Cosmic web analysis and Information Theory

Florent Leclercq
www.florent-leclercq.eu

Imperial Centre for Inference and Cosmology
Imperial College London

July 6th, 2018

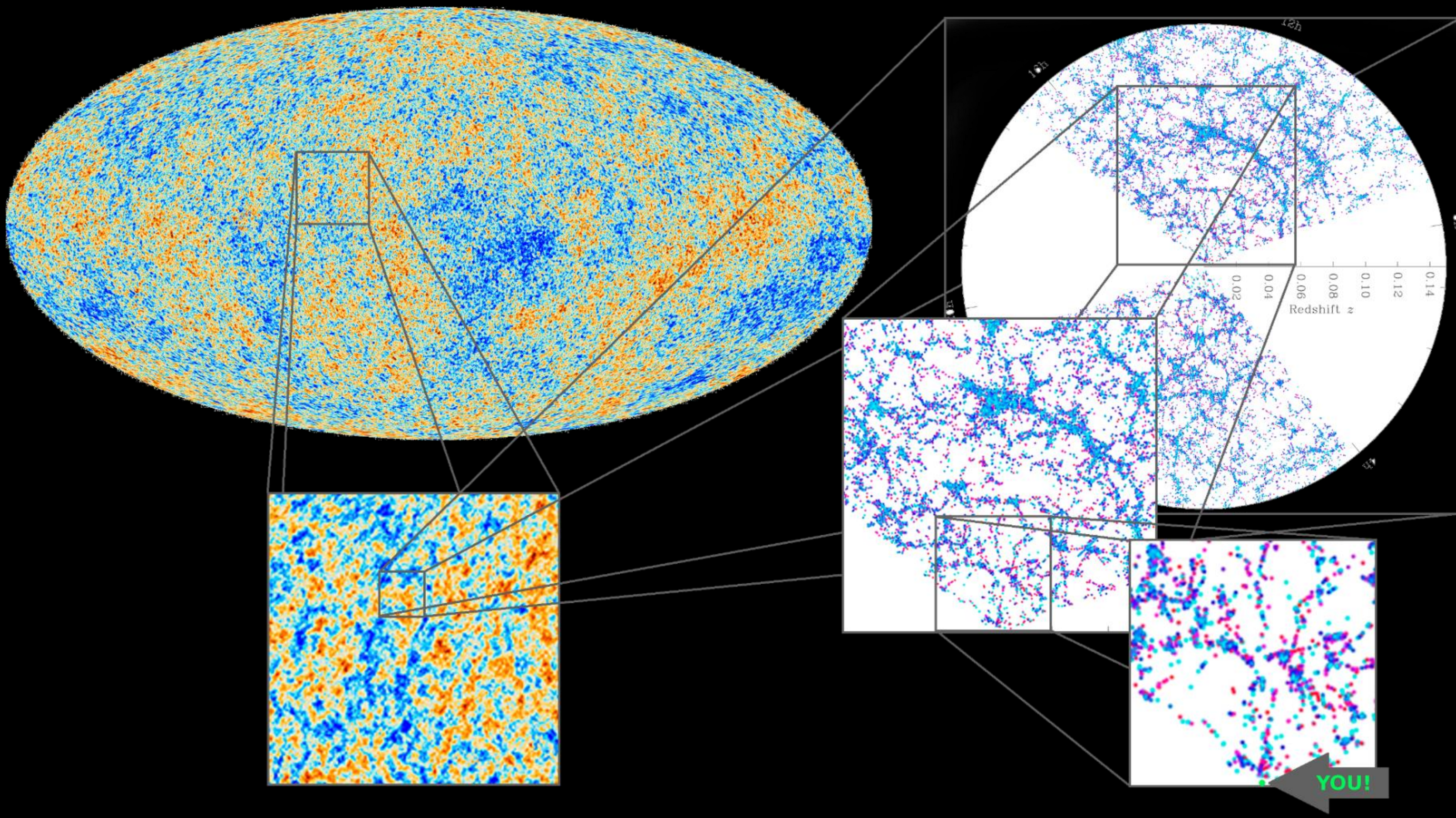
In collaboration with:
Guilhem Lavaux (IAP), Jens Jasche (Stockholm),
Benjamin Wandelt (IAP/CCA), Will Percival (Portsmouth/Waterloo)
and the Aquila Consortium
www.aquila-consortium.org

ICIC
Imperial Centre for Inference & Cosmology

Imperial College London

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)

(Even more) exciting observations coming up!

(see talks by Laureijs, Hoekstra, Meerburg)

The BORG inference framework

Bayesian Origin Reconstruction from Galaxies

- A Bayesian hierarchical model:

$\mathcal{P}(\hat{\delta}) \propto \exp\left(\frac{1}{2}\sum_k \lvert\hat{\delta}_k\rvert^2/P_k\right)$	initial conditions
$\rho_\mathrm{m} = \mathcal{F}(\delta)$	total evolved matter density
$\rho_\mathrm{g} \propto \rho_\mathrm{m}^\alpha$	biased galaxy distribution
$\rho_\mathrm{g}^\mathrm{s}(\vec{x}) = S(\vec{x})\,\rho_\mathrm{g}(\vec{x})$	selected sample
$N_\mathrm{g} \curvearrowleft \mathcal{P}(N_\mathrm{g}\lvert\rho_\mathrm{g}^\mathrm{s})$	galaxy number count: random extraction (Poisson)

- The multi-million dimensional posterior distribution is sampled via Hamiltonian Monte Carlo.

(see also talk by Heavens)

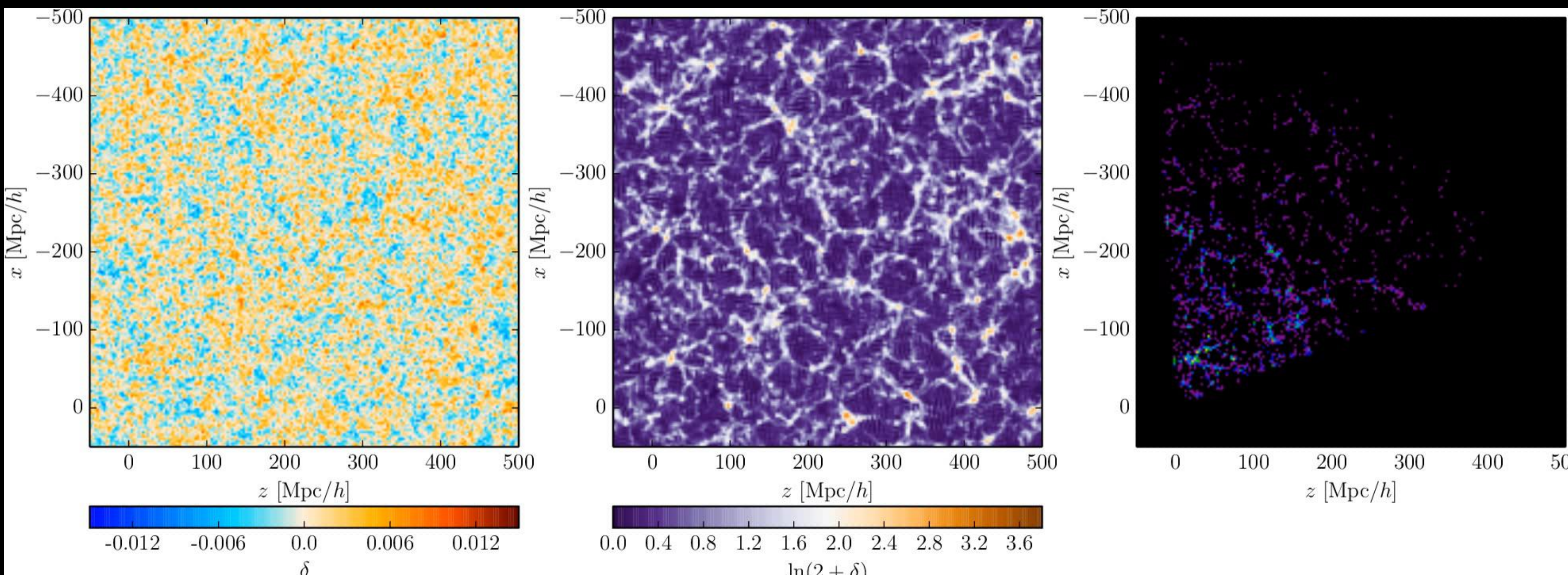
Jasche & Wandelt 2013, 1203.3639 – Jasche, FL & Wandelt 2015, 1409.6308 – Lavaux & Jasche 2016, 1509.05040 – Jasche & Lavaux 2018, 1806.11117

