Cosmic Web Detection and Inference Through Directional Density Ridges

Yikun Zhang[†] (yikun@uw.edu)

Joint work with Yen-Chi Chen[†], Rafael S. de Souza[*], and Alexander Giessing[†]

[†]*Department of Statistics, University of Washington* [*] *Department of Physics, Astronomy and Mathematics, University of Hertfordshire*

Background

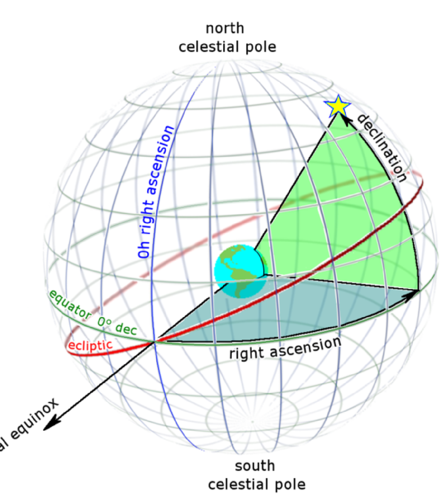
Previous simulation and observational studies have corroborated that on megaparsec scales, matter in the Universe is not uniformly distributed but rather forms a complicated large-scale network structure called **Cosmic Web**. It has four main components:

- 0D massive and overdense *galaxy clusters (or nodes)*,
- 1D interconnected *cosmic filaments*,
- 2D tenuous *cosmic sheets (or walls)*,
- 3D vast and near-empty *cosmic voids*.

Our research objective is to detect the cosmic web structure based on the distribution of observed galaxies by the Sloan Digital Sky Survey (SDSS-IV) and study its effects on the stellar properties of nearby galaxies. In particular, we mainly focus on the one-dimensional cosmic filaments.

Spatial Coordinates of Galaxies: $\{(\alpha_i, \delta_i, z_i)\}_{i=1}^n$:

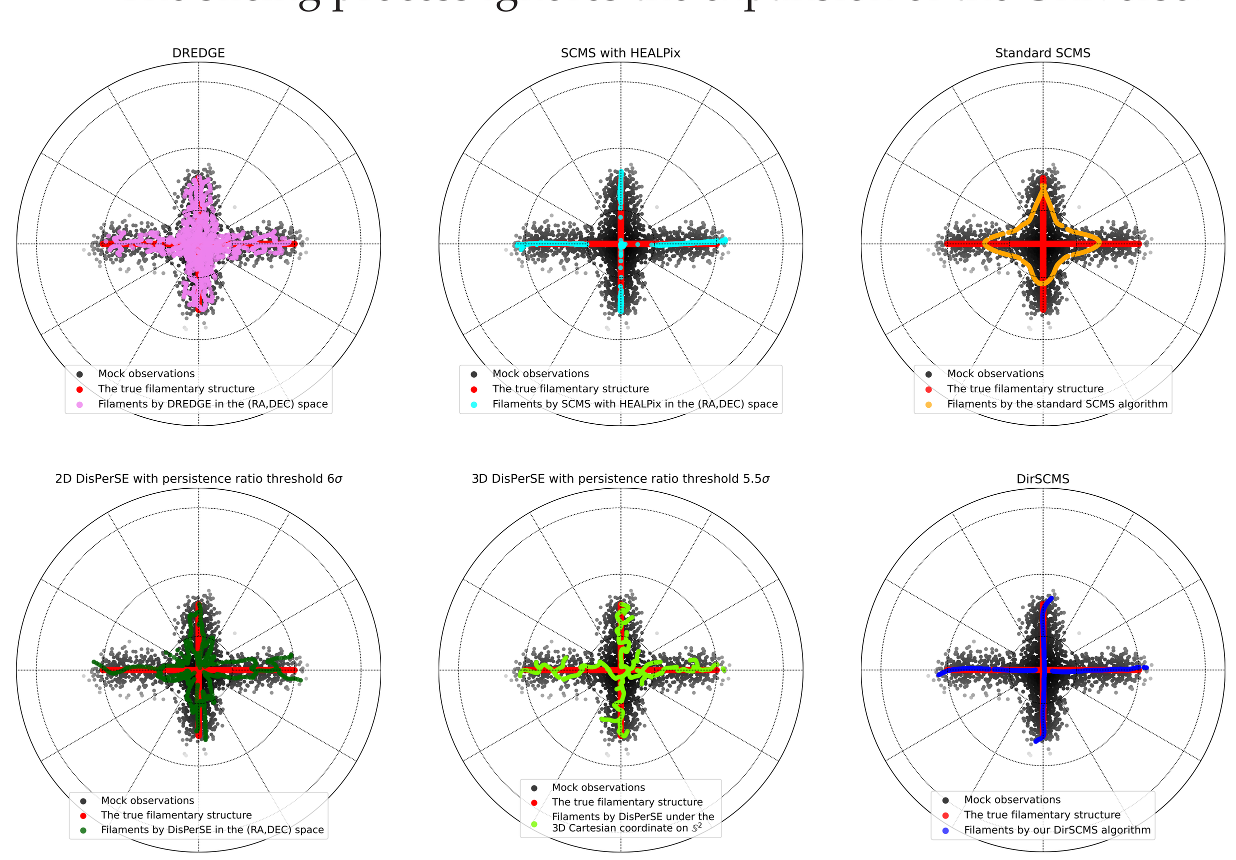
- $\alpha_i \in [0, 360°)$ is the *right ascension* (RA; celestial longitude);
- $\delta_i \in [-90°, 90°]$ is the *declination* (DEC; celestial latitude),
- $z_i \in \mathbb{R}^+$ is the *redshift* value measuring the distance from an observer to the galaxy.



