# 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

February 12, 2015

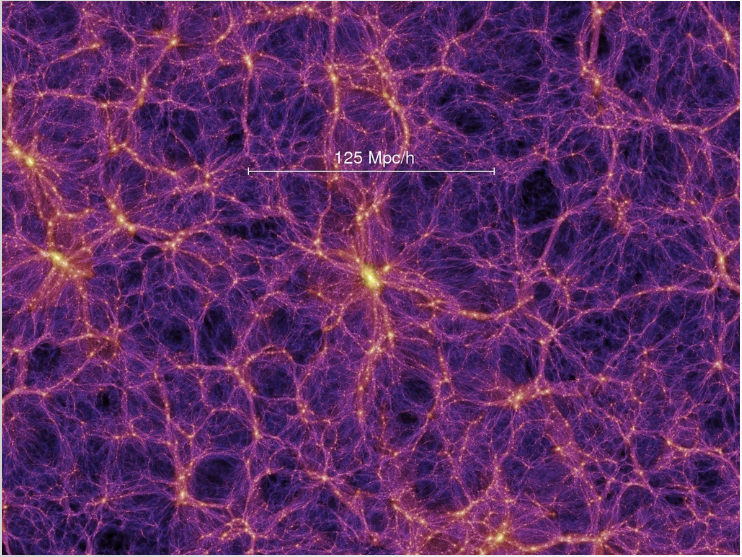
# Outline

- Introduction to Cosmic Web
- Statistical Model and Algorithm
- Filament Coverage and Uncertainty Measures
- Scientific Applications
- Summary

# Outline

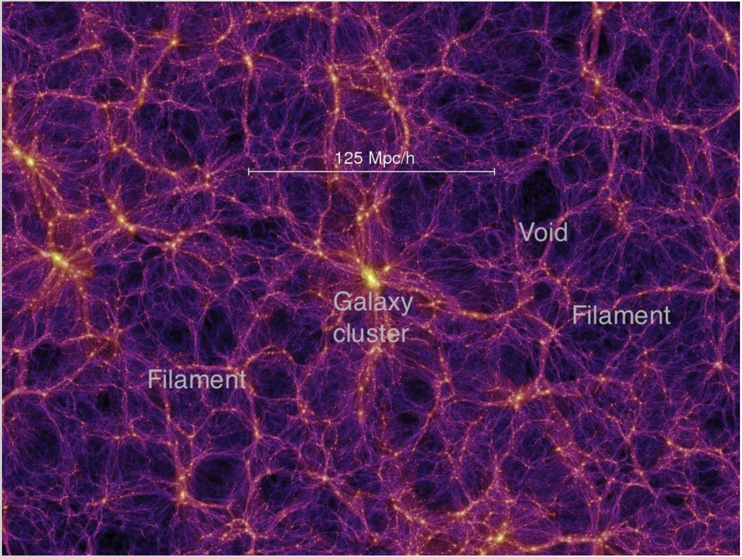
- Introduction to Cosmic Web
- Statistical Model and Algorithm
- Filament Coverage and Uncertainty Measures
- Scientific Applications
- Summary

# Cosmic Web: What Does Our Universe Look Like




Credit: Millennium Simulation

# Cosmic Web: What Does Our Universe Look Like




Credit: Millennium Simulation

# Focus of the Research: Filaments

Why filament?

- Galaxies tend to concentrate around filaments.

# Focus of the Research: Filaments

Why filament?

- Galaxies tend to concentrate around filaments.
- Brightness of galaxies is influenced by filaments.

# Focus of the Research: Filaments

Why filament?

- Galaxies tend to concentrate around filaments.
- Brightness of galaxies is influenced by filaments.
- Shape of galaxies is correlated with filaments.

# Outline

- Introduction to Cosmic Web
- Statistical Model and Algorithm
- Filament Coverage and Uncertainty Measures
- Scientific Applications
- Summary

# A Glance at our Universe

(Loading)

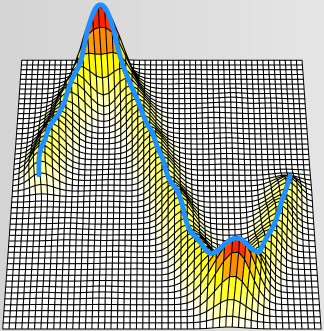
# Statistical Model for Filaments: Density Ridges

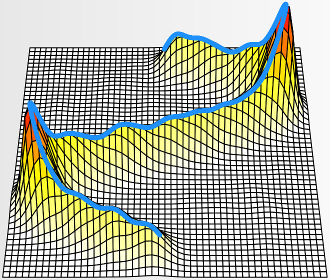
Formally, we define a filament to be a **ridge** of the density.

