Cosmic web analysis and information theory

some recent results

Florent Leclercq

Institute of Cosmology and Gravitation, University of Portsmouth

ICG Portsmouth

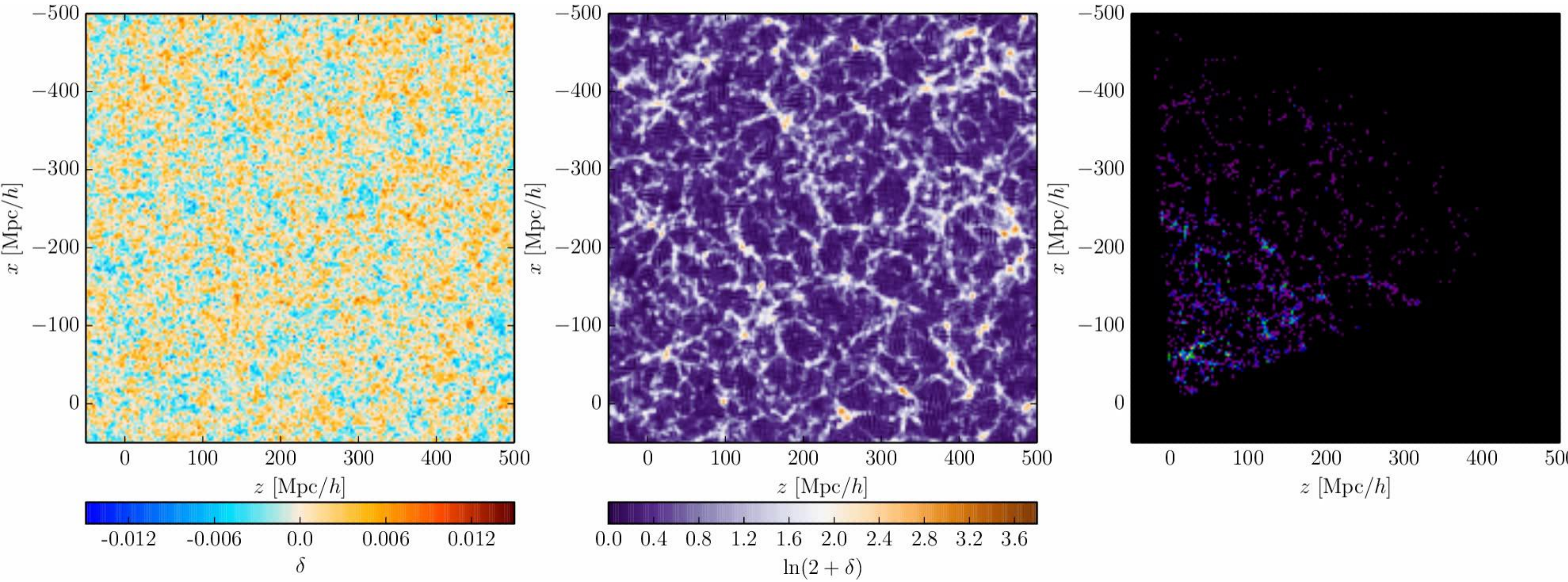
University of Portsmouth

June 10th, 2016

In collaboration with:

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

BORG at work: SDSS chrono-cosmography



Initial conditions | Final conditions | Observations

The BORG SDSS run:

334,074 galaxies, ≈ 17 millions parameters, 12,000 samples, 3 TB, 10 months on 32 cores

