Cosmic web analysis and information theory

some recent results

Florent Leclercq

Institute of Cosmology and Gravitation, University of Portsmouth

May 26th, 2016

In collaboration with:

Jens Jasche (ExC Universe, Garching), Guilhem Lavaux (IAP),

Will Percival (ICG), Benjamin Wandelt (IAP/U. Illinois)

Previously in COSMO21…

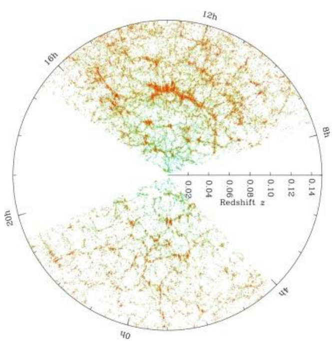
STATISTICAL CHALLENGES in 21st CENTURY COSMOLOGY

IAU SYMPOSIUM 306 Lisbon Portugal 25-29 May 2014

Bayesian, physical large-scale structure inference from galaxy surveys

BORG: *Bayesian Origin Reconstruction from Galaxies*

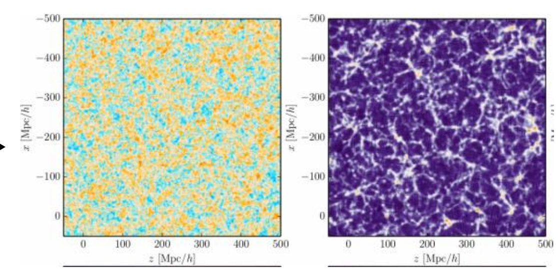
- **Data model**: Gaussian prior – Second-order Lagrangian perturbation theory (2LPT) – Poisson likelihood (and also: luminosity-dependent galaxy bias, automatic noise level calibration)
- **Sampler**: Hamiltonian Markov Chain Monte Carlo method



Observations

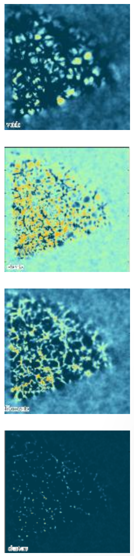
(galaxy catalog + meta-data: selection functions, completeness…)

→ **BORG** →



Inferred dark matter density

→



Cosmic web analysis

Jasche & Wandelt 2013, arXiv:1203.3639

Jasche, FL & Wandelt 2015, arXiv:1409.6308

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)

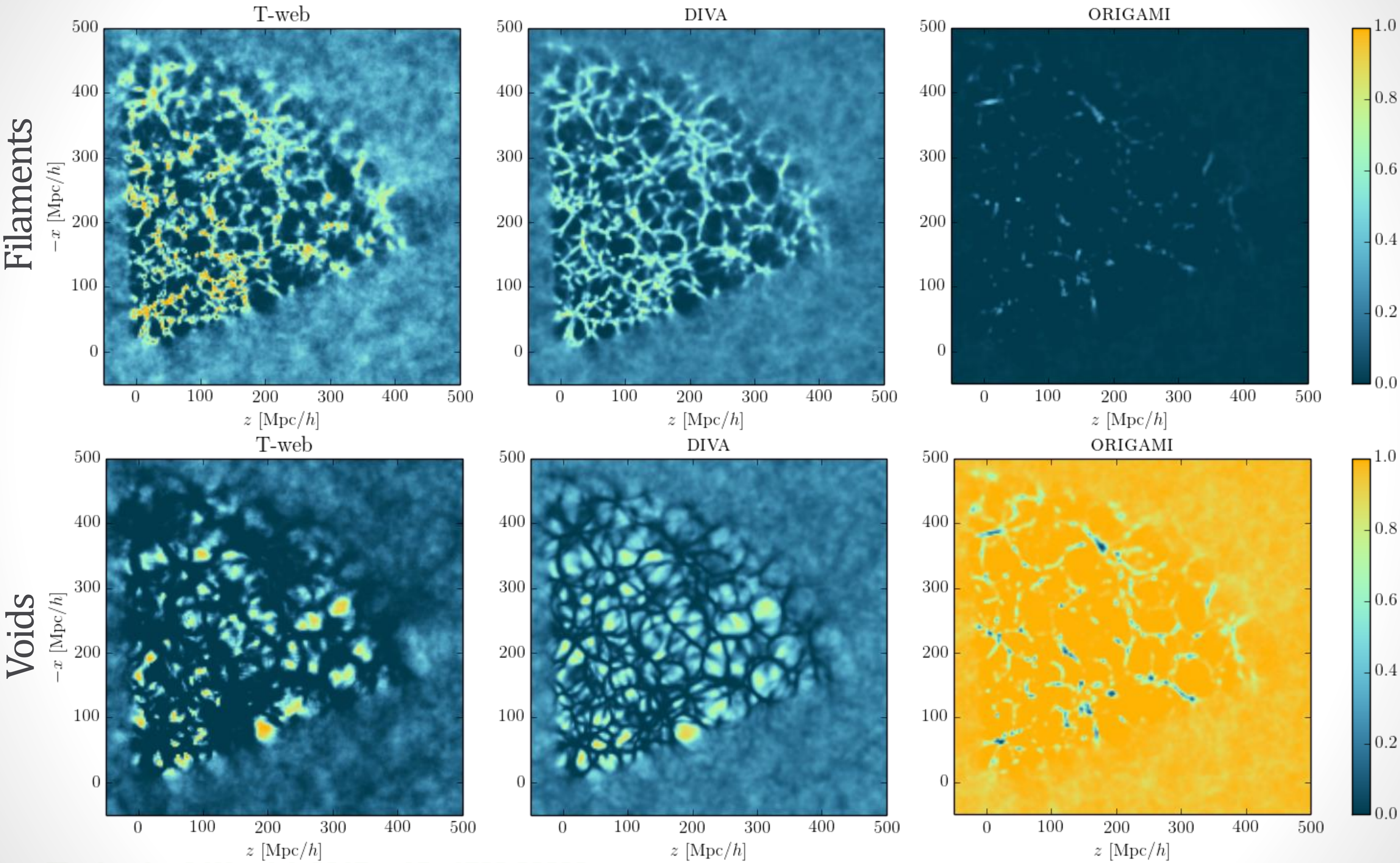
Falck, Neyrinck & Szalay 2012, arXiv:1201.2353

