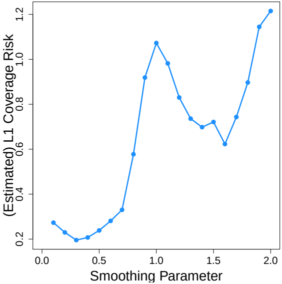


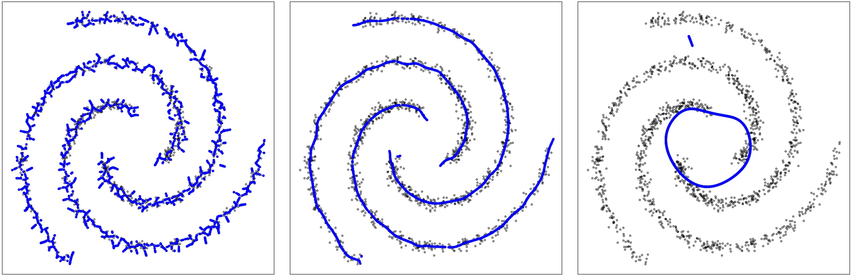
Algorithm

1. Rawdata
2. Density Reconstruction
3. Thresholding
4. Ridge Recovery

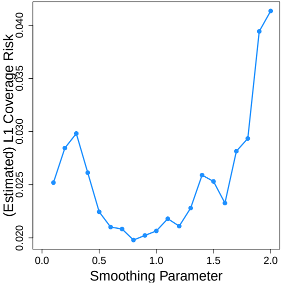


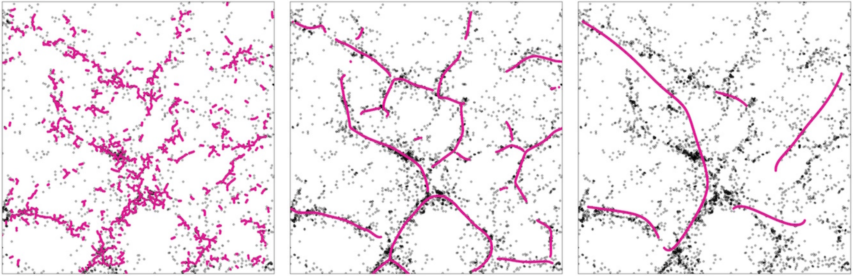
Bandwidth Selection

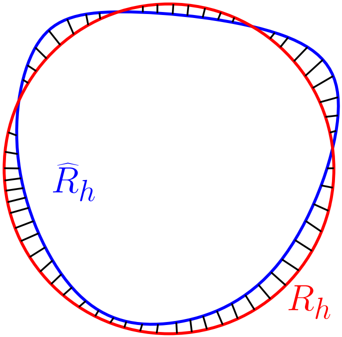




Bandwidth Selection







L_1 distance are like the area of the shady regions.

We estimate this distance by data splitting or the bootstrap.

Reference: **Chen** et al. 'Optimal Ridge Detection using Coverage Risk' (NIPS 2015).

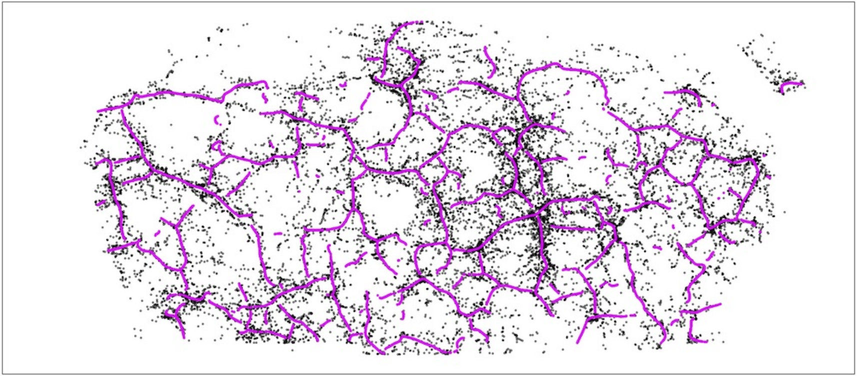
General Ridges

We can generalize ridges to higher dimensions. Pick
$V_r(x) = [v_{r+1}(x), \cdots, v_d(x)]$.
We define