Motivations and Previous Works

The existing filament detection methods lie in two categories:

- **3D Method:** Convert redshift into (comoving) distances and recover the filaments in the 3D space. *Drawback:*
  - The distance transformation function depends on complicated cosmological models.
  - The detected filaments could be spurious due to the finger-of-god effects.
- **2D Method:** Partition the Universe into thin redshift slices and estimate the filaments in each slice. *Drawback:*
  - The existing approaches assume a flat-sky approximation.
  - The slicing process ignores the expansion of the Universe.

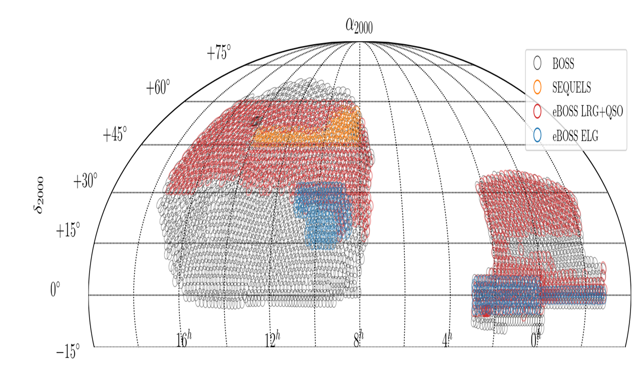


► **Indeed, our proposed method can easily switch between the above two categories!**

Methodology and Theory

Step 1 (Slicing the Universe): We partition the redshift range of observed galaxies into 325 non-overlapping thin slices based on the comoving distance $\Delta L = 25$ Mpc (i.e., tomographic analysis).

- **Caveat:** Any resulting redshift slice is not a flat 2D plane but a *spherical shell* that has a nonlinear curvature!



Step 2 (Restricting to the Region of Interest): Within each slice, we subset the galaxies in the regions:

- *North Galactic Cap*: $\{(\alpha, \delta) : 100° < \alpha < 270°, -5° \leq \delta < 70°\}$,
- *South Galactic Cap*: $\{(\alpha, \delta) : 300° < \alpha < 360°, -15° \leq \delta < 70°\} \cup \{(\alpha, \delta) : 0° \leq \alpha < 100°, -15° \leq \delta < 70°\}$.