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

A decision rule for structure classification

- Space of "input features":

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

- Space of "actions":

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

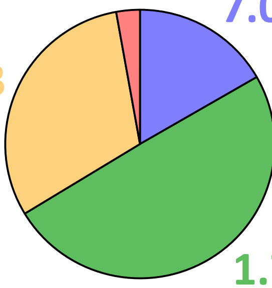
⇨ A problem of **Bayesian decision theory**:
one should take the action that maximizes the utility

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

- How to write down the gain functions?

FL, Jasche & Wandelt 2015, 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

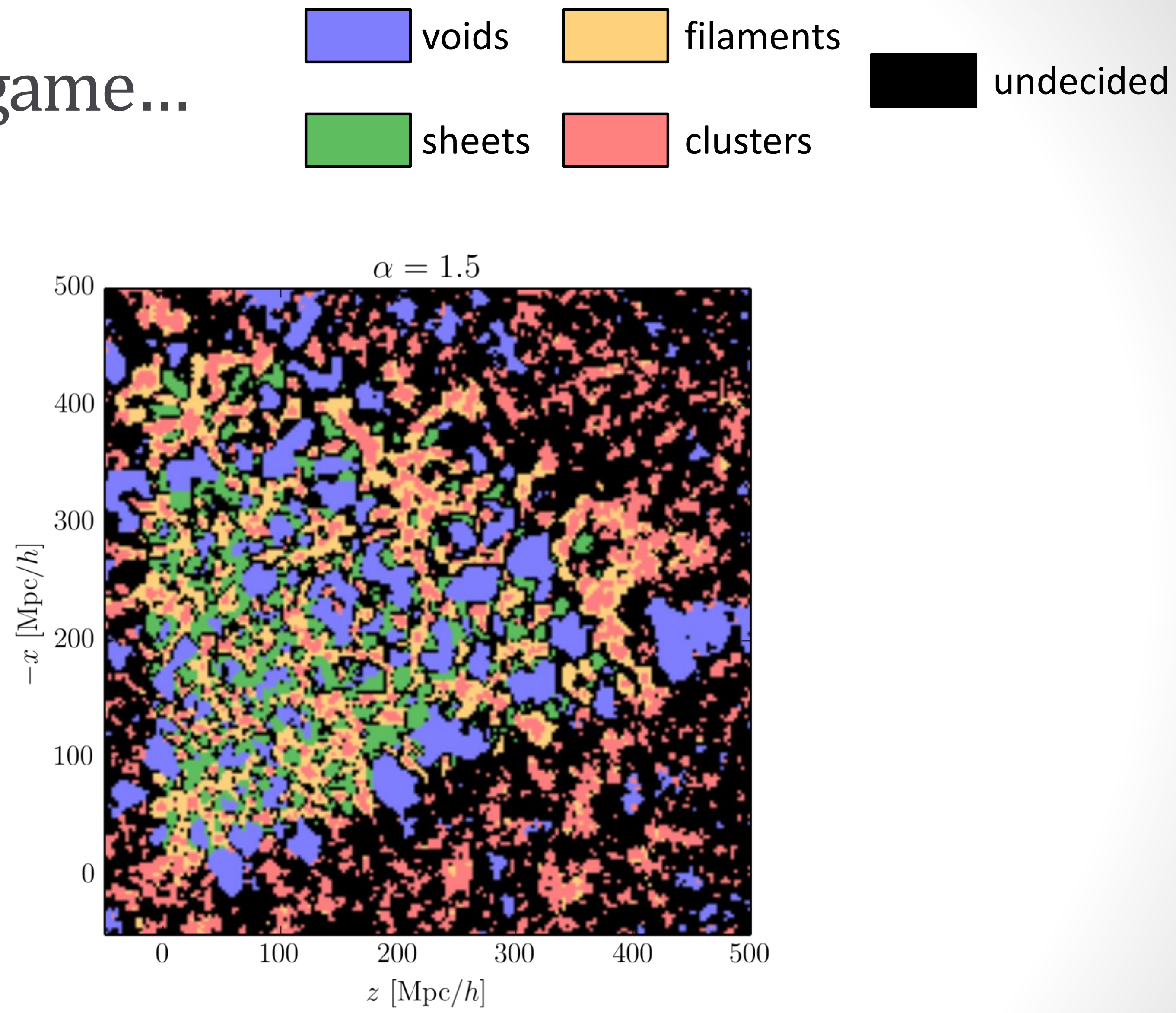
$$U(a_j) = 1 - \alpha \quad \text{if } j \neq 1$$ "Playing the game"