BORG at work: SDSS chrono-cosmography

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

Jasche, FL & Wandelt 2015, 1409.6308

Data products are publicly available at www.florent-leclercq.eu/data.php

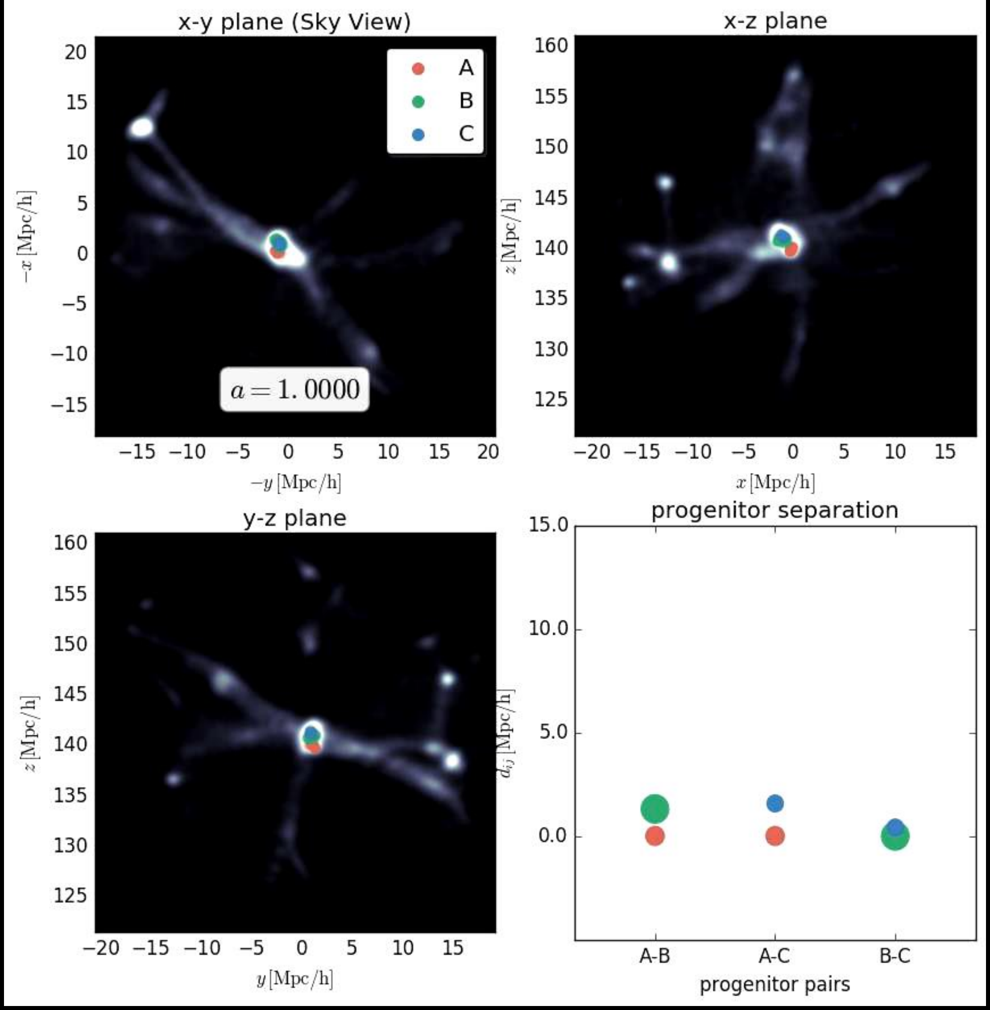


BORG infers the formation history of structures

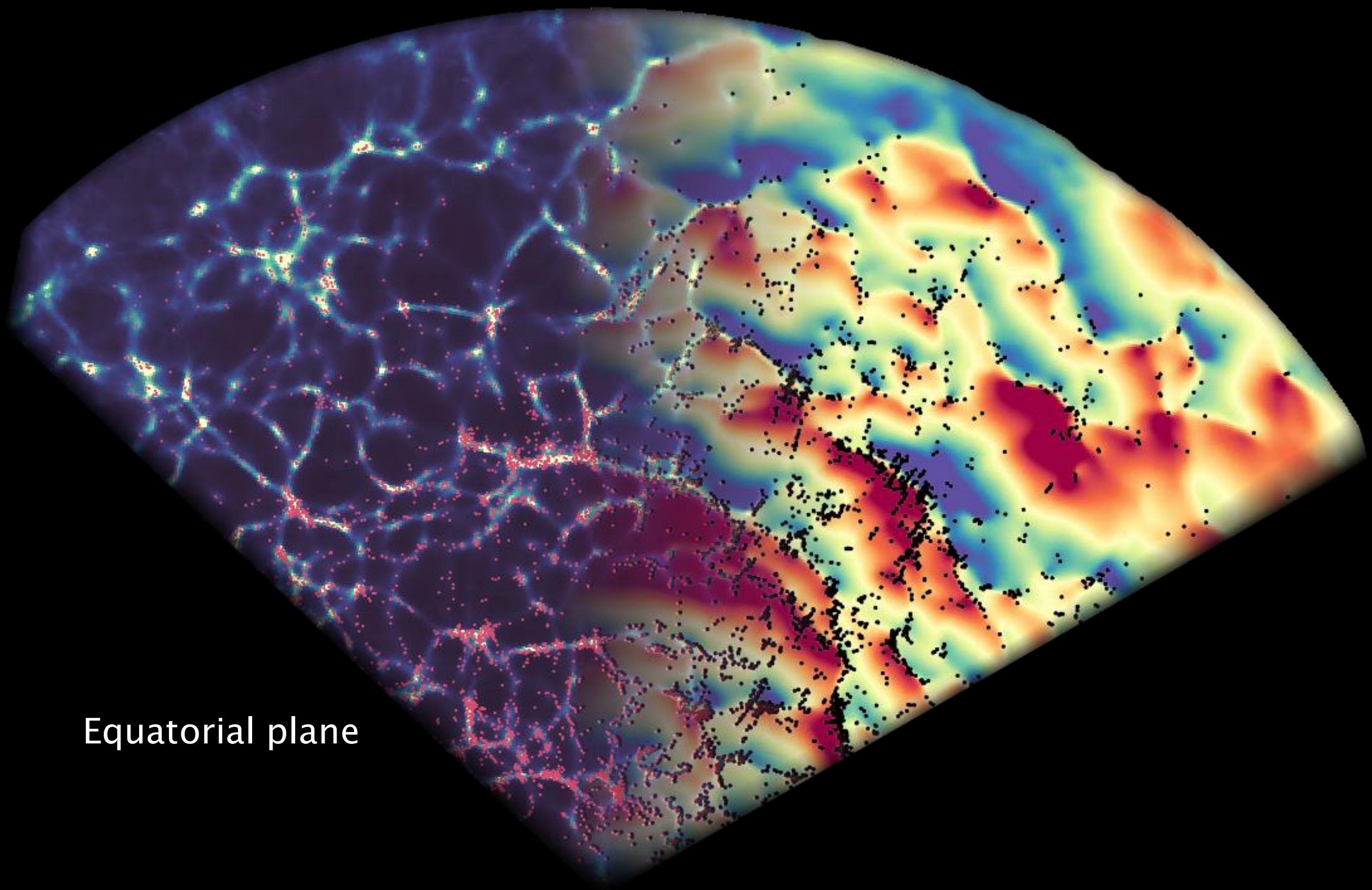
The Coma cluster

(see also talk by Hidding)

