Cosmic web analysis in the SDSS main galaxy sample and implications for galaxy colors

Florent Leclercq

Institute of Cosmology and Gravitation, University of Portsmouth

http://icg.port.ac.uk/~leclercq/

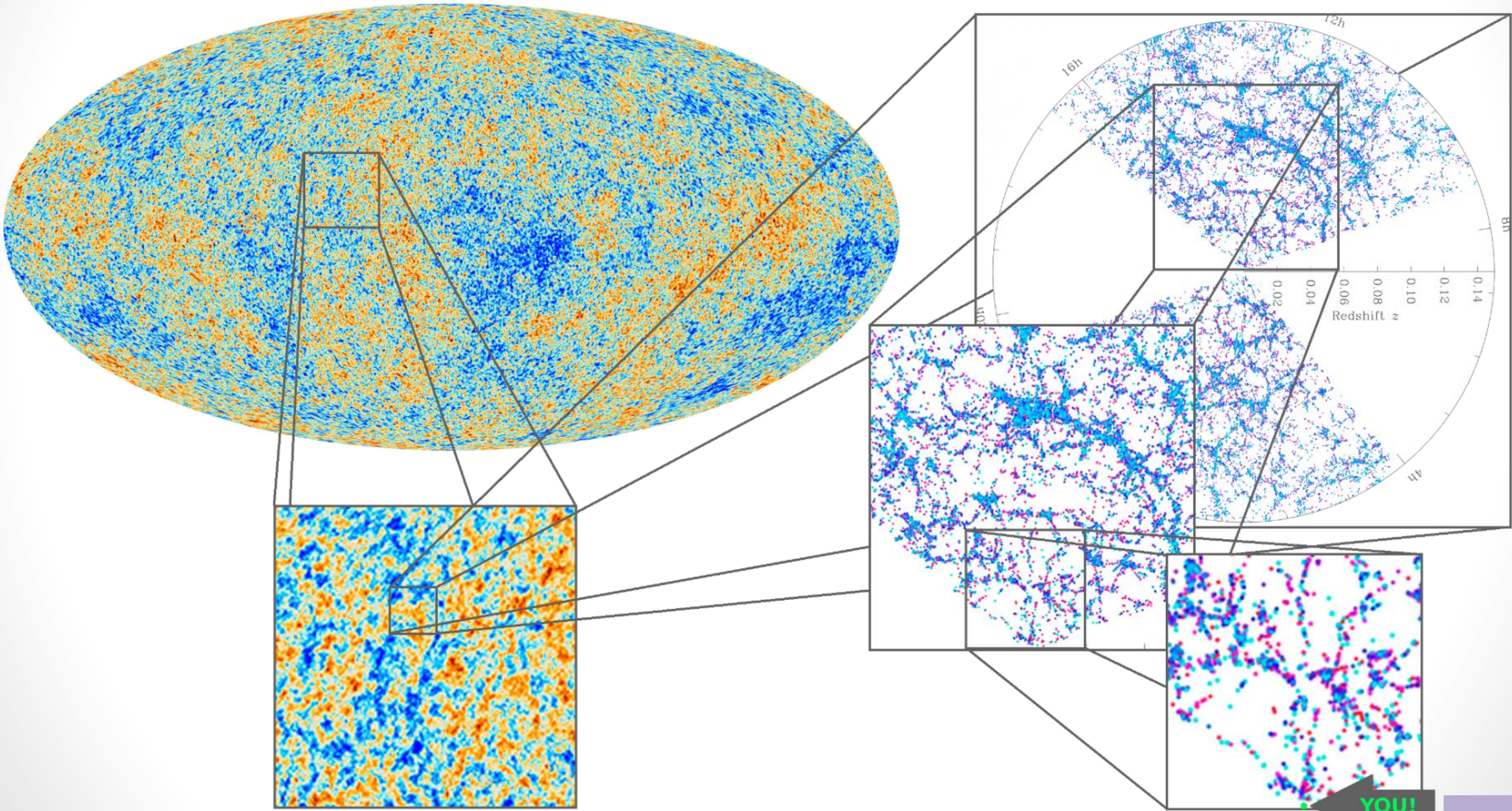
January 26th, 2017

In collaboration with:

Jens Jasche (ExC Universe, Garching), Guilhem Lavaux (IAP),
Will Percival (ICG), Benjamin Wandelt (IAP/U. Illinois)

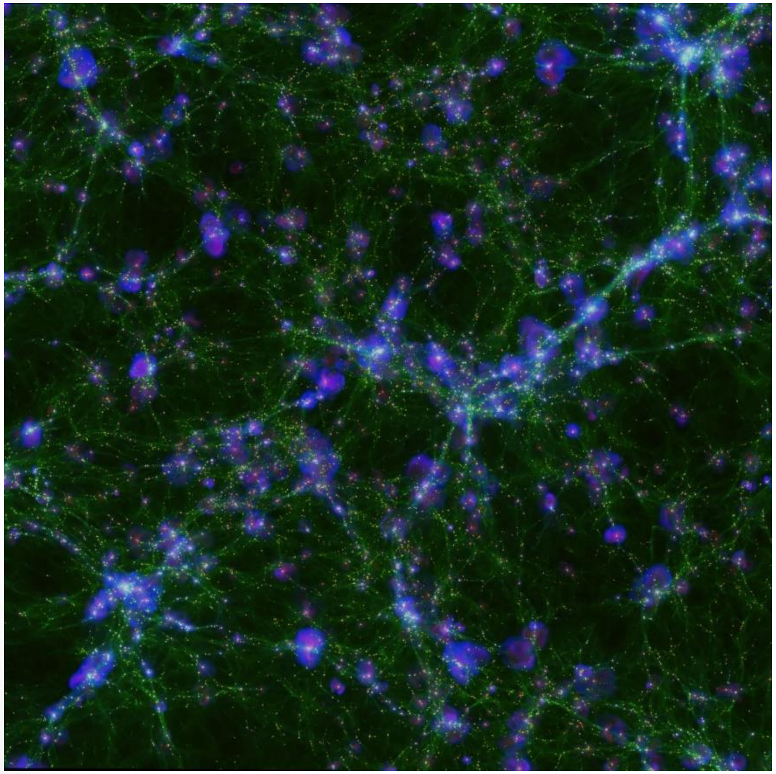
The big picture: the Universe is highly structured

You are here. Make the best of it…



Planck collaboration (2013-2015) M. Blanton and the Sloan Digital Sky Survey (2010-2013)