$$U(a_{-1}) = 0$$ "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 2015, arXiv:1503.00730

Playing the game…

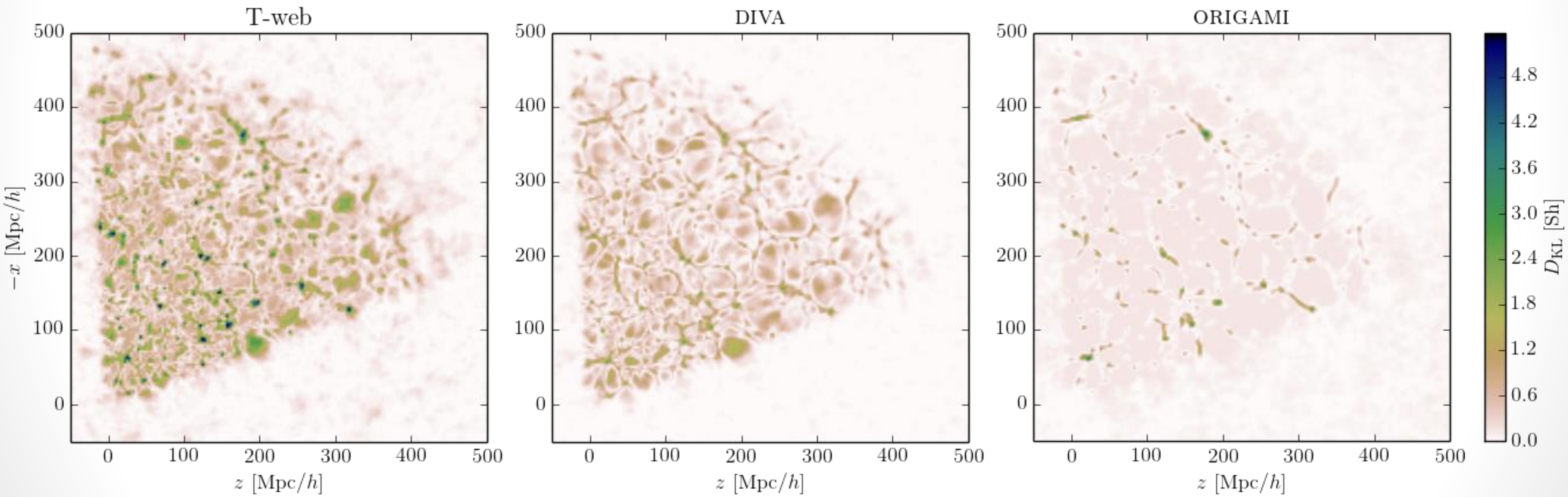


FL, Jasche & Wandelt 2015, arXiv:1503.00730

How much did the data surprise us?

