

Cosmic Web Detection With Directional Density Ridges

Yikun Zhang

*Department of Statistics, University of Chicago
NSF-Simons AI Institute for the Sky (SkAI Institute)*

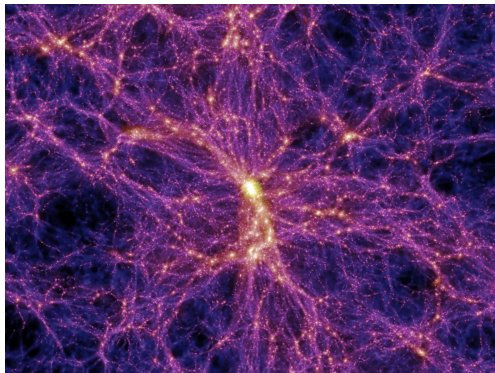
CIERA–Northwestern
October 15, 2026

Background and Introduction



What Is Cosmic Web?

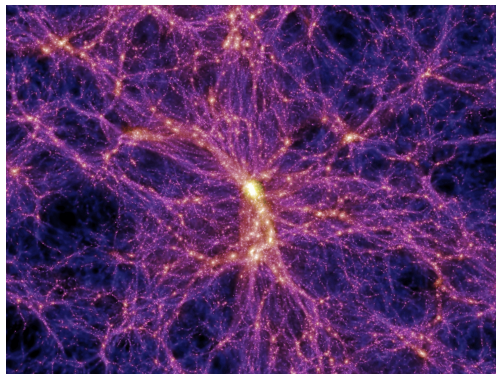
Cosmic Web: Large-scale network structure revealing the non-uniformity of matter distribution in our Universe ([Bond et al., 1996](#)).



Cited from the millennium simulation project ([Springel et al., 2005](#)).

What Is Cosmic Web?

Cosmic Web: Large-scale network structure revealing the non-uniformity of matter distribution in our Universe ([Bond et al., 1996](#)).

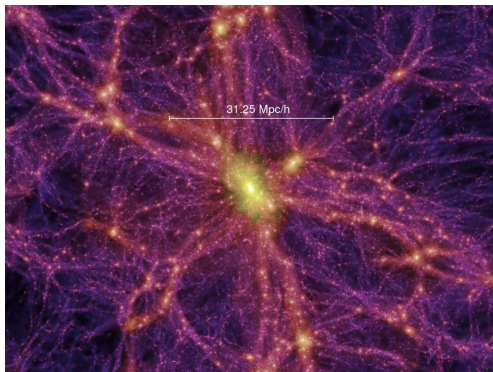


Cited from the millennium simulation project ([Springel et al., 2005](#)).

- **Cause:** Gravitational instability $\Rightarrow$ Anisotropic matter collapse ([Zel'Dovich, 1970](#)).
- **Large-Scale:** $1 \text{ Mpc} \approx 3.26 \times 10^6 \text{ light-years}$.

What Is Cosmic Web?

Cosmic Web: Large-scale network structure revealing the non-uniformity of matter distribution in our Universe ([Bond et al., 1996](#)).



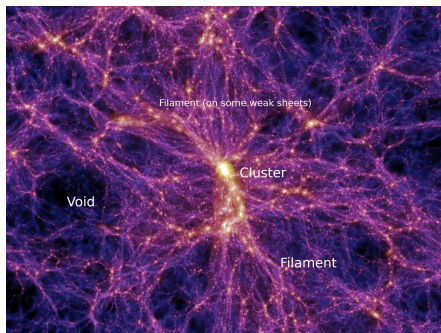
Cited from the millennium simulation project ([Springel et al., 2005](#)).

- **Cause:** Gravitational instability $\Rightarrow$ Anisotropic matter collapse ([Zel'Dovich, 1970](#)).
- **Large-Scale:** $1 \text{ Mpc} \approx 3.26 \times 10^6 \text{ light-years}$.

Key Characteristics of the Cosmic Web

The *cosmic web* consists of four distinct components ([Libeskind et al., 2018](#)):

- Massive *galaxy clusters* (a.k.a., *cosmic nodes*),
 - 1D interconnected *filaments*,
 - 2D tenuous *sheets/walls*,
 - Vast and near-empty *voids*.
- } on which matter concentrates.

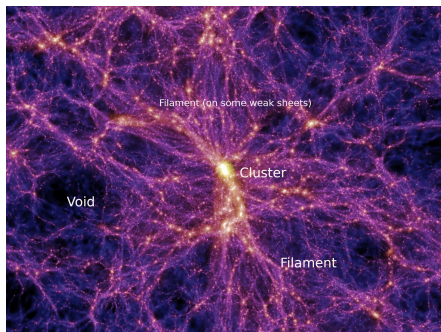


Cited from the millennium simulation project ([Springel et al., 2005](#)).

Key Characteristics of the Cosmic Web

The *cosmic web* consists of four distinct components ([Libeskind et al., 2018](#)):

- Massive **galaxy clusters** (a.k.a., **cosmic nodes**),
 - 1D interconnected **filaments**,
 - 2D tenuous *sheets/walls*,
 - Vast and near-empty *voids*.
- } on which matter concentrates.



Cited from the millennium simulation project ([Springel et al., 2005](#)).

Significance of Cosmic Nodes and Filaments

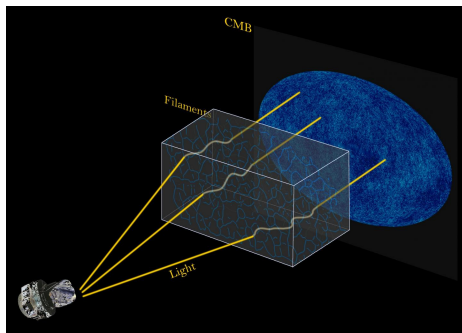
- Trace the regions where matter is mostly concentrated.

Significance of Cosmic Nodes and Filaments

- Trace the regions where matter is mostly concentrated.
- Influence the stellar properties (*e.g.*, mass and size) of nearby galaxies and reveal the nature of dark matter ([Tempel et al., 2014](#); [Malavasi et al., 2022](#)).

Significance of Cosmic Nodes and Filaments

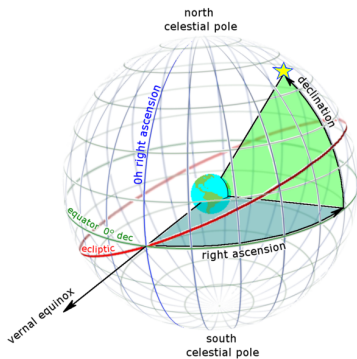
- Trace the regions where matter is mostly concentrated.
- Influence the stellar properties (*e.g.*, mass and size) of nearby galaxies and reveal the nature of dark matter ([Tempel et al., 2014](#); [Malavasi et al., 2022](#)).
- Weak Lensing Effects:** Distort the trajectory of the cosmic microwave background (CMB) light.



An illustration of the bending trajectory of the CMB light ([He et al., 2018](#)).

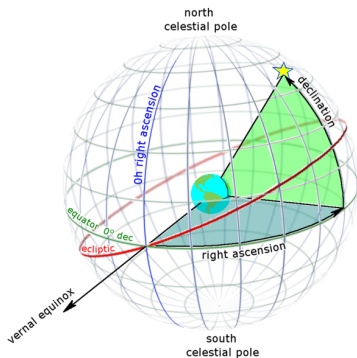
Observational Data For Node and Filament Detections

Data: Galaxy observations $\{(\alpha_1, \delta_1, z_1), \dots, (\alpha_n, \delta_n, z_n)\}$ in astronomical surveys.



Observational Data For Node and Filament Detections

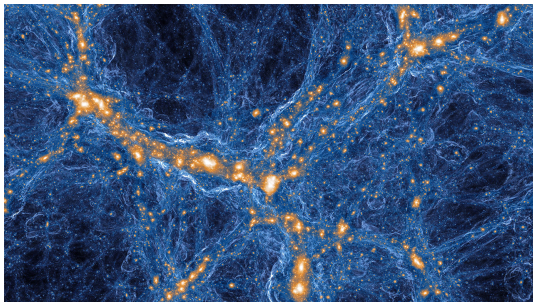
Data: Galaxy observations $\{(\alpha_1, \delta_1, z_1), \dots, (\alpha_n, \delta_n, z_n)\}$ in astronomical surveys.



- $\alpha_i \in [0, 360^\circ]$: *Right Ascension* (RAsc), *i.e.*, celestial longitude;
- $\delta_i \in [-90^\circ, 90^\circ]$: *Declination* (DEC), *i.e.*, celestial latitude;
 - (α_i, δ_i) encodes a location on $\Omega_2 = \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x}\|_2 = 1\}$.
- $z_i \in (0, \infty)$: *Redshift*, *i.e.*, a proxy for its distance to the Earth.

Challenges in Detecting Cosmic Nodes and Filaments

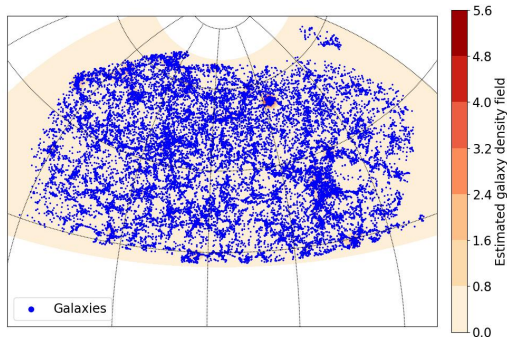
- 1 The cosmic web structures are overwhelmingly complex ([Cautun et al., 2013](#)):
 - Structurally asymmetric, highly uncertain in connectivity, and multi-scale.



The “present-day” cosmic web by the IllustrisTNG simulation ([Vogelsberger et al., 2014](#)).

Challenges in Detecting Cosmic Nodes and Filaments

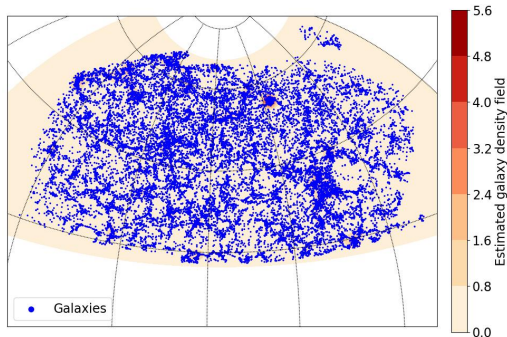
- 1 The cosmic web structures are overwhelmingly complex (Cautun et al., 2013):
 - Structurally asymmetric, highly uncertain in connectivity, and multi-scale.
- 2 Observed data are *discrete* and *noisy* samples of the matter/galaxy density field.



Observed galaxies by the Sloan Digital Sky Survey (SDSS-IV) in a thin redshift slice.

Challenges in Detecting Cosmic Nodes and Filaments

- 1 The cosmic web structures are overwhelmingly complex (Cautun et al., 2013):
 - Structurally asymmetric, highly uncertain in connectivity, and multi-scale.
- 2 Observed data are *discrete* and *noisy* samples of the matter/galaxy density field.



Observed galaxies by the Sloan Digital Sky Survey (SDSS-IV) in a thin redshift slice.

- 3 $\{(\alpha_i, \delta_i)\}_{i=1}^n$ lie not on a flat Euclidean plane but on the **celestial sphere** Ω_2 !

Overview of Today's Talk

Theme: Reconstruct the Cosmic Web Under Spherical Geometry.

Theme: Reconstruct the Cosmic Web Under Spherical Geometry.

- 1 **Model:** Directional density ridge $\mathcal{R}_d$ on $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : \|\mathbf{x}\|_2 = 1\}$.