What we want to know from the LSS

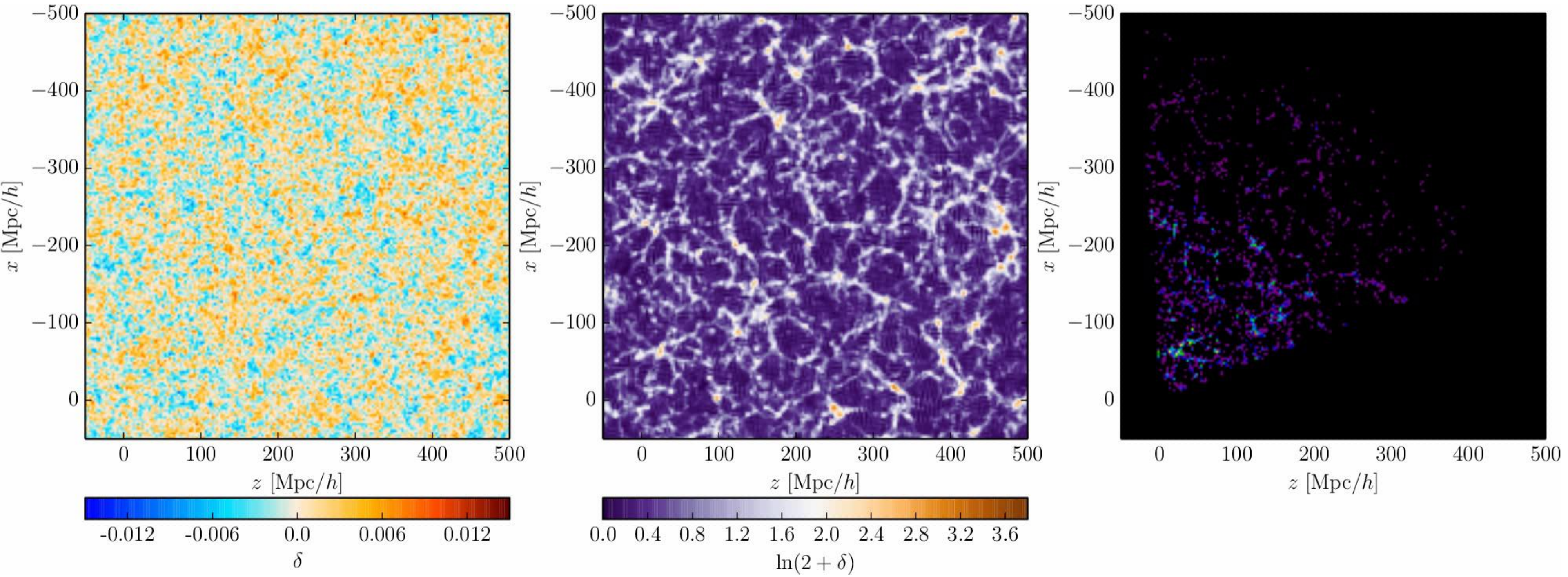


Y. Dubois (PI), Horizon AGN simulation (2014-2016)

The LSS is a vast source of knowledge:

- **Cosmology**:
  - Cosmological parameters and tests of ΛCDM,
  - Physical nature of the dark components,
  - Geometry of the Universe,
  - Tests of General Relativity,
  - Initial conditions and link to high energy physics
- **Astrophysics**: galaxy formation and evolution as a function of their environment
  - Galaxy properties (colors, chemical composition, shapes),
  - Intrinsic alignments