Jasche & Lavaux 2018, 1806.11117

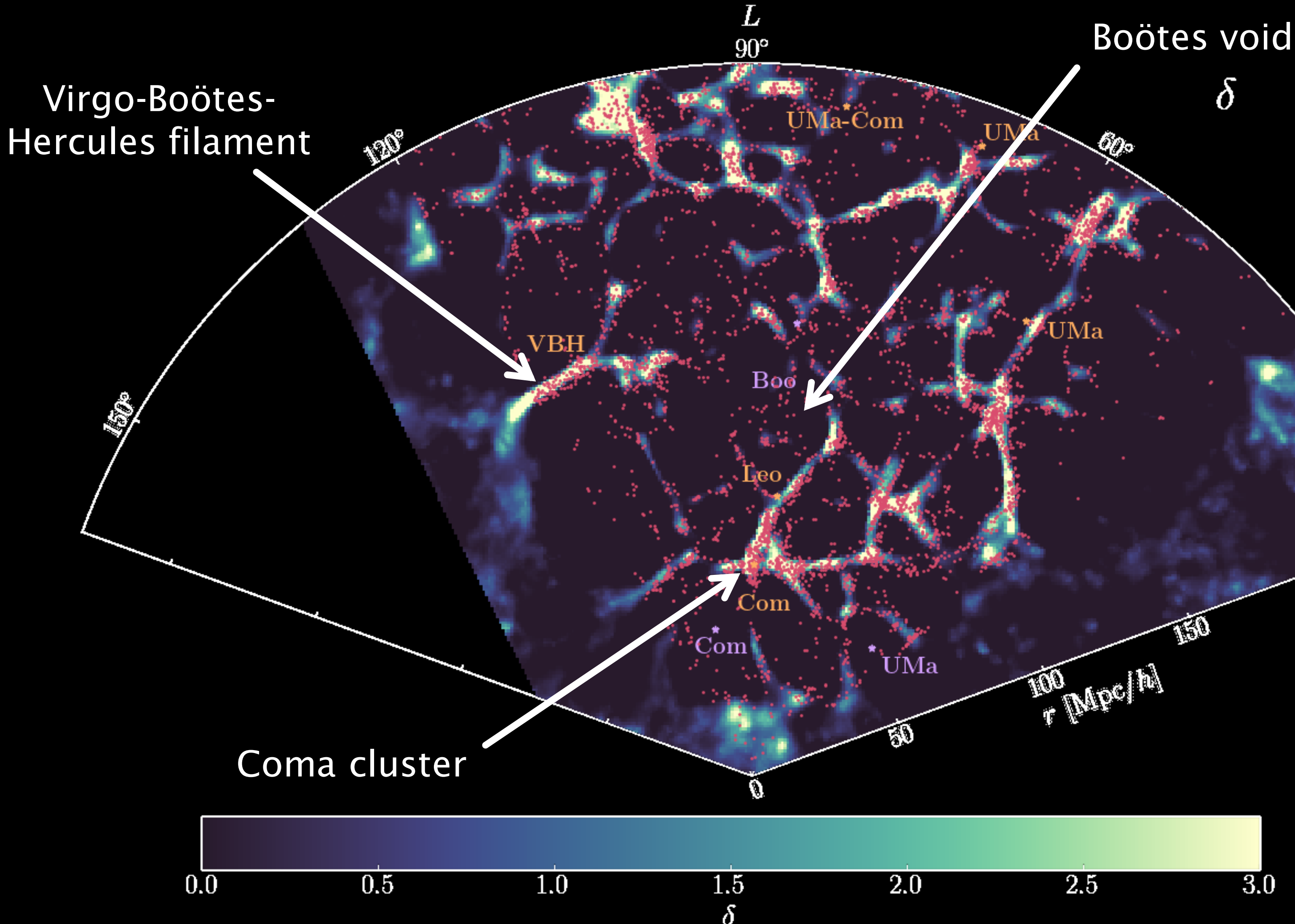


BORG unveils the dynamic cosmic web



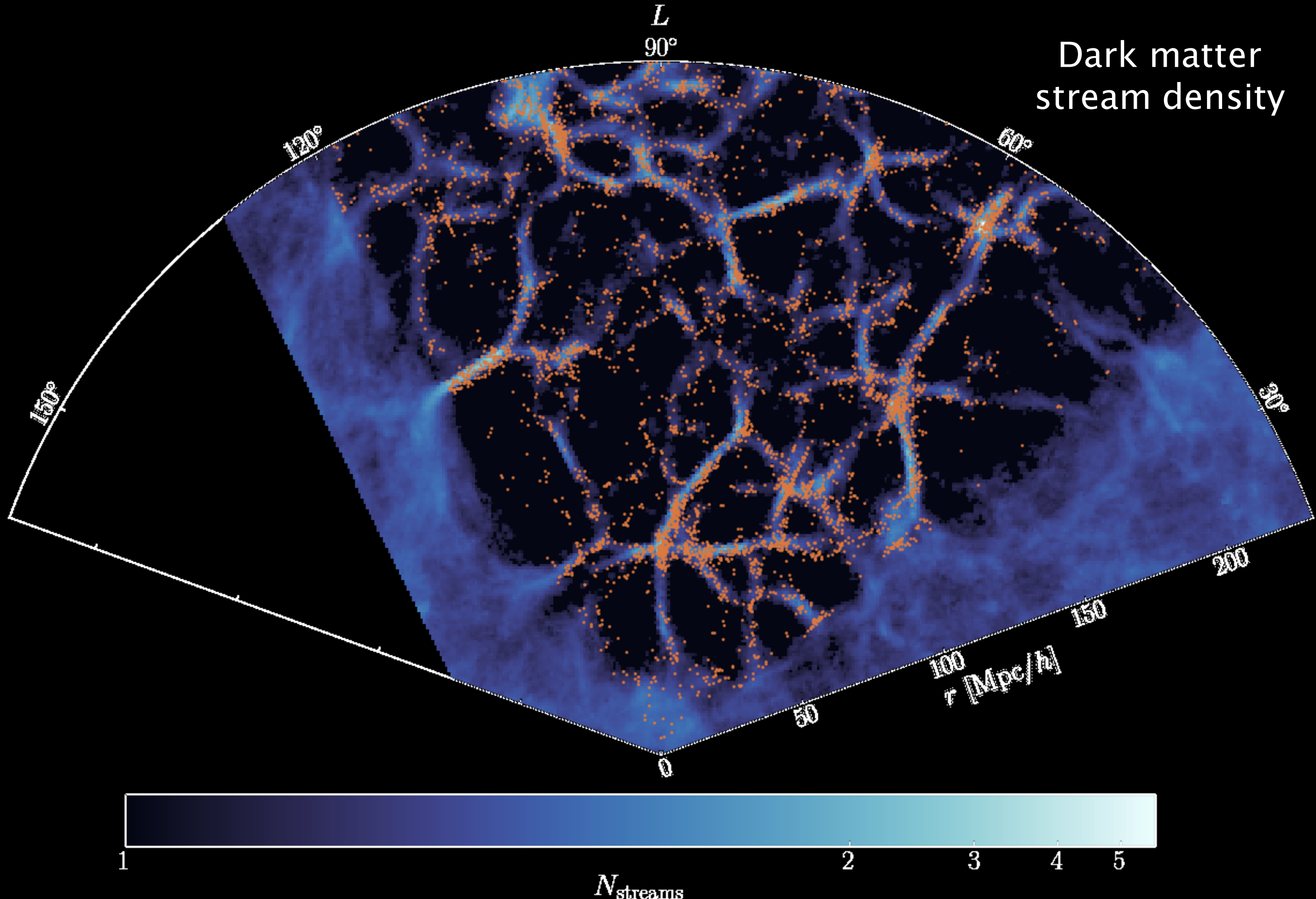
Figures from FL, Jasche, Lavaux, Wandelt & Percival 2017, 1601.00093 – Montage by Guilhem Lavaux

Cosmography in the supergalactic plane



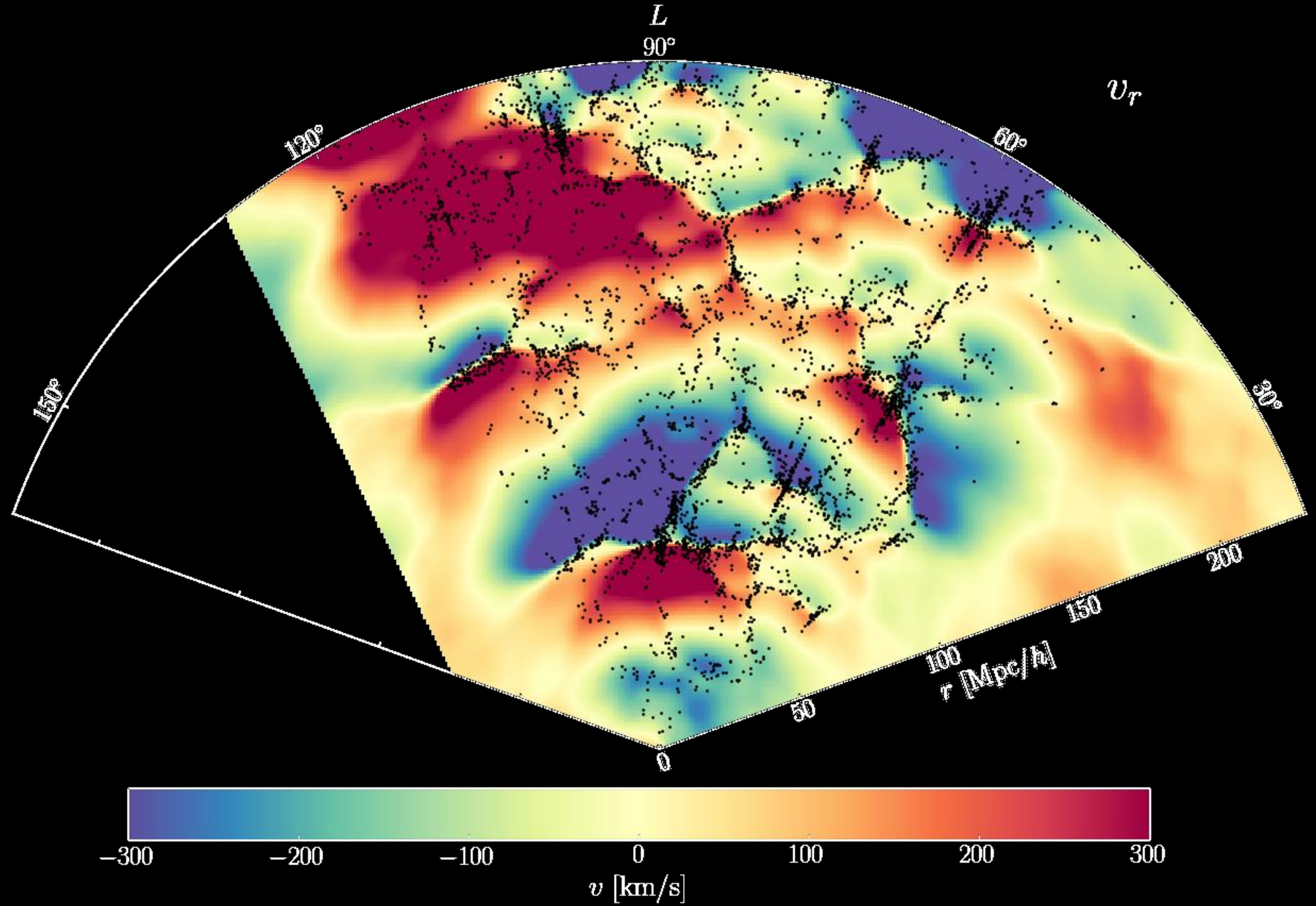
FL, Jasche, Lavaux, Wandelt & Percival 2017, 1601.00093

