Cosmic web analysis with the BORG SDSS run

Florent Leclercq

Institute of Cosmology and Gravitation, University of Portsmouth

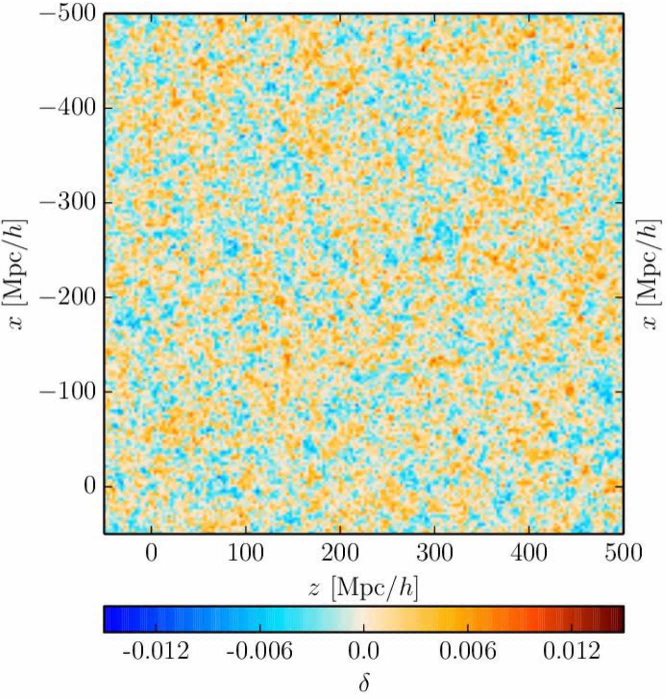
http://icg.port.ac.uk/~leclercq/

October 31st, 2016

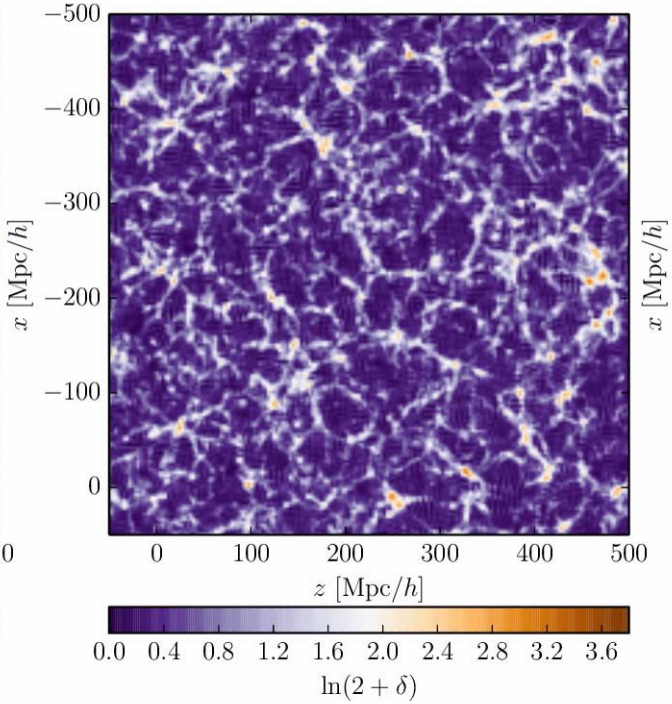
In collaboration with:

Nico Hamaus (LMU), Jens Jasche (ExC Universe, Garching), Guilhem Lavaux (IAP),
Will Percival (ICG), Paul M. Sutter (Ohio State U.), Benjamin Wandelt (IAP/U. Illinois)

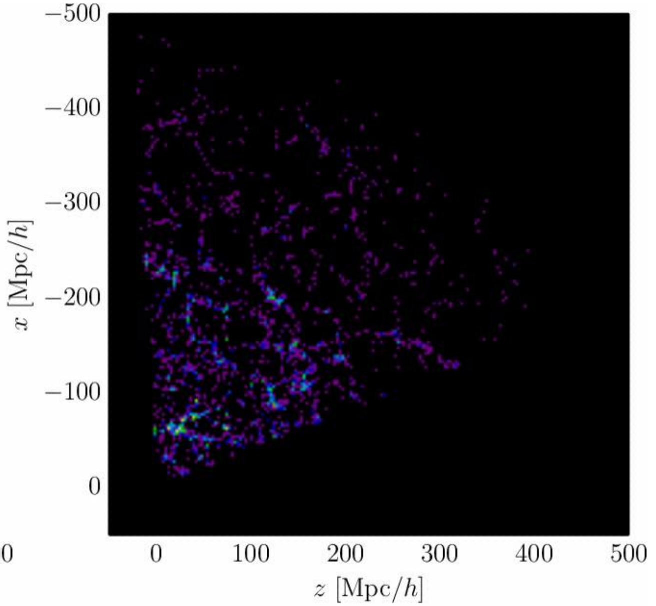
The BORG SDSS run



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

COLA: *COmoving Lagrangian Acceleration*

- Write the displacement vector as: $\mathbf{s} = \mathbf{s}_{\mathrm{LPT}} + \mathbf{s}_{\mathrm{MC}}$

Tassev & Zaldarriaga 2012, arXiv:1203.5785

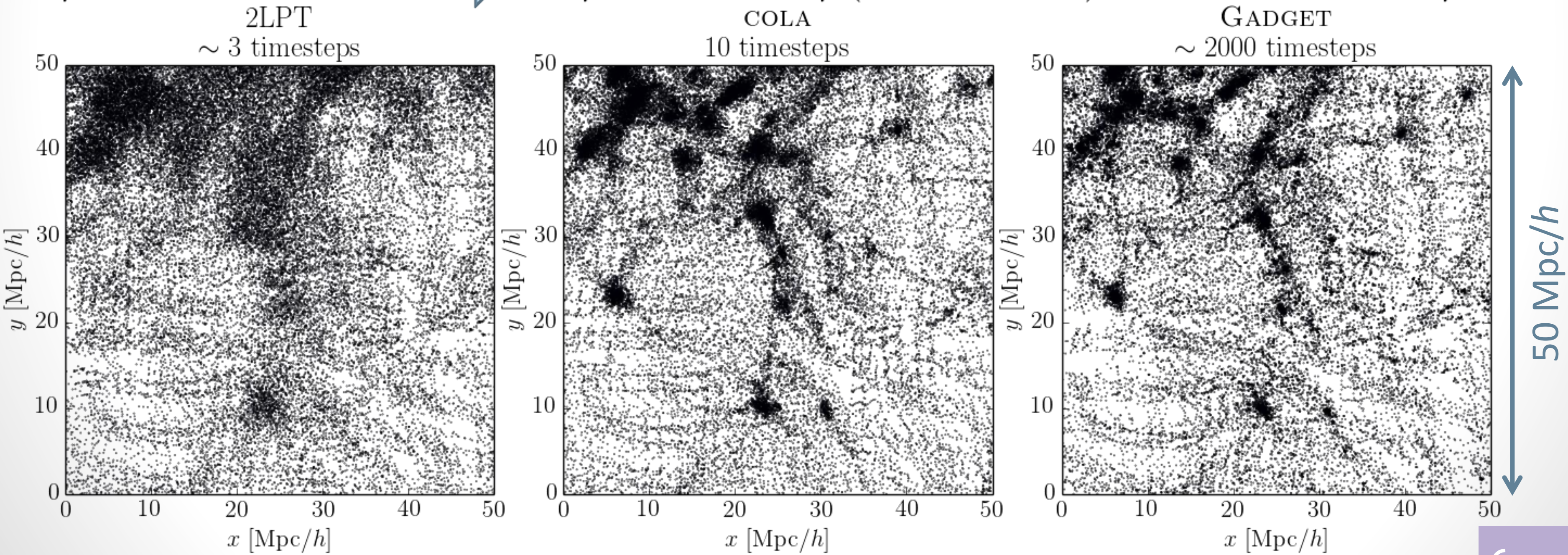
- Time-stepping (omitted constants and Hubble expansion):

Standard:

$$\partial_\tau^2 \mathbf{s} = -\nabla\Phi$$

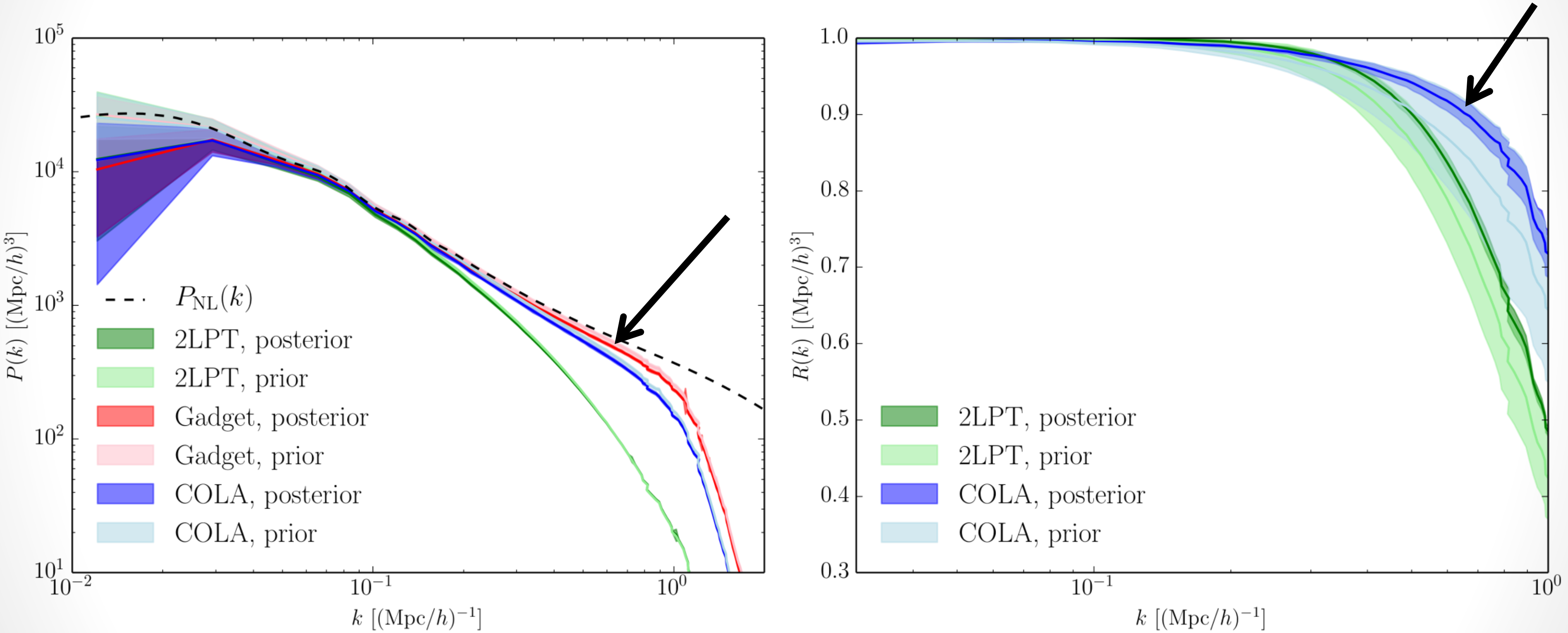
Modified:

$$\partial_\tau^2 \mathbf{s}_{\mathrm{MC}} = \partial_\tau^2(\mathbf{s} - \mathbf{s}_{\mathrm{LPT}}) = -\nabla\Phi - \partial_\tau^2 \mathbf{s}_{\mathrm{LPT}}$$



Tassev, Zaldarriaga & Einsenstein 2013, arXiv:1301.0322

Non-linear filtering improves the fit



FL 2015, arXiv:1512.04985 (chapter 7)

DARK MATTER VOIDS

Dark matter voids: pipeline

Why BORG?

Sparsity & **Bias**

Sutter *et al.* 2013, arXiv:1309.5087

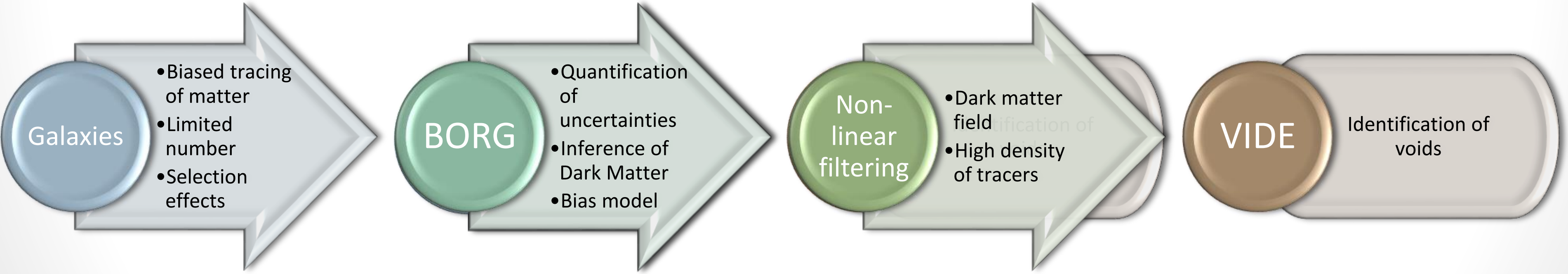
Sutter *et al.* 2013, arXiv:1311.3301

How?