The BORG SDSS run

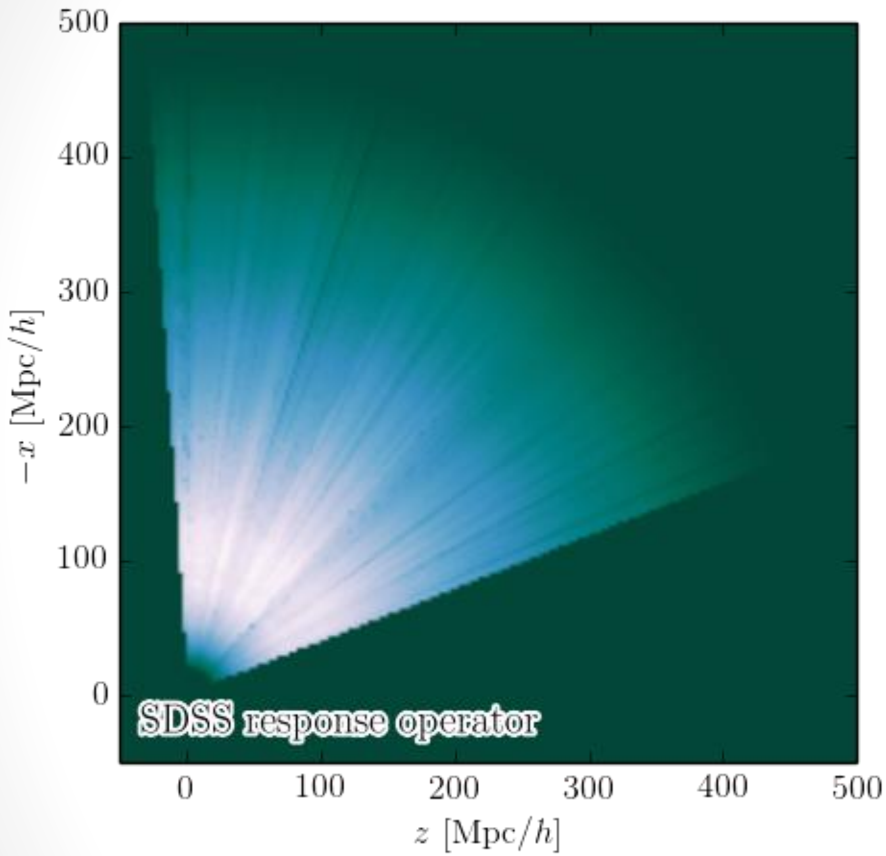


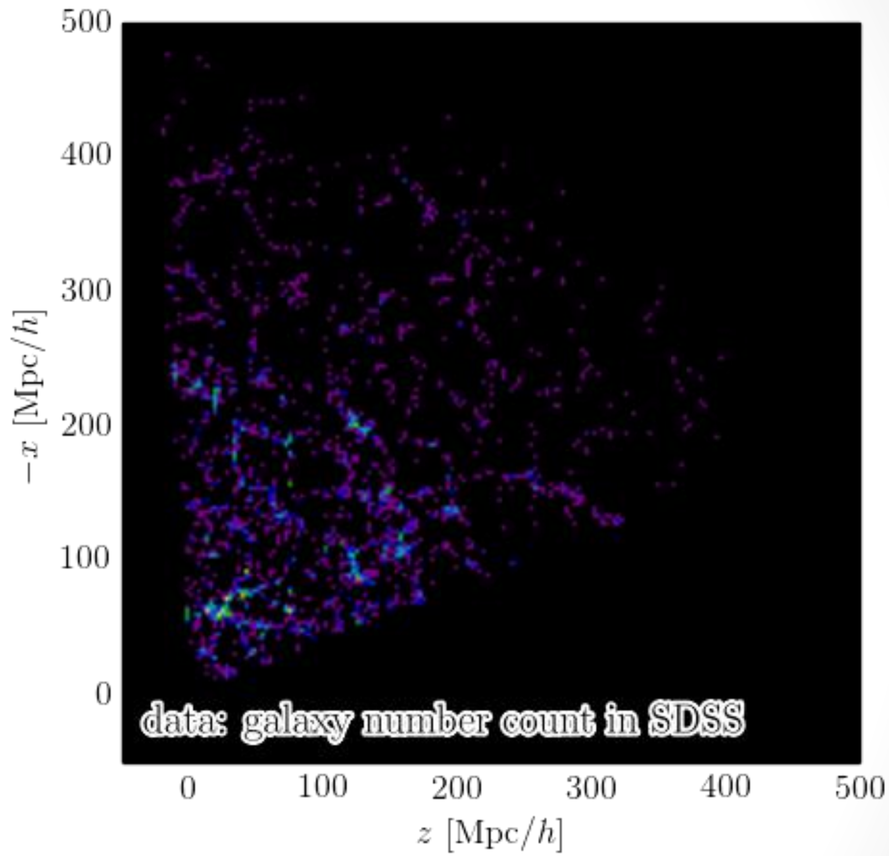
Initial conditions | Final conditions | Observations

334,074 galaxies, ≈ 17 millions parameters, 3 TB of primary data products, 12,000 samples, ≈ 250,000 data model evaluations, 10 months on 32 cores

