Cosmic web analysis and information theory

some recent results

Florent Leclercq

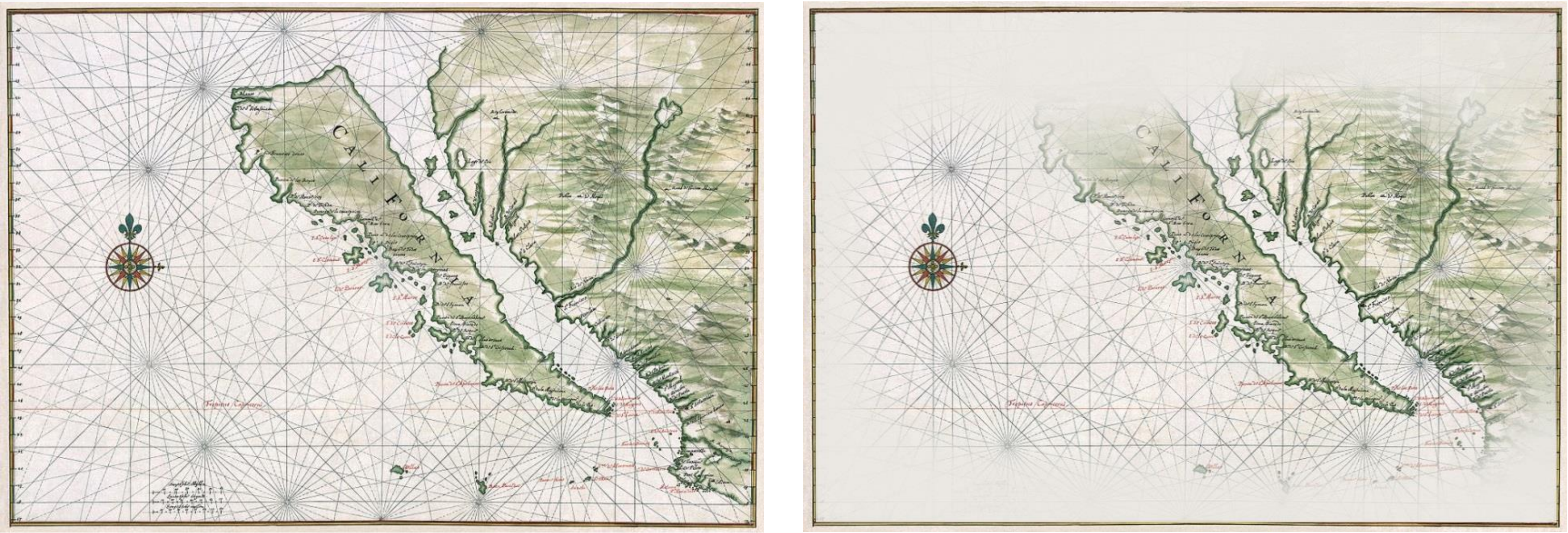
Institute of Cosmology and Gravitation, University of Portsmouth

January 5th, 2016

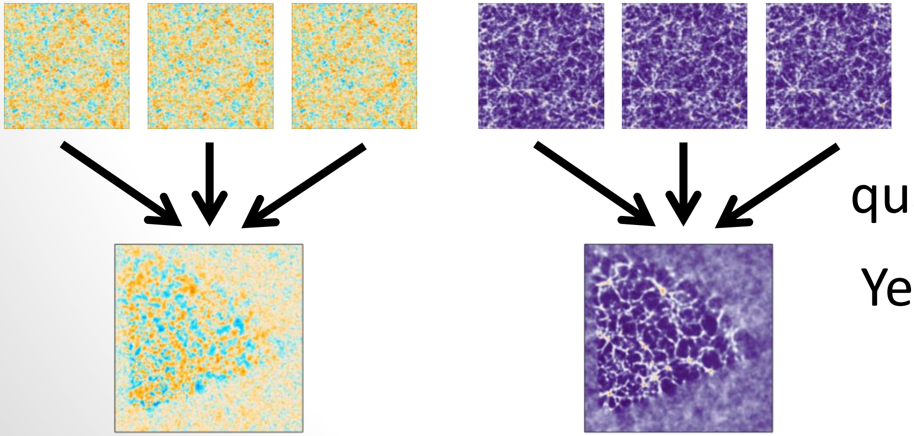
In collaboration with:

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

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?