VIDE toolkit: Sutter *et al.* 2015, arXiv:1406.1191

www.cosmicvoids.net

based on ZOBOV: Neyrinck 2007, arXiv:0712.3049



Galaxies

- Biased tracing of matter
- Limited number
- Selection effects

BORG

- Quantification of uncertainties
- Inference of Dark Matter
- Bias model

Non-linear filtering

- Dark matter field
- High density of tracers

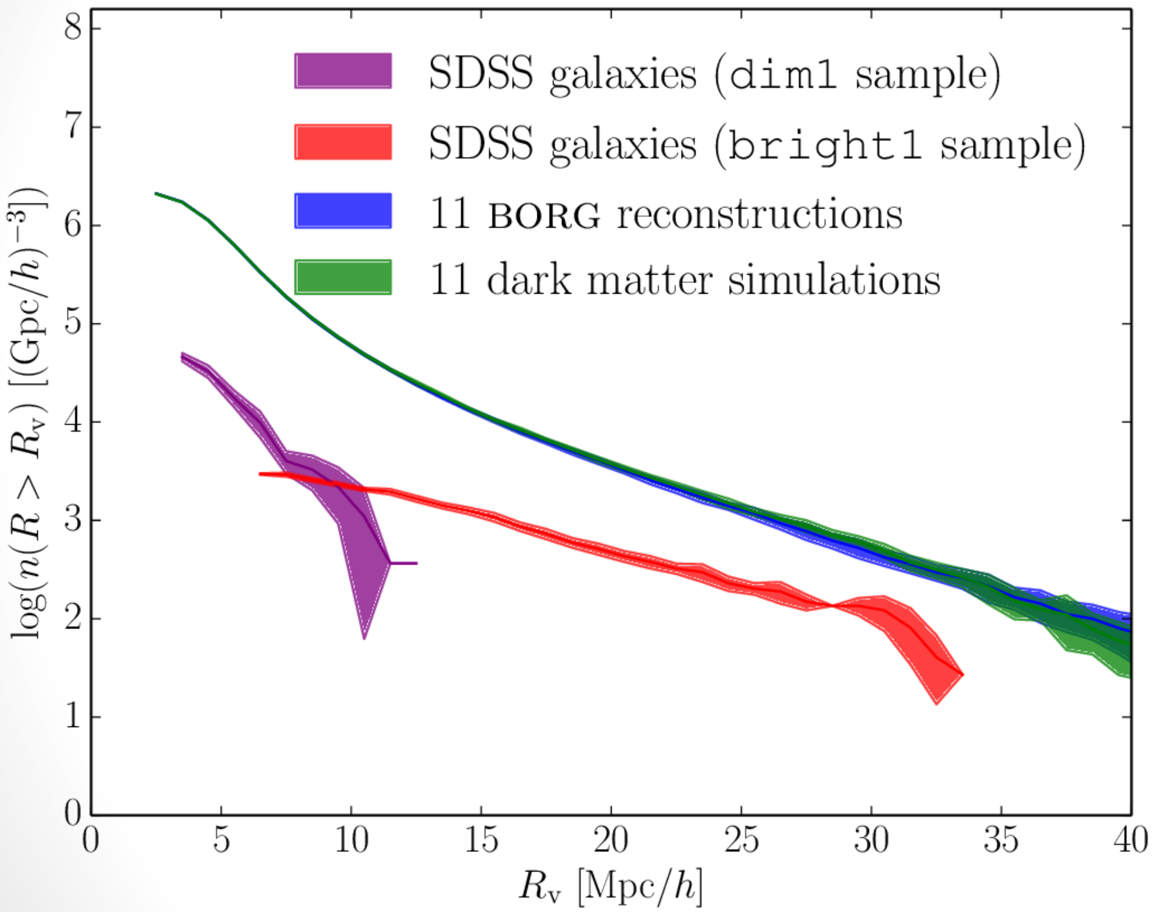
VIDE

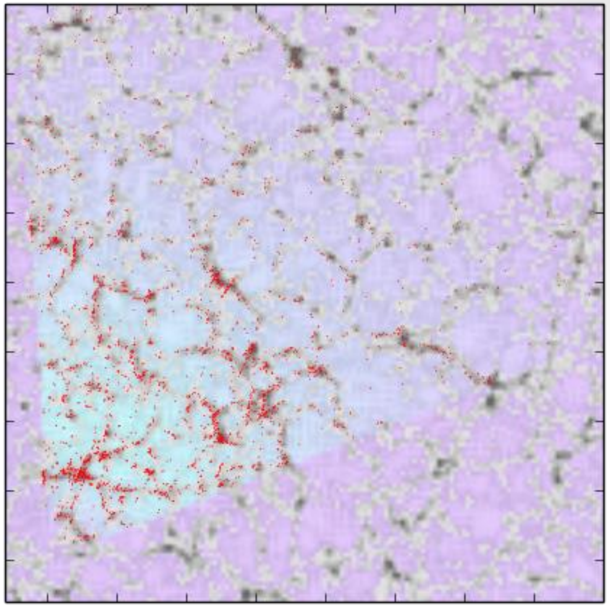
Identification of voids

FL, Jasche, Sutter, Hamaus & Wandelt 2015, arXiv:1410.0355

BORG unveils many more voids

Void number function





Voids in constrained regions only

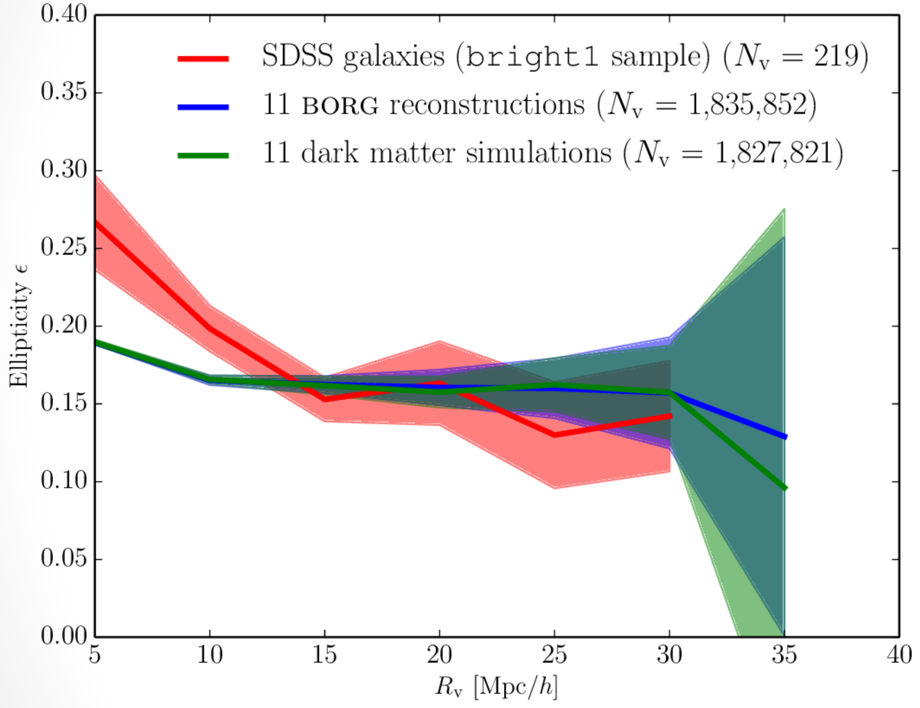
Voids are **Poisson-dominated** objects:

10x more voids require 100x more galaxies!

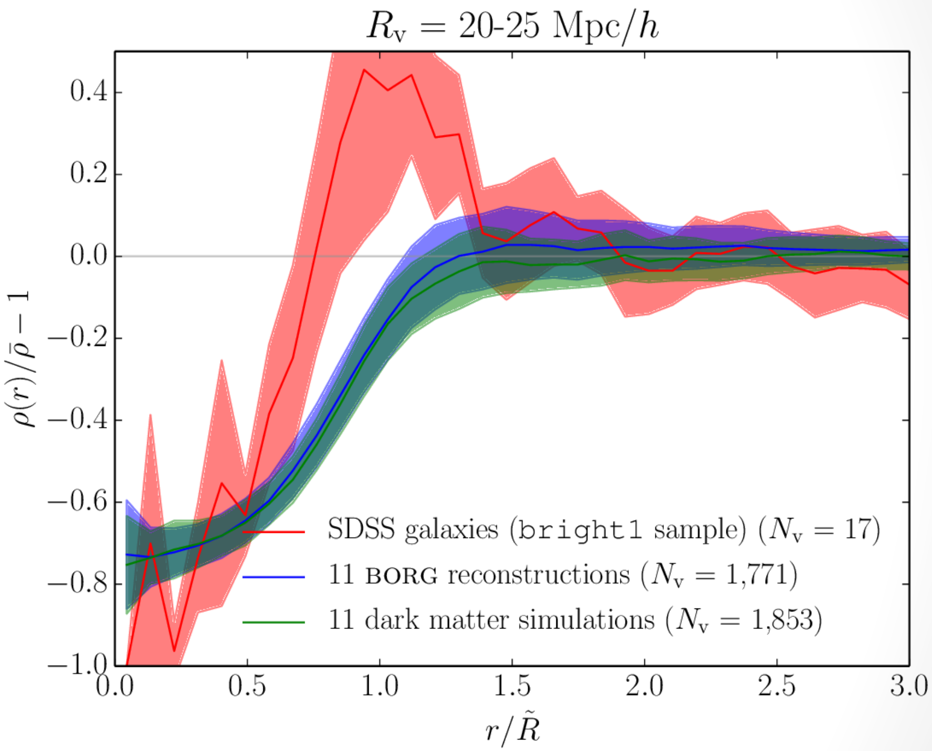
FL, Jasche, Sutter, Hamaus & Wandelt 2015, arXiv:1410.0355