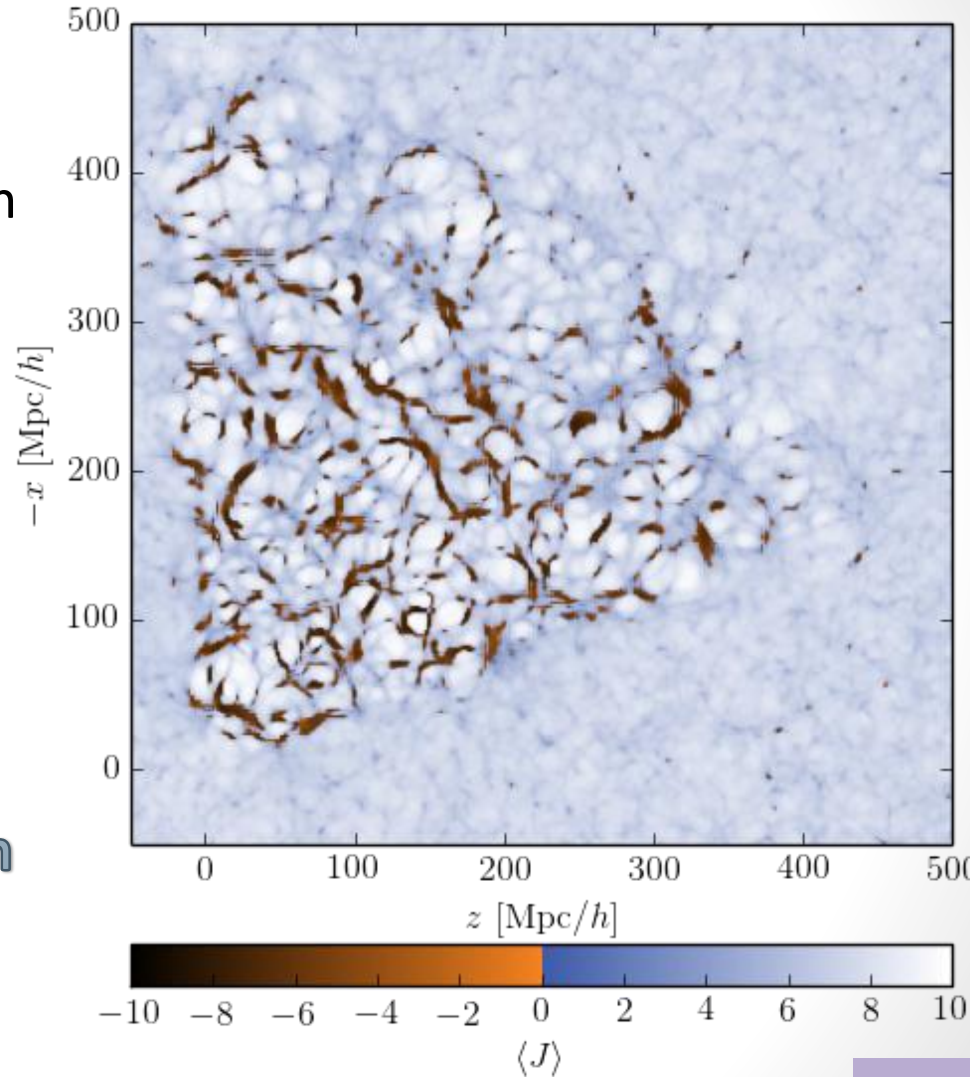
Jasche, FL & Wandelt 2015, arXiv:1409.6308