Jasche, FL & Wandelt 2015, arXiv:1409.6308

Bayesian chrono-cosmography from SDSS DR7

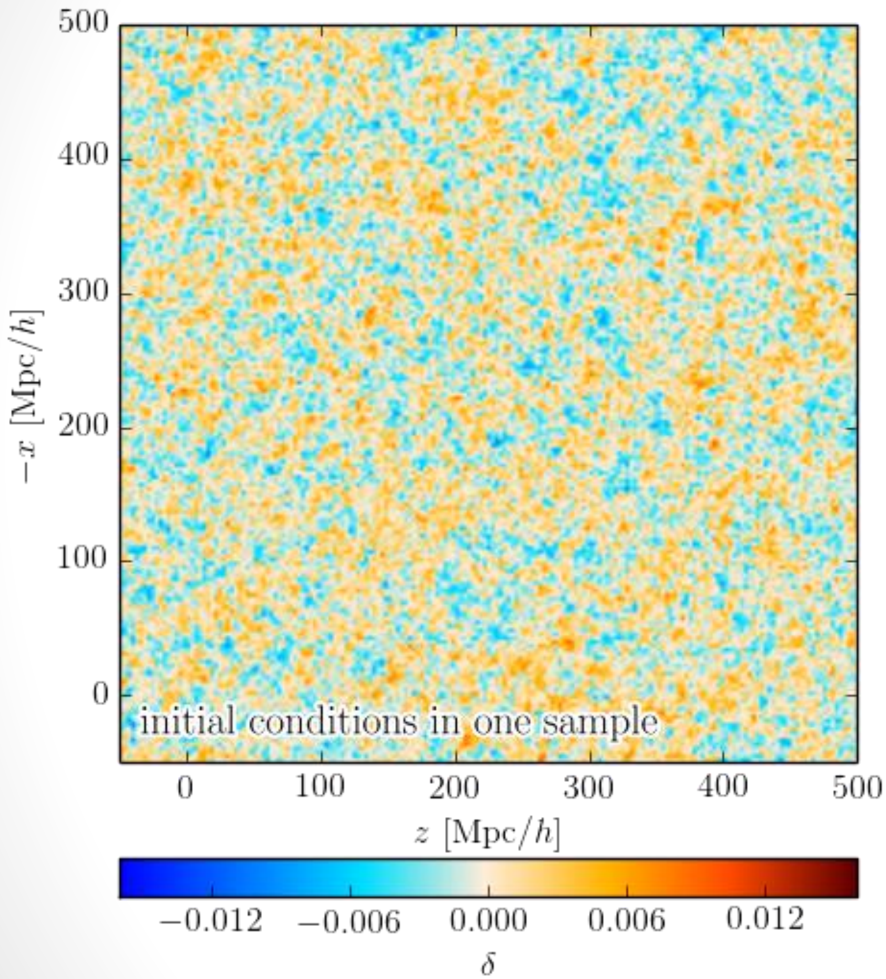


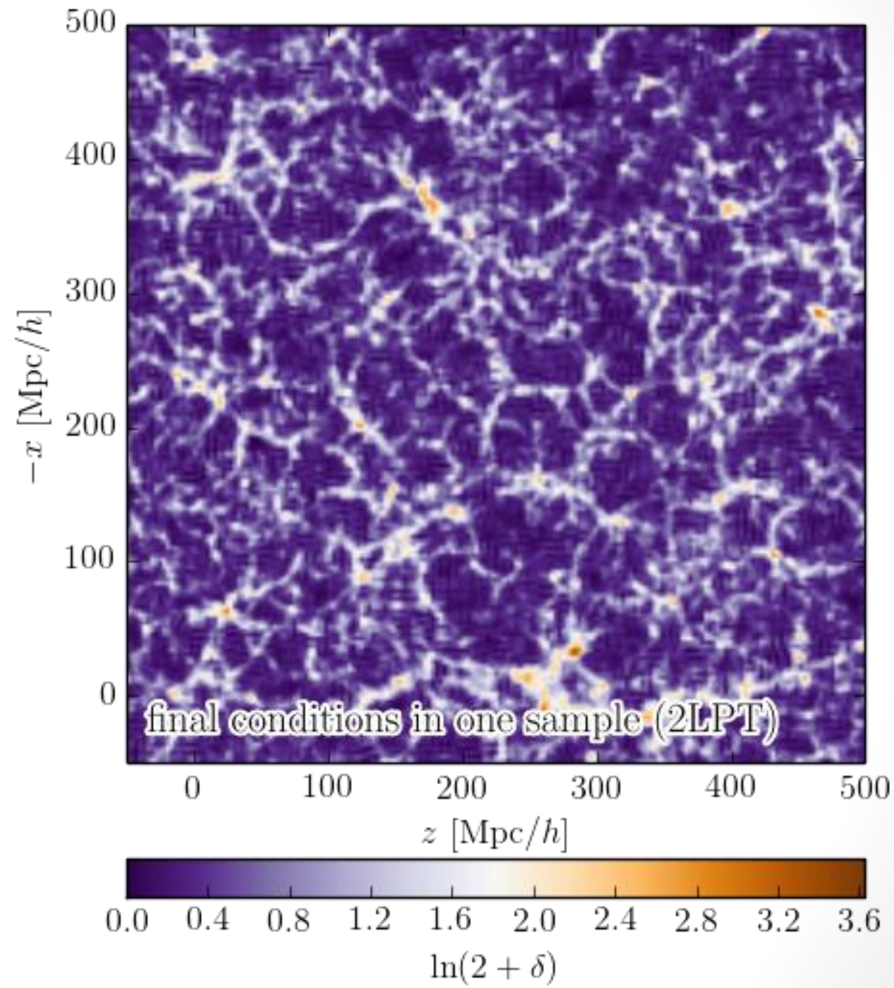


Data

Jasche, FL & Wandelt 2015, arXiv:1409.6308

Bayesian chrono-cosmography from SDSS DR7

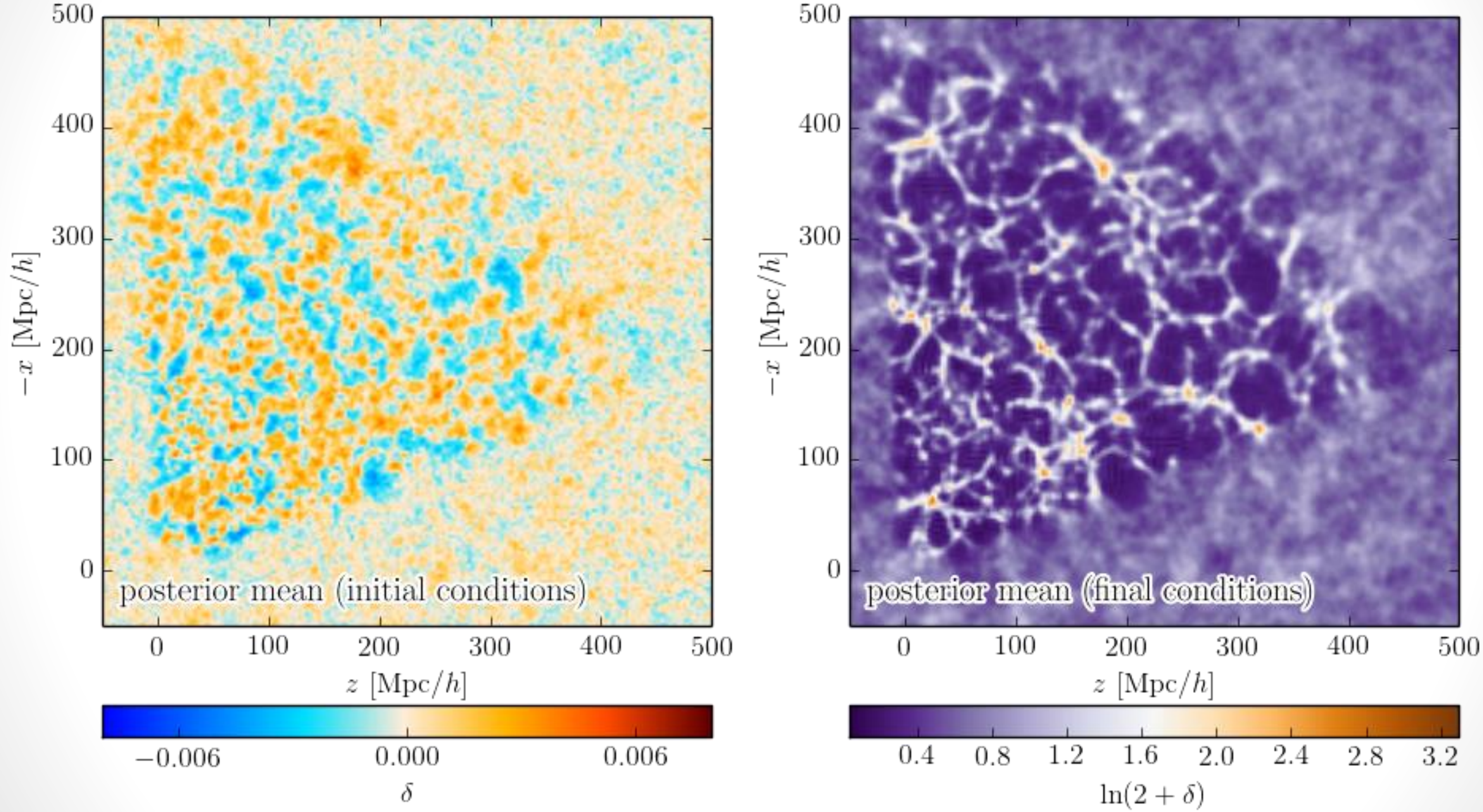




One sample

Jasche, FL & Wandelt 2015, arXiv:1409.6308

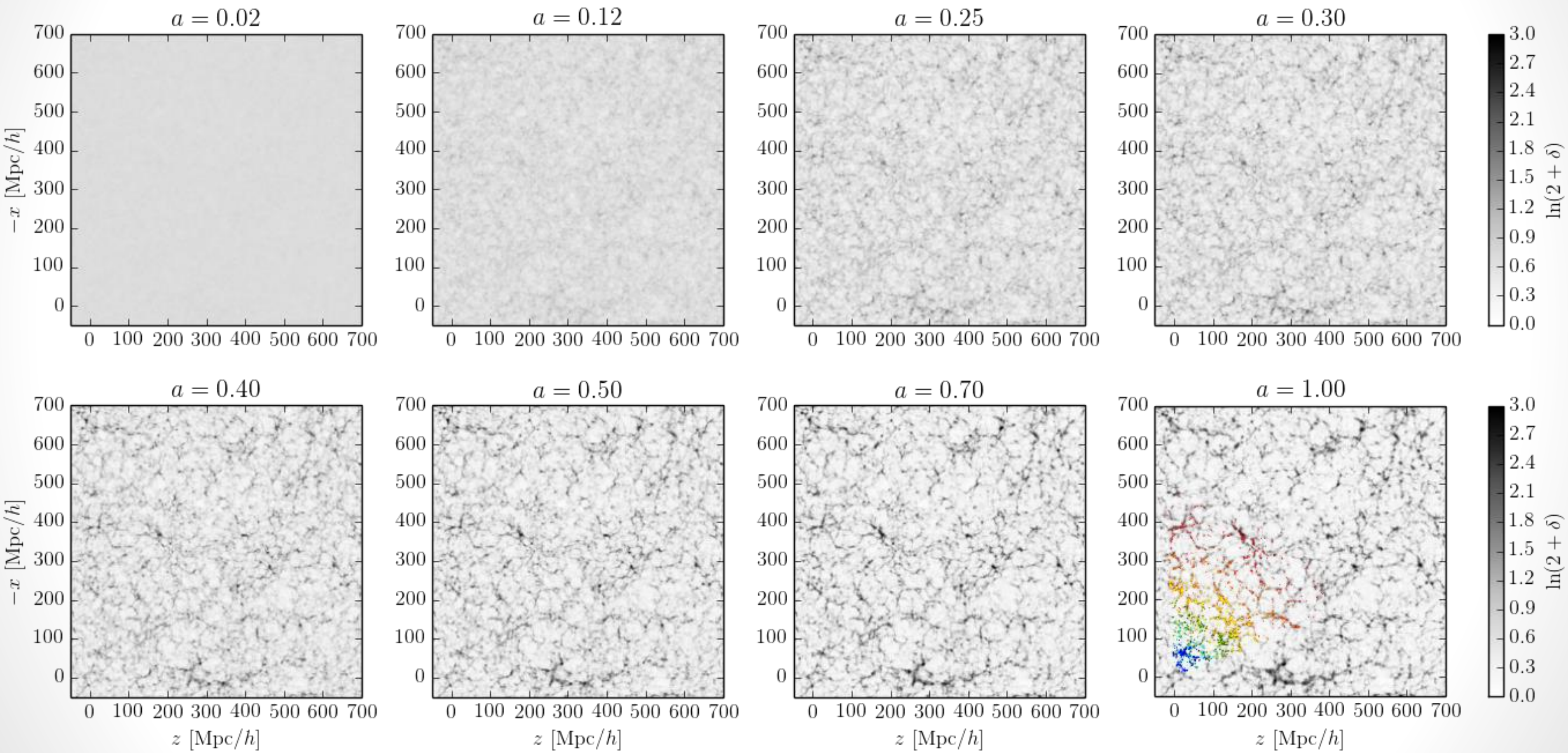
Bayesian chrono-cosmography from SDSS DR7