Theme: Reconstruct the Cosmic Web Under Spherical Geometry.

- 1 **Model:** Directional density ridge $\mathcal{R}_d$ on $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : \|\mathbf{x}\|_2 = 1\}$.
- 2 **Estimation:** Estimate $\mathcal{R}_d$ through directional kernel density estimator (KDE).

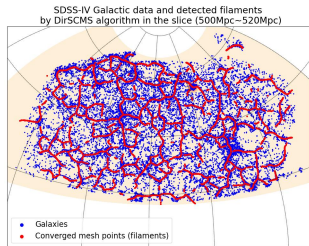
Theme: Reconstruct the Cosmic Web Under Spherical Geometry.

- 1 **Model:** Directional density ridge $\mathcal{R}_d$ on $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : \|\mathbf{x}\|_2 = 1\}$.
- 2 **Estimation:** Estimate $\mathcal{R}_d$ through directional kernel density estimator (KDE).
- 3 **Computation:** Directional subspace constrained mean shift (DirSCMS) algorithm.

Overview of Today's Talk

Theme: Reconstruct the Cosmic Web Under Spherical Geometry.

- 1 **Model:** Directional density ridge $\mathcal{R}_d$ on $\Omega_q = \{x \in \mathbb{R}^{q+1} : \|x\|_2 = 1\}$.
- 2 **Estimation:** Estimate $\mathcal{R}_d$ through directional kernel density estimator (KDE).
- 3 **Computation:** Directional subspace constrained mean shift (DirSCMS) algorithm.
- 4 **Data Product:** A new cosmic web catalog based on SDSS-IV galaxies.



Detected cosmic filaments based on SDSS-IV galaxies in a redshift slice.

Published June 10, 2022 | Version 0.0.1

[Dataset](#) [Open](#)

SDSS-IV Cosmic Web Catalog

Zhang, Yikun¹

[Show affiliations](#)

Contributors

Supervisors: Chen, Yen-Chi¹ [ORCID](#), de Souza, Rafael S.² [ORCID](#)

[Show affiliations](#)

This repository contains the cosmic web catalog data released in the paper "Cosmic Web Catalog on SDSS-IV Data with SCONE" (preparing).

The catalog is constructed on the SDSS-IV galaxies and quasars (QSO) using our proposed Directional Subspace Constrained Mean Shift (DirSCMS) algorithm. We release both the cosmic filaments and local modes (i.e., local maxima of the estimated galaxy/QSO density field, which serves as candidates of galaxy clusters) within 325 thin redshift slices, each of which spans 20Mpc under the Planck15 cosmology. The entire catalog covers the redshift range from $z = 0$ to $z = 3$.

Final catalog data (Zhang, 2022):

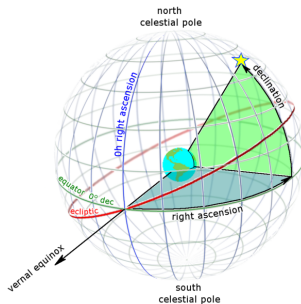
<https://doi.org/10.5281/zenodo.6244866>.

Previous Works on Cosmic Filament Detection



2D vs. 3D Methods for Cosmic Web Detection

The observed galaxies are recorded as $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}$.



The observed galaxies are recorded as $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}$.

- **3D Method:** Convert $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n$ to Cartesian coordinates as

$$\tilde{X}_i = d(z_i) \cos \alpha_i \cos \delta_i, \quad \tilde{Y}_i = d(z_i) \sin \alpha_i \cos \delta_i, \quad \tilde{Z}_i = d(z_i) \sin \delta_i,$$

where $d(\cdot)$ is a distance transformation function ([Tempel et al., 2014](#)).

2D vs. 3D Methods for Cosmic Web Detection

The observed galaxies are recorded as $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}$.

- **3D Method:** Convert $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n$ to Cartesian coordinates as

$$\tilde{X}_i = d(z_i) \cos \alpha_i \cos \delta_i, \quad \tilde{Y}_i = d(z_i) \sin \alpha_i \cos \delta_i, \quad \tilde{Z}_i = d(z_i) \sin \delta_i,$$

where $d(\cdot)$ is a distance transformation function ([Tempel et al., 2014](#)).

- **2D Method:** Slice the Universe into thin redshift slices ([Chen et al., 2015b](#); [Duque et al., 2022](#)).

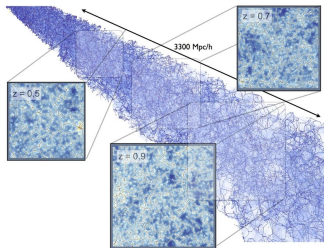
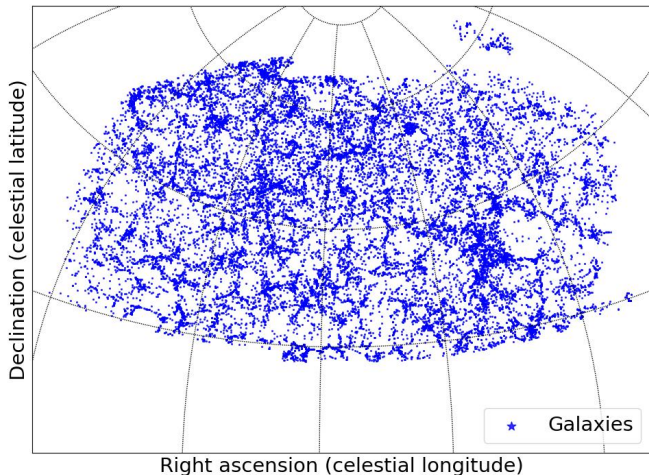


Illustration of slicing the Universe ([Laigle et al., 2018](#)).

Caveats of Slicing the Universe

Each redshift slice (e.g., $\Delta z = 0.005$) is *not* a flat 2D plane, but a **spherical shell** with *nonlinear* curvature!



Galaxies in each slice $\{(\alpha_i, \delta_i)\}_{i=1}^n \subset [0, 360^\circ) \times [-90^\circ, 90^\circ]$ are on Ω_2 .

Caveats of Slicing the Universe

Each redshift slice (*e.g.*, $\Delta z = 0.005$) is *not* a flat 2D plane, but a **spherical shell** with *nonlinear* curvature!

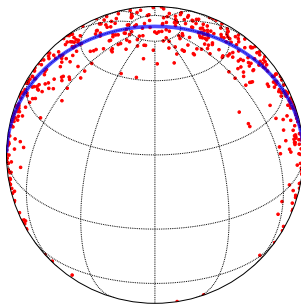
Question: Why can't we ignore the spherical geometry?

Caveats of Slicing the Universe

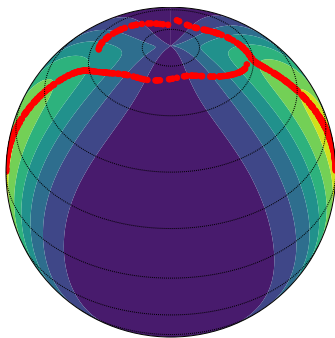
Each redshift slice (e.g., $\Delta z = 0.005$) is *not* a flat 2D plane, but a **spherical shell** with *nonlinear* curvature!

Question: Why can't we ignore the spherical geometry?

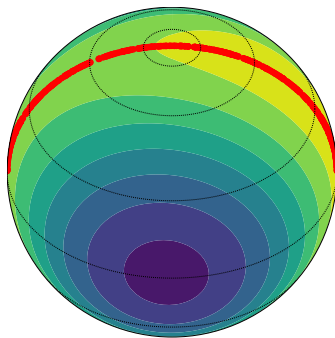
- **Blue curve:** True filament structure across the north and south poles of Ω_2 .
- **Red points:** Noisy data around the true filament.
- **Task:** Recover the true filament (**blue curve**) based on the noisy data (**red points**).



Why can't we ignore the spherical geometry?



(a) Euclidean SCMS method (**ignoring** the spherical geometry).

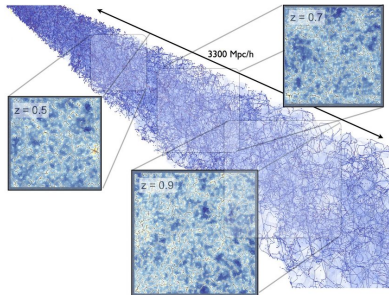


(b) Our proposed DirSCMS method (that **accounts for** the spherical geometry).

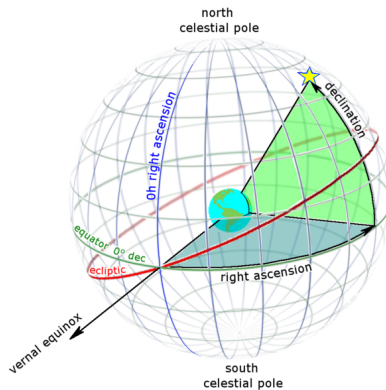
The background contour plots are kernel density estimators on the flat plane $[-90^\circ, 90^\circ] \times [0^\circ, 360^\circ]$ and the unit sphere $\Omega_2 = \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$, respectively.

- SCMS: Subspace Constrained Mean Shift ([Ozertem and Erdogmus, 2011](#)).

One Central Research Question



(a) Slicing the Universe.



(b) Positioning of the observed galaxy.

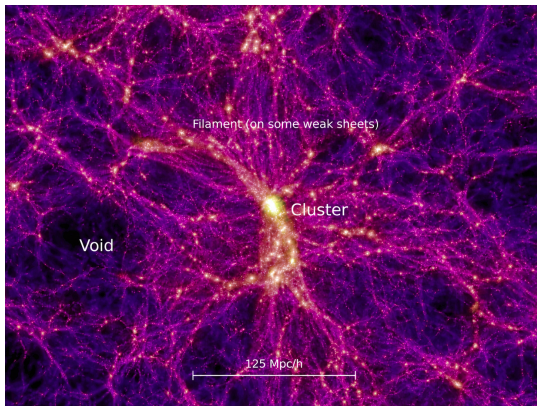
► **Research Question:** How do we model and estimate the cosmic filaments based on the observed galaxies in each (redshift) **spherical** slice?

Cosmic Web Model: Directional Density Ridges



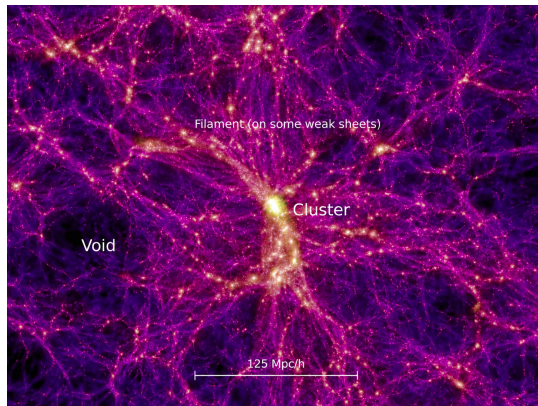
Cosmic Node and Filament Models

Recap: Cosmic nodes and filaments are **0D and 1D high-density** structures.



Cosmic Node and Filament Models

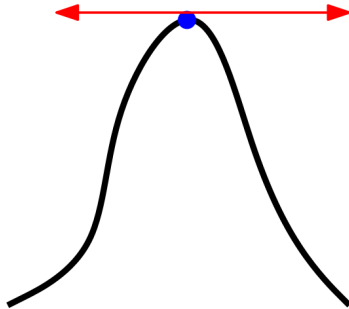
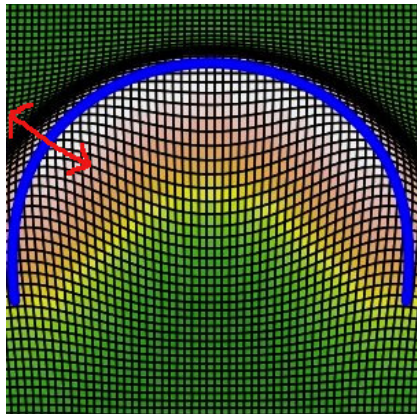
Recap: Cosmic nodes and filaments are *0D and 1D high-density* structures.