Reduction of statistical uncertainty in voids catalogs

Ellipticity distribution



Radial density profile



FL, Jasche, Sutter, Hamaus & Wandelt 2015, arXiv:1410.0355

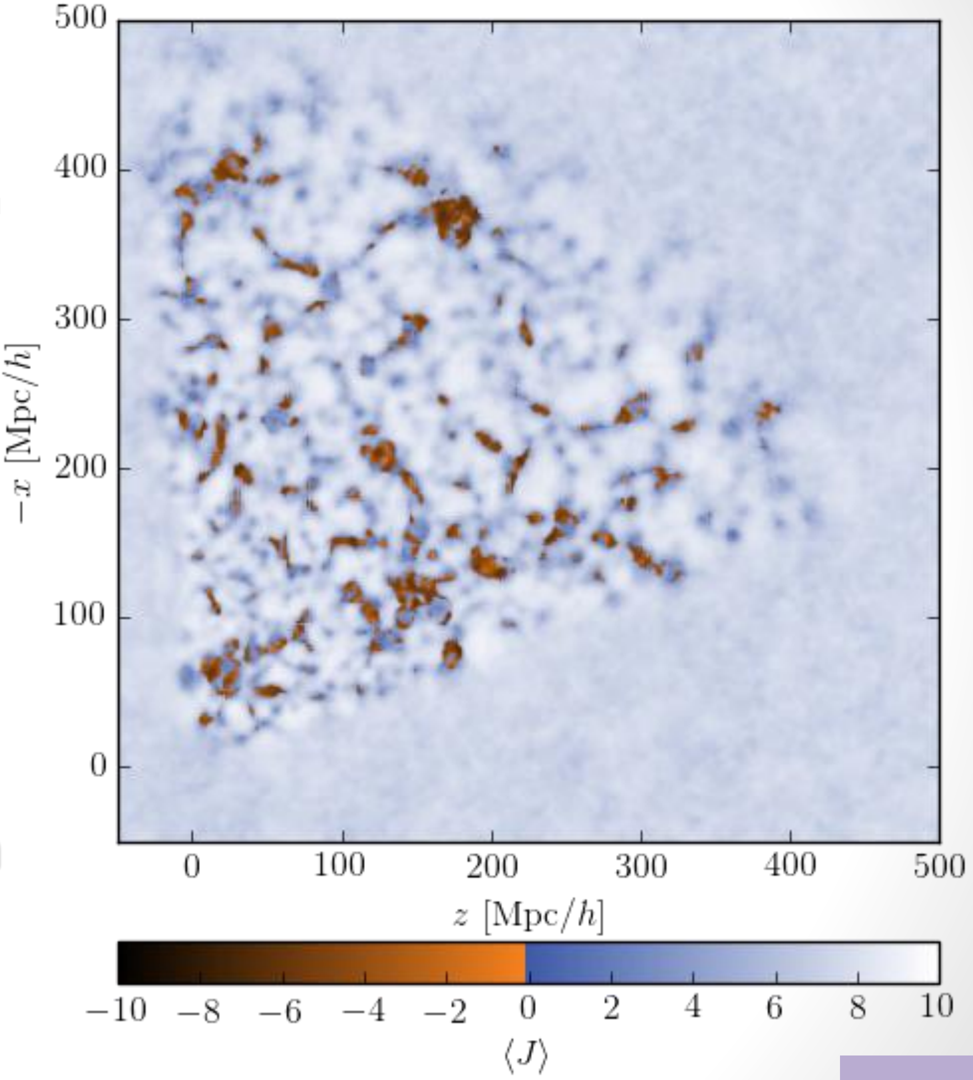
PHASE-SPACE PROPERTIES OF DARK MATTER

Inference of the dark matter phase-space sheet

- The dark matter phase-space sheet has been studied so far in simulations

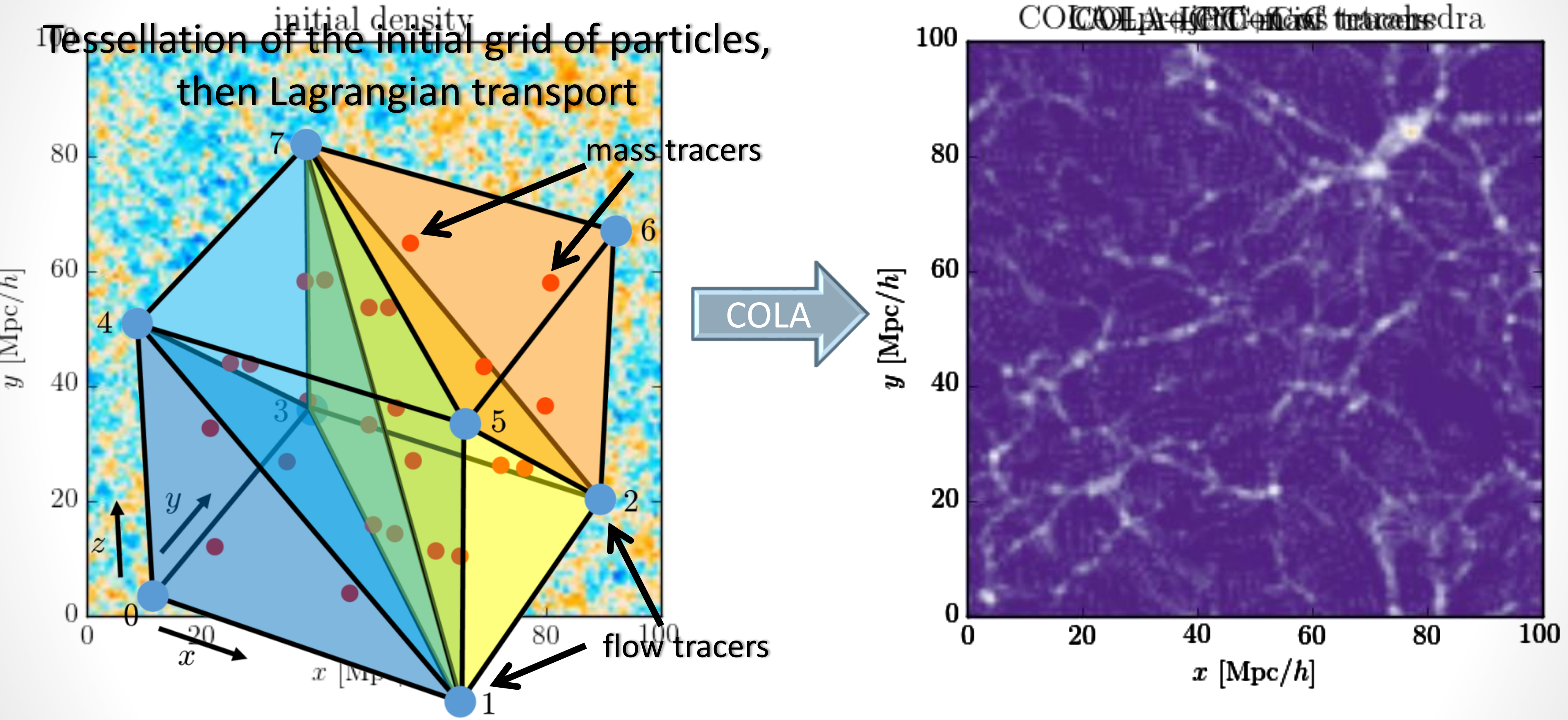
e.g. Neyrinck 2012, arXiv:1202.3364
Abel, Hahn & Kaehler 2012, arXiv:1111.3944
Shandarin, Habib & Heitmann 2012, arXiv:1111.2366

- BORG infers **Lagrangian dynamics** in real data
- Identified structures have a direct **physical interpretation**



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

Non-linear filtering improves density samples


Tessellation of the initial grid of particles, then Lagrangian transport

mass tracers

flow tracers

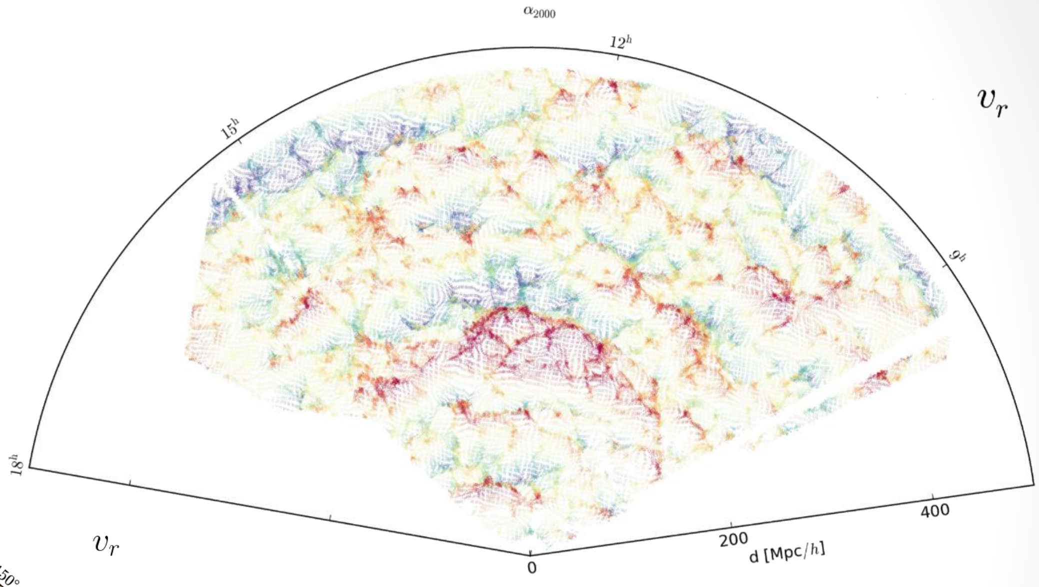
Abel, Hahn & Kaehler 2012, arXiv:1111.3944
Hahn, Abel & Khaeler 2013, arXiv:1210.6652
Hahn, Angulo & Abel 2015, arXiv:1404.2280

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

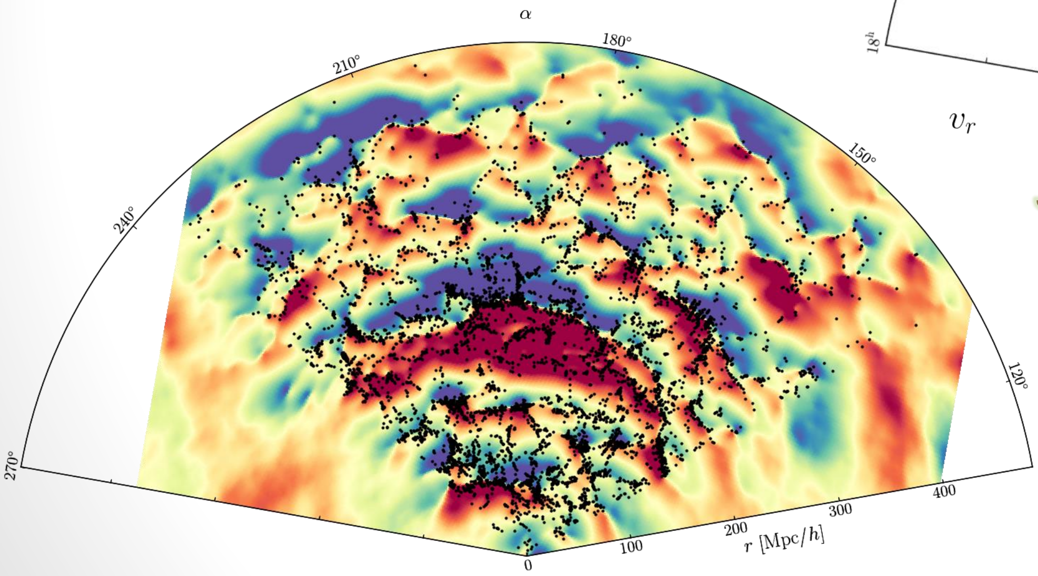
Lagrangian transport of the velocity field

see Hahn, Angulo & Abel 2015, arXiv:1404.2280

Velocity field: 2014 vs 2016



Jasche, FL & Wandelt 2015, arXiv:1409.6308



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

BAYESIAN ANALYSIS OF THE DYNAMIC COSMIC WEB

Cosmic web classification procedures

void, sheet, filament, cluster?

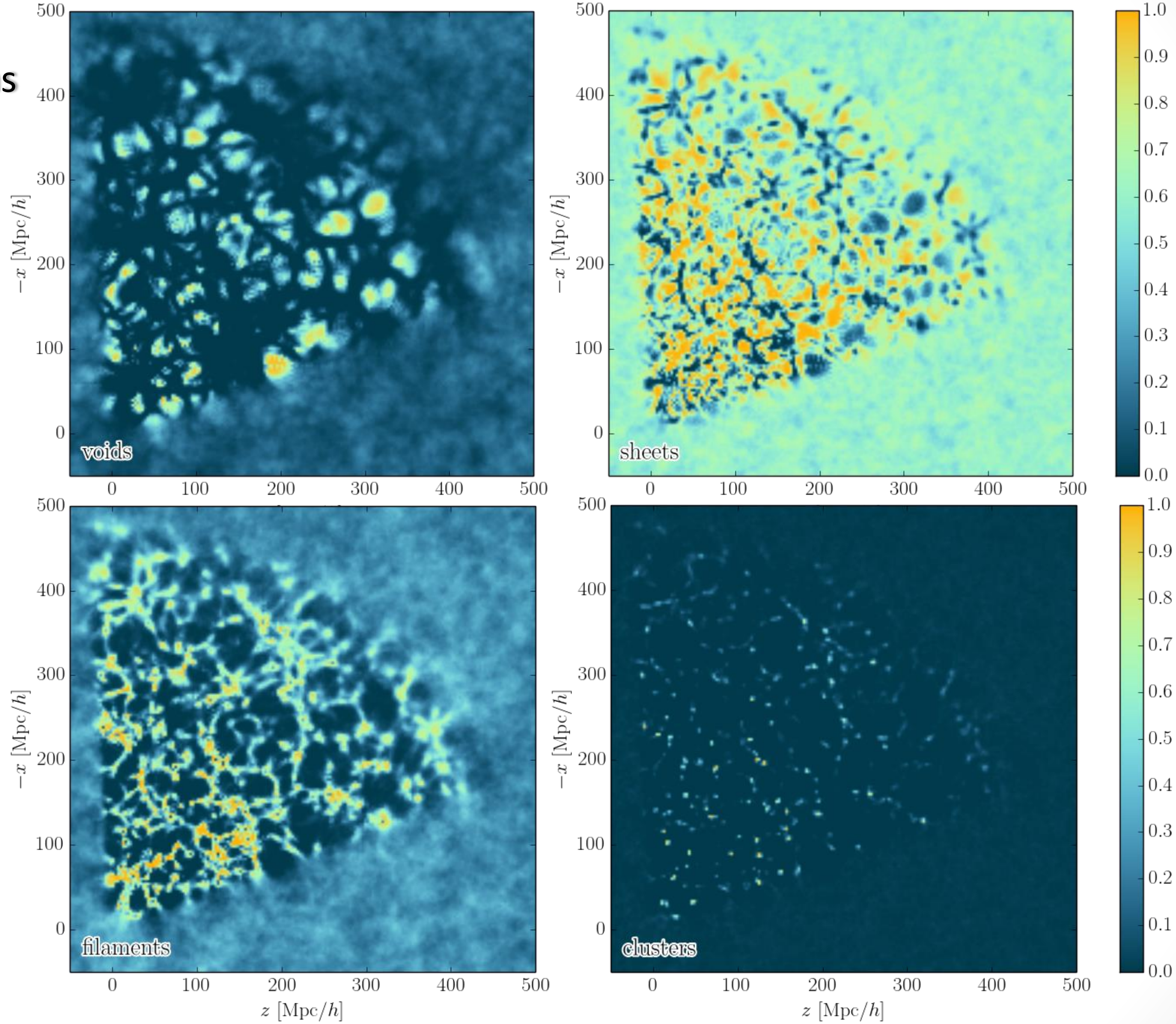
- The **T-web**:

uses the sign of μ_1, μ_2, μ_3: eigenvalues of the tidal field tensor, Hessian of the gravitational potential: $T_{ij}(\mathbf{x}) = \partial_i \partial_j \Phi(\mathbf{x})$

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

T-web structures inferred by BORG

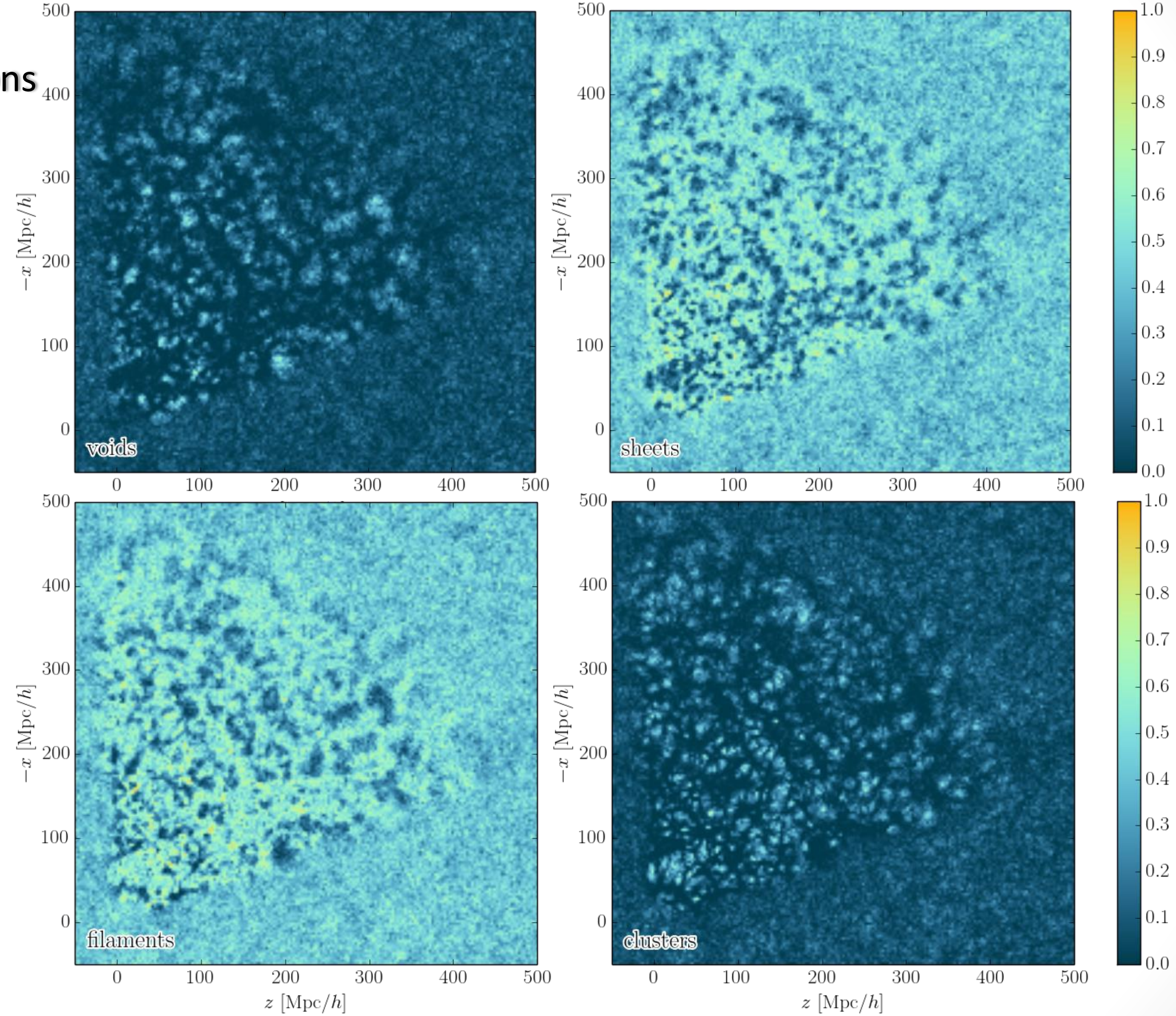
Final conditions



FL, Jasche & Wandelt 2015a, arXiv:1502.02690

T-web structures inferred by BORG

Initial conditions



FL, Jasche & Wandelt 2015a, arXiv:1502.02690

Cosmic web classification procedures

void, sheet, filament, cluster?

- The **T-web**:

uses the sign of μ_1, μ_2, μ_3: eigenvalues of the tidal field tensor, Hessian of the gravitational potential: $T_{ij}(\mathbf{x}) = \partial_i \partial_j \Phi(\mathbf{x})$

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

- **DIVA**:

uses the sign of $\lambda_1, \lambda_2, \lambda_3$: eigenvalues of the shear of the Lagrangian displacement field: $R_{\ell m}(\mathbf{q}) = \partial_m \Psi_\ell(\mathbf{q})$

Lavaux & Wandelt 2010, arXiv:0906.4101

- **ORIGAMI** :

uses the dark matter “phase-space sheet” (number of orthogonal axes along which there is shell-crossing)

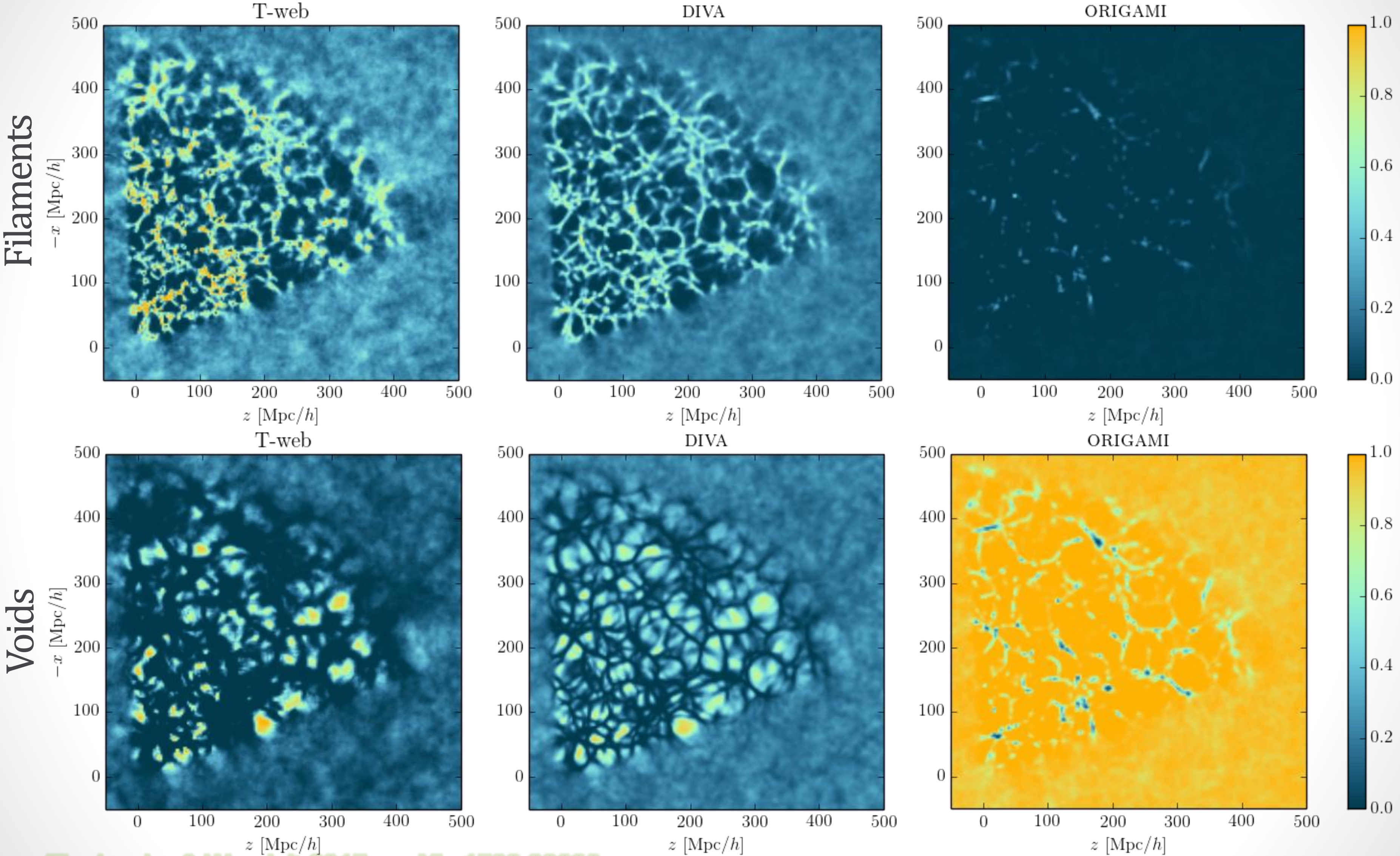
Falck, Neyrinck & Szalay 2012, arXiv:1201.2353