Step 3 (Estimating the Galaxy Distribution): We convert the angular coordinates $(\alpha_i, \delta_i) \in [0, 360°) \times [-90°, 90°], i = 1, ..., n$ of galaxies in each slice to their Cartesian ones on $\mathbb{S}^2$ as:

$$\boldsymbol{X}_i = (\cos\delta_i \cos\alpha_i, \cos\delta_i \sin\alpha_i, \sin\delta_i) \quad \text{for} \quad i = 1, ..., n,$$

and estimate the galaxy density field on each slice by directional kernel density estimator (DirKDE):

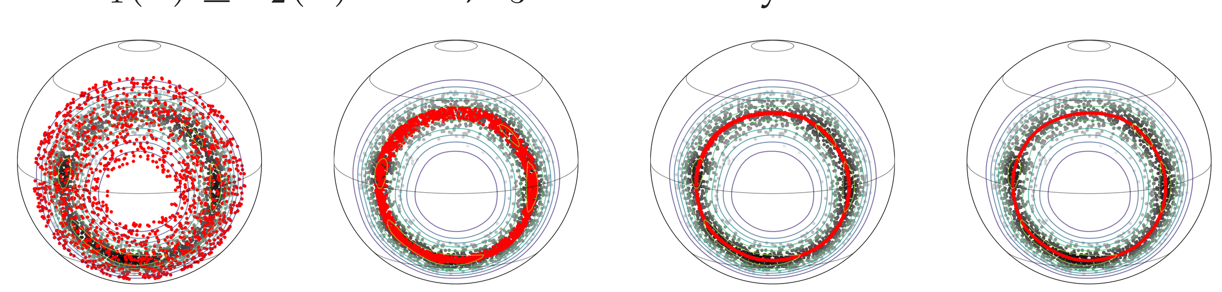
$$f(\boldsymbol{x}) = \frac{C(b)}{n} \sum_{i=1}^n \exp\left(\frac{\boldsymbol{x}^T \boldsymbol{X}_i}{b^2}\right), \tag{1}$$

where $b > 0$ is the smoothing bandwidth parameter and $C(b) > 0$ is a normalizing constant that depends on b.

Step 4 (Directional Density Ridge and SCONCE Algorithm): Given the (estimated) density field f defined in (1), we model the 1D cosmic filament on $\mathbb{S}^2$ by directional density ridge:

$$R_1(f) = \left\{\boldsymbol{x} \in \mathbb{S}^2 : \boldsymbol{v}_2(\boldsymbol{x})^T \mathtt{grad}\, f(\boldsymbol{x}) = \boldsymbol{0}, \lambda_2(\boldsymbol{x}) < 0\right\}, \tag{2}$$

where $\mathtt{grad}\, f(\boldsymbol{x}) = (\boldsymbol{I}_3 - \boldsymbol{x}\boldsymbol{x}^T)\nabla f(\boldsymbol{x}) \in \mathbb{R}^3$ is the Riemannian gradient of f at $\boldsymbol{x} \in \mathbb{S}^2$ and $\boldsymbol{v}_1(\boldsymbol{x}), \boldsymbol{v}_2(\boldsymbol{x})$ are unit eigenvectors of the Riemannian Hessian matrix $\mathcal{H}f(\boldsymbol{x}) = (\boldsymbol{I}_3 - \boldsymbol{x}\boldsymbol{x}^T)\left[\nabla\nabla f(\boldsymbol{x}) - \boldsymbol{x}^T \nabla f(\boldsymbol{x}) \boldsymbol{I}_3\right](\boldsymbol{I}_3 - \boldsymbol{x}\boldsymbol{x}^T) \in \mathbb{R}^{3\times 3}$ that lie within the tangent space $T_{\boldsymbol{x}}$ at $\boldsymbol{x} \in \mathbb{S}^2$ with associated eigenvalues $\lambda_1(\boldsymbol{x}) \geq \lambda_2(\boldsymbol{x})$. Here, $\boldsymbol{I}_3$ is an identity matrix in $\mathbb{R}^{3\times 3}$.



The panels from left to right present the DirSCMS algorithm applied to the mesh points (red dots) at iterations 0, 1, 2, and 8, where the contour lines indicate DirKDE (1) based on the input observations (gray dots).

To estimate the cosmic filaments (2), we employ Directional Subspace Constrained Mean Shift (DirSCMS) algorithm from our filament finder `SCONCE` with an iterative formula ($t = 0, 1, ...$):