Proposal: Model these geometric features through the *directional density ridge* of a density function f on $\Omega_q = \{ \mathbf{x} \in \mathbb{R}^{q+1} : ||\mathbf{x}||_2 = 1 \}$.

Directional Density Ridge

The *directional density ridge* is a set of **generalized** local maxima of $f : \Omega_q \rightarrow \mathbb{R}_{\geq 0}$ within **certain subspaces** on Ω_q .



An illustration of the (directional) density ridge ([Chen et al., 2015a](#)).

Local Maxima of f on Ω_q :

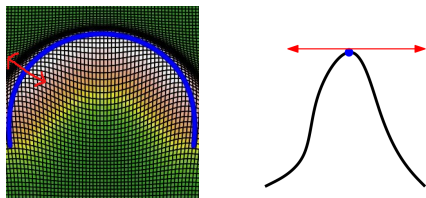
$$\mathcal{M} := \{x \in \Omega_q : \text{grad}f(x) = \mathbf{0}, \lambda_1(x) < 0\}.$$

- $\text{grad}f(x)$ is the *Riemannian* gradient on Ω_q .
- $\lambda_1(x) \geq \dots \geq \lambda_q(x)$ are eigenvalues of the *Riemannian* Hessian $\mathcal{H}f(x)$.
- $v_1(x), \dots, v_q(x)$ are associated eigenvectors of $\mathcal{H}f(x)$ within the tangent space T_x .

Local Maxima of f on Ω_q :

$$\mathcal{M} := \{x \in \Omega_q : \text{grad} f(x) = \mathbf{0}, \lambda_1(x) < 0\}.$$

- $\text{grad} f(x)$ is the *Riemannian* gradient on Ω_q .
- $\lambda_1(x) \geq \dots \geq \lambda_q(x)$ are eigenvalues of the *Riemannian* Hessian $\mathcal{H}f(x)$.
- $v_1(x), \dots, v_q(x)$ are associated eigenvectors of $\mathcal{H}f(x)$ within the tangent space T_x .



(Order- d) Directional density ridge of f on Ω_q :

$$\mathcal{R}_d := \left\{ x \in \Omega_q : V_d(x) V_d(x)^T \text{grad} f(x) = \mathbf{0}, \lambda_{d+1}(x) < 0 \right\}.$$

- $V_d(x) = [v_{d+1}(x), \dots, v_q(x)] \in \mathbb{R}^{(q+1) \times (q-d)}$ contains the last $q-d$ eigenvectors of $\mathcal{H}f(x)$.

$$\mathcal{R}_d := \left\{ \boldsymbol{x} \in \Omega_q : V_d(\boldsymbol{x})V_d(\boldsymbol{x})^T \text{grad} f(\boldsymbol{x}) = \mathbf{0}, \lambda_{d+1}(\boldsymbol{x}) < 0 \right\}.$$

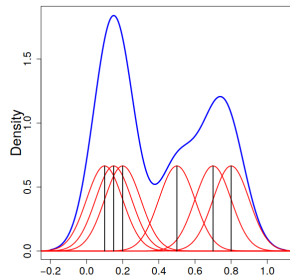
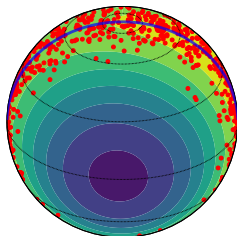
Estimation of Directional Density Ridge

$$\mathcal{R}_d := \left\{ \mathbf{x} \in \Omega_q : V_d(\mathbf{x}) V_d(\mathbf{x})^T \text{grad} f(\mathbf{x}) = \mathbf{0}, \lambda_{d+1}(\mathbf{x}) < 0 \right\}.$$

- ① Estimate $f : \Omega_q \rightarrow \mathbb{R}$ via directional KDE (Hall et al., 1987; Bai et al., 1988; García-Portugués, 2013):

$$\hat{f}_h(\mathbf{x}) = \frac{C_{L,q}(h)}{n} \sum_{i=1}^n L \left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2} \right) \quad \text{with} \quad \mathbf{X}_1, \dots, \mathbf{X}_n \in \Omega_q.$$

- $L : [0, \infty) \rightarrow [0, \infty)$ is a directional kernel, e.g., $L(r) = e^{-r}$.
- $h > 0$ is a bandwidth parameter, and $C_{L,q}(h) > 0$ is a normalizing term.



$$\mathcal{R}_d := \left\{ \mathbf{x} \in \Omega_q : V_d(\mathbf{x})V_d(\mathbf{x})^T \text{grad} f(\mathbf{x}) = \mathbf{0}, \lambda_{d+1}(\mathbf{x}) < 0 \right\}.$$

- ① Estimate $f : \Omega_q \rightarrow \mathbb{R}$ via directional KDE ([Hall et al., 1987](#); [Bai et al., 1988](#); [García-Portugués, 2013](#)):

$$\hat{f}_h(\mathbf{x}) = \frac{C_{L,q}(h)}{n} \sum_{i=1}^n L \left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2} \right) \quad \text{with} \quad \mathbf{X}_1, \dots, \mathbf{X}_n \in \Omega_q.$$

- ② Construct a plug-in estimator of $\mathcal{R}_d$ as ([Zhang and Chen, 2021, 2023](#)):

$$\hat{\mathcal{R}}_d = \left\{ \mathbf{x} \in \Omega_q : \hat{V}_d(\mathbf{x})\hat{V}_d(\mathbf{x})^T \text{grad} \hat{f}_h(\mathbf{x}) = \mathbf{0}, \hat{\lambda}_{d+1}(\mathbf{x}) < 0 \right\}.$$

$$\mathcal{R}_d := \left\{ \mathbf{x} \in \Omega_q : V_d(\mathbf{x})V_d(\mathbf{x})^T \text{grad} f(\mathbf{x}) = \mathbf{0}, \lambda_{d+1}(\mathbf{x}) < 0 \right\}.$$

- 1 Estimate $f : \Omega_q \rightarrow \mathbb{R}$ via directional KDE (Hall et al., 1987; Bai et al., 1988; García-Portugués, 2013):

$$\hat{f}_h(\mathbf{x}) = \frac{C_{L,q}(h)}{n} \sum_{i=1}^n L \left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h^2} \right) \quad \text{with} \quad \mathbf{X}_1, \dots, \mathbf{X}_n \in \Omega_q.$$

- 2 Construct a plug-in estimator of $\mathcal{R}_d$ as (Zhang and Chen, 2021, 2023):

$$\hat{\mathcal{R}}_d = \left\{ \mathbf{x} \in \Omega_q : \hat{V}_d(\mathbf{x})\hat{V}_d(\mathbf{x})^T \text{grad} \hat{f}_h(\mathbf{x}) = \mathbf{0}, \hat{\lambda}_{d+1}(\mathbf{x}) < 0 \right\}.$$

Statistical Consistency:

$$\text{Haus} \left(\mathcal{R}_d, \hat{\mathcal{R}}_d \right) = O(h^2) + O_P \left(\sqrt{\frac{|\log h|}{nh^{q+4}}} \right)$$

$$\text{as } h \rightarrow 0, \frac{nh^{q+6}}{|\log h|} \rightarrow \infty, \text{ where } \text{Haus}(A, B) = \max \left\{ \sup_{\mathbf{x} \in A} d(\mathbf{x}, B), \sup_{\mathbf{y} \in B} d(\mathbf{y}, A) \right\}.$$

Computation of Directional Density Ridge

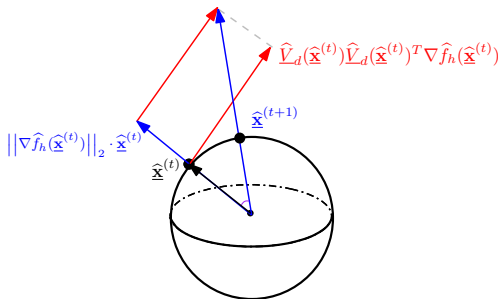
$$\hat{\mathcal{R}}_d = \left\{ \mathbf{x} \in \Omega_q : \hat{V}_d(\mathbf{x}) \hat{V}_d(\mathbf{x})^T \text{grad} \hat{f}_h(\mathbf{x}) = \mathbf{0}, \hat{\lambda}_{d+1}(\mathbf{x}) < 0 \right\}.$$

Computation of Directional Density Ridge

$$\hat{\mathcal{R}}_d = \left\{ \mathbf{x} \in \Omega_q : \hat{V}_d(\mathbf{x}) \hat{V}_d(\mathbf{x})^T \text{grad} \hat{f}_h(\mathbf{x}) = \mathbf{0}, \hat{\lambda}_{d+1}(\mathbf{x}) < 0 \right\}.$$

- 3 Compute $\hat{\mathcal{R}}_d$ via our Directional Subspace Constrained Mean Shift (DirSCMS) algorithm (Zhang and Chen, 2023):

$$\hat{\mathbf{x}}^{(t+1)} \leftarrow \hat{\mathbf{x}}^{(t)} + \hat{V}_d(\hat{\mathbf{x}}^{(t)}) \hat{V}_d(\hat{\mathbf{x}}^{(t)})^T \cdot \frac{\nabla \hat{f}_h(\hat{\mathbf{x}}^{(t)})}{\left\| \nabla \hat{f}_h(\hat{\mathbf{x}}^{(t)}) \right\|_2} \quad \text{and} \quad \hat{\mathbf{x}}^{(t+1)} \leftarrow \frac{\hat{\mathbf{x}}^{(t+1)}}{\left\| \hat{\mathbf{x}}^{(t+1)} \right\|_2} \quad \text{for } t = 0, 1, \dots$$

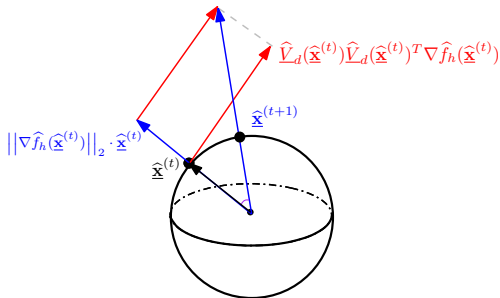


Computation of Directional Density Ridge

$$\widehat{\mathcal{R}}_d = \left\{ \mathbf{x} \in \Omega_q : \widehat{V}_d(\mathbf{x}) \widehat{V}_d(\mathbf{x})^T \text{grad} \widehat{f}_h(\mathbf{x}) = \mathbf{0}, \widehat{\lambda}_{d+1}(\mathbf{x}) < 0 \right\}.$$

- 3 Compute $\widehat{\mathcal{R}}_d$ via our Directional Subspace Constrained Mean Shift (DirSCMS) algorithm (Zhang and Chen, 2023):

$$\widehat{\mathbf{x}}^{(t+1)} \leftarrow \widehat{\mathbf{x}}^{(t)} + \widehat{V}_d(\widehat{\mathbf{x}}^{(t)}) \widehat{V}_d(\widehat{\mathbf{x}}^{(t)})^T \cdot \frac{\nabla \widehat{f}_h(\widehat{\mathbf{x}}^{(t)})}{\left\| \nabla \widehat{f}_h(\widehat{\mathbf{x}}^{(t)}) \right\|_2} \quad \text{and} \quad \widehat{\mathbf{x}}^{(t+1)} \leftarrow \frac{\widehat{\mathbf{x}}^{(t+1)}}{\left\| \widehat{\mathbf{x}}^{(t+1)} \right\|_2} \quad \text{for } t = 0, 1, \dots$$