# Example: Ridges in Mountians

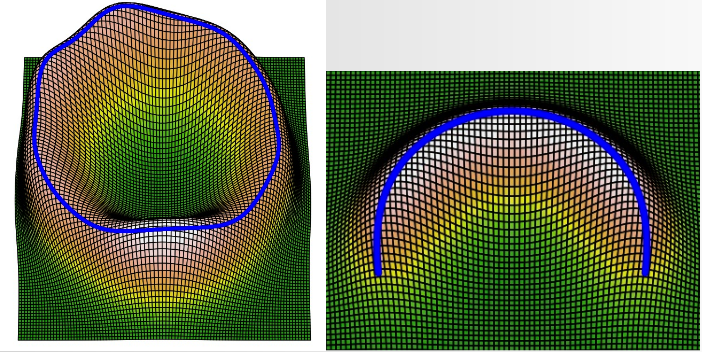

Credit: Google

# Example: Ridges in Smooth Functions

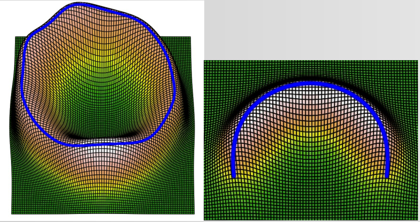




# Example: Ridges in Smooth Functions

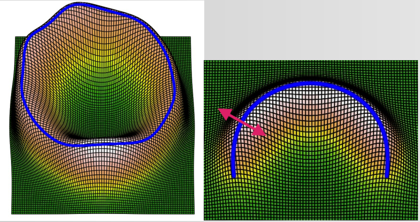


# Ridges: Local Modes in Subspace



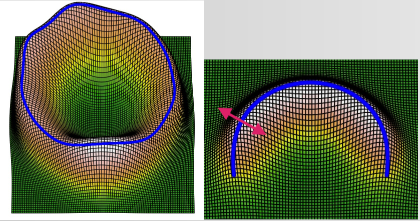
- A generalized local mode in a specific ‘subspace’.

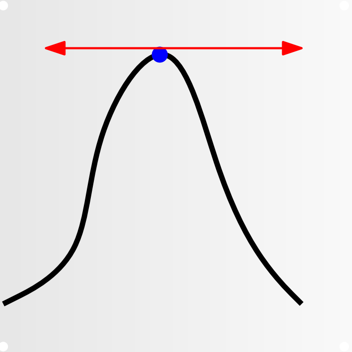
# Ridges: Local Modes in Subspace



- 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

- $p(x)$: a density function.

# Formal Definition of Density Ridges

- $p(x)$: a density function.
- $(\lambda_j(x), v_j(x))$: $j$th eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.

# Formal Definition of Density Ridges

- $p(x)$: a density function.
- $(\lambda_j(x), v_j(x))$: $j$th eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of 2nd to last eigenvectors

# Formal Definition of Density Ridges

- $p(x)$: a density function.
- $(\lambda_j(x), v_j(x))$: $j$th eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of 2nd to last eigenvectors
- $V(x)V(x)^T$: a projection

# Formal Definition of Density Ridges

- $p(x)$: a density function.
- $(\lambda_j(x), v_j(x))$: $j$th eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of 2nd to last eigenvectors
- $V(x)V(x)^T$: a projection
- Ridges:

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

# Formal Definition of Density Ridges

- $p(x)$: a density function.
- $(\lambda_j(x), v_j(x))$: $j$th eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of 2nd to last eigenvectors
- $V(x)V(x)^T$: a projection
- Ridges:

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

- Local modes:

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

# Formal Definition of Density Ridges

- $p(x)$: a density function.
- $(\lambda_j(x), v_j(x))$: $j$th eigenvalue/vector of $H(x) = \nabla\nabla p(x)$.
- $V(x) = [v_2(x), \cdots, v_d(x)]$: matrix of 2nd to last eigenvectors
- $V(x)V(x)^T$: a projection
- Ridges:

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

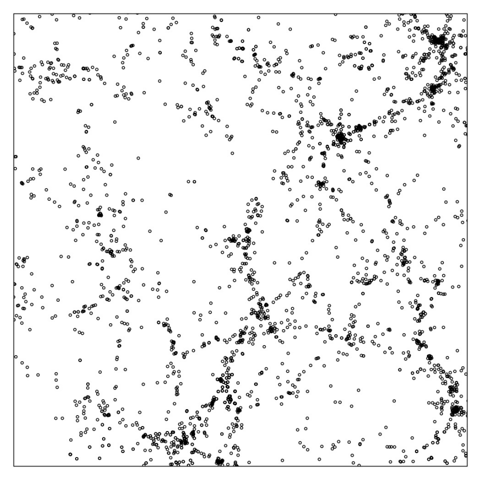
- Local modes:

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

- In practice, we estimate $p$ by the kernel density estimator $\widehat{p}_n$.