Posterior mean

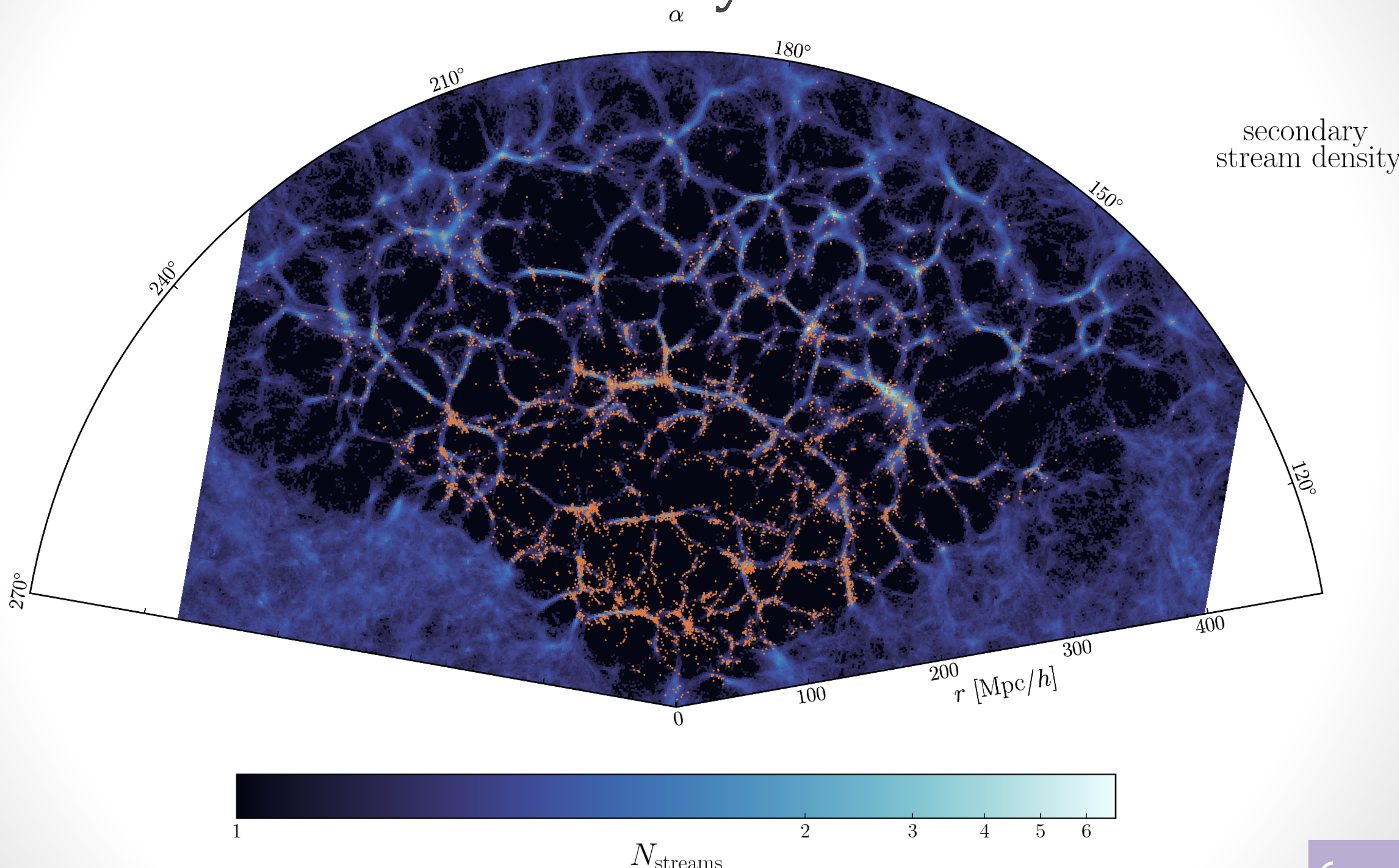
Jasche, FL & Wandelt 2015, arXiv:1409.6308