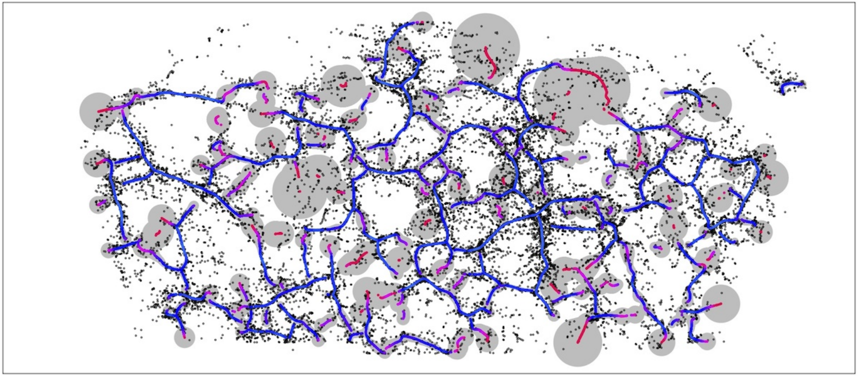
$$r\text{-Ridge}(p) = \{x : V_r(x)V_r(x)^T \nabla p(x) = 0, \lambda_{r+1}(x) < 0\}.$$

$V_r(x)$ is a $d \times (d - r)$ matrix. There are $d - r$ constraints.
By Implicit Function Theorem, r-ridges are r-manifolds.
In Astronomy, $r = 2$ can be used to model 'Cosmic Sheets (Walls)'.
$r = 0$ coincides with the definition of local modes.