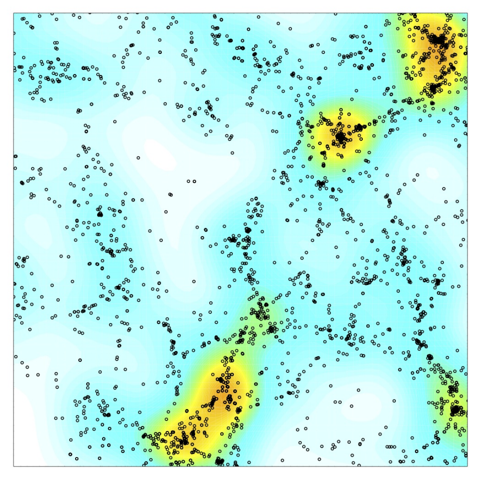
# Algorithm

1. Rawdata



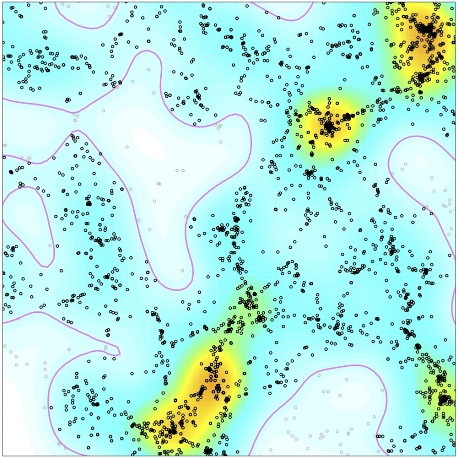
# Algorithm

1. Rawdata
2. Density Reconstruction



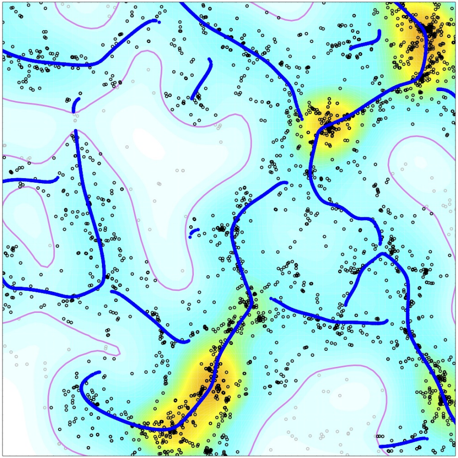
# Algorithm

1. Rawdata
2. Density Reconstruction
3. Thresholding



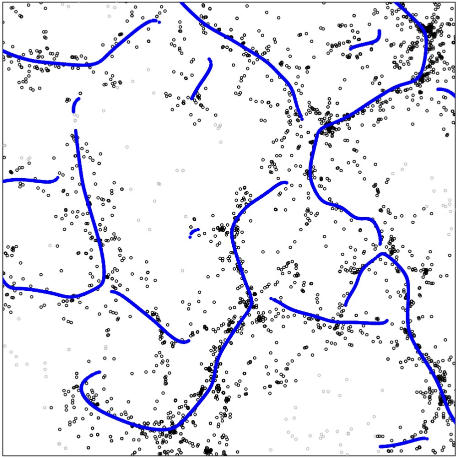
# Algorithm

1. Rawdata
2. Density Reconstruction
3. Thresholding
4. Ridge Recovery (Ozertem and Erdogmus 2011)

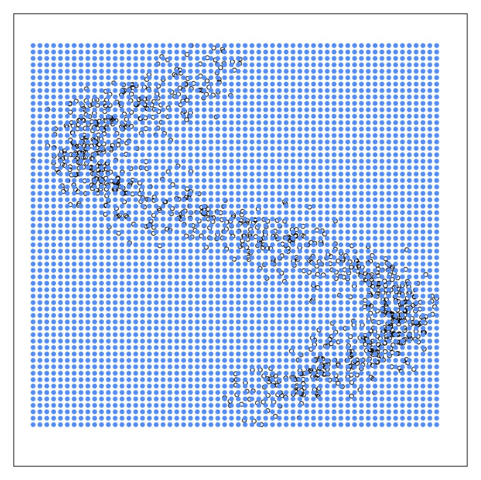


# Algorithm

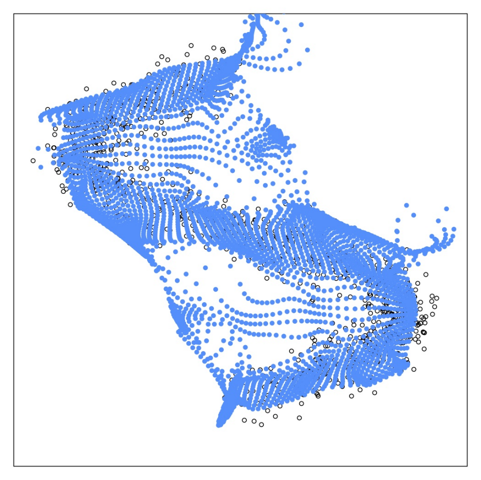
1. Rawdata
2. Density Reconstruction
3. Thresholding
4. Ridge Recovery (Ozertem and Erdogmus 2011)