Evolution of cosmic structure



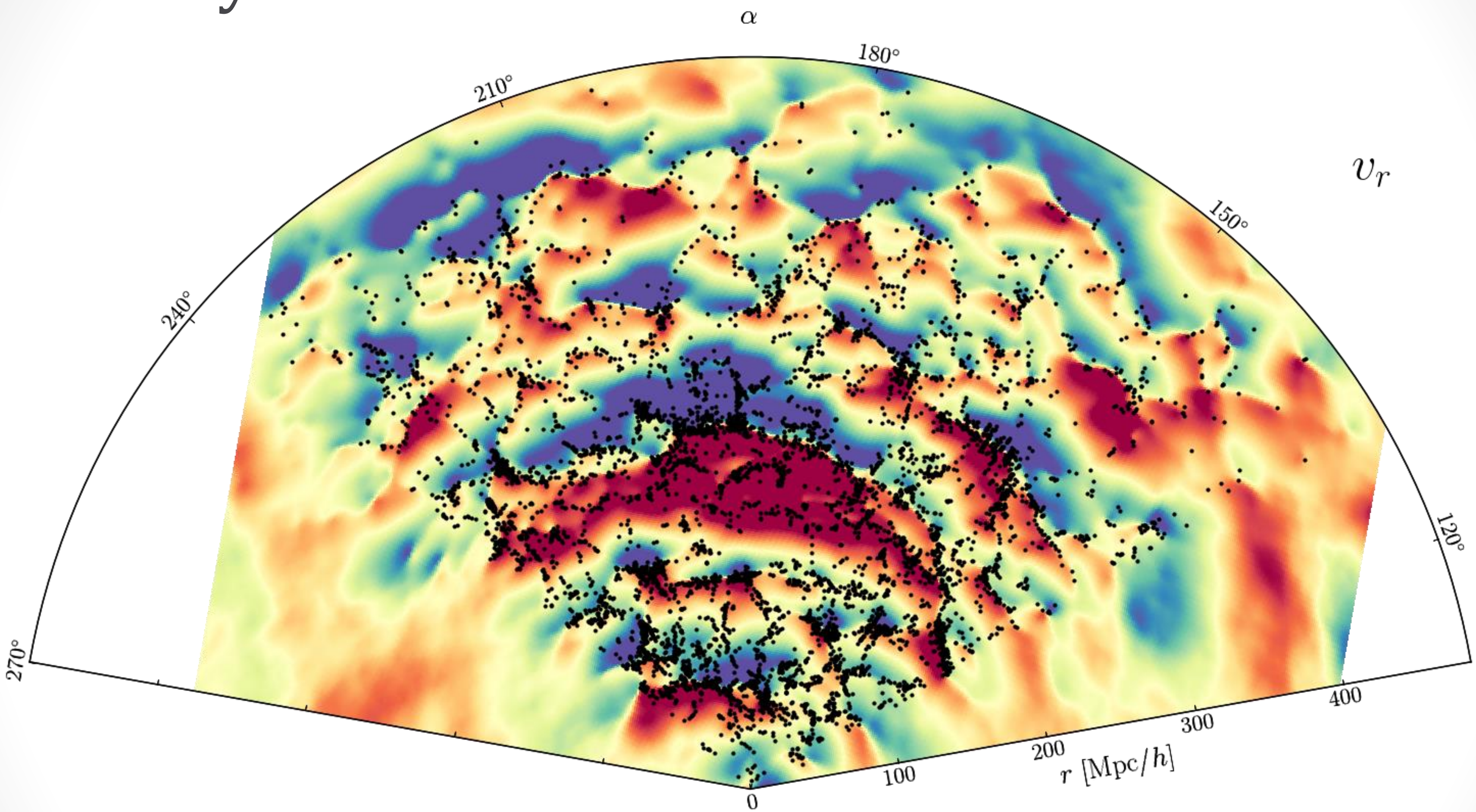
Jasche, FL & Wandelt 2015, arXiv:1409.6308

Dark matter stream density



FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093

Velocity field

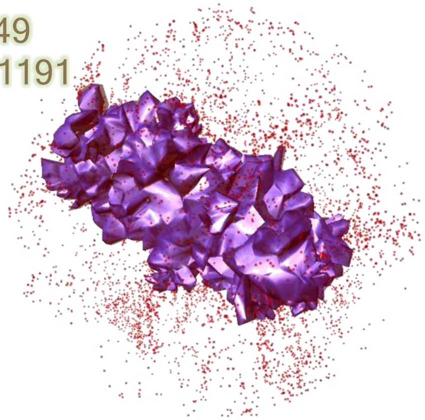


FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093

Cosmic web elements: some algorithms

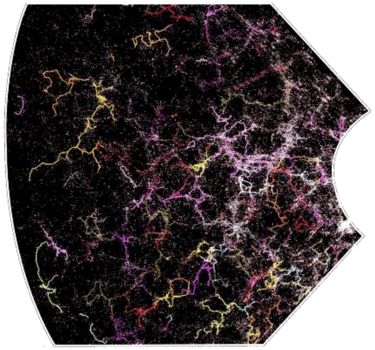
- "**Structure finders**" focus on one element at a time
  - **ZOBOV/VIDE**

Neyrinck 2008, arXiv:0712.3049
Sutter *et al.* 2015, arXiv:1406.1191



  - **DisPerSE**

Sousbie 2011, arXiv:1009.4015
Sousbie *et al.* 2011, arXiv:1009.4014



- "**Classifiers**" dissect the cosmic web all at once
  - The **T-web** (tidal field tensor)

Hahn *et al.* 2007, arXiv:astro-ph/0610280

  - **DIVA** (Lagrangian displacement field, potential structures)

Lavaux & Wandelt 2010, arXiv:0906.4101

  - **ORIGAMI** (particle crossings)

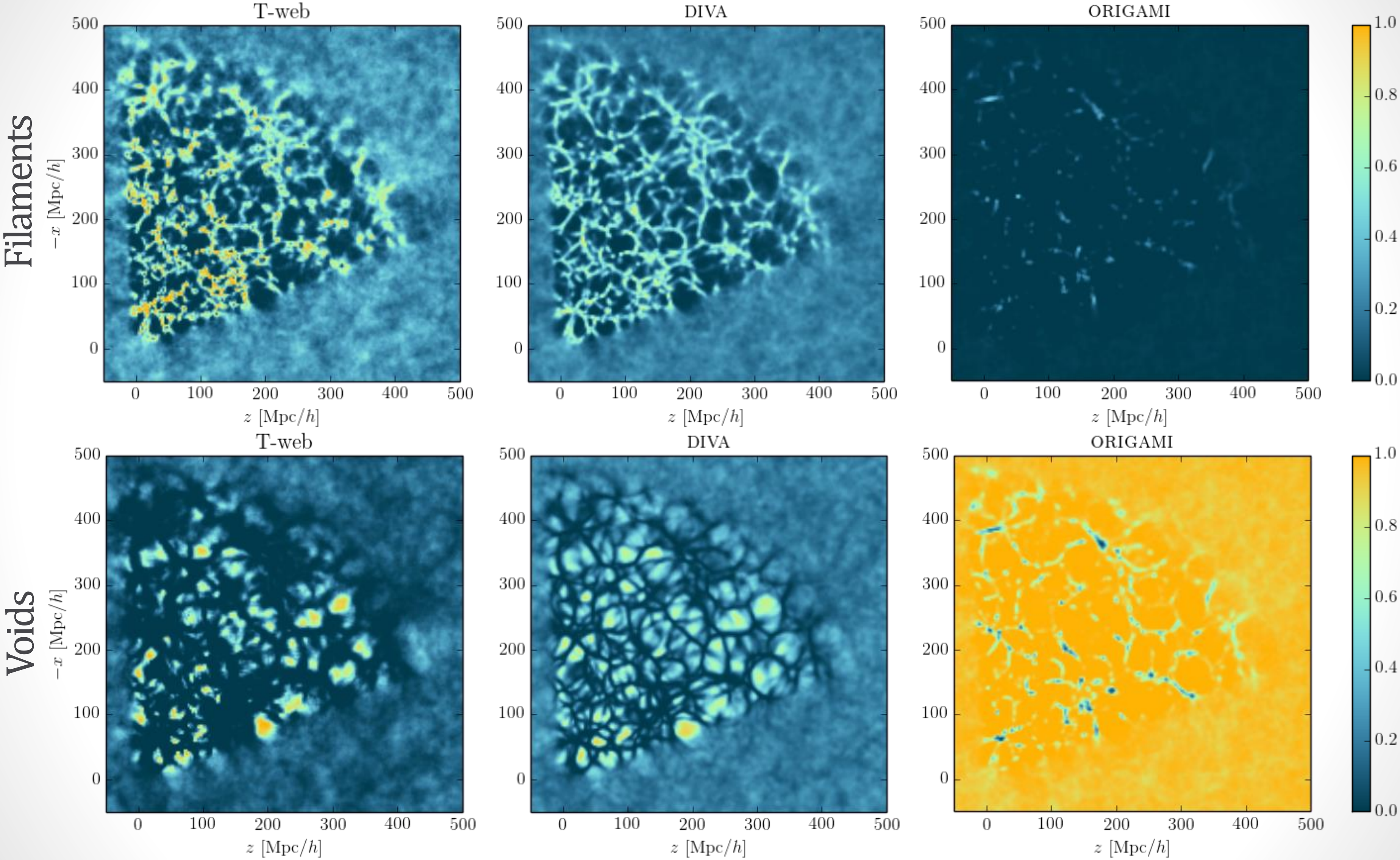
Falck, Neyrinck & Szalay 2012, arXiv:1201.2353