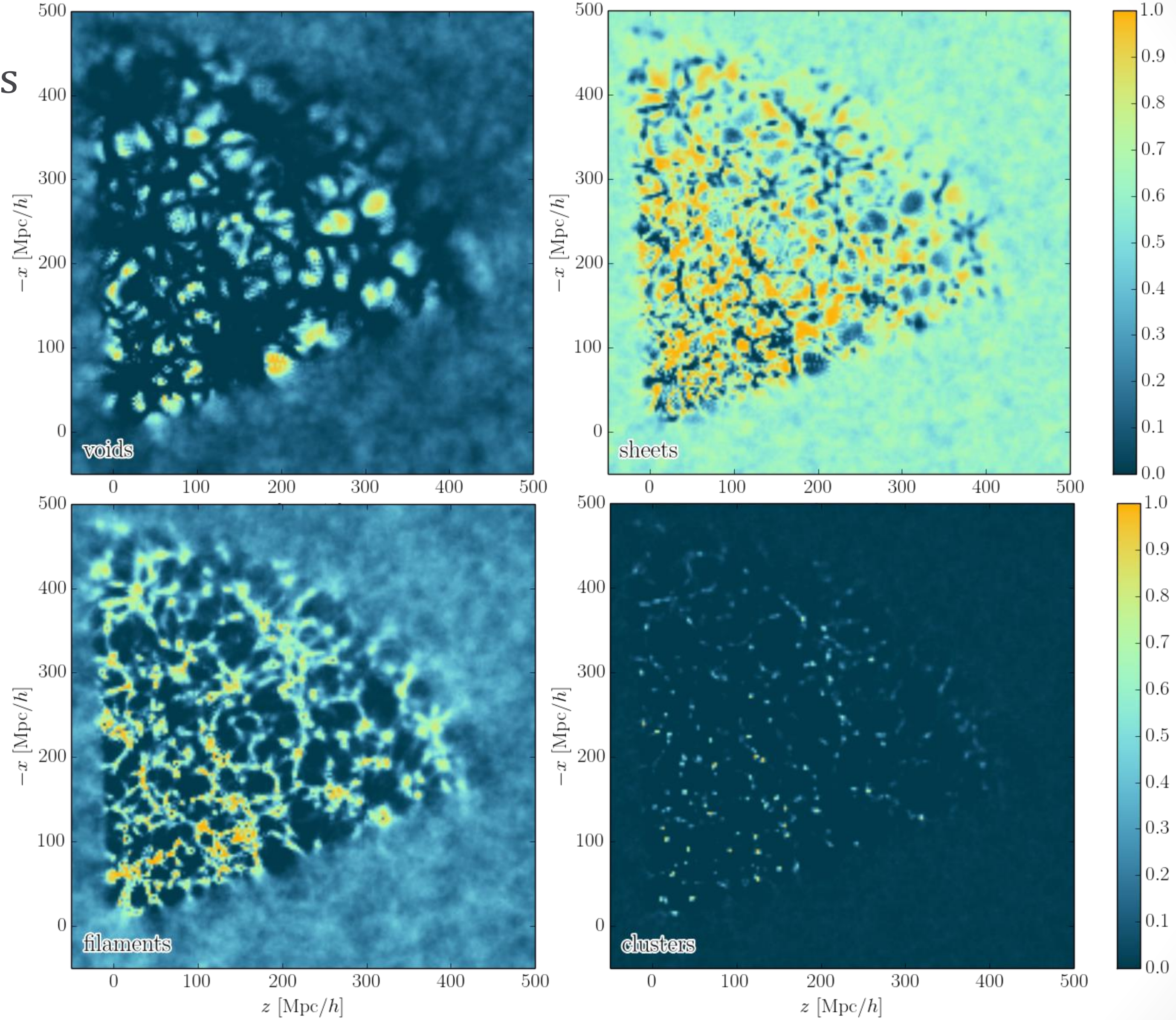
- The **T-web**:

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

T-web structures inferred by BORG

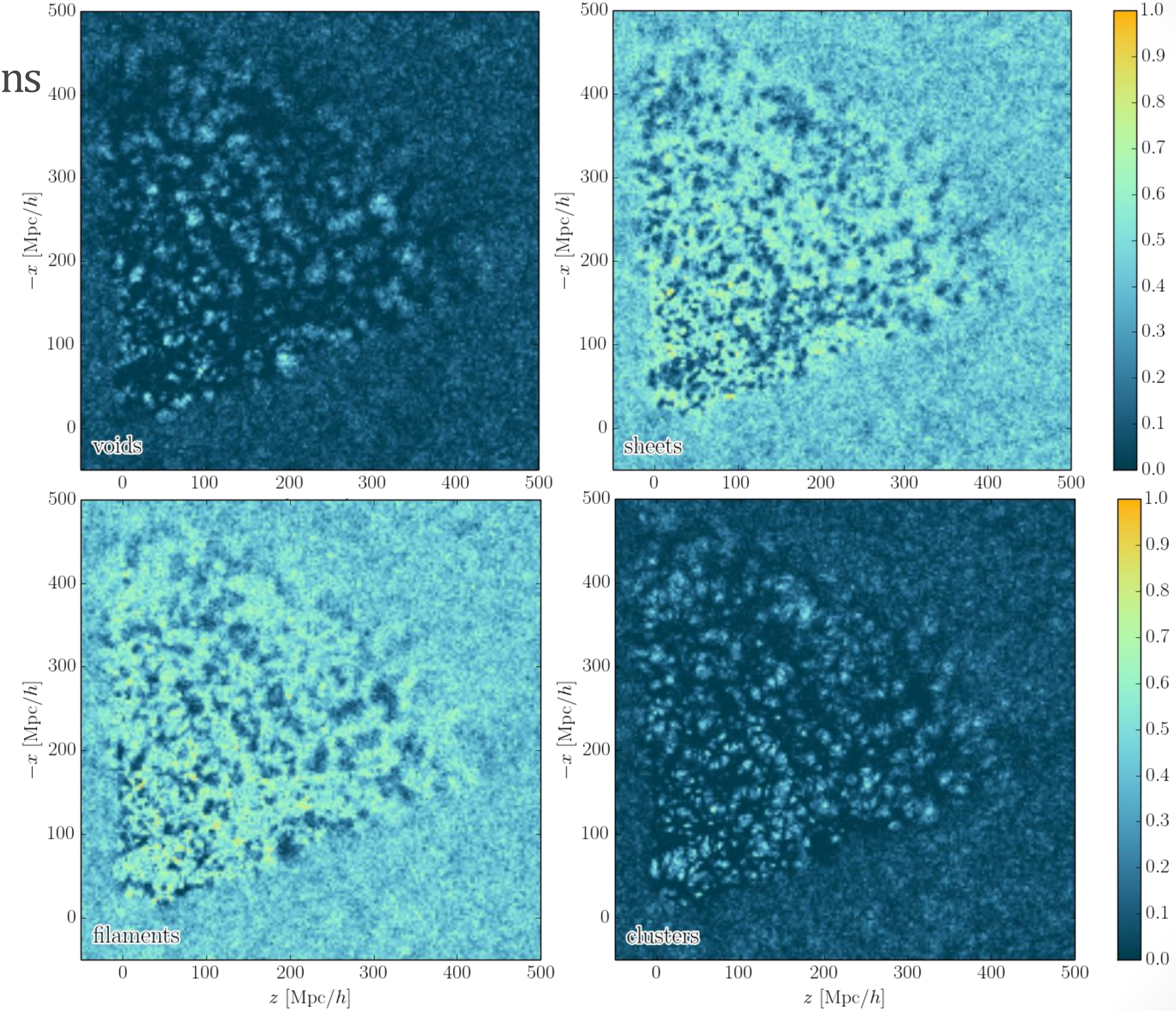
Final conditions



FL, Jasche & Wandelt 2015, arXiv:1502.02690