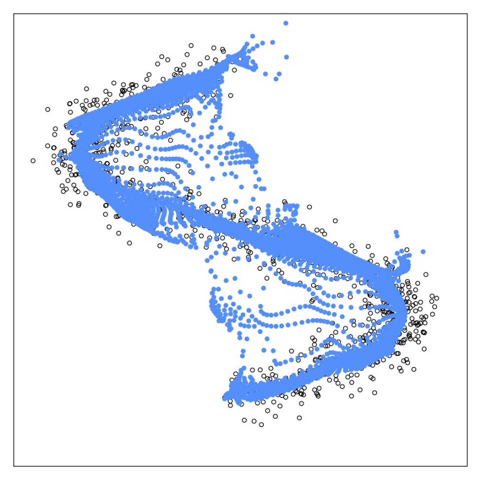
# SCMS: Ridge Recovery Algorithm



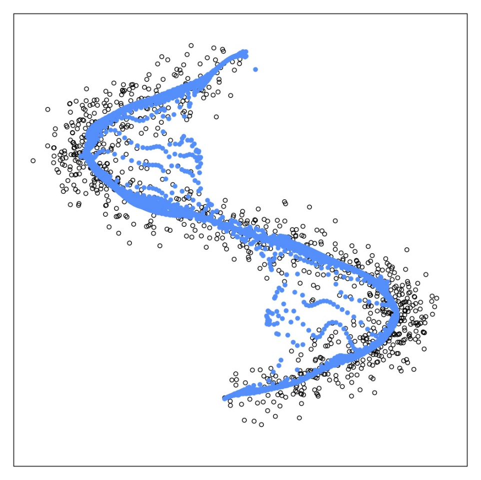
# SCMS: Ridge Recovery Algorithm



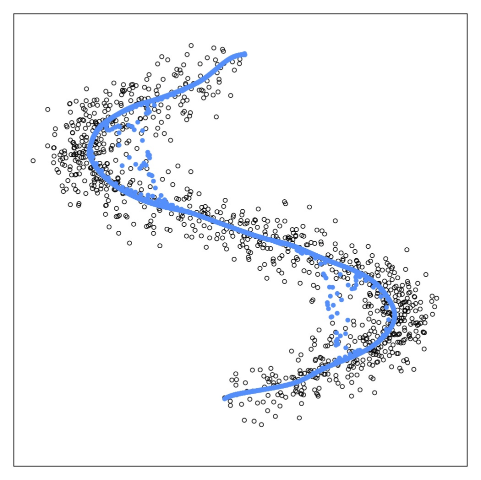
# SCMS: Ridge Recovery Algorithm



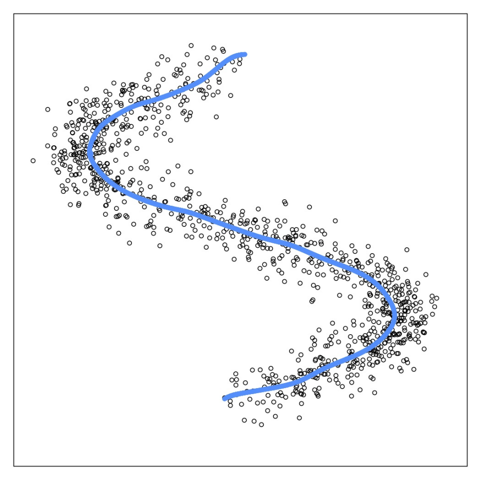
# SCMS: Ridge Recovery Algorithm



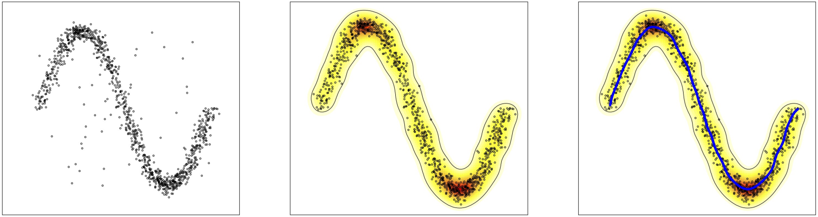
# SCMS: Ridge Recovery Algorithm



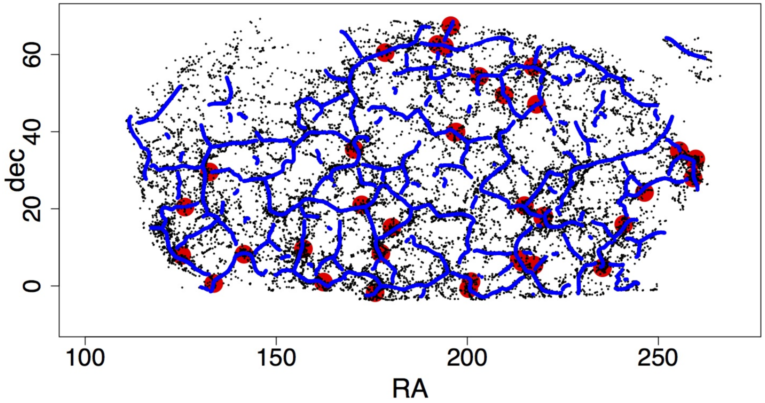
# SCMS: Ridge Recovery Algorithm



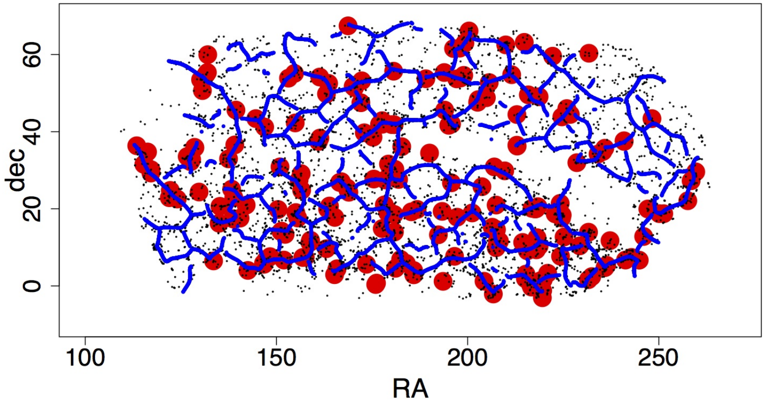
# Summary for the Algorithm



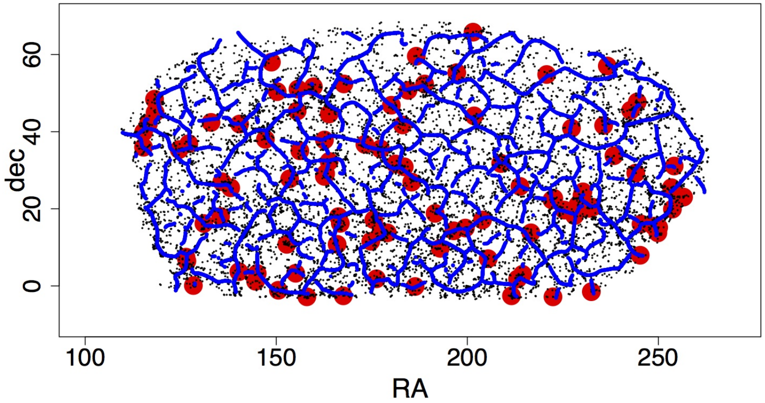
# Density Ridges on the SDSS data



# Density Ridges on the SDSS data



# Density Ridges on the SDSS data



# Outline

- Introduction to Cosmic Web
- Statistical Model and Algorithm
- Filament Coverage and Uncertainty Measures
- Scientific Applications
- Summary