Lagrangian classifiers

now usable in real data!

and many others…

Comparing classifiers



FL, Jasche & Wandelt 2015a, arXiv:1502.02690

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

COSMIC WEB CLASSIFICATION USING DECISION THEORY

A decision rule for structure classification

- Space of “input features”:

$$\{\mathrm{T}_0 = \text{void}, \mathrm{T}_1 = \text{sheet}, \mathrm{T}_2 = \text{filament}, \mathrm{T}_3 = \text{cluster}\}$$

- Space of “actions”:

$$\{a_0 = \text{“decide void”}, a_1 = \text{“decide sheet”}, a_2 = \text{“decide filament”}, a_3 = \text{“decide cluster”}, a_{-1} = \text{“do not decide”}\}$$

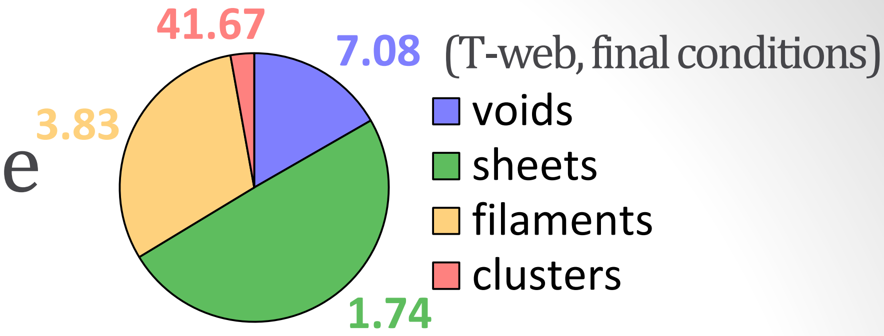
⇨ A problem of **Bayesian decision theory**:
one should take the action that maximizes the utility

$$U(a_j(\vec{x}_k)|d) = \sum_{i=0}^{3} G(a_j|\mathrm{T}_i)\,\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d)$$

- How to write down the gain functions?

FL, Jasche & Wandelt 2015b, arXiv:1503.00730

Gambling with the Universe



- One proposal:

$$G(a_j|\mathrm{T}_i) = \begin{cases} \frac{1}{\mathcal{P}(\mathrm{T}_i)} - \alpha & \text{if } j \in [\![0,3]\!] \text{ and } i = j \\ -\alpha & \text{if } j \in [\![0,3]\!] \text{ and } i \neq j \\ 0 & \text{if } j = -1. \end{cases}$$

"Winning"

"Loosing"

"Not playing"

- Without data, the expected utility is

$$U(a_j) = 1 - \alpha \quad \text{if } j \neq 1$$

"Playing the game"

$$U(a_{-1}) = 0$$

"Not playing the game"

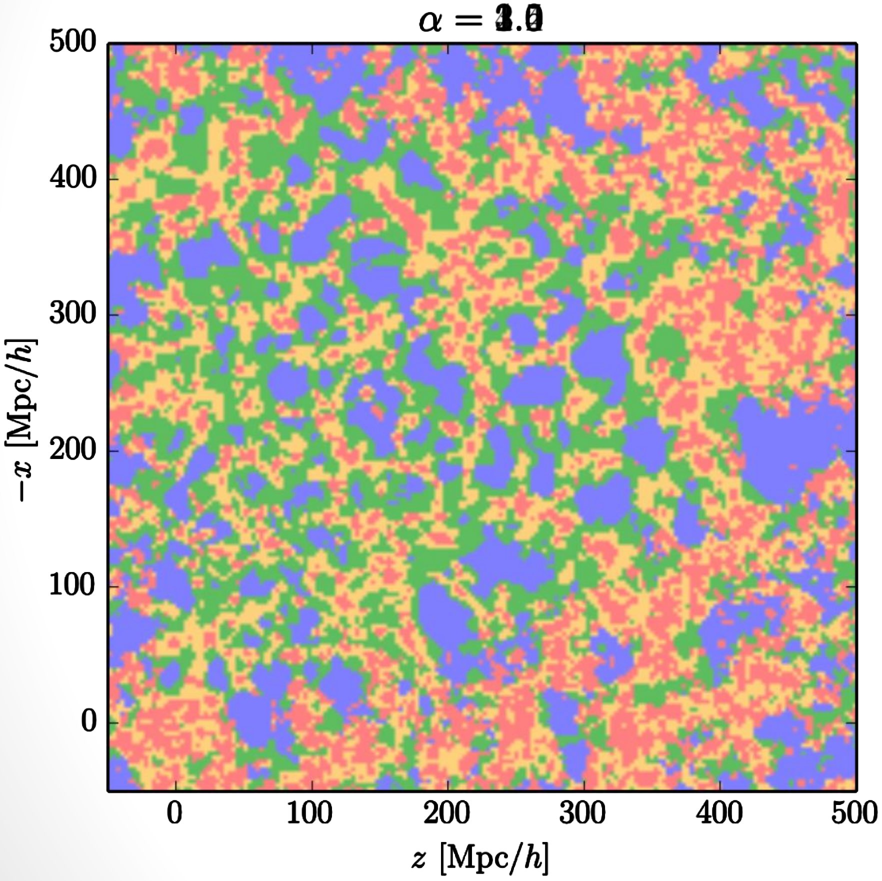
- With $\alpha = 1$, it's a *fair game* ⇨ always play
  ⇨ "**speculative map**" of the LSS
- Values $\alpha > 1$ represent an *aversion for risk*
  ⇨ increasingly "**conservative maps**" of the LSS