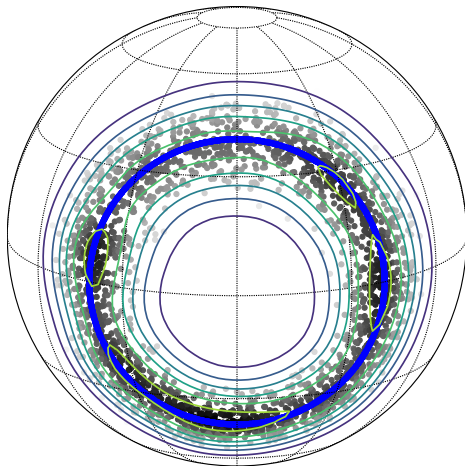
Linear Convergence: For a constant rate $\Upsilon \in (0, 1)$,

$$d\left(\widehat{\mathbf{x}}^{(t)}, \mathbf{x}^{(\infty)}\right) \leq \Upsilon^t \cdot d\left(\widehat{\mathbf{x}}^{(0)}, \mathbf{x}^{(\infty)}\right) + O(h^2) + O_P\left(\sqrt{\frac{|\log h|}{nh^{q+2}}}\right),$$

where $\mathbf{x}^{(\infty)} \in \mathcal{R}_d$ and $h \rightarrow 0$, $\frac{nh^{q+2}}{|\log h|} \rightarrow \infty$.

DirSCMS Algorithm: Simulation Study

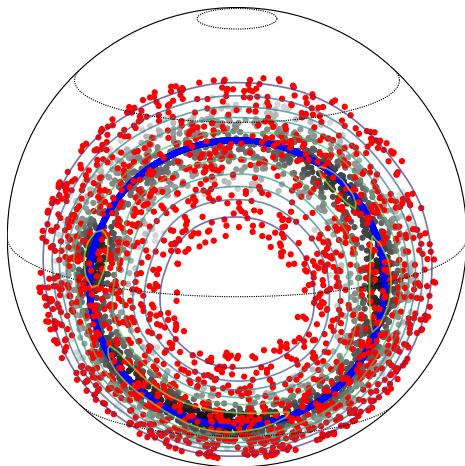
Simulate 2000 data points from a circle on Ω_2 , add Gaussian noise $\mathcal{N}(0, 0.1^2)$ to their Cartesian coordinates, and then apply L_2 normalization.



The underlying circle (blue curve) and sampled points (gray dots) on Ω_2 .

DirSCMS Algorithm: Simulation Study

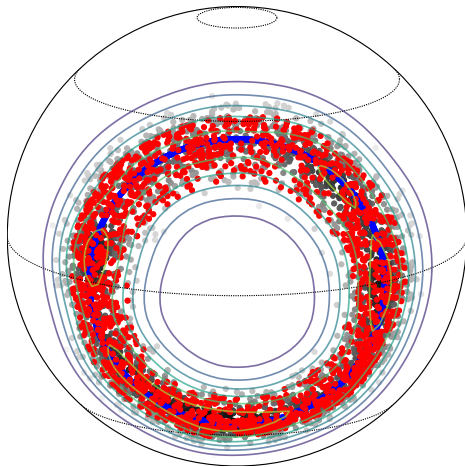
Simulate 2000 data points from a circle on Ω_2 , add Gaussian noise $\mathcal{N}(0, 0.1^2)$ to their Cartesian coordinates, and then apply L_2 normalization.



Directional SCMS at Step 0.

DirSCMS Algorithm: Simulation Study

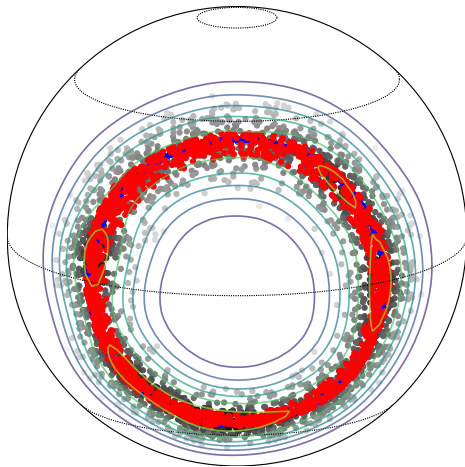
Simulate 2000 data points from a circle on Ω_2 , add Gaussian noise $\mathcal{N}(0, 0.1^2)$ to their Cartesian coordinates, and then apply L_2 normalization.



Directional SCMS at Step 1.

DirSCMS Algorithm: Simulation Study

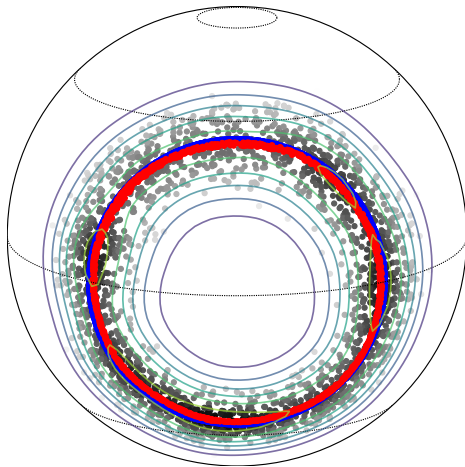
Simulate 2000 data points from a circle on Ω_2 , add Gaussian noise $\mathcal{N}(0, 0.1^2)$ to their Cartesian coordinates, and then apply L_2 normalization.



Directional SCMS at Step 2.

DirSCMS Algorithm: Simulation Study

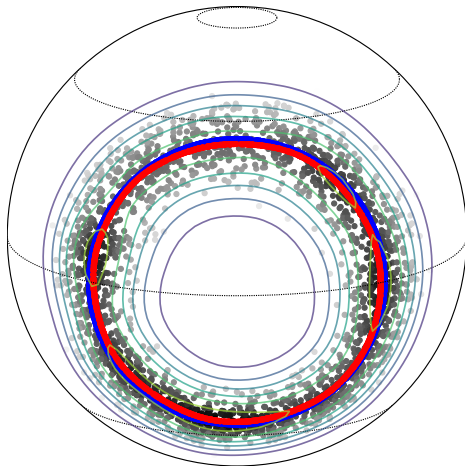
Simulate 2000 data points from a circle on Ω_2 , add Gaussian noise $\mathcal{N}(0, 0.1^2)$ to their Cartesian coordinates, and then apply L_2 normalization.



Directional SCMS at Step 4.

DirSCMS Algorithm: Simulation Study

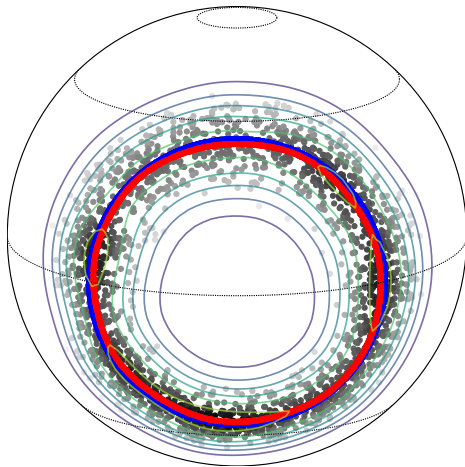
Simulate 2000 data points from a circle on Ω_2 , add Gaussian noise $\mathcal{N}(0, 0.1^2)$ to their Cartesian coordinates, and then apply L_2 normalization.



Directional SCMS at Step 8.

DirSCMS Algorithm: Simulation Study

Simulate 2000 data points from a circle on Ω_2 , add Gaussian noise $\mathcal{N}(0, 0.1^2)$ to their Cartesian coordinates, and then apply L_2 normalization.



Directional SCMS at Step 24 (converged).

Proposed DirSCMS and related methods are encapsulated in a Python package called **SCONCE-SCMS** (Spherical and CONic Cosmic wEb finder with the extended SCMS algorithms; [Zhang et al. 2022](#)).

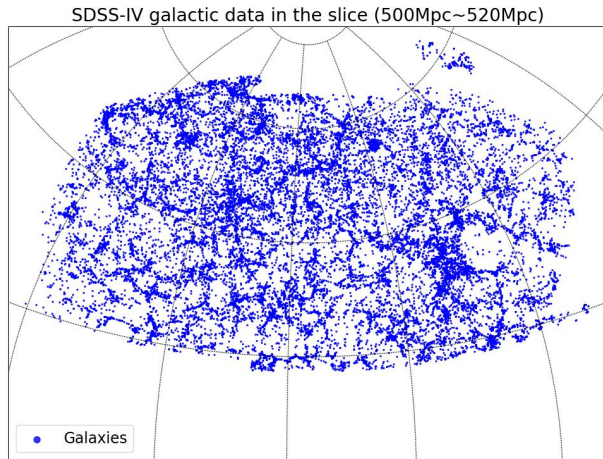


- **Python Package Index:** <https://pypi.org/project/sconce-scms/>.
- **Documentation:** <https://sconce-scms.readthedocs.io/en/latest/>.

SDSS-IV Cosmic Web Catalog

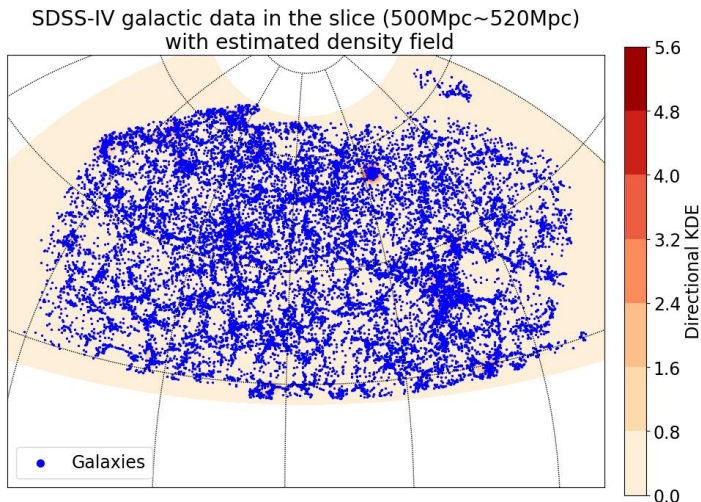


Step 1 (Slicing the Universe): Partition the redshift range into 325 spherical slices based on the comoving distance $\Delta L = 20$ Mpc (under Planck15 Λ CDM cosmology).

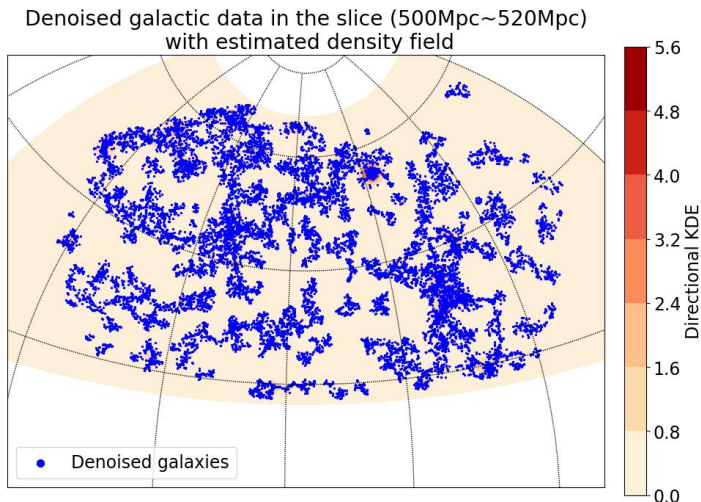


Within each slice, observed galaxies lie on Ω_2

Step 2 (Density Estimation): Estimate the galaxy density field f on each spherical slice Ω_2 by directional KDE.