T-web structures inferred by BORG

Initial conditions



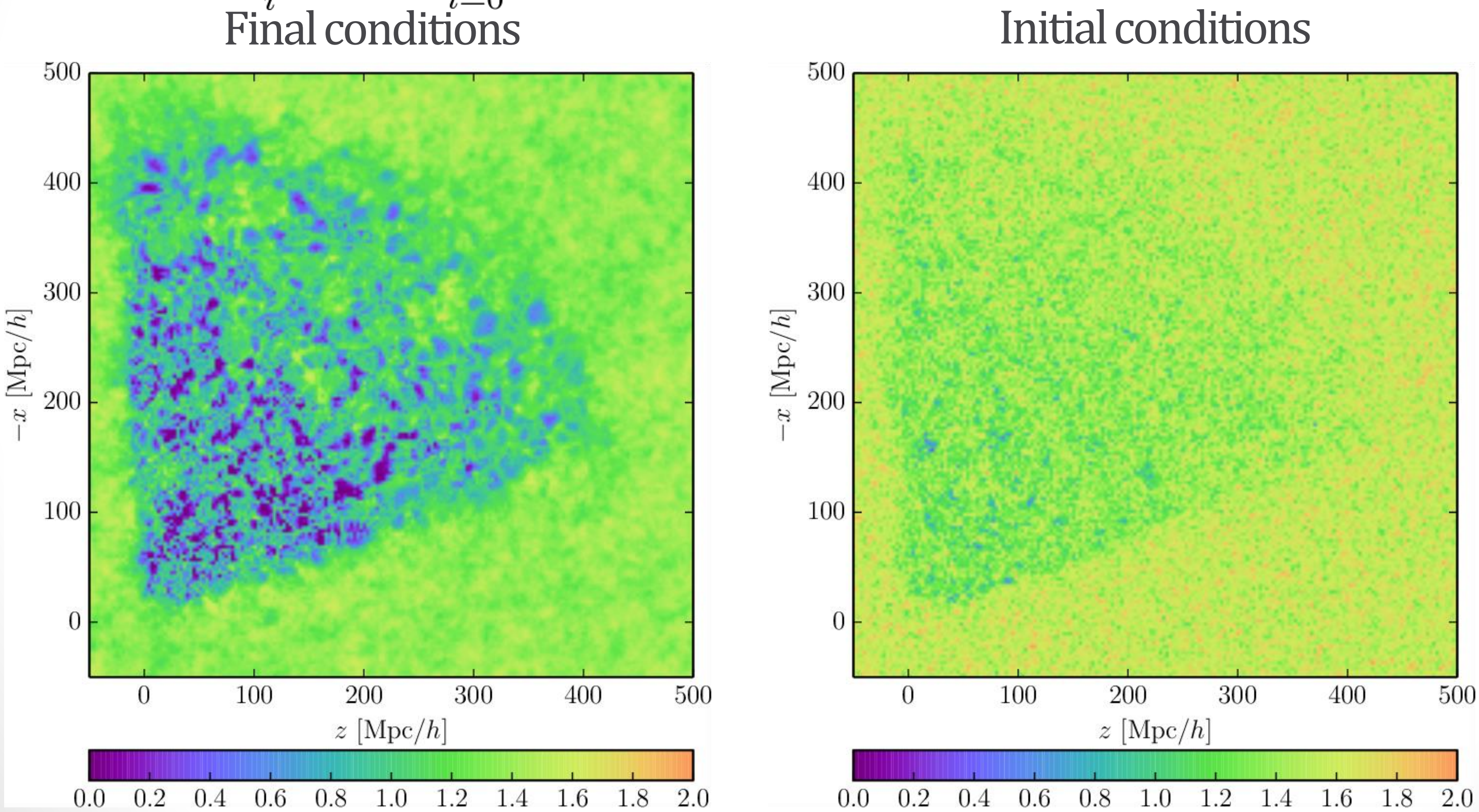
FL, Jasche & Wandelt 2015, arXiv:1502.02690

What is the information content of these maps?