Cosmography in the supergalactic plane



FL, Jasche, Lavaux, Wandelt & Percival 2017, 1601.00093

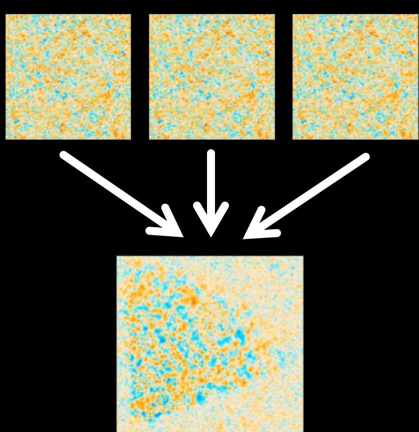
Cosmography in the supergalactic plane

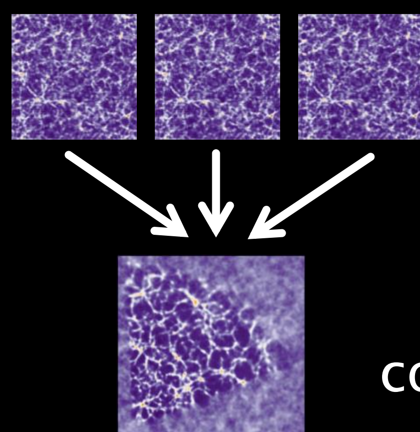


FL, Jasche, Lavaux, Wandelt & Percival 2017, 1601.00093

Uncertainty quantification

Uncertainty quantification is crucial!





Can we propagate uncertainties to cosmic web analysis?

Yes, and this is what yields a connection with information theory!

Cosmic web classification procedures

void, sheet, filament, cluster? (see also talk by Neyrinck)

- 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, 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, 0906.4101

- ORIGAMI:

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

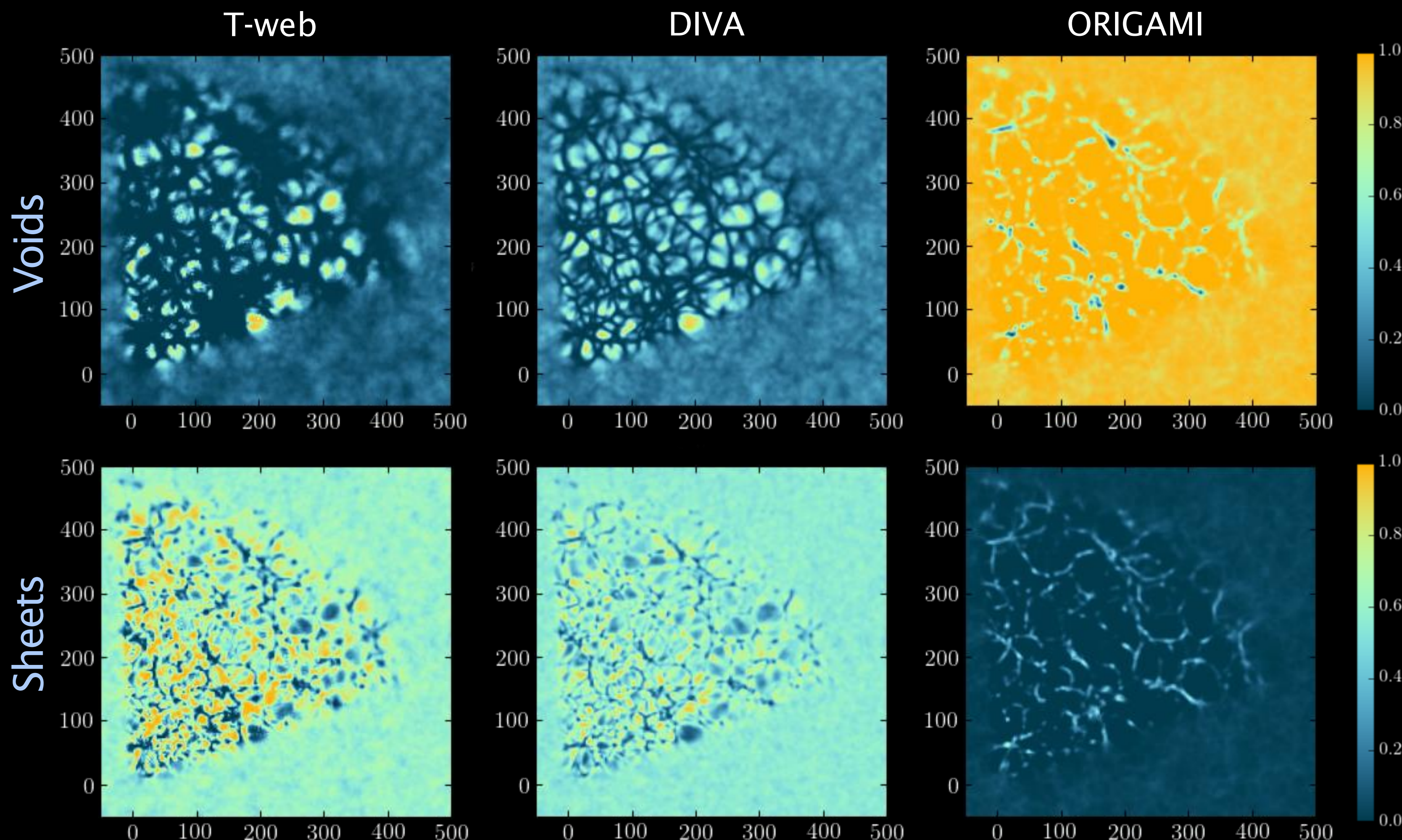
Falck, Neyrinck & Szalay 2012, 1201.2353

Lagrangian classifiers

now usable in real data!