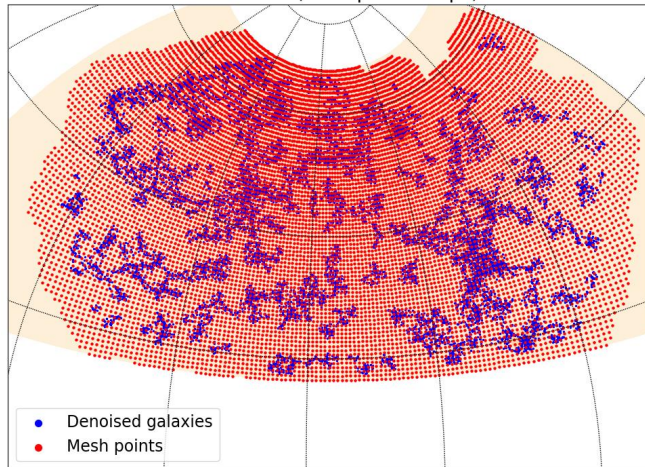


Step 3 (Denoising): Remove 20% of the galactic observations with low estimated density values.

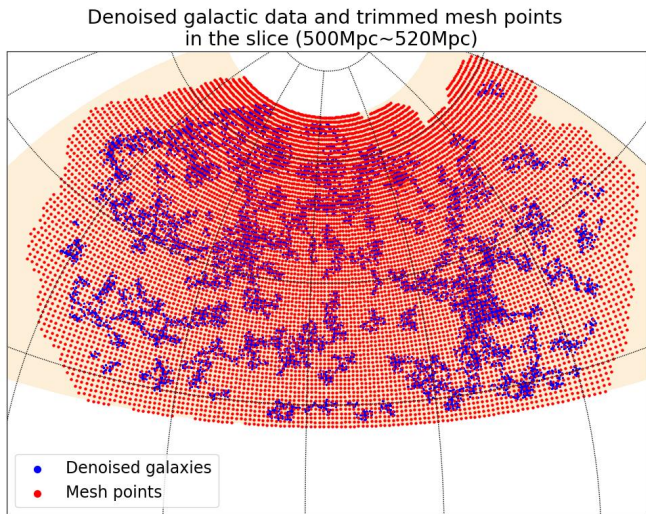


Step 4 (Laying Down Mesh Points): Initialize our DirSCMS algorithm with dense mesh points over the region of interest.

Denoised galactic data and trimmed mesh points
in the slice (500Mpc~520Mpc)

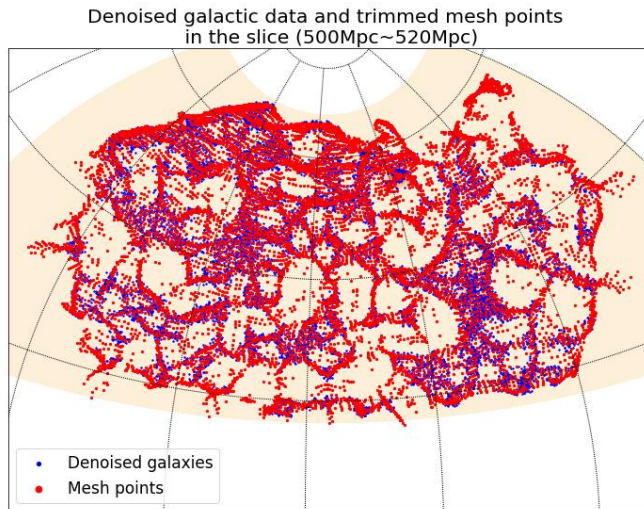


Step 5 (DirSCMS Iterations): Iterate our DirSCMS algorithm on each mesh point until convergence.



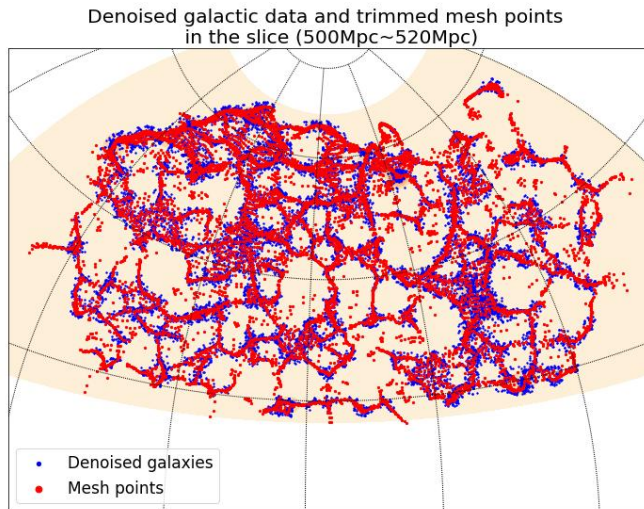
DirSCMS Iterations (Step 0).

Step 5 (DirSCMS Iterations): Iterate our DirSCMS algorithm on each mesh point until convergence.



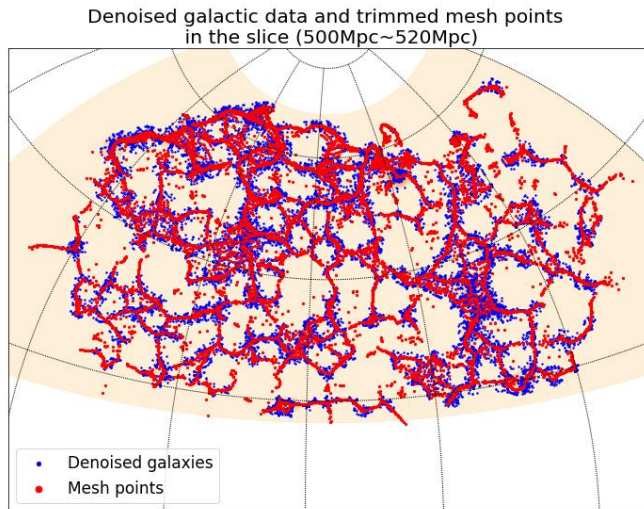
DirSCMS Iterations (Step 1).

Step 5 (DirSCMS Iterations): Iterate our DirSCMS algorithm on each mesh point until convergence.



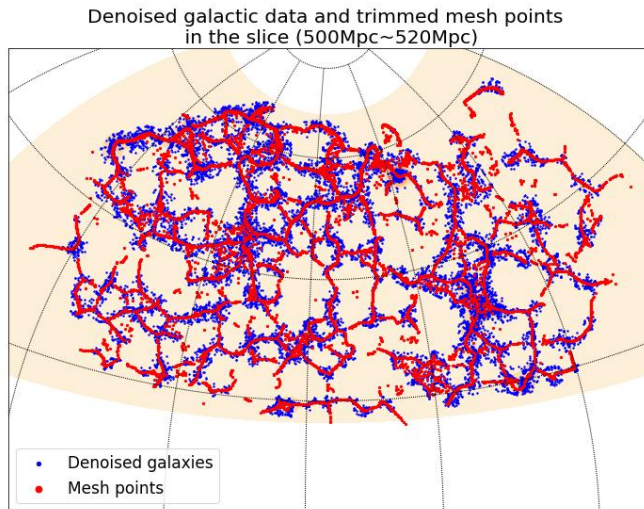
DirSCMS Iterations (Step 2).

Step 5 (DirSCMS Iterations): Iterate our DirSCMS algorithm on each mesh point until convergence.



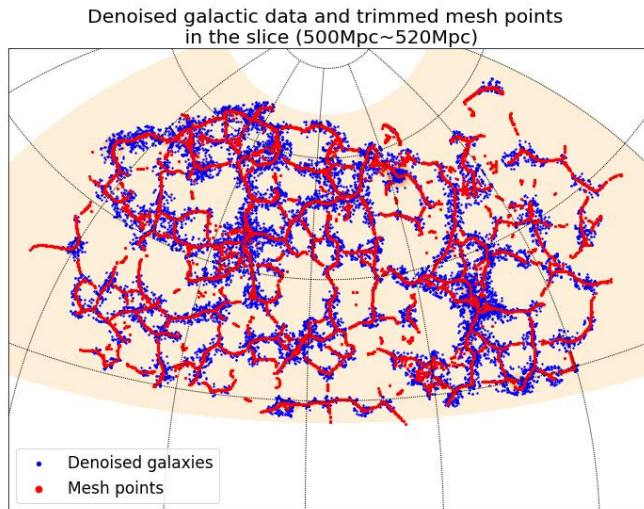
DirSCMS Iterations (Step 3).

Step 5 (DirSCMS Iterations): Iterate our DirSCMS algorithm on each mesh point until convergence.



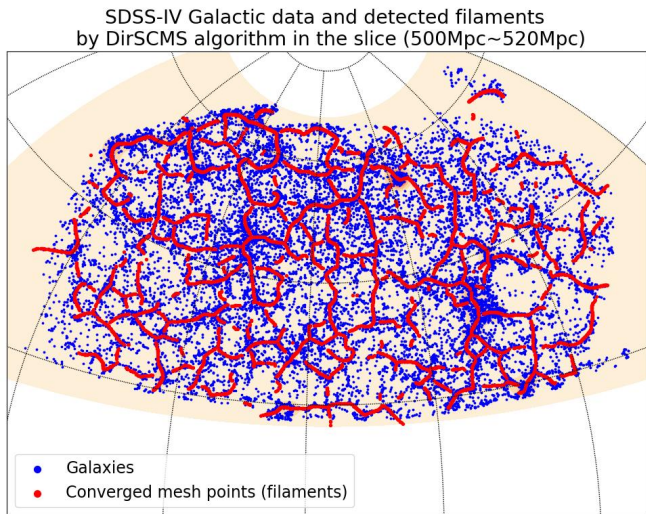
DirSCMS Iterations (Step 5).

Step 5 (DirSCMS Iterations): Iterate our DirSCMS algorithm on each mesh point until convergence.



DirSCMS Iterations (Step 8).

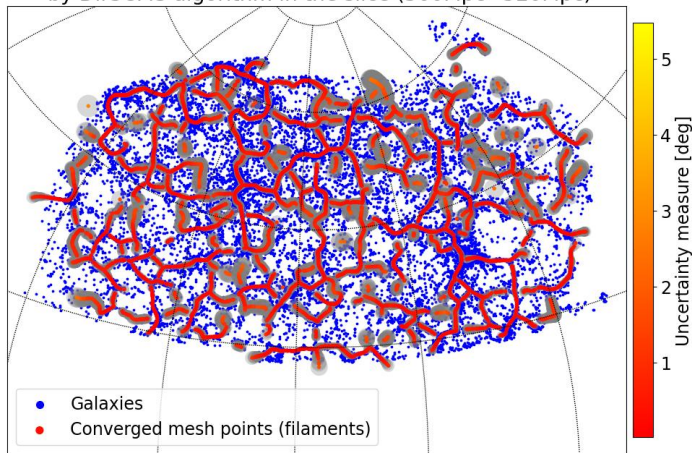
Step 5 (DirSCMS Iterations): Iterate our DirSCMS algorithm on each mesh point until convergence.



DirSCMS Iterations (Final).