FL, Jasche & Wandelt 2015b, arXiv:1503.00730

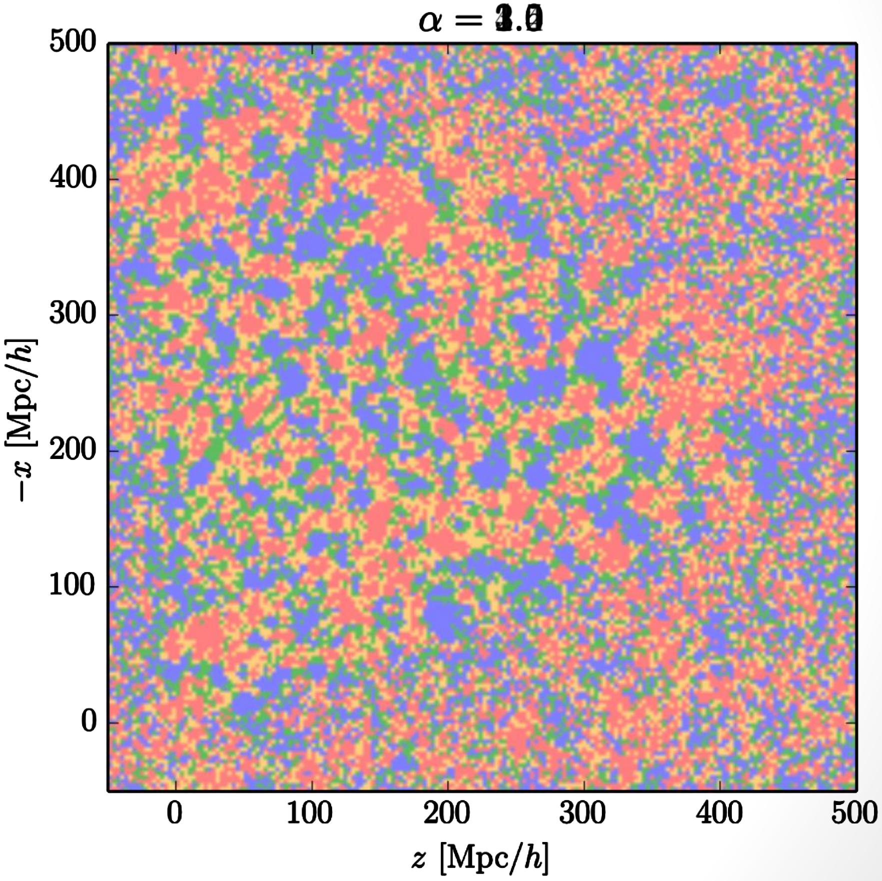
Playing the game…

🟦 voids
🟨 filaments
⬛ undecided
🟩 sheets
🟥 clusters

Final conditions



Initial conditions



FL, Jasche & Wandelt 2015b, arXiv:1503.00730

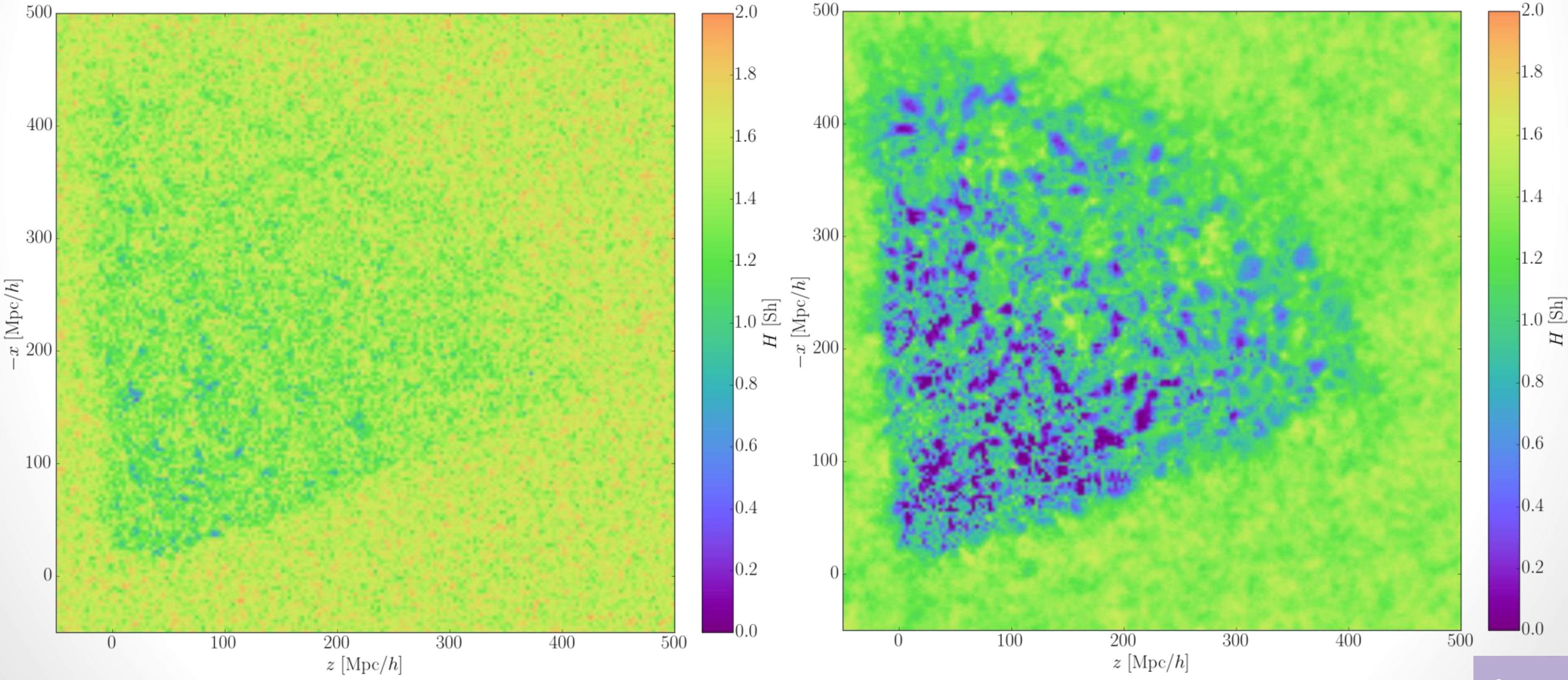
COSMIC WEB ANALYSIS AND INFORMATION THEORY

What is the information content of these maps?

Shannon entropy

$$H\left[\mathcal{P}\right](\mathrm{T}(\vec{x}_k)|d) = -\sum_{i} p_i \log_2 p_i = -\sum_{i=0}^{3} \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d)\log_2(\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d))$$

in shannons (Sh)


Initial conditions

Final conditions

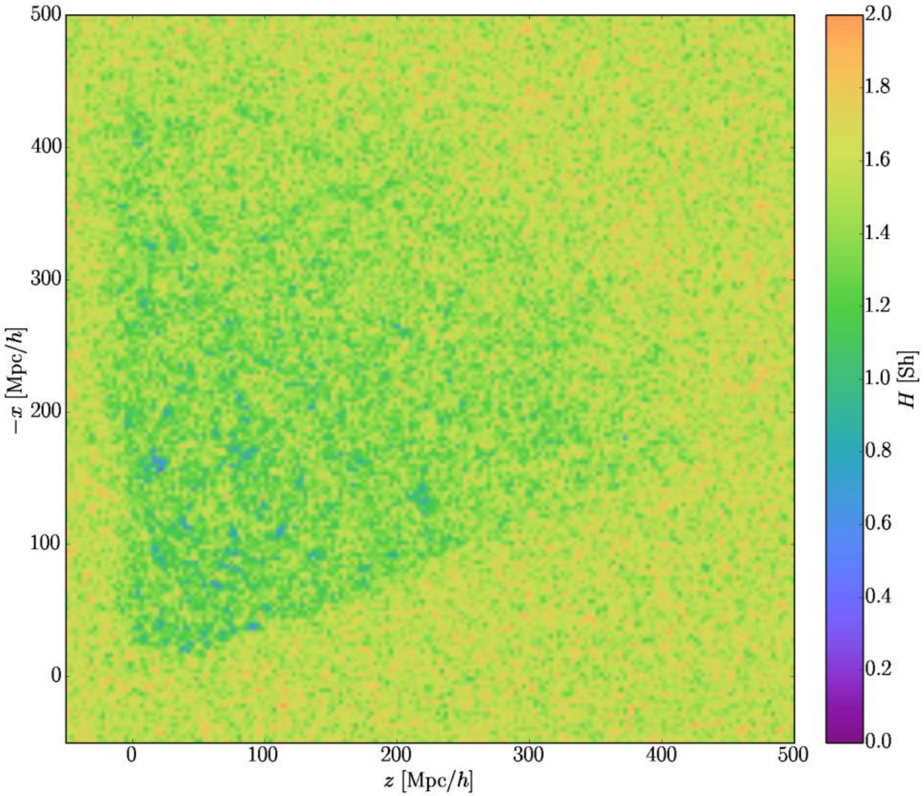
FL, Jasche & Wandelt 2015a, arXiv:1502.02690

How is information propagated?

Shannon entropy

$$H\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d)\right] \equiv -\sum_{i=0}^{3} \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d) \log_2(\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d))$$

in shannons (Sh)



FL, Jasche & Wandelt 2015a, arXiv:1502.02690

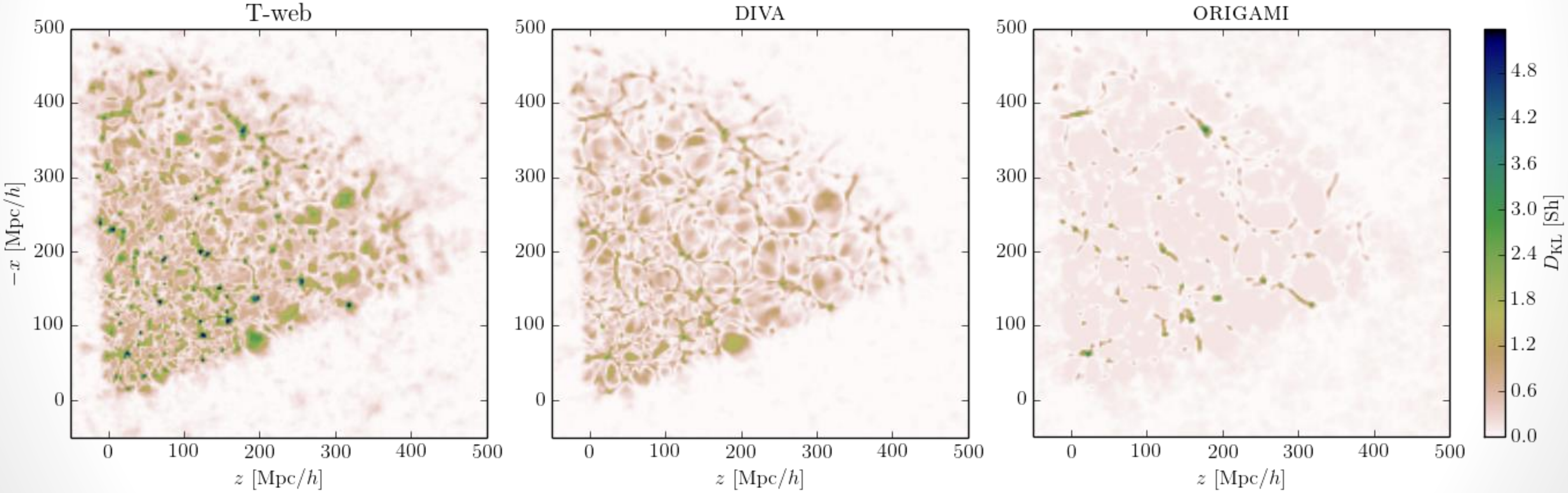
How much did the data surprise us?

information gain a.k.a. relative entropy or Kullback-Leibler divergence posterior/prior

$$D_{\mathrm{KL}}\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d)||\mathcal{P}(\mathrm{T})\right] = \sum_i \mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d) \log_2 \left(\frac{\mathcal{P}(\mathrm{T}_i(\vec{x}_k)|d)}{\mathcal{P}(\mathrm{T}_i)}\right)$$

$$D_{\mathrm{KL}}\left[P||Q\right] = \sum_i p_i \log_2 \left(\frac{p_i}{q_i}\right)$$

in Sh



(more about the Kullback-Leibler divergence later)