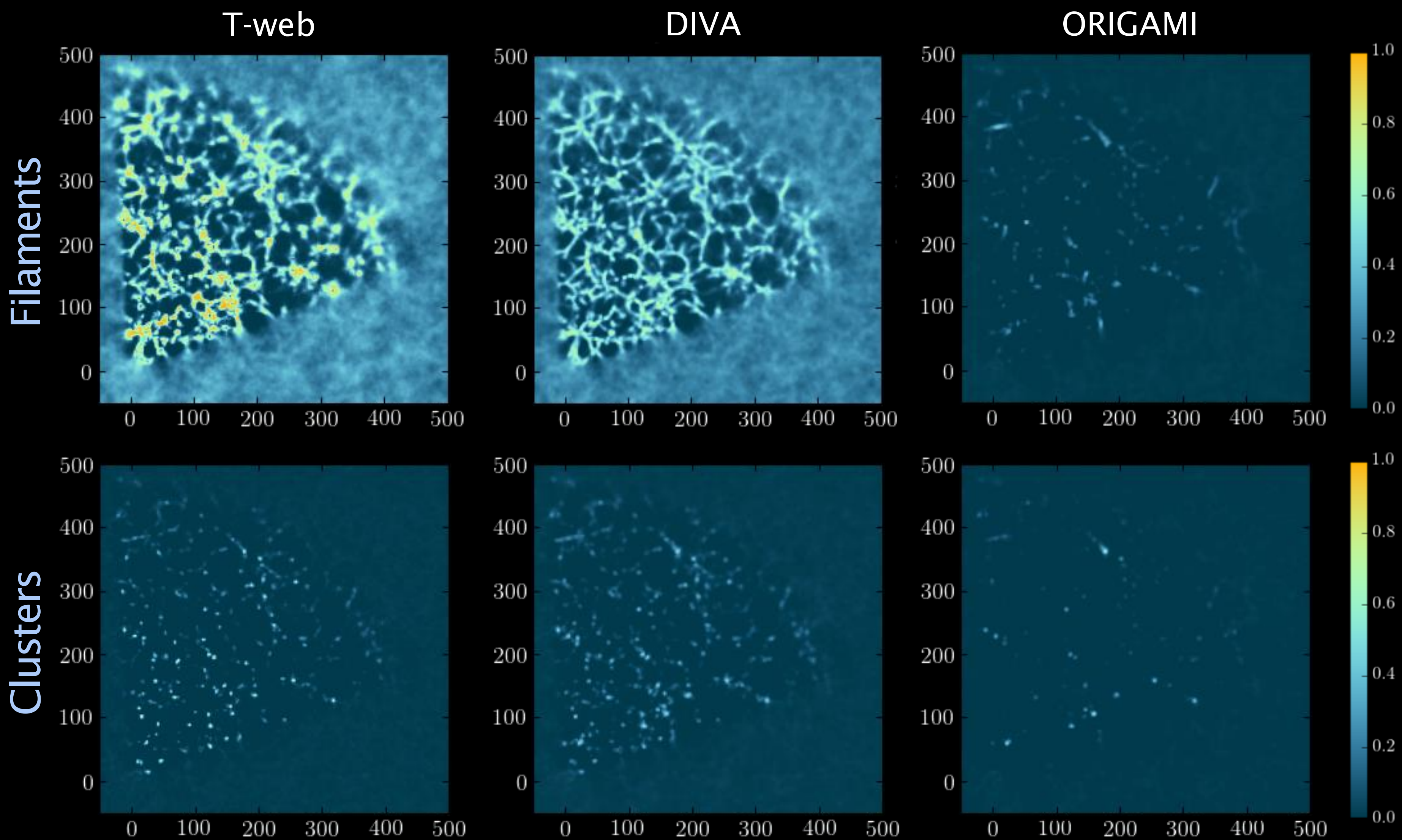
and many others…

Comparing classifiers



FL, Jasche & Wandelt 2015, 1502.02690
FL, Jasche, Lavaux & Wandelt 2016, 1601.00093

Comparing classifiers



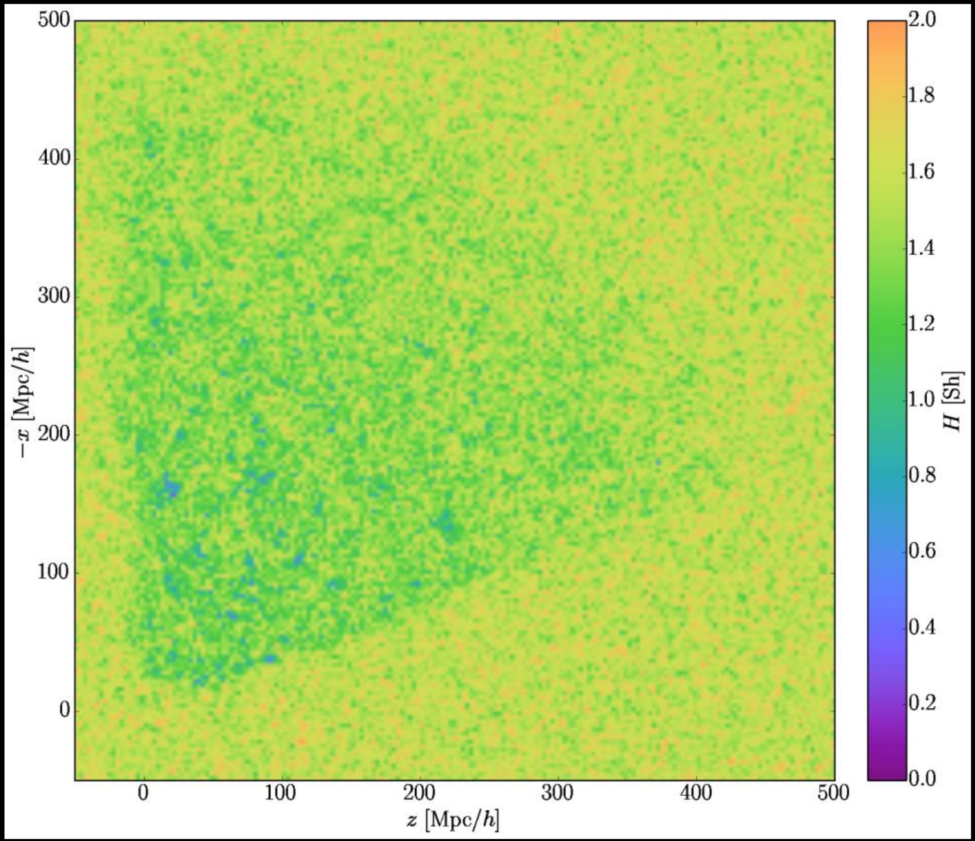
FL, Jasche & Wandelt 2015, 1502.02690
FL, Jasche, Lavaux & Wandelt 2016, 1601.00093

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)

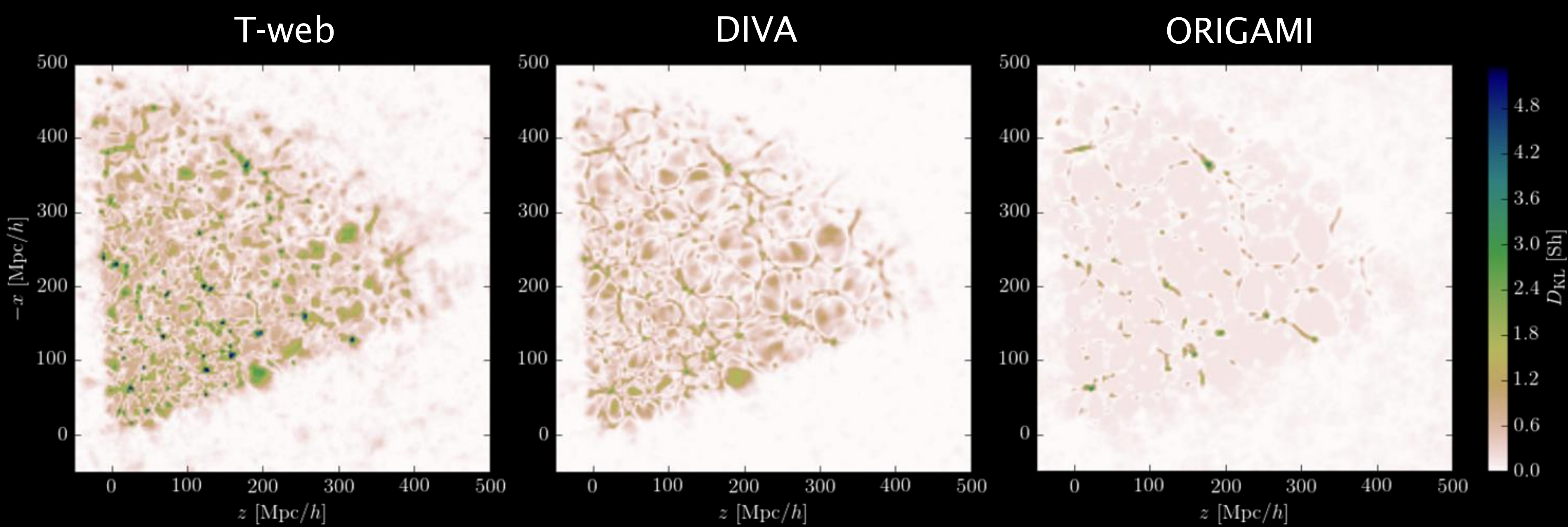
FL, Jasche & Wandelt 2015, 1502.02690



How much did the data surprise us?