Step 6 (Uncertainty Quantification): Compute the uncertainty measures for detected filamentary points via nonparametric bootstrap.

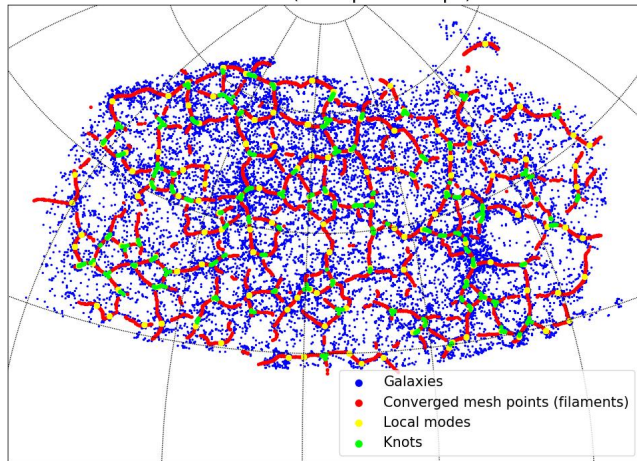
SDSS-IV Galactic data and detected filaments
by DirSCMS algorithm in the slice (500Mpc~520Mpc)



Detected filaments with uncertainty measures.

Step 7 (Mode and Knot Estimation): Identify the local modes and knots on the detected filaments as cosmic nodes.

SDSS-IV Galactic data and detected filaments by DirSCMS algorithm
in the slice (500Mpc~520Mpc)



Cosmic nodes on the detected filaments.

Cosmic Web Catalog on SDSS-IV Data

The final cosmic node and filament catalog data ([Zhang, 2022](#)) are available at <https://doi.org/10.5281/zenodo.6244866>.

Published June 10, 2022 | Version 0.0.1

Dataset

Open

SDSS-IV Cosmic Web Catalog

Zhang, Yikun¹ 

Show affiliations

Contributors

Supervisors: Chen, Yen-Chi¹ ; de Souza, Rafael S.² 

Show affiliations

This repository contains the cosmic web catalog data released in the paper "Cosmic Web Catalog on SDSS-IV Data with SCONCE" (preparing).

The catalog is constructed on the SDSS-IV galaxies and quasars (QSO) using our proposed Directional Subspace Constrained Mean Shift (DirSCMS) algorithm. We release both the cosmic filaments and local modes (i.e., local maxima of the estimated galaxy/QSO density field, which serves as candidates of galaxy clusters) within 325 thin redshift slices, each of which spans 20Mpc under the Planck15 cosmology. The entire catalog covers the redshift range from $z = 0$ to $z = 3$.

Conclusion and Future Works



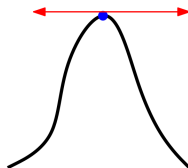
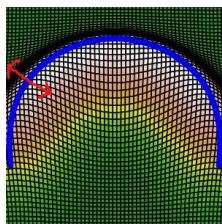
We discuss both statistical and computational learning theory for cosmic filament detection with SDSS-IV galaxy data on Ω_q .

Conclusions

We discuss both statistical and computational learning theory for cosmic filament detection with SDSS-IV galaxy data on Ω_q .

① From the statistical perspective:

- Model the cosmic filament through **directional density ridge** and consistently estimate it through directional KDE.



Conclusions

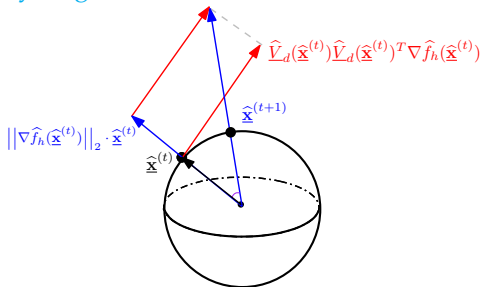
We discuss both statistical and computational learning theory for cosmic filament detection with SDSS-IV galaxy data on Ω_q .

1 From the statistical perspective:

- Model the cosmic filament through **directional density ridge** and consistently estimate it through directional KDE.

2 From the computational perspective:

- Propose a **geometry-aware** and **linearly convergent** algorithm (DirSCMS) to compute the **directional density ridge**.



We discuss both statistical and computational learning theory for cosmic filament detection with SDSS-IV galaxy data on Ω_q .

① **From the statistical perspective:**

- Model the cosmic filament through **directional density ridge** and consistently estimate it through directional KDE.

② **From the computational perspective:**

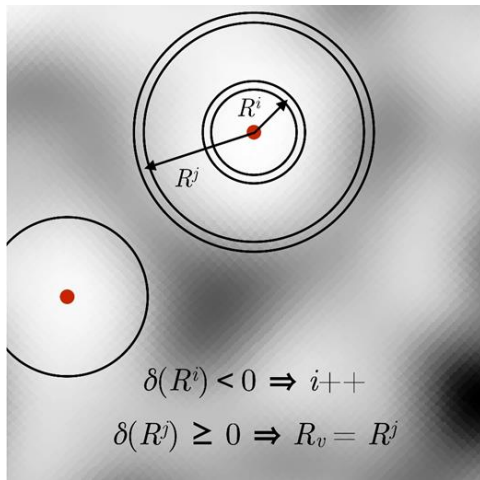
- Propose a **geometry-aware** and **linearly convergent** algorithm (DirSCMS) to compute the **directional density ridge**.

③ **For downstream scientific analyses:**

- Publish a public cosmic node and filament catalog.

Future Work: Cosmic Void Detection

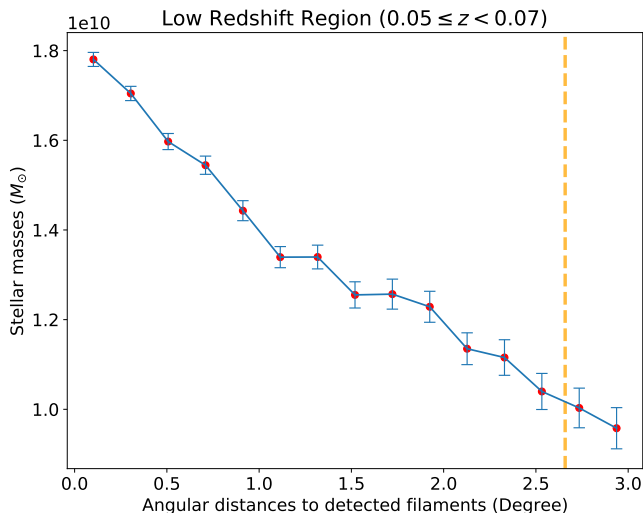
- 1 Leverage our cosmic filament catalog to identify cosmic voids and infer the precise cosmology (Sánchez et al., 2016).



Simple void-finding algorithm (Sánchez et al., 2016).

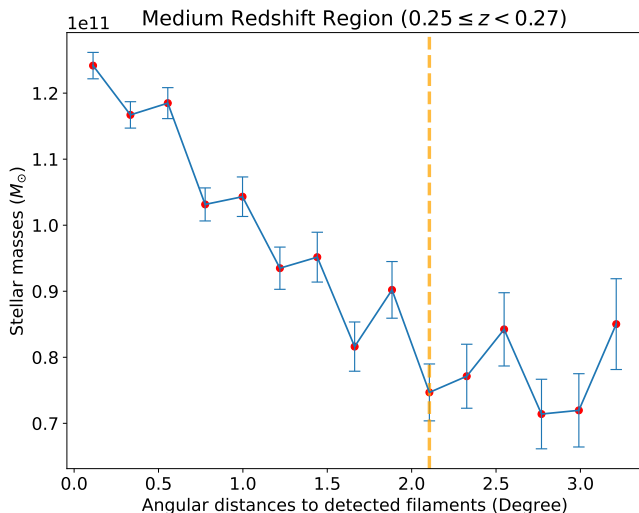
Future Work: Filament Effects of Galaxy Properties

- Analyze if galaxy properties, such as stellar mass, color, and star formation rate, are correlated with our detected cosmic web structures ([Chen et al., 2017](#); [Kotecha, 2020](#))...



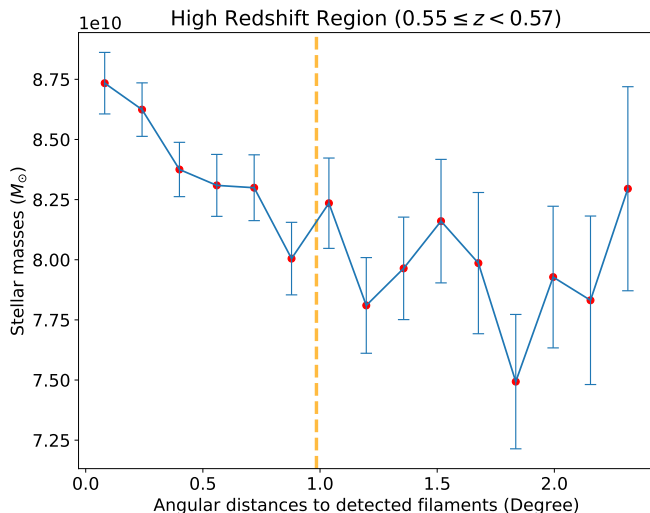
Future Work: Filament Effects of Galaxy Properties

- Analyze if galaxy properties, such as stellar mass, color, and star formation rate, are correlated with our detected cosmic web structures ([Chen et al., 2017](#); [Kotecha, 2020](#))...



Future Work: Filament Effects of Galaxy Properties

- Analyze if galaxy properties, such as stellar mass, color, and star formation rate, are correlated with our detected cosmic web structures ([Chen et al., 2017](#); [Kotecha, 2020](#))...



Thank you!

More details can be found in <https://zhangyk8.github.io/>.



- [1] Y. Zhang and Y.-C. Chen. Kernel Smoothing, Mean Shift, and Their Learning Theory with Directional Data. *Journal of Machine Learning Research*, 22(154):1–92, 2021. <https://arxiv.org/abs/2010.13523>
- [2] Y. Zhang and Y.-C. Chen. The EM Perspective of Directional Mean Shift Algorithm. arXiv preprint, 2021. <https://arxiv.org/abs/2101.10058>
- [3] Y. Zhang and Y.-C. Chen. Linear Convergence of the Subspace Constrained Mean Shift Algorithm: From Euclidean to Directional Data. *Information and Inference: A Journal of the IMA*, 12(1): 210–311, 2022. <https://arxiv.org/abs/2104.14977>
- [4] Y. Zhang and Y.-C. Chen. Mode and Ridge Estimation in Euclidean and Directional Product Spaces: A Mean Shift Approach. *Journal of Computational and Graphical Statistics*, 35(1): 101–110, 2026. <https://arxiv.org/abs/2110.08505>
- [5] Y. Zhang, R. S. de Souza, and Y.-C. Chen. SCONE: A Cosmic Web Finder for Spherical and Conic Geometries. *Monthly Notices of the Royal Astronomical Society*, 517(1): 1197–1217, 2022. <https://arxiv.org/abs/2207.07001>

Reference

- P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, 2008.
- P. A. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Banday, R. Barreiro, J. Bartlett, N. Bartolo, et al. Planck 2015 results-xiii. cosmological parameters. *Astronomy & Astrophysics*, 594:A13, 2016.
- Z. Bai, C. Rao, and L. Zhao. Kernel estimators of density function of directional data. *Journal of Multivariate Analysis*, 27(1):24 – 39, 1988.
- J. R. Bond, L. Kofman, and D. Pogosyan. How filaments of galaxies are woven into the cosmic web. *Nature*, 380 (6575):603–606, 1996.
- M. Cautun, R. van de Weygaert, and B. J. Jones. Nexus: tracing the cosmic web connection. *Monthly Notices of the Royal Astronomical Society*, 429(2):1286–1308, 2013.
- Y.-C. Chen, C. R. Genovese, and L. Wasserman. Asymptotic theory for density ridges. *The Annals of Statistics*, 43(5): 1896–1928, 2015a.
- Y.-C. Chen, S. Ho, P. E. Freeman, C. R. Genovese, and L. Wasserman. Cosmic web reconstruction through density ridges: method and algorithm. *Monthly Notices of the Royal Astronomical Society*, 454(1):1140–1156, 2015b.
- Y.-C. Chen, S. Ho, R. Mandelbaum, N. A. Bahcall, J. R. Brownstein, P. E. Freeman, C. R. Genovese, D. P. Schneider, and L. Wasserman. Detecting effects of filaments on galaxy properties in the sloan digital sky survey iii. *Monthly Notices of the Royal Astronomical Society*, 466(2):1880–1893, 2017.
- J. C. Duque, M. Migliaccio, D. Marinucci, and N. Vittorio. A novel cosmic filament catalogue from sdss data. *Astronomy & Astrophysics*, 659:A166, 2022.