FL, Jasche & Wandelt 2015a, arXiv:1502.02690

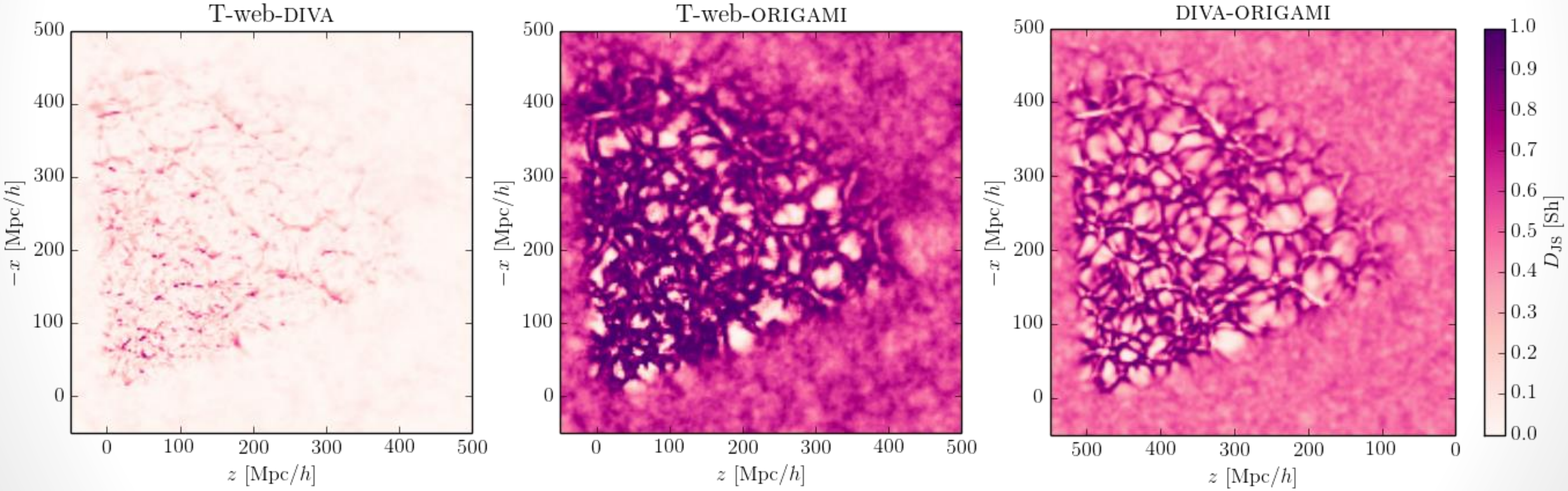
FL, Lavaux, Jasche & Wandelt 2016, arXiv:1606.06758

How similar are different classifications?

Jensen-Shannon divergence

$$D_{\mathrm{JS}}[\mathcal{P}:\mathcal{Q}] \equiv \frac{1}{2} D_{\mathrm{KL}}\left[\mathcal{P}||\frac{\mathcal{P}+\mathcal{Q}}{2}\right] + \frac{1}{2} D_{\mathrm{KL}}\left[\mathcal{Q}||\frac{\mathcal{P}+\mathcal{Q}}{2}\right]$$

in Sh,
between 0 and 1



(more about the Jensen-Shannon divergence later)

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

Which is the best classifier?

- **Decision theory**: a framework to classify structures in the presence of uncertainty.
  Can we extend the decision problem to the space of classifiers?
- As before, 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

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

- Property: $\boxed{I\,[X:Y] = \langle D_{\mathrm{KL}}[\mathcal{P}(x|y)||\mathcal{P}(x)] \rangle_{\mathcal{P}(Y)}}$

Mutual information is the expectation of the Kullback-Leibler divergence of the conditional from the unconditional distribution.

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?*

- In analogy with the formalism of **Bayesian experimental design**: maximize the **expected information gain** for cosmic web maps

$$U_1(d,\xi)(\vec{x}_k) = D_{\mathrm{KL}}\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d,\xi)||\mathcal{P}(\mathrm{T}|\xi)\right]$$

$$\boxed{U_1(\xi) = I\,[\underset{\substack{\nearrow\\ \text{classification}}}{\mathrm{T}} : \underset{\substack{\nwarrow\\ \text{data}}}{d}|\xi]}$$

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

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?

- The **Jensen-Shannon divergence** between posterior predictive distributions can be used as an approximate **predictor for the change in the Bayes factor**

Vanlier *et al.* 2014, BMC Syst Biol 8, 20 (2014)

- In analogy: $U_2(d,\xi)(\vec{x}_k) = D_{\mathrm{JS}}\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d,\mathcal{M}_1) : \mathcal{P}(\mathrm{T}(\vec{x}_k)|d,\mathcal{M}_2)|\xi\right]$

$$U_2(\xi) = I\left[\mathcal{M} : \mathcal{R}(d)|\xi\right]$$

model classifier mixture distribution

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

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** for some new quantity

$$U_3(d, \mathrm{T}, \xi) = D_{\mathrm{KL}} \left[\mathcal{P}(c|d, \mathrm{T}, \xi) || \mathcal{P}(c|\xi) \right]$$

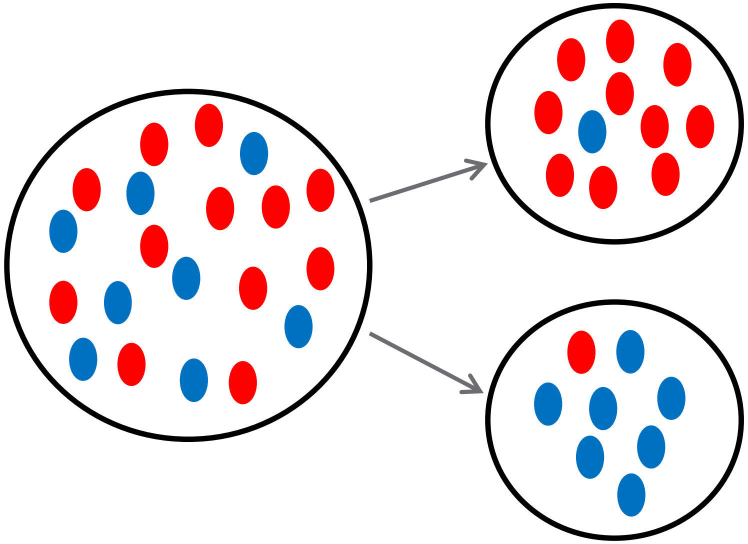
$$\boxed{U_3(\xi) = I[\underset{\text{predicted data}}{\underset{\nearrow}{c}} : \underset{\text{classification}}{\underset{\nwarrow}{\mathrm{T}}} | \xi]}$$

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:

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

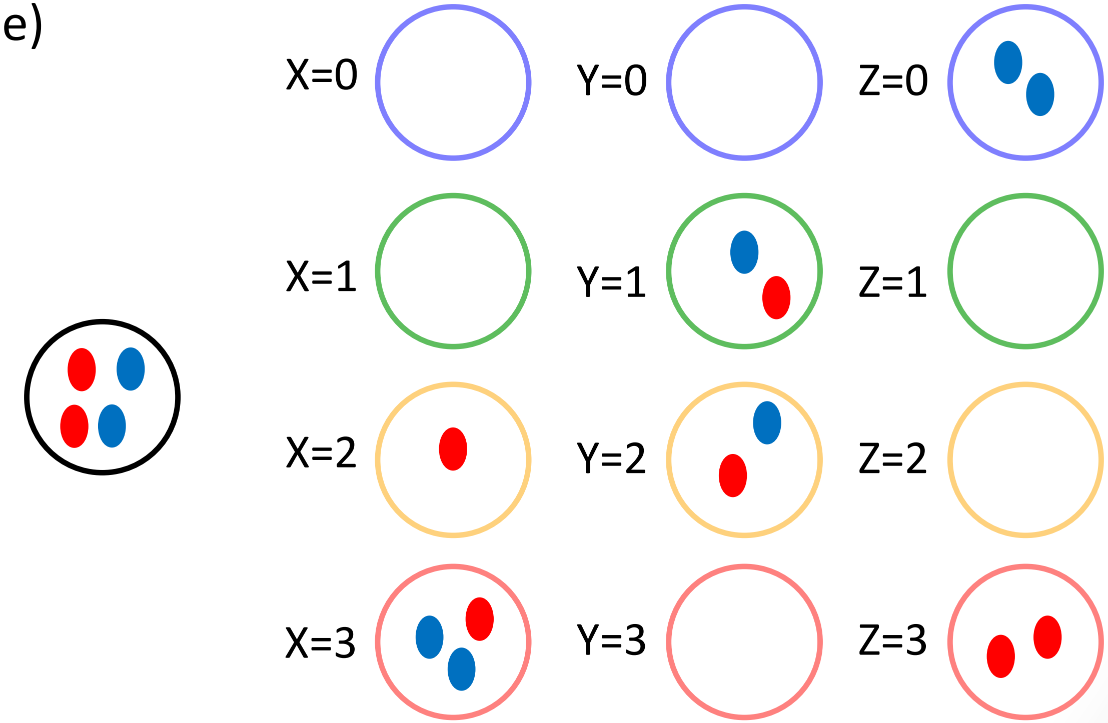
information gain for this split: $0.9709 - 0.4682 = 0.5027\ \mathrm{Sh}$

3. Utility for prediction of new data:

example: galaxy colors

- A **supervised machine learning** problem!
  - 3 **features** = classifications (T-web, DIVA, ORIGAMI) with
  - 4 **possible values** (void, sheet, filament, cluster)
  - 2 **classes** (red, blue)

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



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

no gain: worst

best!

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

Concluding thoughts

- We did so many great things with the BORG SDSS run…
  - Dark matter voids
  - T-web inference
  - DIVA, ORIGAMI and phase-space properties of dark matter
  - A decision rule for cosmic web classifications
  - Comparison of classifiers using decision theory

- Hopefully it's not the end, since the data models are now much better!