Information gain a.k.a. relative entropy or Kullback-Leibler divergence

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



(more about the Kullback-Leibler divergence later)

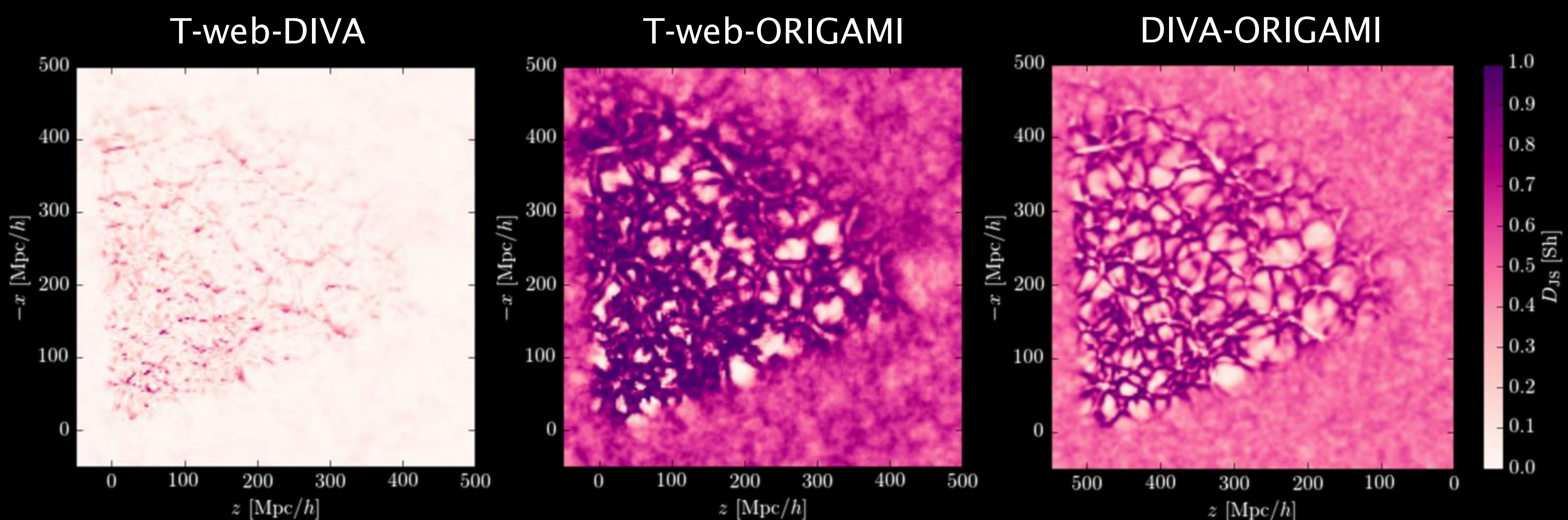
FL, Jasche & Wandelt 2015, 1502.02690

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, 1606.06758

Which is the best classifier?

- Decision theory: a framework to classify structures in the presence of uncertainty. FL, Jasche & Wandelt 2015, 1503.00730
  Can we extend the decision problem to the space of classifiers?

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

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

FL, Lavaux, Jasche & Wandelt 2016, 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: maximise the expected information gain for cosmic web maps

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

$$U_1(\xi) = I[\mathrm{T} : d|\xi]$$

classification data

FL, Lavaux, Jasche & Wandelt 2016, 1606.06758

2. Utility for model selection

example: dark energy equation of state

(see also talk by Silvestri)

- 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, 1606.06758

3. Utility for prediction of new data

example: galaxy colours

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

- Maximize the expected information gain for some new quantity

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

$$U_3(\xi) = I[c : \mathrm{T}|\xi]$$

predicted data classification

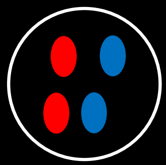
FL, Lavaux, Jasche & Wandelt 2016, 1606.06758

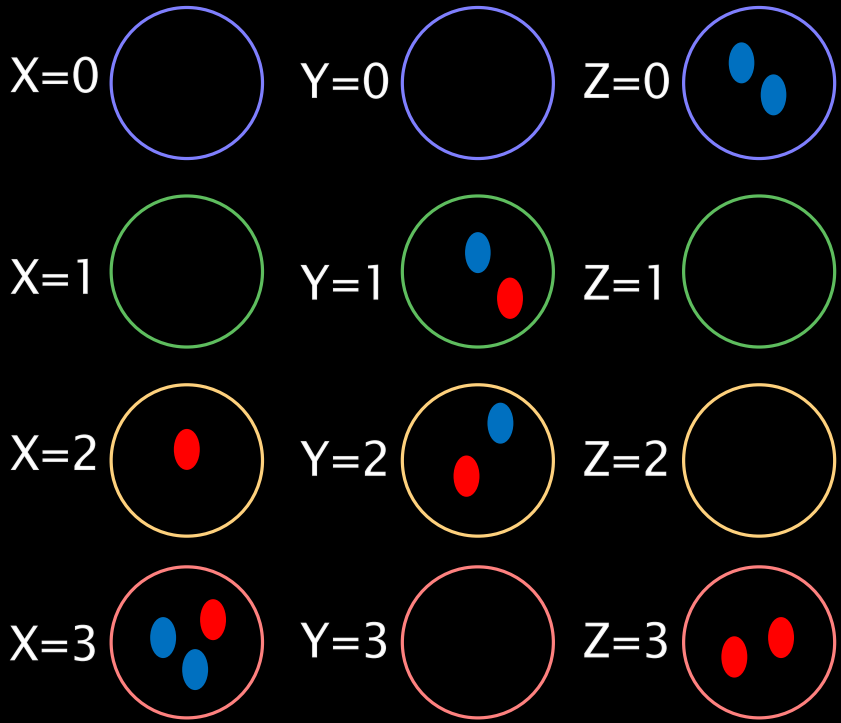
3. Utility for prediction of new data

example: galaxy colours

- 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!