Inference of the dark matter phase-space sheet

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

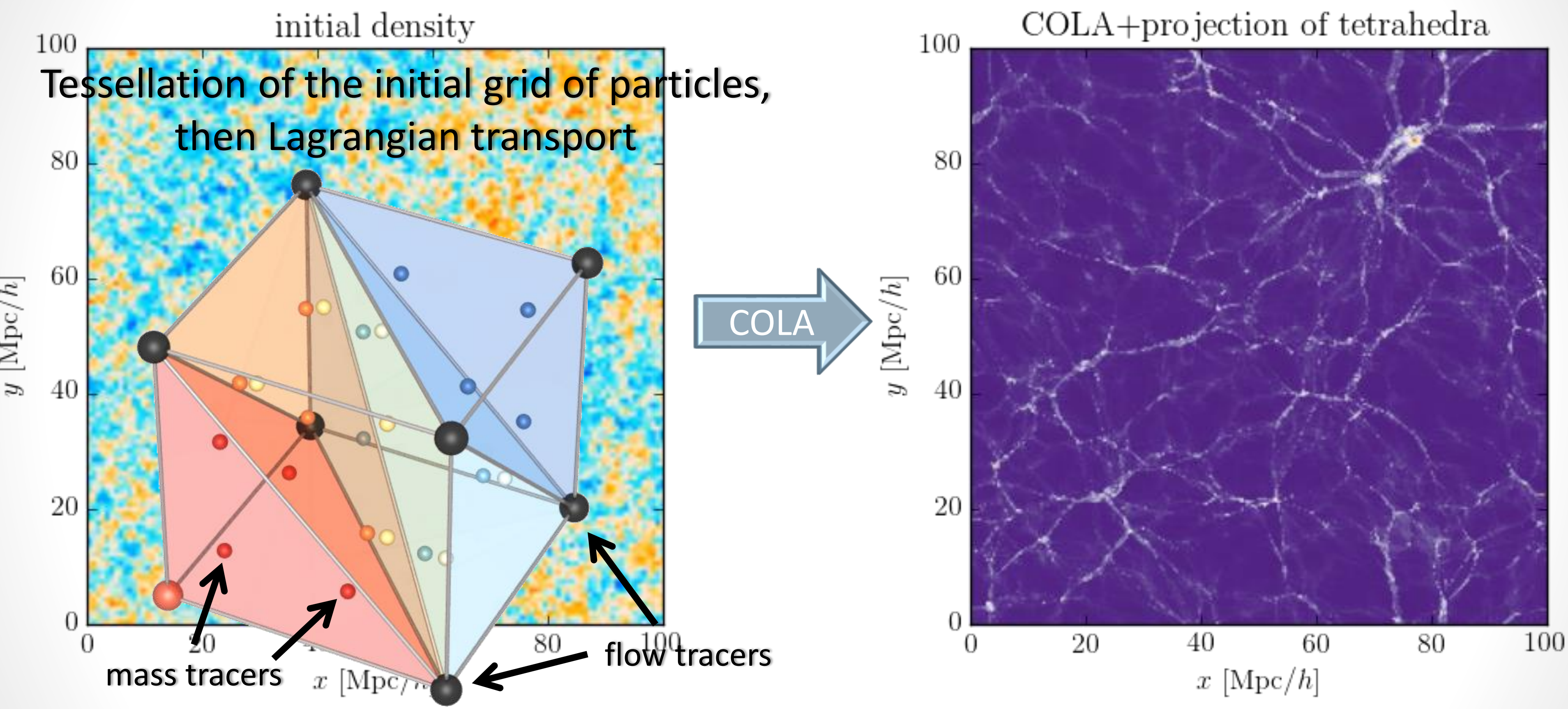
e.g. Neyrinck 2012, arXiv:1202.3364
Abel, Hahn & Kaehler 2012, arXiv:1111.3944
Shandarin, Habib & Heitmann 2012, arXiv:1111.2366

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



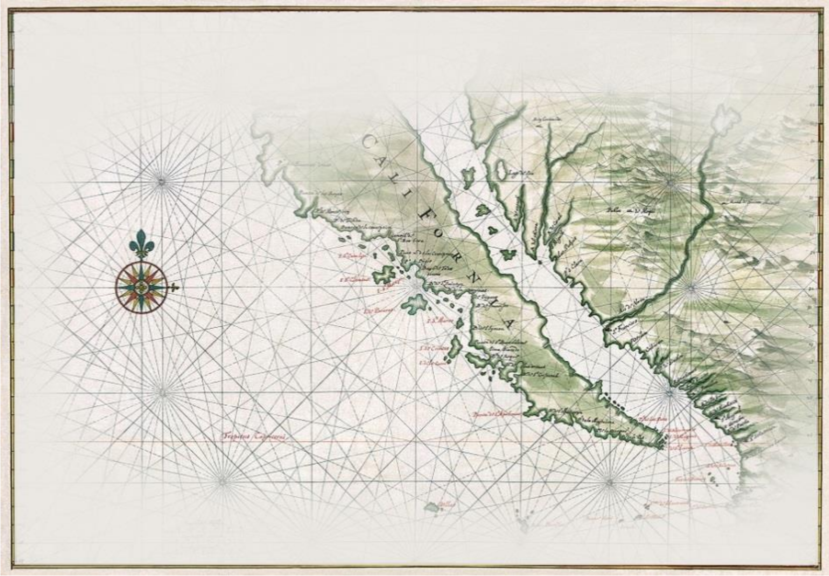
FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093