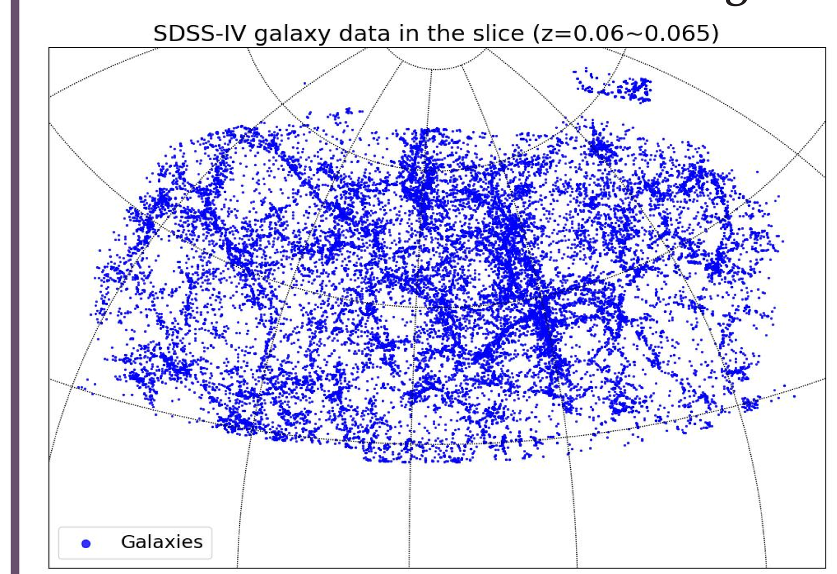
$$\boldsymbol{x}^{(t+1)} \leftarrow \boldsymbol{x}^{(t)} + \boldsymbol{v}_2(\boldsymbol{x}^{(t)})\boldsymbol{v}_2(\boldsymbol{x}^{(t)})^T \frac{\sum_{i=1}^n w_i \boldsymbol{X}_i \exp\left(\frac{\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{b^2}\right)}{\left\|\sum_{i=1}^n w_i \boldsymbol{X}_i \exp\left(\frac{\boldsymbol{X}_i^T \boldsymbol{x}^{(t)}}{b^2}\right)\right\|_2}.$$

Step 5 (Bootstrap Inference): We quantify the local uncertainty of each detected filamentary point via nonparametric bootstrap.

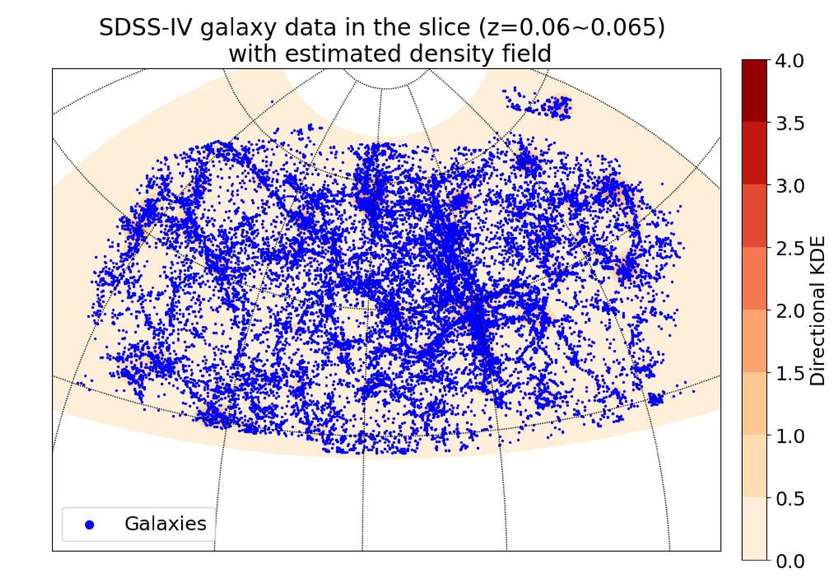
Step 6 (Estimating Local Modes and Filament Knots): We identify local modes of the (estimated) galaxy density field f via directional mean shift algorithm and intersection points of detected filaments via metric graph reconstruction algorithm.