  - **LICH** (Lagrangian displacement field, potential and vortical structures)

FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093

and many others…

Comparing classifiers



FL, Jasche & Wandelt 2015a, arXiv:1502.02690
FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093

Which is the best classifier?

- The framework is provided by **Bayesian decision theory** and **information theory**
- The idea is to maximize a utility function

$$U(\xi) = \langle U(d, \mathrm{T}, \xi) \rangle_{\mathcal{P}(d, \mathrm{T}|\xi)}$$

- An important notion: the **mutual information** between two random variables (or **expected information gain**)

$$\begin{aligned} I\,[X:Y] &\equiv D_{\mathrm{KL}}[\mathcal{P}(x,y)||\mathcal{P}(x)\mathcal{P}(y)] \\ &= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \mathcal{P}(x,y) \log_2 \left(\frac{\mathcal{P}(x,y)}{\mathcal{P}(x)\mathcal{P}(y)} \right) \end{aligned}$$

FL, Lavaux, Jasche & Wandelt 2016, arXiv:1606.06758

1. Utility for parameter inference:

example: cosmic web analysis

- **Example:** *Which classifier produces the most "surprising" cosmic web maps when looking at the data?*

$$U_1(\xi) = I[\underbrace{\mathrm{T}}_{\text{classification}} : \underbrace{d}_{\text{data}} | \xi]$$

2. Utility for model selection:

example: dark energy equation of state

- **Example:** *Let us consider three dark energy models with*

$$w = -0.9, w = -1, w = -1.1$$

Which classifier separates them better?

$$U_2(\xi) = I\left[\underbrace{\mathcal{M}}_{\text{model classifier}} : \underbrace{\mathcal{R}(d)}_{\text{mixture distribution}} | \xi\right]$$

$$\mathcal{R}(d) \equiv \frac{\mathcal{P}(\mathrm{T}(\vec{x}_k)|d, \mathcal{M}_1) + \mathcal{P}(\mathrm{T}(\vec{x}_k)|d, \mathcal{M}_2)}{2}$$

FL, Lavaux, Jasche & Wandelt 2016, arXiv:1606.06758

3. Utility for prediction of new data:

example: galaxy colors

- **Example:** *So far we have not used galaxy colors. Which classifier predicts them best?*

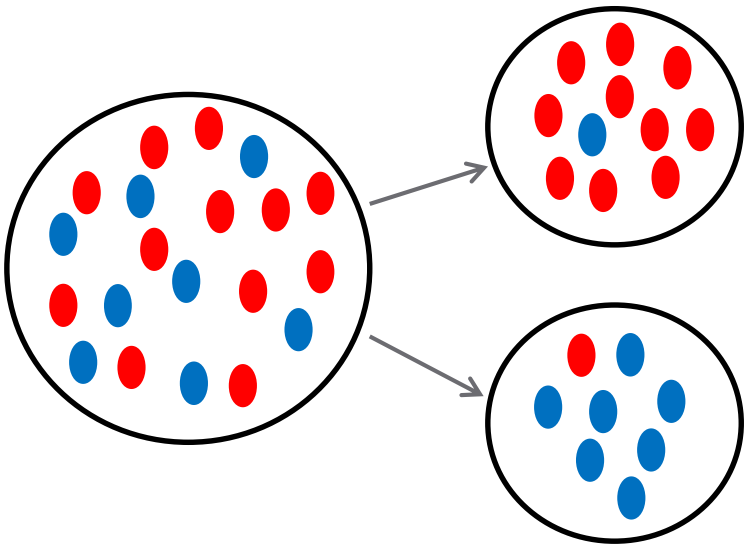
- Maximize the **expected information gain** on some new quantity

$$U_3(\xi) = I[\underset{\text{predicted data}}{c} : \underset{\text{classification}}{\mathrm{T}} | \xi]$$

FL, Lavaux, Jasche & Wandelt 2016, arXiv:1606.06758

3. Utility for prediction of new data:
example: galaxy colors

- How to compute the information gain?



child1 entropy:

$$H = -\frac{10}{11}\log_2\left(\frac{10}{11}\right) - \frac{1}{11}\log_2\left(\frac{1}{11}\right) = 0.4395$$

child2 entropy:

$$H = -\frac{8}{9}\log_2\left(\frac{8}{9}\right) - \frac{1}{9}\log_2\left(\frac{1}{9}\right) = 0.5033$$

parent entropy:

$$H = -\frac{8}{20}\log_2\left(\frac{8}{20}\right) - \frac{12}{20}\log_2\left(\frac{12}{20}\right) = 0.9709$$

weighted average entropy of children:

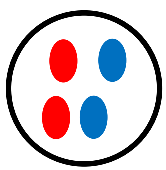
$$\frac{11}{20} \times 0.4395 + \frac{9}{20} \times 0.5033 = 0.4682$$