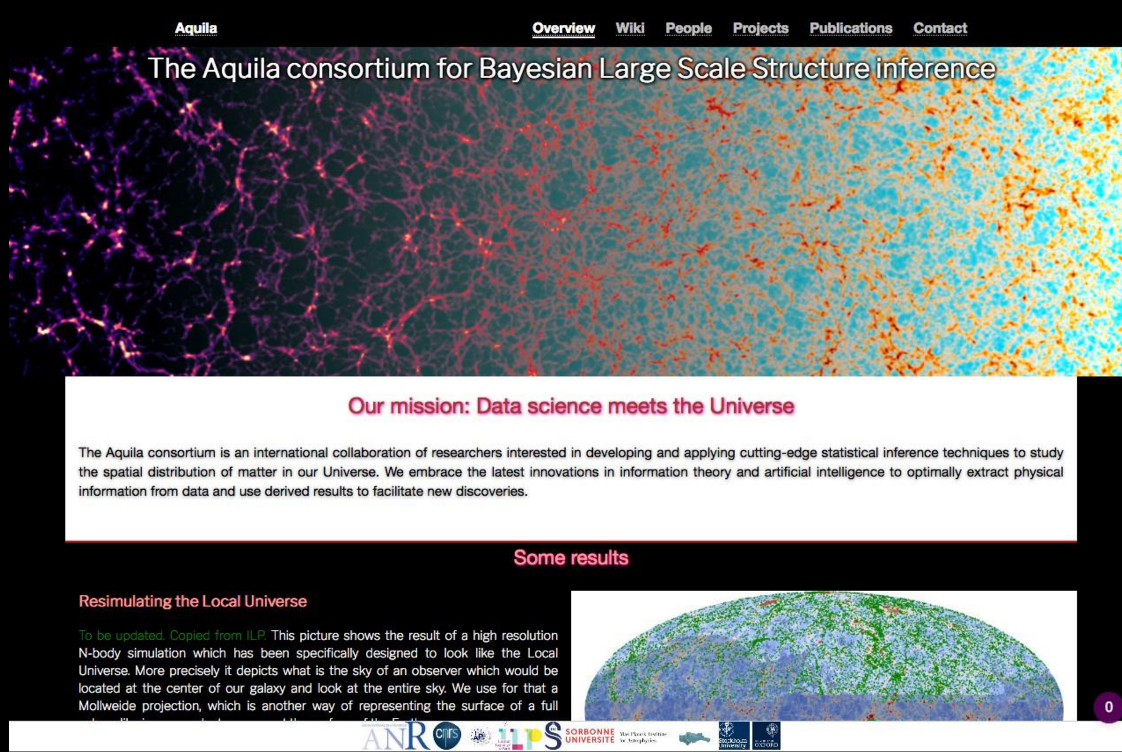
FL, Lavaux, Jasche & Wandelt 2016, 1606.06758

The Aquila Consortium

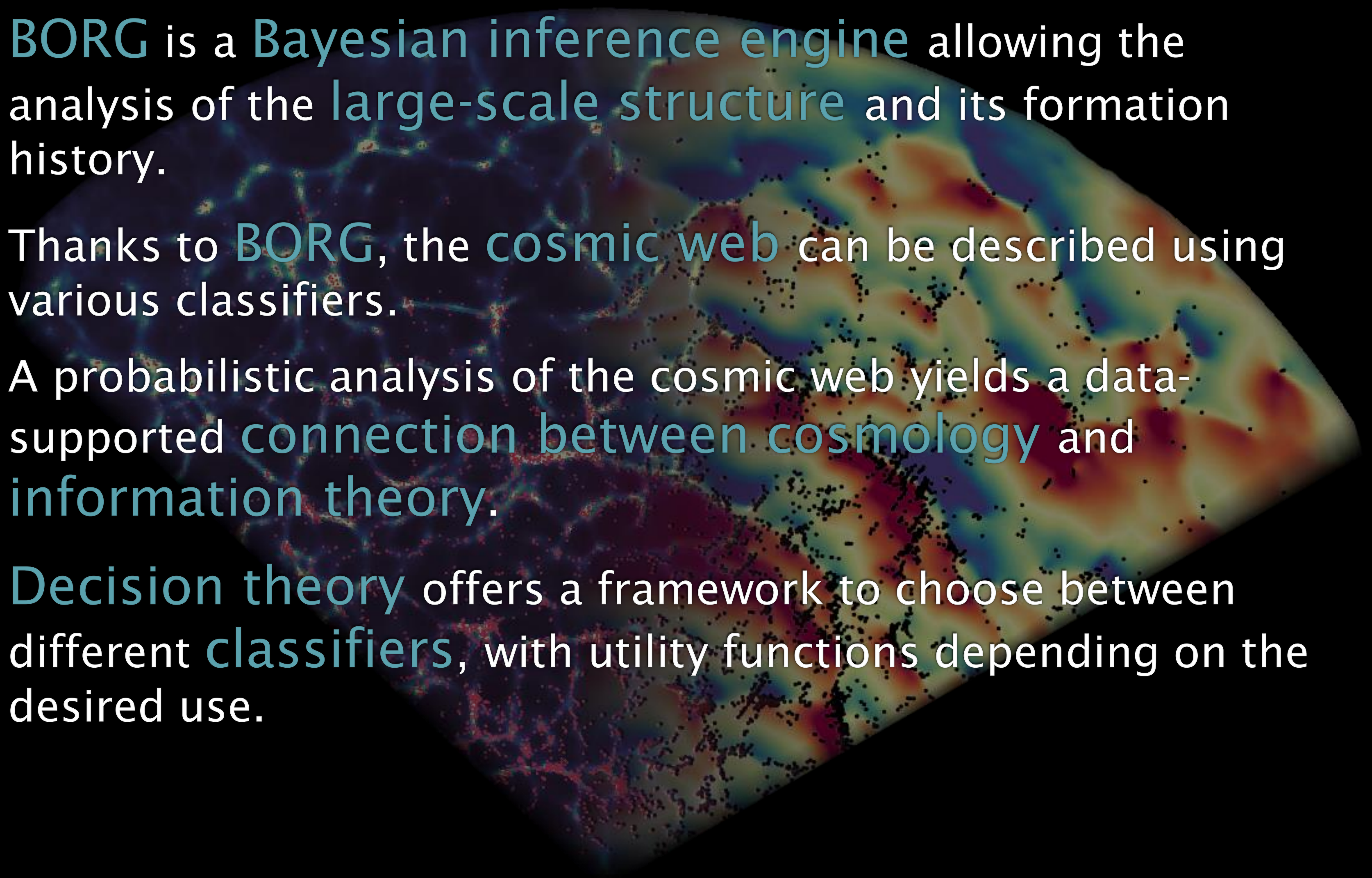
for Bayesian large-scale structure inference

- Created in 2016. Members from the UK, France, Germany & Sweden.
- Gathers people interested in developing the Bayesian pipelines and running analyses on cosmological data.

www.aquila-consortium.org



Conclusions



- BORG is a Bayesian inference engine allowing the analysis of the large-scale structure and its formation history.
- Thanks to BORG, the cosmic web can be described using various classifiers.
- A probabilistic analysis of the cosmic web yields a data-supported connection between cosmology and information theory.
- Decision theory offers a framework to choose between different classifiers, with utility functions depending on the desired use.