Cosmic Web Catalog From SDSS Observed Galaxies

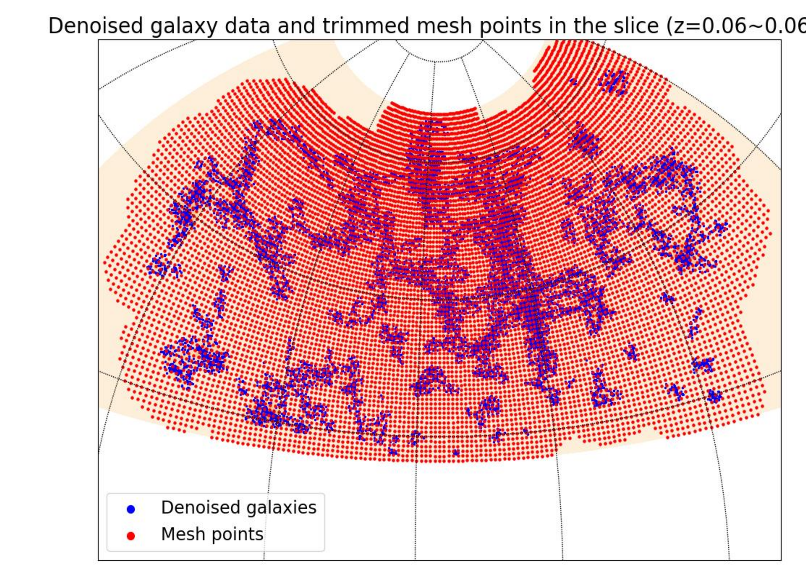
Detailed Illustration of Constructing Our Cosmic Web Catalog[a] on SDSS-IV Galaxy Data:



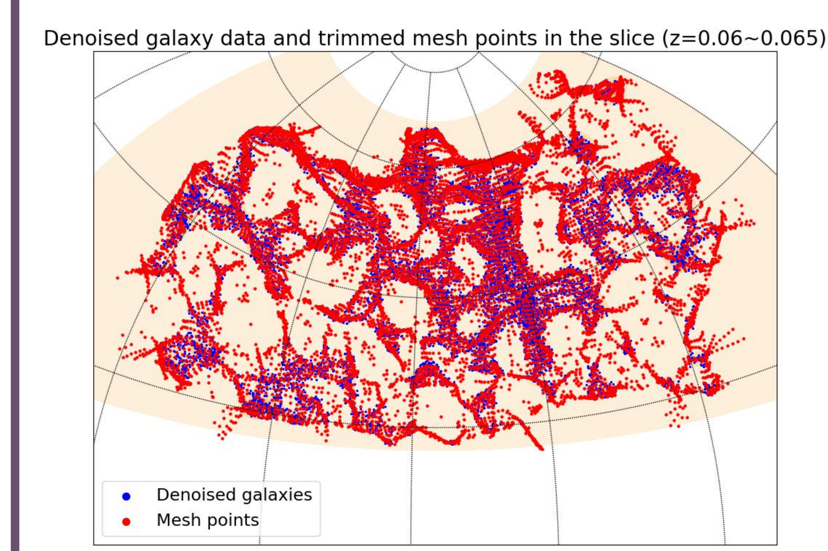
Steps 1 and 2: Slice the Universe and Restrict to the North Galactic Cap.



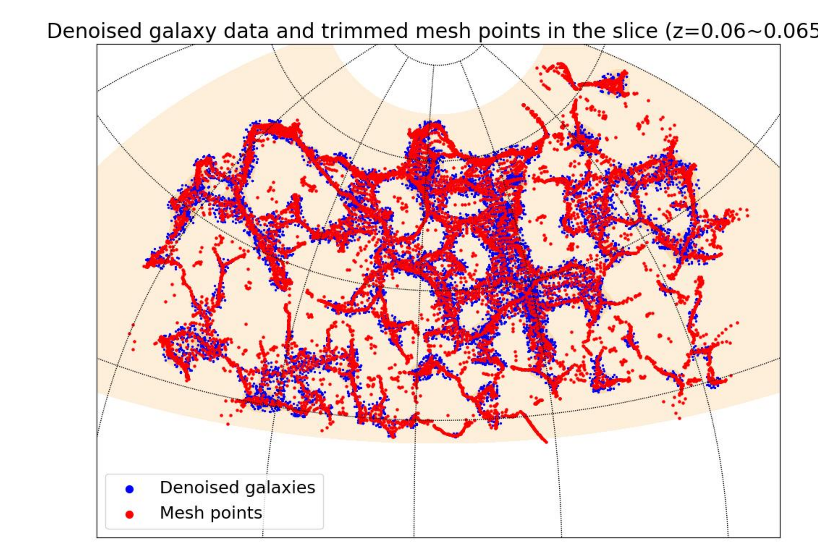
Step 3: Estimate the Galaxy Density Field.