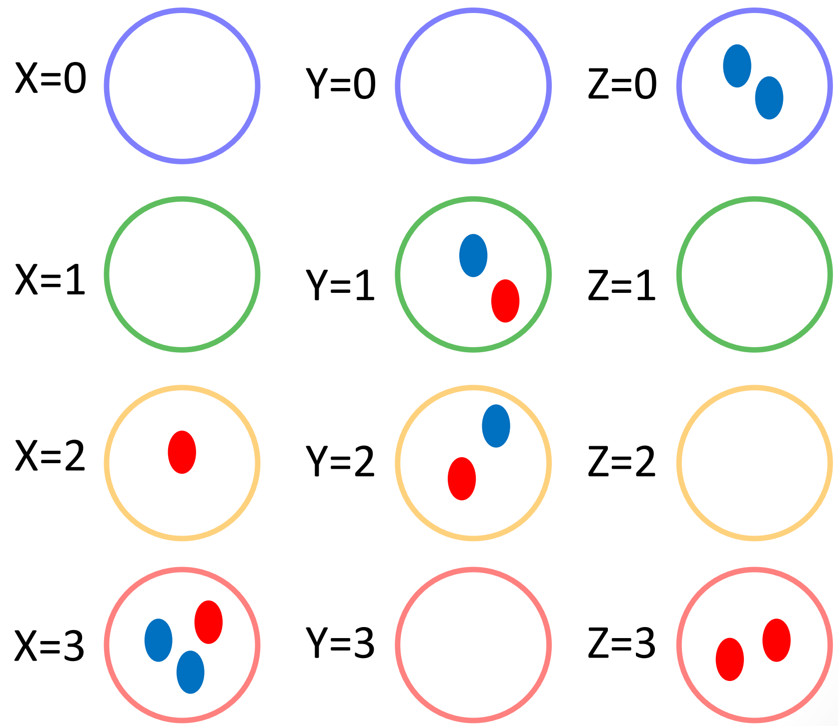
information gain for this split: $0.9709 - 0.4682 = 0.5027\ \mathrm{Sh}$

3. Utility for prediction of new data:
example: galaxy colors

- A **supervised machine learning** problem!
  - 3 **features** = classifications (X=T-web, Y=DIVA, Z=ORIGAMI)
  - 4 **possible values** (0=void, 1=sheet, 2=filament, 3=cluster)
  - 2 **classes** (I=blue, II=red)

X	Y	Z	C
3	2	0	I
3	1	0	I
2	2	3	II
3	1	3	II

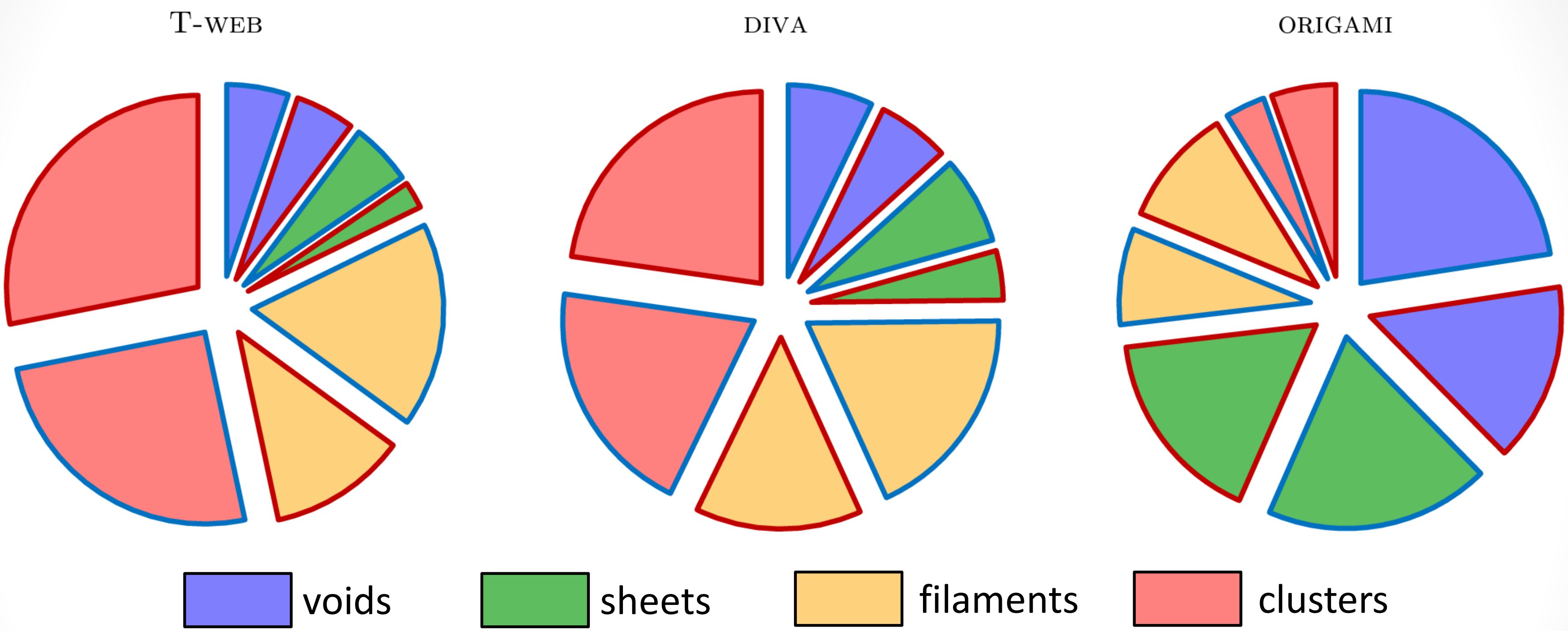




$\mathrm{IG} = 0.3112$ $\mathrm{IG} = 0.$ $\mathrm{IG} = 1.$

no gain: worst best!

Application to SDSS main sample galaxies
(367,157 galaxies)



	T-web	DIVA	ORIGAMI
Utility [Sh]	0.0152	0.0101	0.0143

FL, Lavaux, Jasche & Wandelt 2016, arXiv:1606.06758

Summary & Conclusions

- **BORG** allows a rich description of the large-scale structure of the Universe (density field, velocity field, stream density…)
- The **cosmic web** can be described using various classifiers (T-web, DIVA, ORIGAMI…)
- **Decision theory** and **information theory** offer a framework to rank classifiers, with utility functions depending on the desired use
- Potential **applications**: galaxy properties, intrinsic alignments…

All maps, catalogs & scripts are publicly available at http://icg.port.ac.uk/~leclercq/

References

Jasche & Wandelt 2013, arXiv:1203.3639 (BORG proof of concept)
Jasche, FL & Wandelt 2015, arXiv:1409.6308 (BORG SDSS analysis)
FL, Jasche & Wandelt 2015a, arXiv:1502.02690 (T-web, entropy, relative entropy)
FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093 (DIVA, ORIGAMI, phase-space properties)
FL, Lavaux, Jasche & Wandelt 2016, arXiv:1606.06758 (mutual information, classifier utilities)