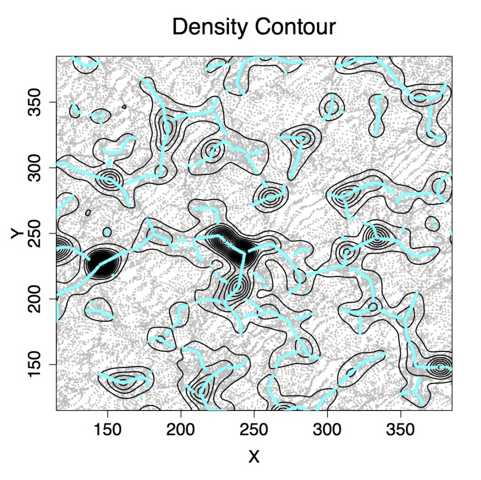
## Simulation: Consistency for Density Ridges

- To evaluate the quality of our method, we use the N-body simulation.

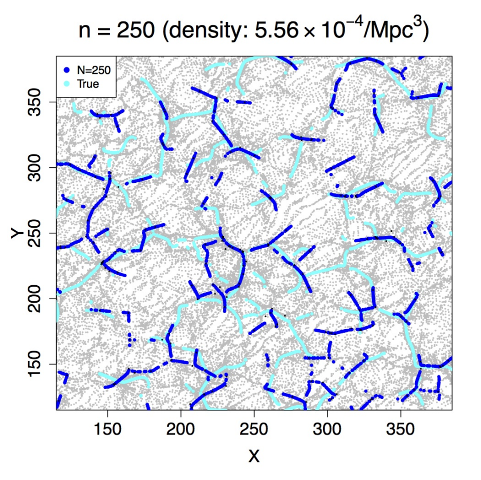
# Simulation: Consistency for Density Ridges

- To evaluate the quality of our method, we use the N-body simulation.
- We define ‘true’ filaments as applying our method to ‘all’ galaxies in the simulation.
- We subsample part of the galaxies from the simulation.

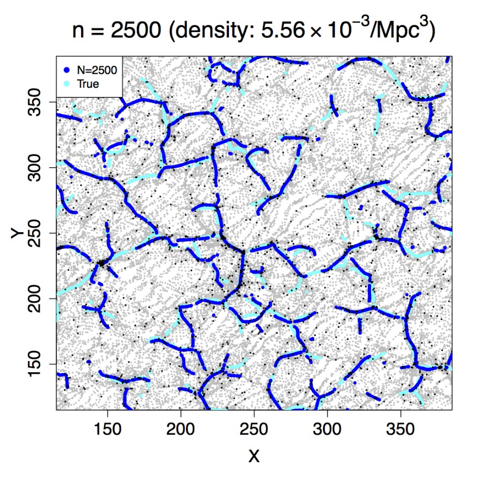
# Simulation: Consistency



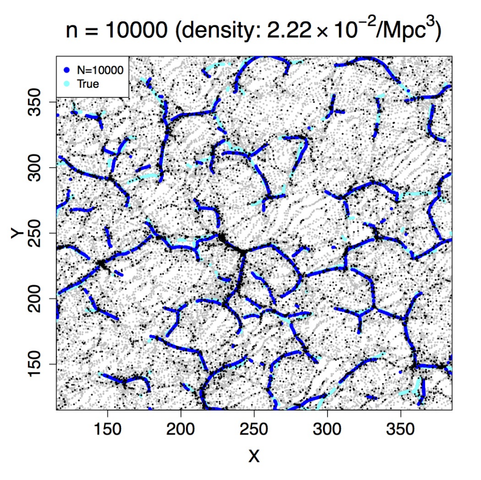
# Simulation: Consistency



# Simulation: Consistency



# Simulation: Consistency



# Filament Coverage

- Can we quantify the process of convergence?

# Filament Coverage

- Can we quantify the process of convergence?
- *Filament coverage* provides a simple way to attain this.

# Filament Coverage

- Can we quantify the process of convergence?
- *Filament coverage* provides a simple way to attain this.
- **True positive coverage**:

$$TP(r) = \frac{\text{length}(R \cap \widehat{R}_n \oplus r)}{\text{length}(R)}.$$

- **False positive coverage**:

$$FP(r) = 1 - \frac{\text{length}(\widehat{R}_n \cap R \oplus r)}{\text{length}(\widehat{R}_n)}.$$

- $R$ and $\widehat{R}_n$ are the 'true' filaments and estimated filaments.