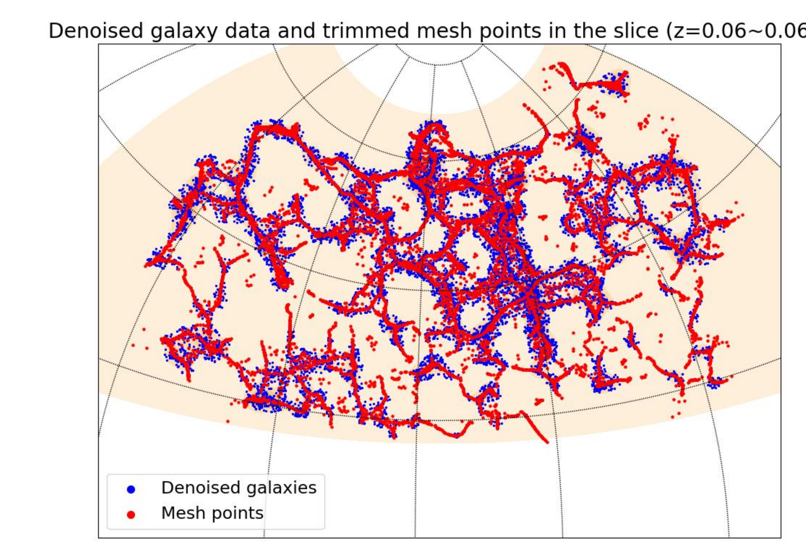
Step 4-1: Denoise the Estimated Density Field and Lay Down Mesh Points.



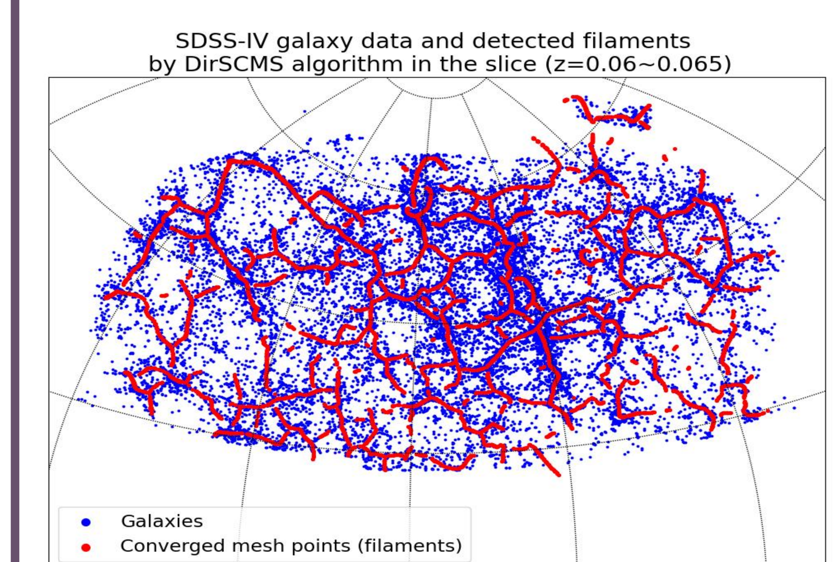
Step 4-2: Apply Our DirSCMS Algorithm (Iteration 1).



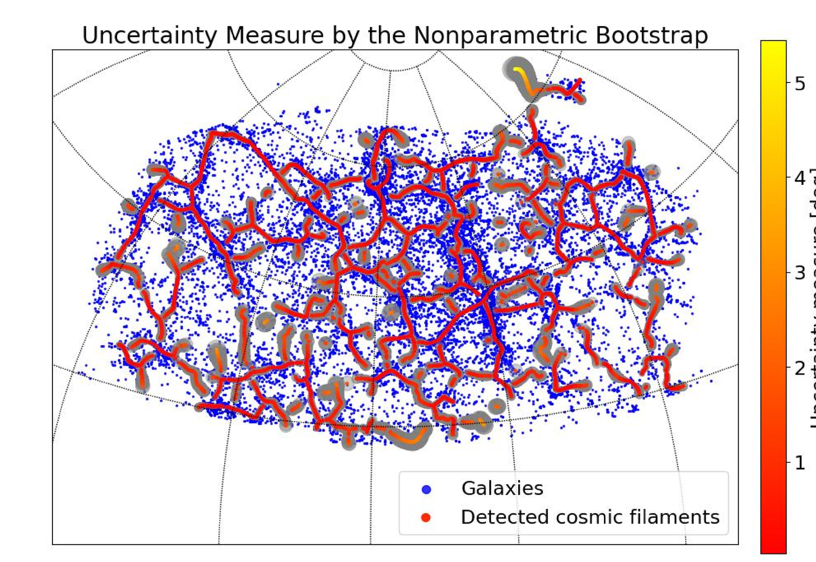
Step 4-2: Apply Our DirSCMS Algorithm (Iteration 2).