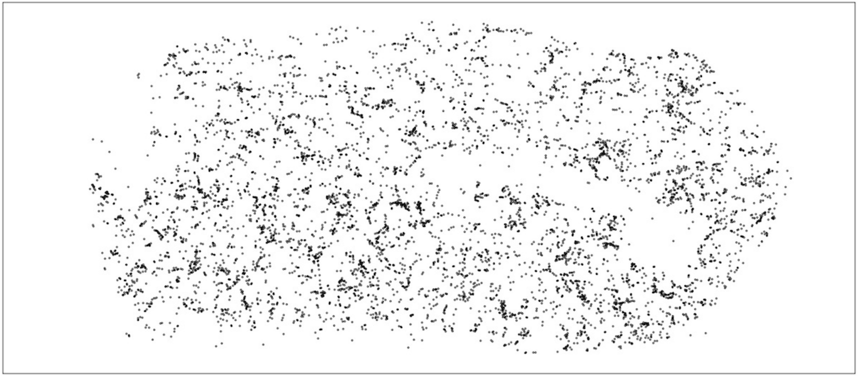
Density Ridges on the SDSS data



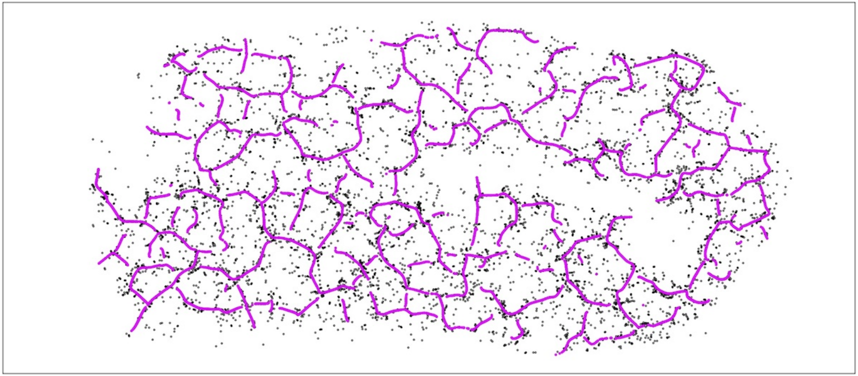
Density Ridges on the SDSS data



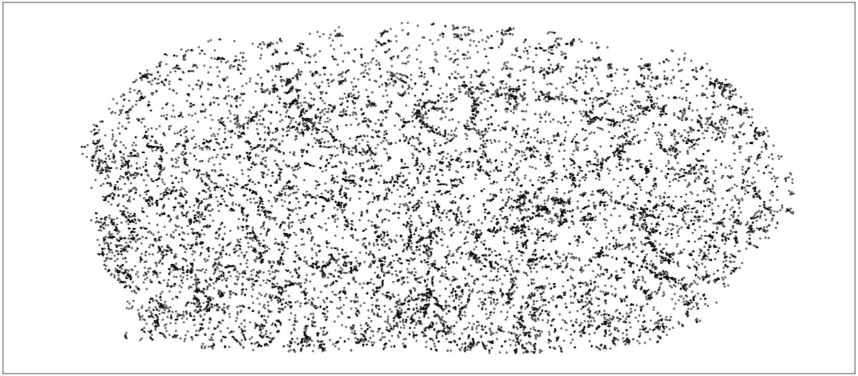
Curse of Number Density

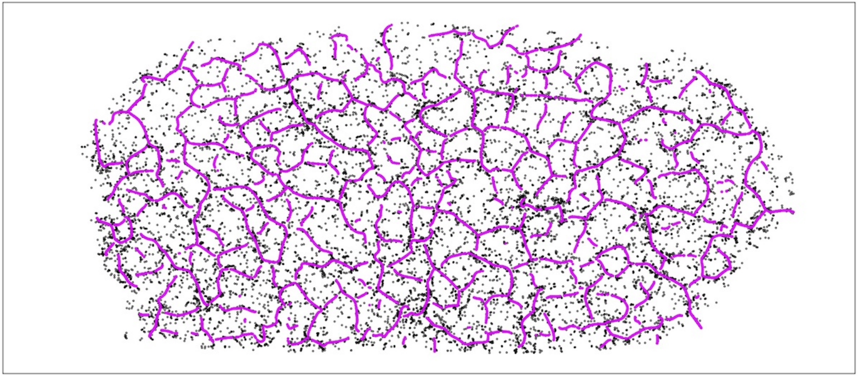


Curse of Number Density



Curse of Number Density

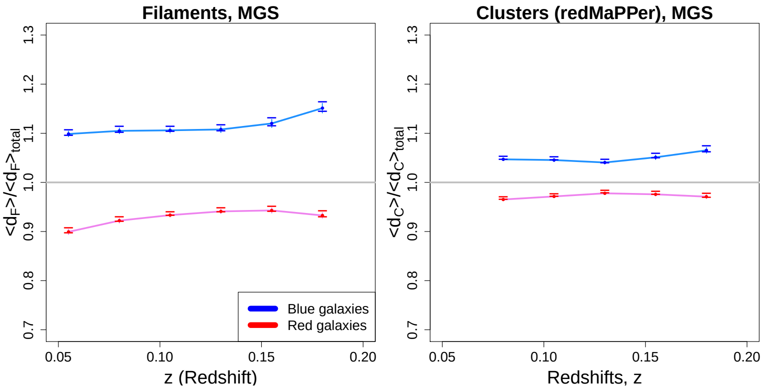




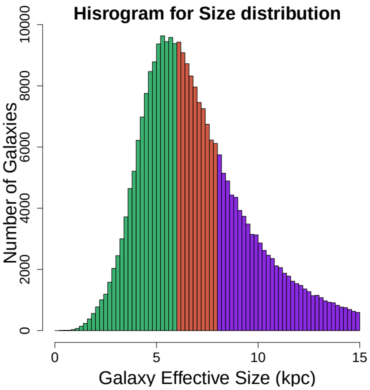
SDSS: Red and Blue Galaxies

- Redshift range: $0.05 < z < 0.20$ (main sample galaxy).
- Color cut: $(g - r) = 0.8$.

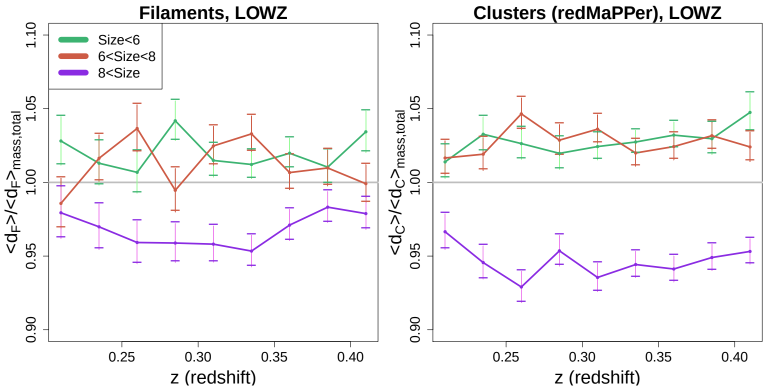
SDSS: Red and Blue Galaxies



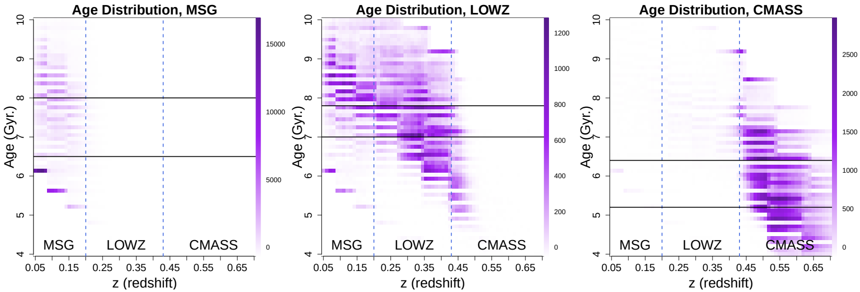
1. Size: Effective Radii.
2. Data: LOWZ ($0.20 < z < 0.43$)
3. Partitioning galaxies into three groups according to their size.