Non-linear filtering improves density samples

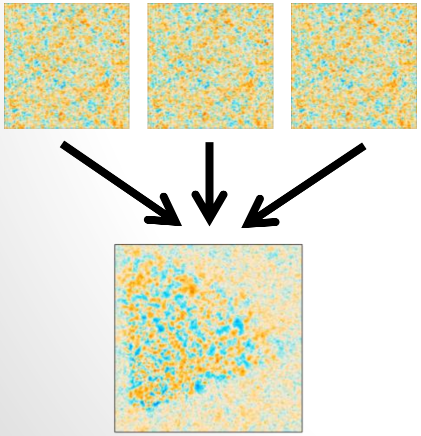


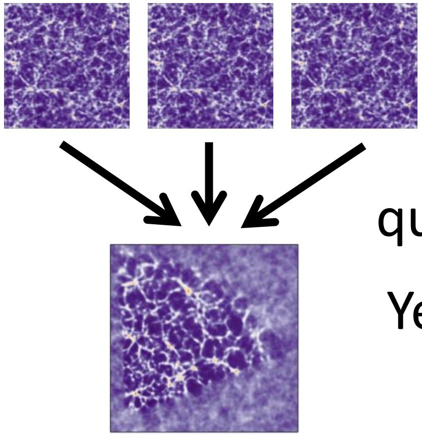
Hahn, Abel & Khaeler, arXiv:1210.6652

Uncertainty quantification



Uncertainty quantification is crucial!





Can we **propagate** uncertainty quantification to **cosmic web analysis**?

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

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)

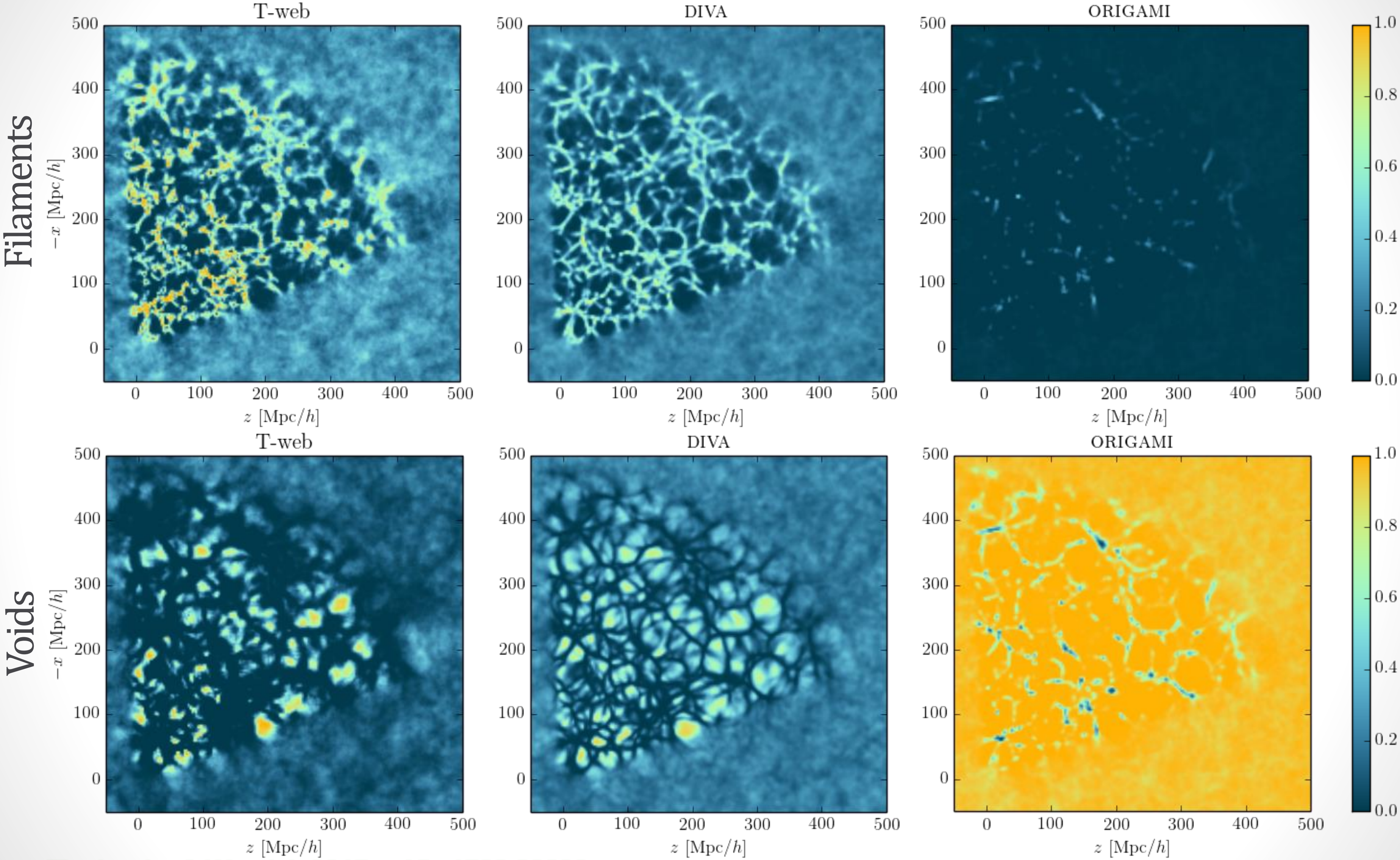
Falck, Neyrinck & Szalay 2012, arXiv:1201.2353