- E. García-Portugués. Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electronic Journal of Statistics*, 7:1655–1685, 2013.
- E. García-Portugués, R. M. Crujeiras, and W. González-Manteiga. Central limit theorems for directional and linear random variables with applications. *Statistica Sinica*, pages 1207–1229, 2015.
- P. Hall, G. S. Watson, and J. Cabrara. Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762, 12 1987. ISSN 0006-3444. URL <https://doi.org/10.1093/biomet/74.4.751>.
- S. He, S. Alam, S. Ferraro, Y.-C. Chen, and S. Ho. The detection of the imprint of filaments on cosmic microwave background lensing. *Nature Astronomy*, 2(5):401–406, 2018.
- G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. Bennett, J. Dunkley, M. Nolta, M. Halpern, R. Hill, N. Odegard, et al. Nine-year wilkinson microwave anisotropy probe (wmap) observations: cosmological parameter results. *The Astrophysical Journal Supplement Series*, 208(2):19, 2013.
- N. Kaiser. Clustering in real space and in redshift space. *Monthly Notices of the Royal Astronomical Society*, 227(1):1–21, 1987.
- S. Kotecha. *The Effect of Cosmic Web Filaments on Quenching in Galaxy Clusters*. PhD thesis, McMaster University, 2020.
- U. Kuchner, A. Aragón-Salamanca, A. Rost, F. R. Pearce, M. E. Gray, W. Cui, A. Knebe, E. Rasia, and G. Yepes. Cosmic filaments in galaxy cluster outskirts: quantifying finding filaments in redshift space. *Monthly Notices of the Royal Astronomical Society*, 503(2):2065–2076, 2021.

- C. Laigle, C. Pichon, S. Arnouts, H. J. McCracken, Y. Dubois, J. Devriendt, A. Slyz, D. Le Borgne, A. Benoit-Levy, H. S. Hwang, et al. Cosmological photometric redshifts probe the impact of filaments on galaxy properties. *Monthly Notices of the Royal Astronomical Society*, 474(4):5437–5458, 2018.
- N. I. Libeskind, R. Van De Weygaert, M. Cautun, B. Falck, E. Tempel, T. Abel, M. Alpaslan, M. A. Aragón-Calvo, J. E. Forero-Romero, R. Gonzalez, et al. Tracing the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 473(1):1195–1217, 2018.
- N. Malavasi, M. Langer, N. Aghanim, D. Galárraga-Espinosa, and C. Gouin. Relative effect of nodes and filaments of the cosmic web on the quenching of galaxies and the orientation of their spin. *Astronomy & Astrophysics*, 658:A113, 2022.
- U. Ozertem and D. Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12(34):1249–1286, 2011.
- C. Sánchez, J. Clampitt, A. Kovacs, B. Jain, J. García-Bellido, S. Nadathur, D. Gruen, N. Hamaus, D. Huterer, P. Vielzeuf, et al. Cosmic voids and void lensing in the dark energy survey science verification data. *Monthly Notices of the Royal Astronomical Society*, 465(1):746–759, 10 2016.
- W. Sargent and E. Turner. A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies. *The Astrophysical Journal*, 212:L3–L7, 1977.
- V. Springel, S. D. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *nature*, 435(7042):629–636, 2005.

- E. Tempel, R. Stoica, V. J. Martinez, L. Liivamägi, G. Castellan, and E. Saar. Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the sdss. *Monthly Notices of the Royal Astronomical Society*, 438(4): 3465–3482, 2014.
- M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, and L. Hernquist. Introducing the illustris project: simulating the coevolution of dark and visible matter in the universe. *Monthly Notices of the Royal Astronomical Society*, 444(2):1518–1547, 2014.
- Y. B. Zel'Dovich. Gravitational instability: An approximate theory for large density perturbations. *Astronomy and astrophysics*, 5:84–89, 1970.
- Y. Zhang. SDSS-IV cosmic web catalog, June 2022. URL <https://doi.org/10.5281/zenodo.6244866>.
- Y. Zhang and Y.-C. Chen. Kernel smoothing, mean shift, and their learning theory with directional data. *Journal of Machine Learning Research*, 22(154):1–92, 2021.
- Y. Zhang and Y.-C. Chen. Linear convergence of the subspace constrained mean shift algorithm: From euclidean to directional data. *Information and Inference: A Journal of the IMA*, 12(1):210–311, 2023. URL <https://doi.org/10.1093/imaiai/iaac005>.
- Y. Zhang and Y.-C. Chen. Mode and ridge estimation in euclidean and directional product spaces: A mean shift approach. *Journal of Computational and Graphical Statistics*, 35(1):101–110, 2026.
- Y. Zhang, R. S. de Souza, and Y.-C. Chen. Sconce: a cosmic web finder for spherical and conic geometries. *Monthly Notices of the Royal Astronomical Society*, 517(1):1197–1217, 2022.

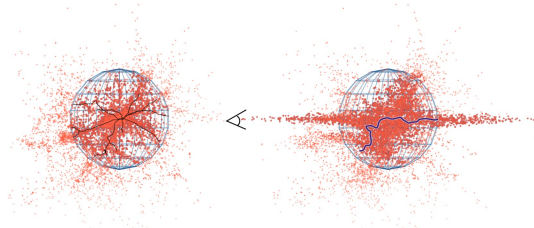
Drawbacks of 3D Cosmic Web Detection Methods

3D method: Convert $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n$ to Cartesian coordinates as

$$\tilde{X}_i = d(z_i) \cos \alpha_i \cos \delta_i, \quad \tilde{Y}_i = d(z_i) \sin \alpha_i \cos \delta_i, \quad \tilde{Z}_i = d(z_i) \sin \delta_i,$$

where $d(\cdot)$ is a distance transformation function.

- The determination of $d(\cdot)$ requires complex cosmological models.
- The galaxy distribution is distorted along the line of sight, *i.e.*, the so-called *finger-of-god* ([Sargent and Turner, 1977](#)) and *Kaiser* ([Kaiser, 1987](#)) effects.



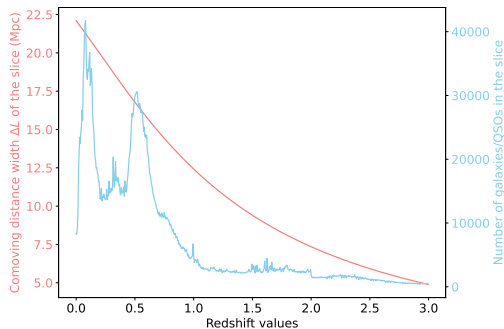
Redshift distortions along the line of sight ([Kuchner et al., 2021](#)).

- The number of galaxies differs across redshift, so applying 3D approaches is computationally intensive.

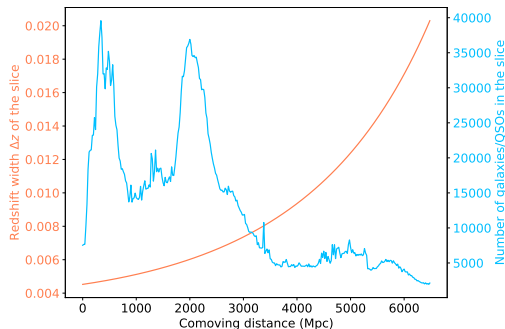
Slicing the Universe According to Comoving Distances

We slice the Universe using a Λ CDM cosmology, such as Planck15 (Ade et al., 2016) or WMAP9 (Hinshaw et al., 2013), but not in the original redshift space.

- It helps balance the number of galaxies across slices.
- The resulting cosmic web catalog is relatively insensitive to the choice of cosmological model.



(a) By redshift $\Delta z = 0.005$.



(b) By comoving distance $\Delta L = 20$ Mpc.

Assume tentatively that the directional function f is well-defined and smooth in $\mathbb{R}^{q+1} \setminus \{\mathbf{0}\}$ (or at least in an open neighborhood $U \supset \Omega_q$).

- *Riemannian gradient* $\text{grad} f(\mathbf{x})$ on Ω_q :

$$\text{grad} f(\mathbf{x}) = (\mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T) \nabla f(\mathbf{x}),$$

where $\mathbf{I}_{q+1}$ is the identity matrix in $\mathbb{R}^{(q+1) \times (q+1)}$ and $\nabla f(\mathbf{x})$ is the total gradient of f in $\mathbb{R}^{q+1}$.

- *Riemannian Hessian* $\mathcal{H}f(\mathbf{x})$ on Ω_q ([Absil et al., 2008](#); [Zhang and Chen, 2021](#)):

$$\mathcal{H}f(\mathbf{x}) = (\mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T) \left[\nabla \nabla f(\mathbf{x}) - \nabla f(\mathbf{x})^T \mathbf{x} \cdot \mathbf{I}_{q+1} \right] (\mathbf{I}_{q+1} - \mathbf{x}\mathbf{x}^T),$$

where $\nabla \nabla f(\mathbf{x})$ is the total Hessian of f in $\mathbb{R}^{q+1}$

Input:

- Directional data sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ and a suitable mesh $\mathcal{M}_D \subset \Omega_q$ of initial points;
- Order d of directional density ridge, bandwidth $h > 0$, and tolerance level $\epsilon > 0$.

Step 1: Compute the directional KDE $\hat{f}_h(\mathbf{x}) = \frac{c_{L,q}(h)}{n} \sum_{i=1}^n L\left(\frac{1-\mathbf{x}^T \mathbf{X}_i}{h^2}\right)$ on the mesh $\mathcal{M}_D$.