# Illustration: Filament Coverage

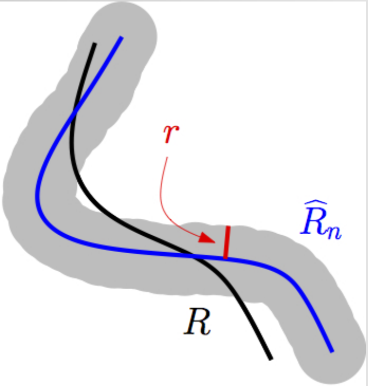



Figure: TP(r)

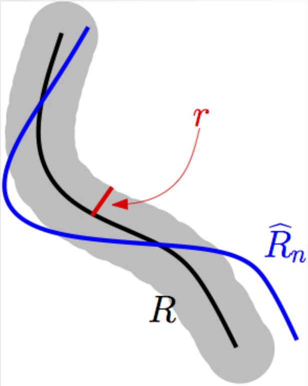
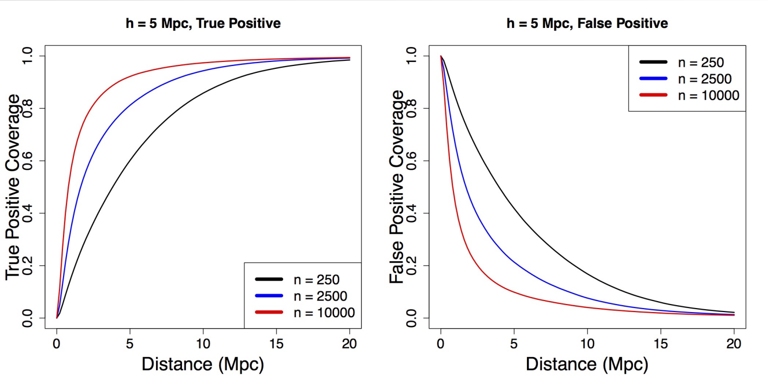



Figure: 1 - FP(r)

# Filament Coverage



# The Need for Uncertainty Measure

- Filament coverage gives a (global) evaluation for filaments.
- We have no idea about the local uncertainty along filaments.
- Moreover, filament coverage requires the knowledge of truth.

# Uncertianty Measures

Let $R$ and $\widehat{R}_n$ be the true filaments and the estimated filaments. For each $x \in R$, we define the **(local) uncertainty measure** as

$$\rho_n^2(x) = \mathbb{E}\left(d^2(x, \widehat{R}_n)\right),$$

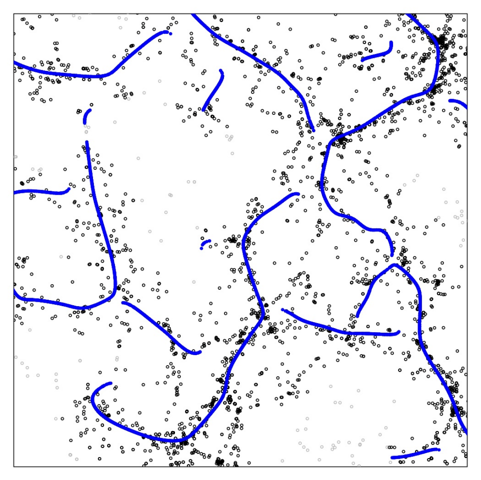
where $d(x, A)$ is the projection distance from point $x$ to a set $A$.
*Remark*:

- This is analogous to the mean square error.

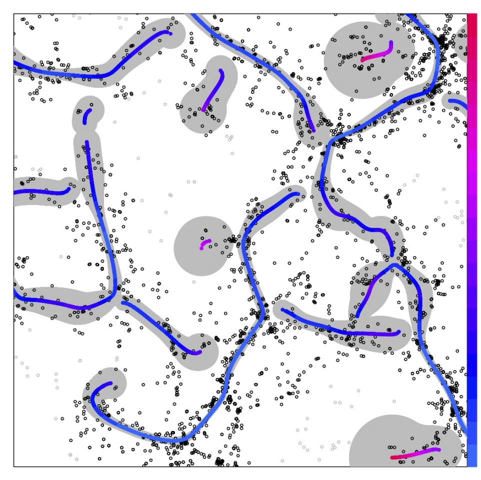
# Estimating Uncertainty Measures

We apply the local uncertainty measure to our estimated filaments and use the *bootstrap* to evaluate the errors.