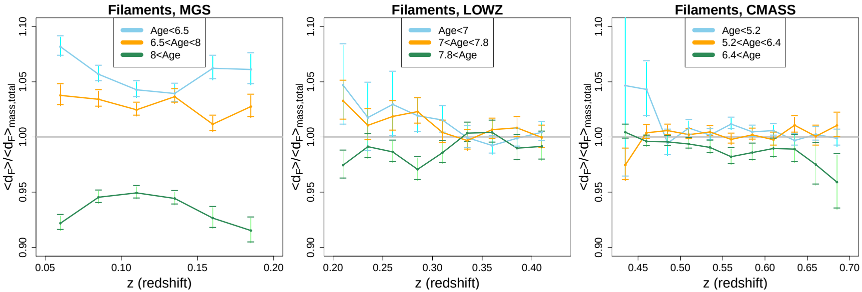
SDSS: Size for Galaxies

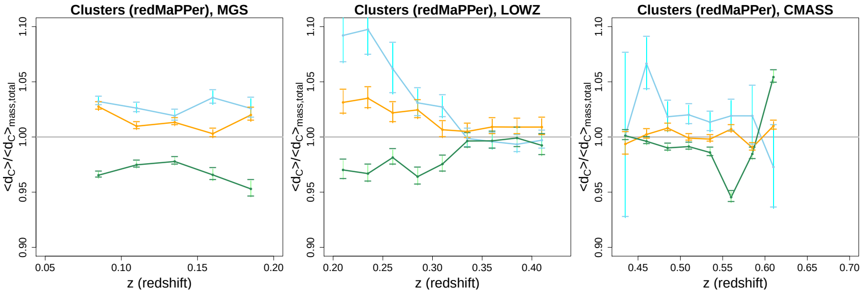


Age for Galaxies

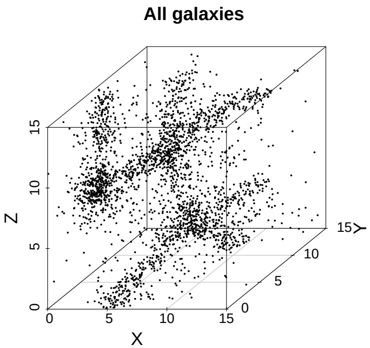


Age for Galaxies



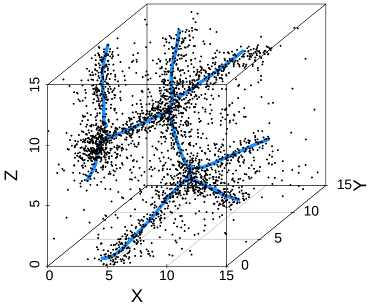

Clusters (redMaPPer), MGS
Clusters (redMaPPer), LOWZ
Clusters (redMaPPer), CMASS
<dC>/<dC>mass,total
z (redshift)
0.90
0.95
1.00
1.05
1.10
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70

Comparison: Voronoi Model

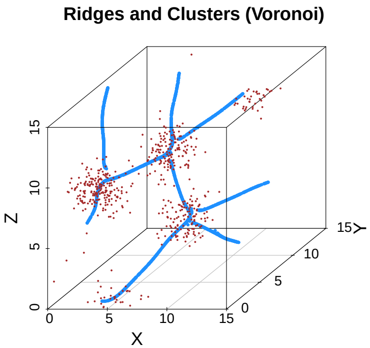


Comparison: Voronoi Model

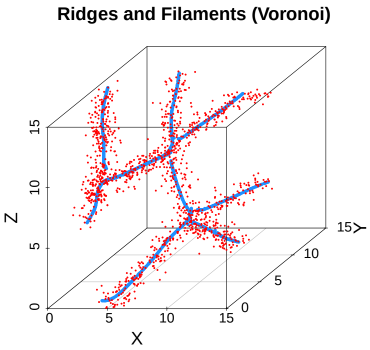
Ridges and all galaxies



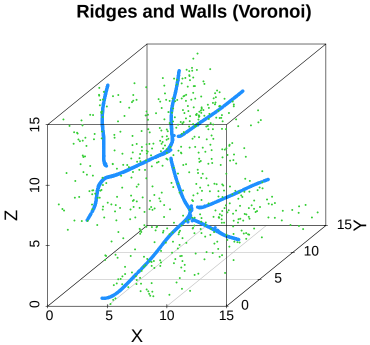
Comparison: Voronoi Model



Comparison: Voronoi Model



Comparison: Voronoi Model



Comparison: Voronoi Model