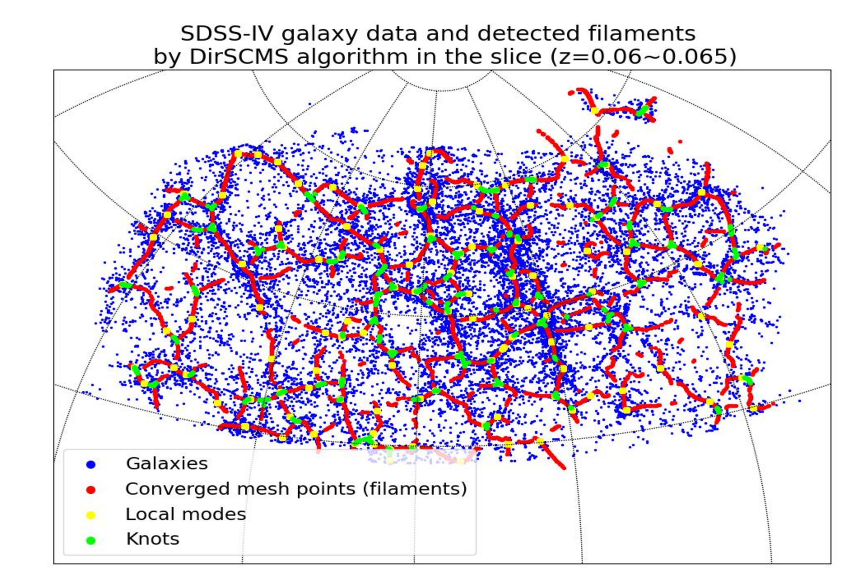
Step 4-2: Apply Our DirSCMS Algorithm (Iteration 3).



Step 4-2: Apply Our DirSCMS Algorithm (Final Convergence).



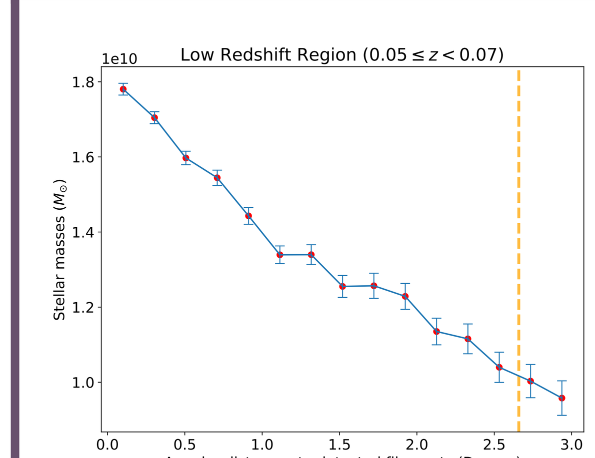
Step 5: Quantify Filament Uncertainties via Nonparametric Bootstrap.