Lagrangian classifiers

now usable in real data!

and many others…

Comparing classifiers



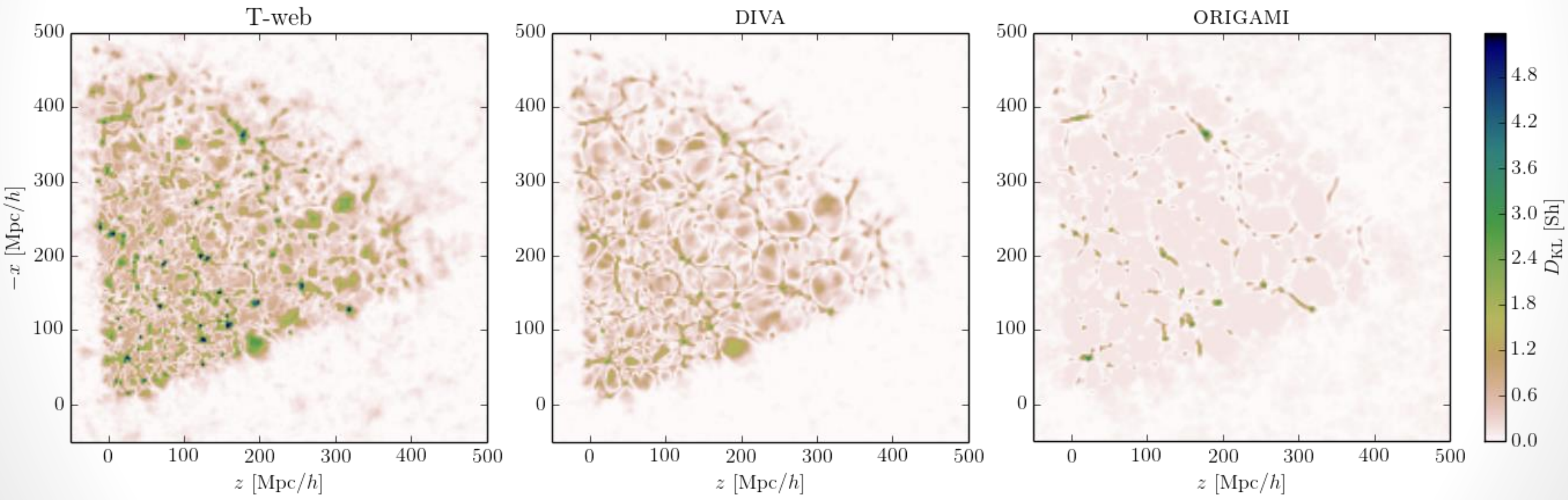
FL, Jasche & Wandelt 2015, arXiv:1502.02690

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

How much did the data surprise us?