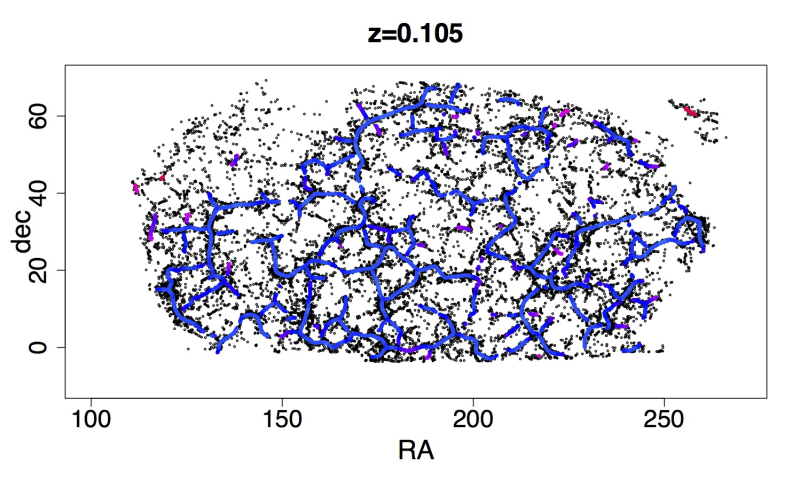
# Real Data Evaluation



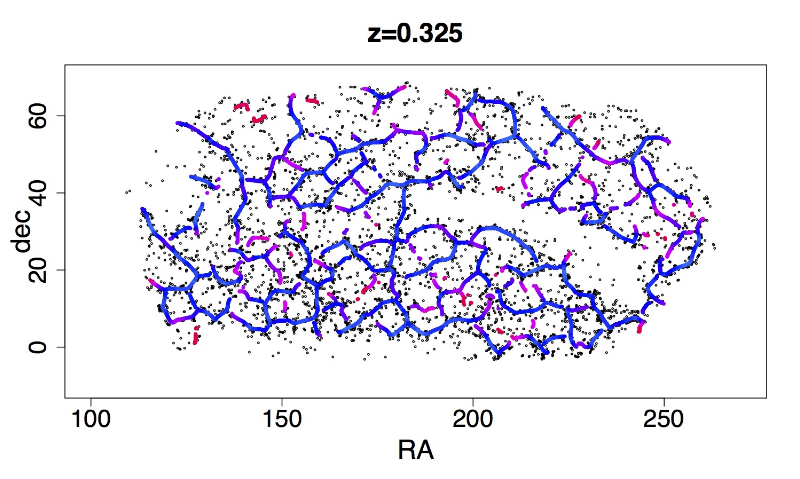
# Real Data Evaluation



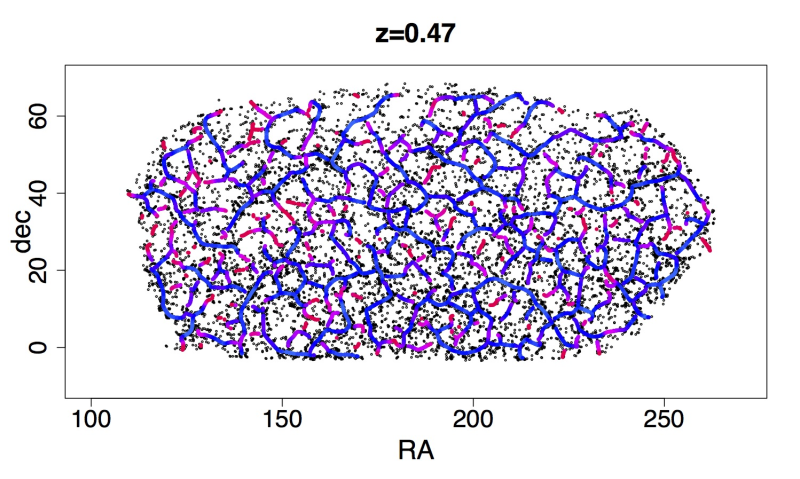
# Density Ridges on the SDSS data



# Density Ridges on the SDSS data



# Density Ridges on the SDSS data



# Outline

- Introduction to Cosmic Web
- Statistical Model and Algorithm
- Filament Coverage and Uncertainty Measures
- Scientific Applications
- Summary

# Filaments and Galaxy Luminosity

- Goal: We want to see if galaxies close to filament are brighter than those away from filaments.

# Filaments and Galaxy Luminosity

- Goal: We want to see if galaxies close to filament are brighter than those away from filaments.
- For each galaxy, we have two variables.

# Filaments and Galaxy Luminosity

- Goal: We want to see if galaxies close to filament are brighter than those away from filaments.
- For each galaxy, we have two variables.
- Variable 1: distance to filaments.

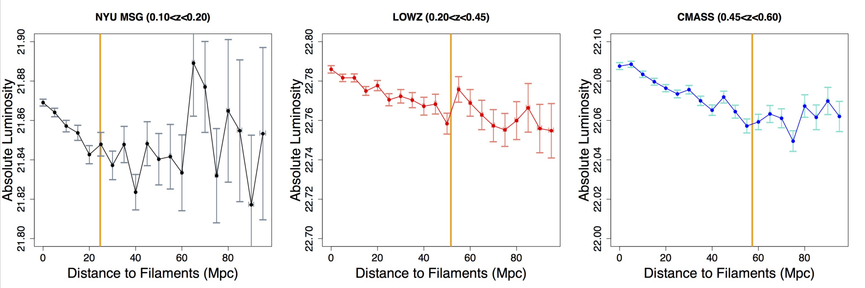
# Filaments and Galaxy Luminosity

- Goal: We want to see if galaxies close to filament are brighter than those away from filaments.
- For each galaxy, we have two variables.
- Variable 1: distance to filaments.
- Variable 2: brightness–absolute luminosity ($-1\times$absolute magnitude).

# Filaments and Galaxy Luminosity

- Goal: We want to see if galaxies close to filament are brighter than those away from filaments.
- For each galaxy, we have two variables.
- Variable 1: distance to filaments.
- Variable 2: brightness–absolute luminosity ($-1\times$absolute magnitude).
- We analyze three datasets (at different ranges of redshifts).

# Filaments and Galaxy Luminosity



# Filaments and Galaxy Intrinsic Alignment

- Goal: We want to see if the orientation of a galaxy is influenced by the orientation of nearby filaments.

# Filaments and Galaxy Intrinsic Alignment

- Goal: We want to see if the orientation of a galaxy is influenced by the orientation of nearby filaments.
- For each galaxy, we have three variables.

# Filaments and Galaxy Intrinsic Alignment

- Goal: We want to see if the orientation of a galaxy is influenced by the orientation of nearby filaments.
- For each galaxy, we have three variables.
- Variable 1, $\mu_F$: the nearby filament orientation (vector).

# Filaments and Galaxy Intrinsic Alignment

- Goal: We want to see if the orientation of a galaxy is influenced by the orientation of nearby filaments.
- For each galaxy, we have three variables.
- Variable 1, $\mu_F$: the nearby filament orientation (vector).
- Variable 2, $\mu_j$: $j$-th principal axes for the galaxy (vector).