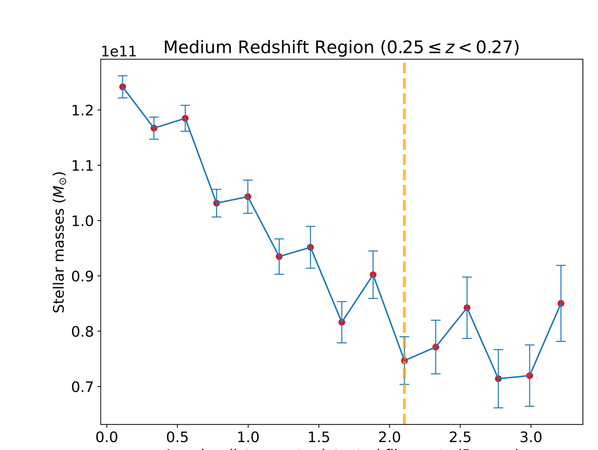


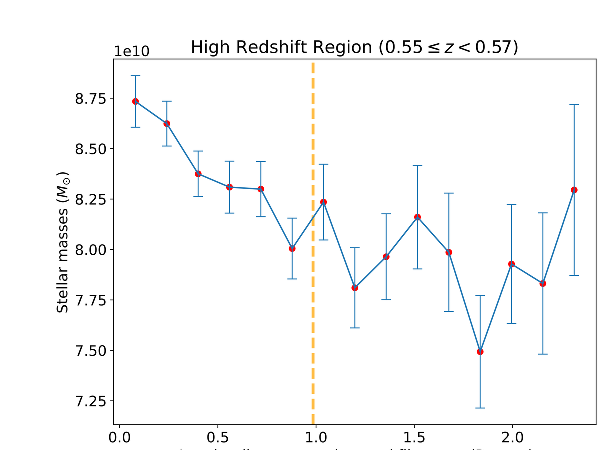
Step 6: Estimate Local Modes and Filament Knots.

[a]The full catalog data can be downloaded at `https://doi.org/10.5281/zenodo.6244866`.

Effects of Cosmic Filaments on Galaxy Properties

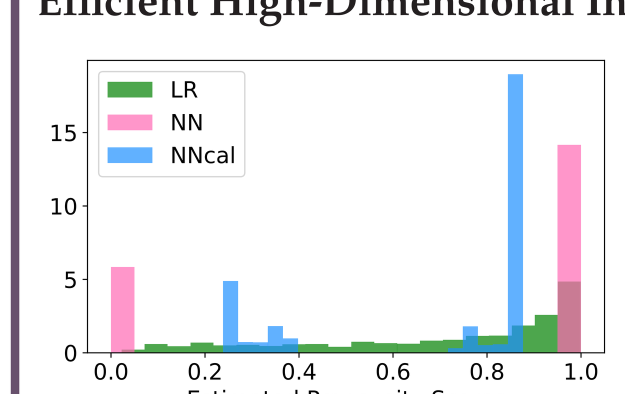


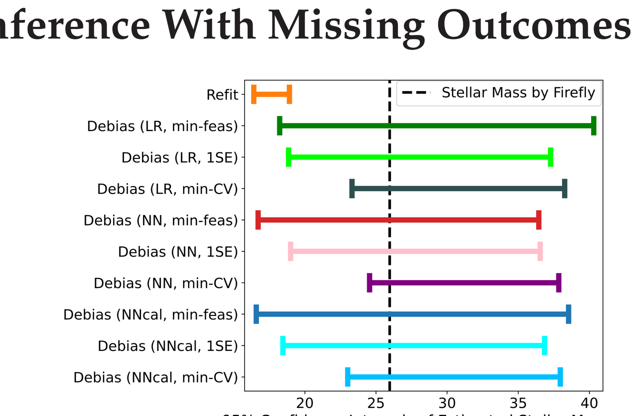


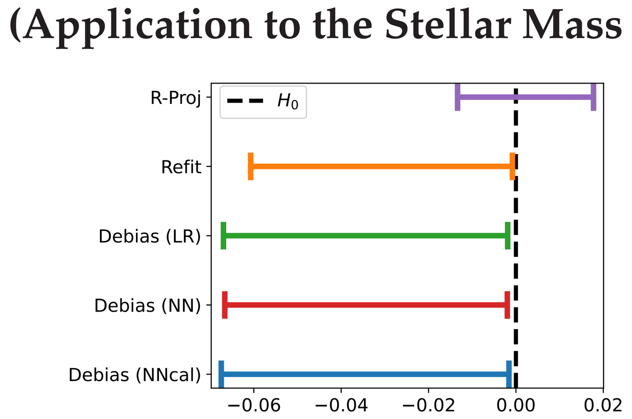


The three figures on the left study the correlations between stellar masses of galaxies in different redshift regions and their distances to our detected filaments. *It demonstrates that galaxies near filaments tend to be more massive than those that are further away from filaments.*

Efficient High-Dimensional Inference With Missing Outcomes (Application to the Stellar Mass Inference of SDSS-IV Galaxies):







Middle Panel: Inference on the stellar mass of a new galaxy based on its spectroscopic and photometric properties.
Right Panel: Testing the negative correlation between the stellar mass and its distance to nearby cosmic filaments.