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

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



(more about the Kullback-Leibler divergence later)

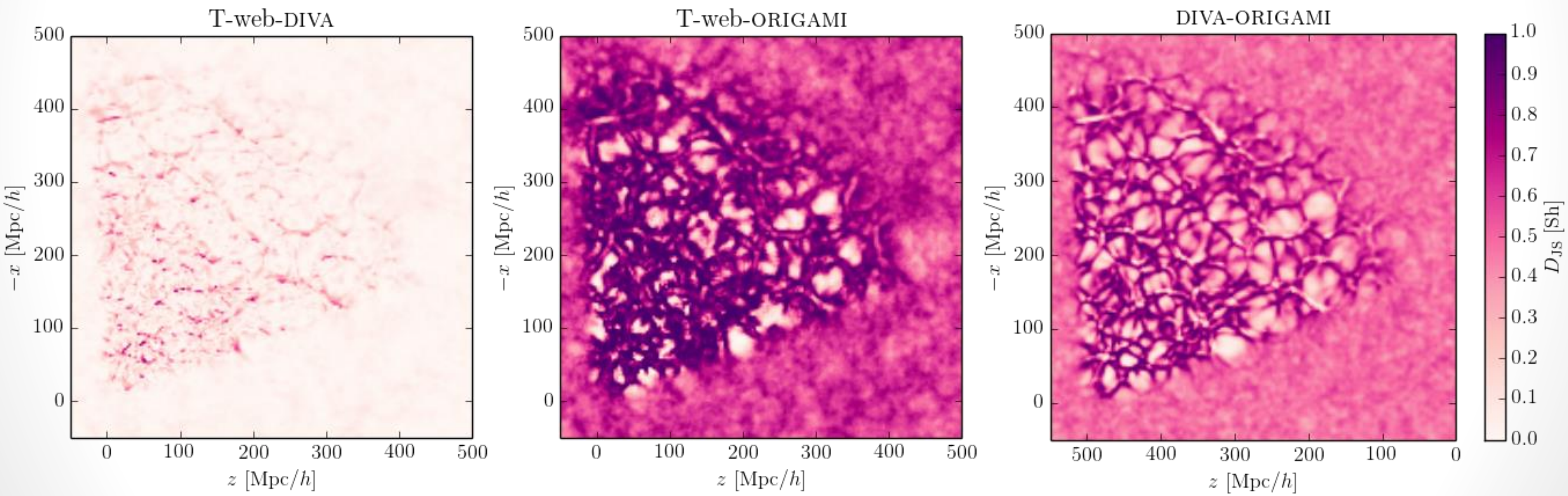
FL, Jasche & Wandelt 2015, arXiv:1502.02690

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, in prep.

Which is the best classifier?

- **Decision theory**: a framework to classify structures in the presence of uncertainty. FL, Jasche & Wandelt 2015, arXiv: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: $\boxed{I\,[X:Y] = \langle D_{\mathrm{KL}}[\mathcal{P}(x|y)||\mathcal{P}(x)]\rangle_{\mathcal{P}(Y)}}$

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

FL, Lavaux, Jasche & Wandelt 2016, in prep.

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

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

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

FL, Lavaux, Jasche & Wandelt 2016, in prep.

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

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

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, in prep.

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

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

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

FL, Lavaux, Jasche & Wandelt 2016, in prep.

Conclusions

- Thanks to **BORG**, the **cosmic web** can be described using various classifiers.
- 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.

(Some numerical results for classifier utilities in the upcoming paper)

References

Jasche & Wandelt 2013, arXiv:1203.3639 (BORG proof of concept)
Jasche, FL & Wandelt 2015, arXiv:1409.6308 (BORG SDSS analysis)
FL, Jasche & Wandelt 2015, arXiv:1502.02690 (T-web, entropy, relative entropy)
FL, Jasche & Wandelt 2015, arXiv:1503.00730 (decision theory)
FL, Jasche, Lavaux & Wandelt 2016, arXiv:1601.00093 (phase-space sheet, DIVA & ORIGAMI)
FL, Lavaux, Jasche & Wandelt 2016, in prep. (very soon) (mutual information, classifier utilities)