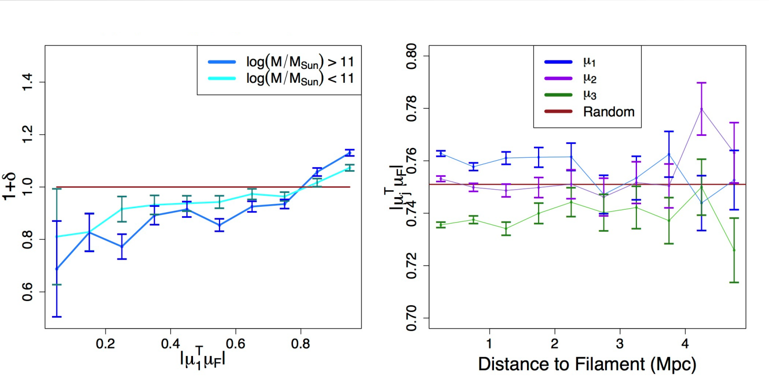
# Filaments and Galaxy Intrinsic Alignment

- Goal: We want to see if the orientation of a galaxy is influenced by the orientation of nearby filaments.
- For each galaxy, we have three variables.
- Variable 1, $\mu_F$: the nearby filament orientation (vector).
- Variable 2, $\mu_j$: $j$-th principal axes for the galaxy (vector).
- Variable 3: distance to filaments.

# Filaments and Galaxy Intrinsic Alignment

- Goal: We want to see if the orientation of a galaxy is influenced by the orientation of nearby filaments.
- For each galaxy, we have three variables.
- Variable 1, $\mu_F$: the nearby filament orientation (vector).
- Variable 2, $\mu_j$: $j$-th principal axes for the galaxy (vector).
- Variable 3: distance to filaments.
- We analyze the massive blackhole dataset (a simulation dataset).

# Filaments and Galaxy Intrinsic Alignment

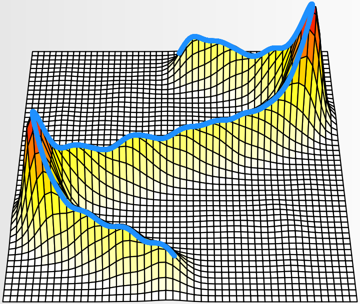


# Outline

- Introduction to Cosmic Web
- Statistical Model and Algorithm
- Filament Coverage and Uncertainty Measures
- Scientific Applications
- Summary

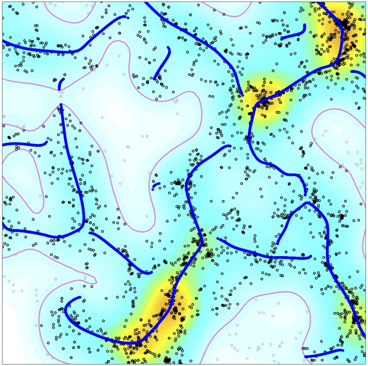
# Summary

1. Model: density ridges.



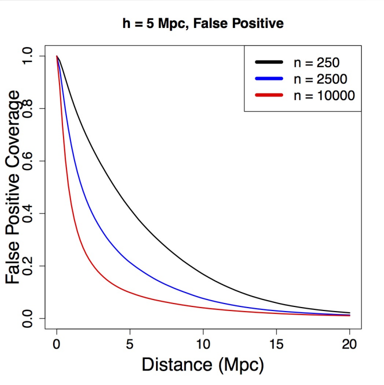
# Summary

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



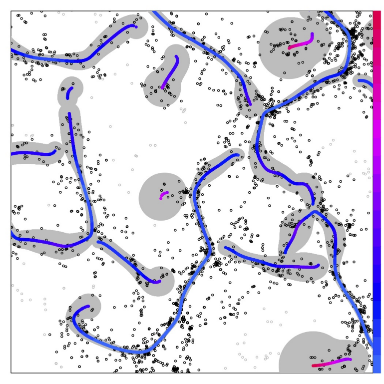
# Summary

1. Model: density ridges.
2. Algorithm: SCMS.
3. Consistency: filament coverage.



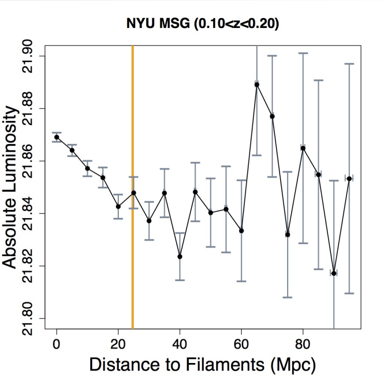
# Summary

1. Model: density ridges.
2. Algorithm: SCMS.
3. Consistency: filament coverage.
4. Errors: uncertainty measures.



# Summary

1. Model: density ridges.
2. Algorithm: SCMS.
3. Consistency: filament coverage.
4. Errors: uncertainty measures.
5. Application: galaxy luminosity, alignment.



Thank you!

# reference


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 preprint arXiv:1501.05303 (2015).
2. Chen, Yen-Chi, Christopher R. Genovese, and Larry Wasserman. "Asymptotic theory for density ridges." arXiv preprint arXiv:1406.5663 (2014).
3. Chen, Yen-Chi, Christopher R. Genovese, and Larry Wasserman. "Generalized mode and ridge estimation." arXiv preprint arXiv:1406.1803 (2014).
4. Eberly, David. Ridges in image and data analysis. Vol. 7. Springer Science & Business Media, 1996.
5. Genovese, Christopher R., et al. "Nonparametric ridge estimation." The Annals of Statistics 42.4 (2014): 1511-1545.
6. Ozertem, Umut, and Deniz Erdogmus. "Locally defined principal curves and surfaces." The Journal of Machine Learning Research 12 (2011): 1249-1286.