Step 2: For each $\hat{\mathbf{x}}^{(0)} \in \mathcal{M}_D$, iterate the following DirSCMS update until convergence:

while $\left\| \sum_{i=1}^n \hat{V}_d(\hat{\mathbf{x}}^{(t)}) \hat{V}_d(\hat{\mathbf{x}}^{(t)})^T \mathbf{X}_i \cdot L'\left(\frac{1-\mathbf{X}_i^T \hat{\mathbf{x}}^{(t)}}{h^2}\right) \right\|_2 > \epsilon$ **do:**

- **Step 2-1:** Compute the scaled version of the estimated Hessian matrix as:

$$\begin{aligned} \frac{nh^2}{c_{L,q}(h)} \mathcal{H} \hat{f}_h(\hat{\mathbf{x}}^{(t)}) &= \left[\mathbf{I}_{q+1} - \hat{\mathbf{x}}^{(t)} (\hat{\mathbf{x}}^{(t)})^T \right] \left[\frac{1}{h^2} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \cdot L''\left(\frac{1-\mathbf{X}_i^T \hat{\mathbf{x}}^{(t)}}{h^2}\right) \right. \\ &\quad \left. + \sum_{i=1}^n \mathbf{X}_i^T \hat{\mathbf{x}}^{(t)} \mathbf{I}_{q+1} \cdot L'\left(\frac{1-\mathbf{X}_i^T \hat{\mathbf{x}}^{(t)}}{h^2}\right) \right] \left[\mathbf{I}_{q+1} - \hat{\mathbf{x}}^{(t)} (\hat{\mathbf{x}}^{(t)})^T \right]. \end{aligned}$$

- **Step 2-2:** Perform the spectral decomposition on $\frac{nh^2}{c_{L,q}(h)} \mathcal{H} \hat{f}_h(\hat{\mathbf{x}}^{(t)})$ and compute $\hat{V}_d(\hat{\mathbf{x}}^{(t)}) = [\mathbf{v}_{d+1}(\hat{\mathbf{x}}^{(t)}), \dots, \mathbf{v}_q(\hat{\mathbf{x}}^{(t)})]$, whose columns are orthonormal eigenvectors corresponding to the smallest $q - d$ eigenvalues inside the tangent space $T_{\hat{\mathbf{x}}^{(t)}}$.
- **Step 2-3:** Update

$$\hat{\mathbf{x}}^{(t+1)} \leftarrow \hat{\mathbf{x}}^{(t)} - \hat{V}_d(\hat{\mathbf{x}}^{(t)}) \hat{V}_d(\hat{\mathbf{x}}^{(t)})^T \left[\frac{\sum_{i=1}^n \mathbf{X}_i L' \left(\frac{1 - \mathbf{X}_i^T \hat{\mathbf{x}}^{(t)}}{h^2} \right)}{\sum_{i=1}^n L' \left(\frac{1 - \mathbf{X}_i^T \hat{\mathbf{x}}^{(t)}}{h^2} \right)} \right].$$

- **Step 2-4:** Standardize $\hat{\mathbf{x}}^{(t+1)}$ as $\hat{\mathbf{x}}^{(t+1)} \leftarrow \frac{\hat{\mathbf{x}}^{(t+1)}}{\|\hat{\mathbf{x}}^{(t+1)}\|_2}$.

Output: An estimated directional d -ridge $\hat{\mathcal{R}}_d$ represented by the collection of convergent points.

Under some regularity conditions, we prove the following properties:

- 1 **R-Linear convergence of $d(\mathbf{x}^{(k)}, \mathcal{R}_d)$ with f .** When the step size $\underline{\eta} > 0$ is sufficiently small and the initial point $\mathbf{x}^{(0)}$ lies within a small neighborhood of its limiting point $\mathbf{x}^*$ in $\mathcal{R}_d$,

$$d(\mathbf{x}^{(k)}, \mathcal{R}_d) \leq \underline{\Upsilon}^k \cdot d(\mathbf{x}^{(0)}, \mathbf{x}^*) \quad \text{with} \quad \underline{\Upsilon} = \sqrt{1 - \frac{\eta\beta_0}{4}},$$

where $\beta_0 > 0$ is the eigengap between the d -th and $(d+1)$ -th eigenvalues of $\mathcal{H}f(\mathbf{x})$.

- 2 **R-Linear convergence of $d(\hat{\mathbf{x}}^{(k)}, \mathcal{R}_d)$ with $\hat{f}_h$.** When the step size $\underline{\eta} > 0$ is sufficiently small and the initial point $\hat{\mathbf{x}}^{(0)}$ lies within a small neighborhood of $\mathbf{x}^*$ in $\mathcal{R}_d$,

$$d(\hat{\mathbf{x}}^{(k)}, \mathcal{R}_d) \leq \underline{\Upsilon}^k \cdot d(\hat{\mathbf{x}}^{(0)}, \mathbf{x}^*) + O(h^2) + O_P\left(\sqrt{\frac{|\log h|}{nh^{q+4}}}\right)$$

with probability tending to 1, as $h \rightarrow 0$ and $\frac{nh^{q+4}}{|\log h|} \rightarrow \infty$.

- The (directional) mean shift and SCMS algorithms can be viewed as variants of the (subspace constrained) gradient ascent methods (on Ω_q) but with adaptive step sizes (Zhang and Chen, 2021, 2023):

$$\hat{\mathbf{x}}^{(t+1)} = \text{Exp}_{\hat{\mathbf{x}}^{(t)}} \left(\underline{\eta}_{n,h}^{(t)} \cdot \hat{V}_d(\hat{\mathbf{x}}^{(t)}) \hat{V}_d(\hat{\mathbf{x}}^{(t)})^T \text{grad} \hat{f}_h(\hat{\mathbf{x}}^{(t)}) \right),$$

where

$$\underline{\eta}_{n,h}^{(t)} = \frac{\theta'_t}{\left\| \nabla \hat{f}_h(\hat{\mathbf{x}}^{(t)}) \right\|_2 \tan \theta'_t} \quad \text{for } t = 0, 1, \dots$$

with θ'_t being the angle between $\hat{\mathbf{x}}^{(t+1)}$ and $\hat{\mathbf{x}}^{(t)}$.

- The step size can be made sufficiently small:

$$\underline{\eta}_{n,h}^{(t)} = O(h^2) \quad \text{as } h \rightarrow 0 \quad \text{and} \quad nh^q \rightarrow \infty,$$

but is also universally bounded away from 0 with respect to the iteration number t .

An Extension to Directional-Linear Product Spaces

When the observed galaxy data $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n \subset \Omega_2 \times \mathbb{R}$ are directional-linear (3D),

$$\mathcal{R}_d := \left\{ (\mathbf{x}, z) \in \Omega_2 \times \mathbb{R} : V_d(\mathbf{x}, z) V_d(\mathbf{x}, z)^T \text{grad} f(\mathbf{x}, z) = \mathbf{0}, \lambda_{d+1}(\mathbf{x}, z) < 0 \right\}.$$

- Density estimation at $(\mathbf{x}, z) \in \Omega_2 \times \mathbb{R}$ (García-Portugués et al., 2015):

$$\hat{f}_h(\mathbf{x}, z) = \frac{C_L(h_1)}{nh_2} \sum_{i=1}^n L\left(\frac{1 - \mathbf{x}^T \mathbf{X}_i}{h_1^2}\right) K\left(\frac{z - z_i}{h_2}\right).$$

- Ridge-Finding via SCMS algorithm on $\mathbf{y}^{(t)} = (\mathbf{x}^{(t)}, z^{(t)})$ as (Zhang and Chen, 2026):

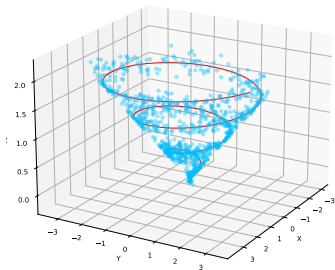
$$\mathbf{y}^{(t+1)} \leftarrow \mathbf{y}^{(t)} + \eta \cdot \hat{V}_d(\mathbf{y}^{(t)}) \hat{V}_d(\mathbf{y}^{(t)})^T \mathbf{H}^{-1} \Xi(\mathbf{y}^{(t)}) \quad \text{with } \mathbf{H} = \text{Diag}(h_1^2, h_1^2, h_2^2) \text{ and}$$

$$\Xi(\mathbf{y}) = (\Xi_x(\mathbf{x}, z), \Xi_z(\mathbf{x}, z))^T = \left(\frac{\sum_{i=1}^n \mathbf{X}_i L'\left(\frac{1 - \mathbf{X}_i^T \mathbf{x}^{(t)}}{h_1^2}\right) K\left(\frac{z^{(t)} - z_i}{h_2}\right)}{\sum_{i=1}^n L'\left(\frac{1 - \mathbf{X}_i^T \mathbf{x}^{(t)}}{h_1^2}\right) K\left(\frac{z^{(t)} - z_i}{h_2}\right)} - \mathbf{x}, \frac{\sum_{i=1}^n z_i L'\left(\frac{1 - \mathbf{X}_i^T \mathbf{x}^{(t)}}{h_1^2}\right) K'\left(\frac{z^{(t)} - z_i}{h_2}\right)}{\sum_{i=1}^n L'\left(\frac{1 - \mathbf{X}_i^T \mathbf{x}^{(t)}}{h_1^2}\right) K'\left(\frac{z^{(t)} - z_i}{h_2}\right)} - z \right).$$

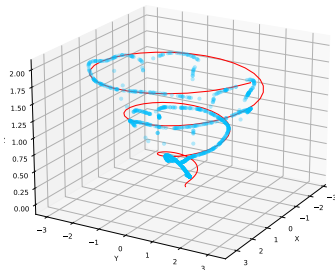
► **Note:** A naive generalization of SCMS algorithm $\mathbf{y}^{(t+1)} \leftarrow \mathbf{y}^{(t)} + \hat{V}_d(\mathbf{y}^{(t)}) \hat{V}_d(\mathbf{y}^{(t)})^T \Xi(\mathbf{y}^{(t)})$ plus standardization as with pure Euclidean/directional data does not work!

Filament Detection in the Directional-Linear Space

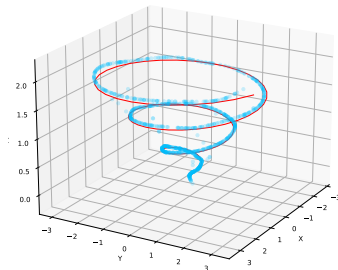
We sample 1000 observations on a spiral curve with additive Gaussian noise $\mathcal{N}(0, 0.2^2)$ in their angular-linear coordinates.



(a) Simulated data points.



(b) Euclidean SCMS.

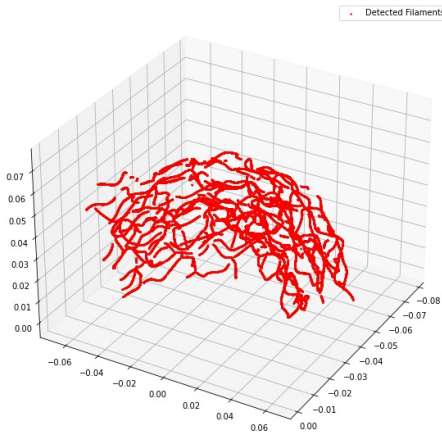


(c) Directional-linear SCMS.

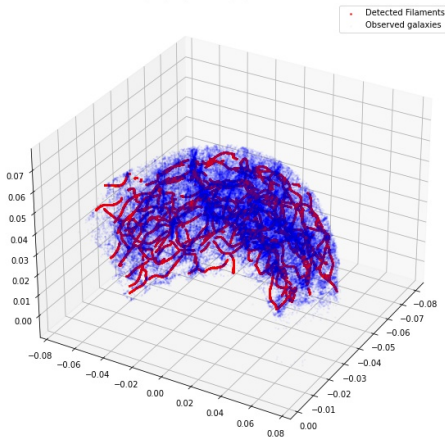
► Our directional-linear SCMS algorithm is more stable than its Euclidean prototype.

Application of DirLinSCMS to SDSS-IV Galaxy Data

Cosmic Filament Detection via Directional-Linear SCMS Algorithm
on the 3D (RA,DEC,Redshift) space with $0.05 \leq z < 0.08$



Cosmic Filament Detection via Directional-Linear SCMS Algorithm
on the 3D (RA,DEC,Redshift) space with $0.05 \leq z < 0.08$



Cosmic filament detection in the 3D (RA, DEC, Redshift) space with our directional-linear SCMS algorithm.