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

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



(more about the Kullback-Leibler divergence later)

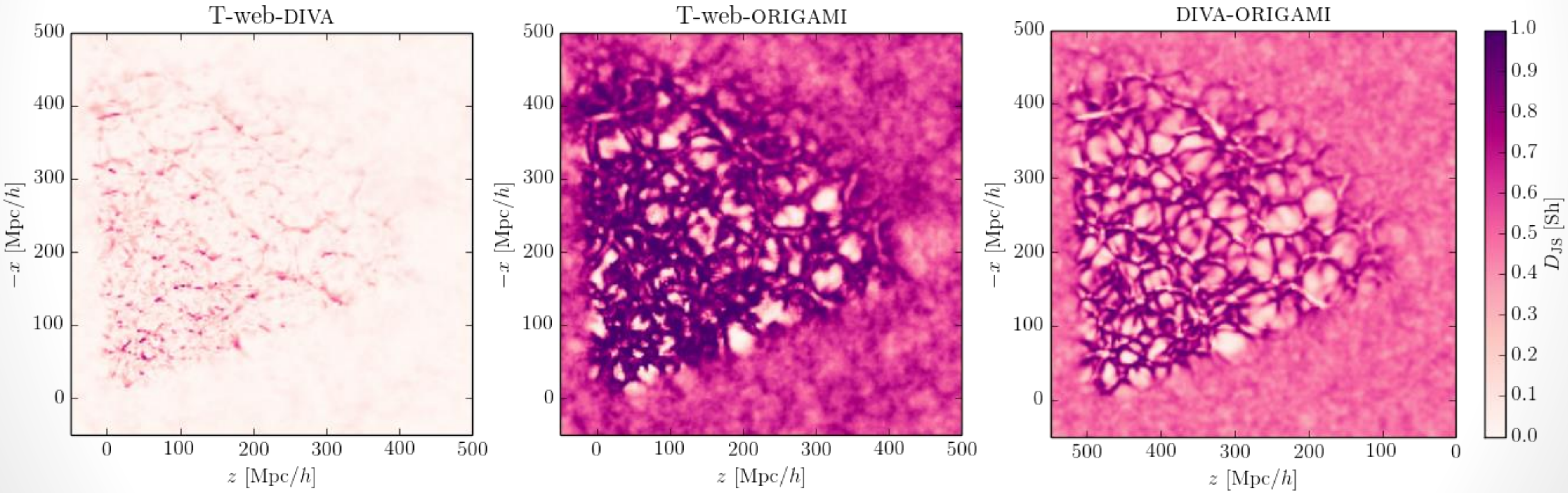
FL, Jasche & Wandelt 2015, arXiv:1502.02690

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?

- 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

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

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

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

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

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

classification (T), data (d)

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

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

$\mathcal{M}$: model classifier; $\mathcal{R}(d)$: 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

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

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

predicted data (c) — classification (T)

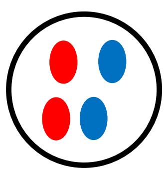
FL, Lavaux, Jasche & Wandelt 2016, in prep.

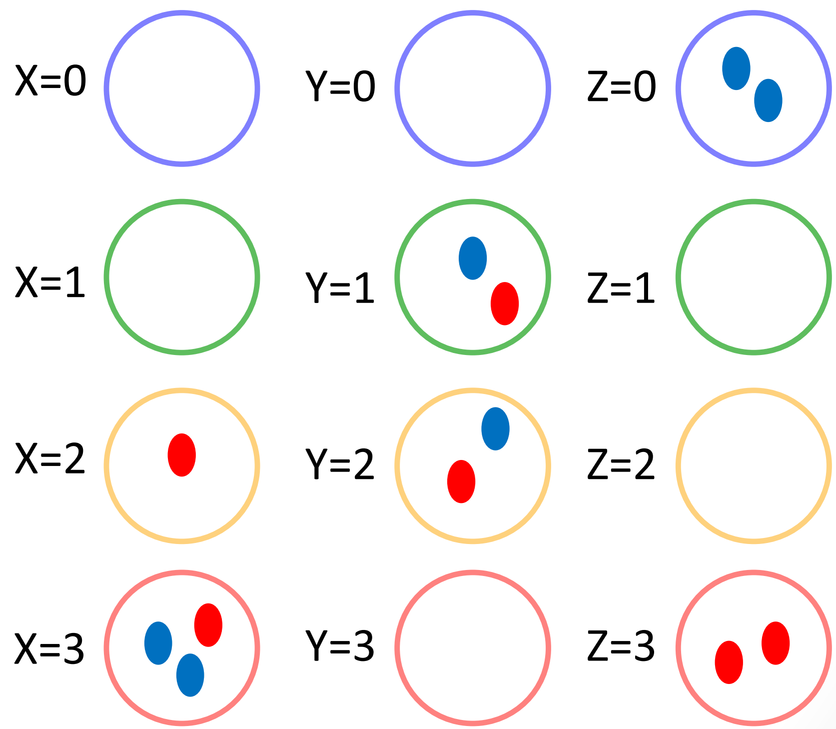
3. Utility for prediction of new data:

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, 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 classify structures in the presence of uncertainty.
- The decision problem can be extended to the **space of 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 (DIVA & ORIGAMI)
FL, Lavaux, Jasche & Wandelt 2016, in prep. (very soon) (mutual information, classifier utilities)