Shannon entropy

$$S = -\sum_{i} p_i \log_2 p_i$$

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



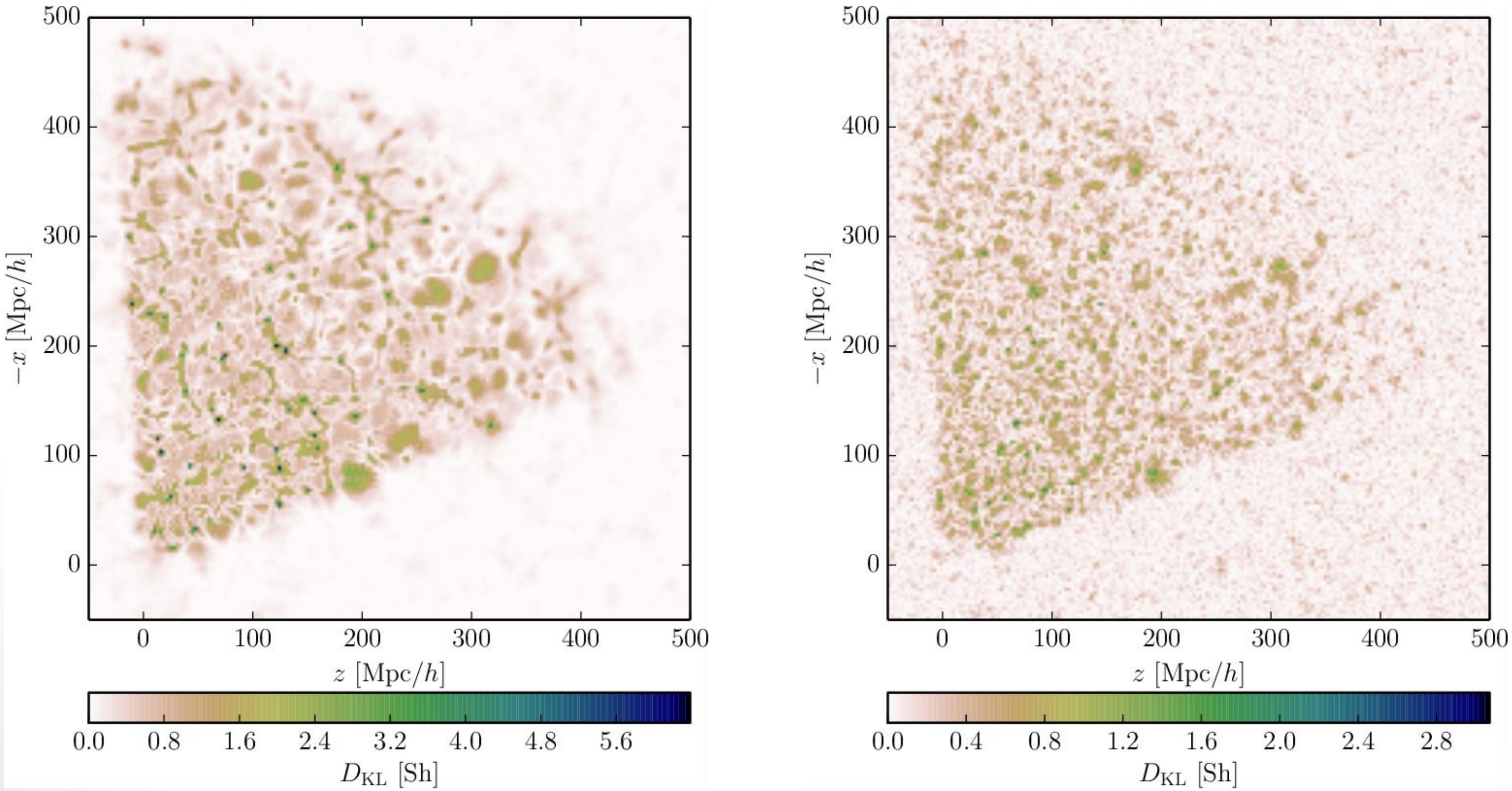
FL, Jasche & Wandelt 2015, arXiv:1502.02690

How much did the data surprise us?

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

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

in Sh


Final conditions

Initial conditions

FL, Jasche & Wandelt 2015, arXiv:1502.02690

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”:

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

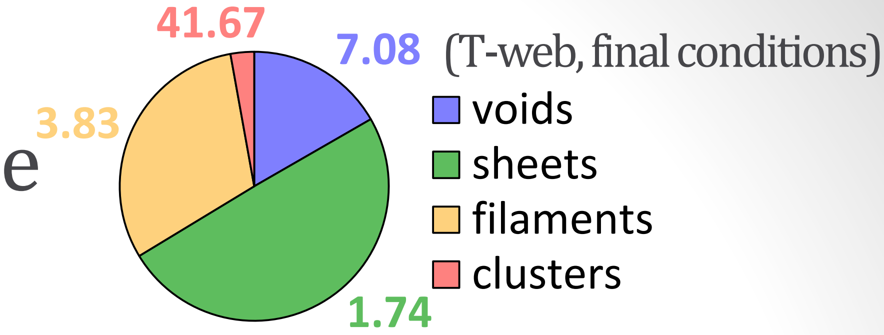
⇨ A problem of **Bayesian decision theory**: one should take the action that maximizes the utility

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

- How to write down the gain functions?

FL, Jasche & Wandelt 2015, arXiv:1503.00730

Gambling with the Universe



- One proposal:

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

"Winning"

"Loosing"

"Not playing"

- Without data, the expected utility is

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

"Playing the game"

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

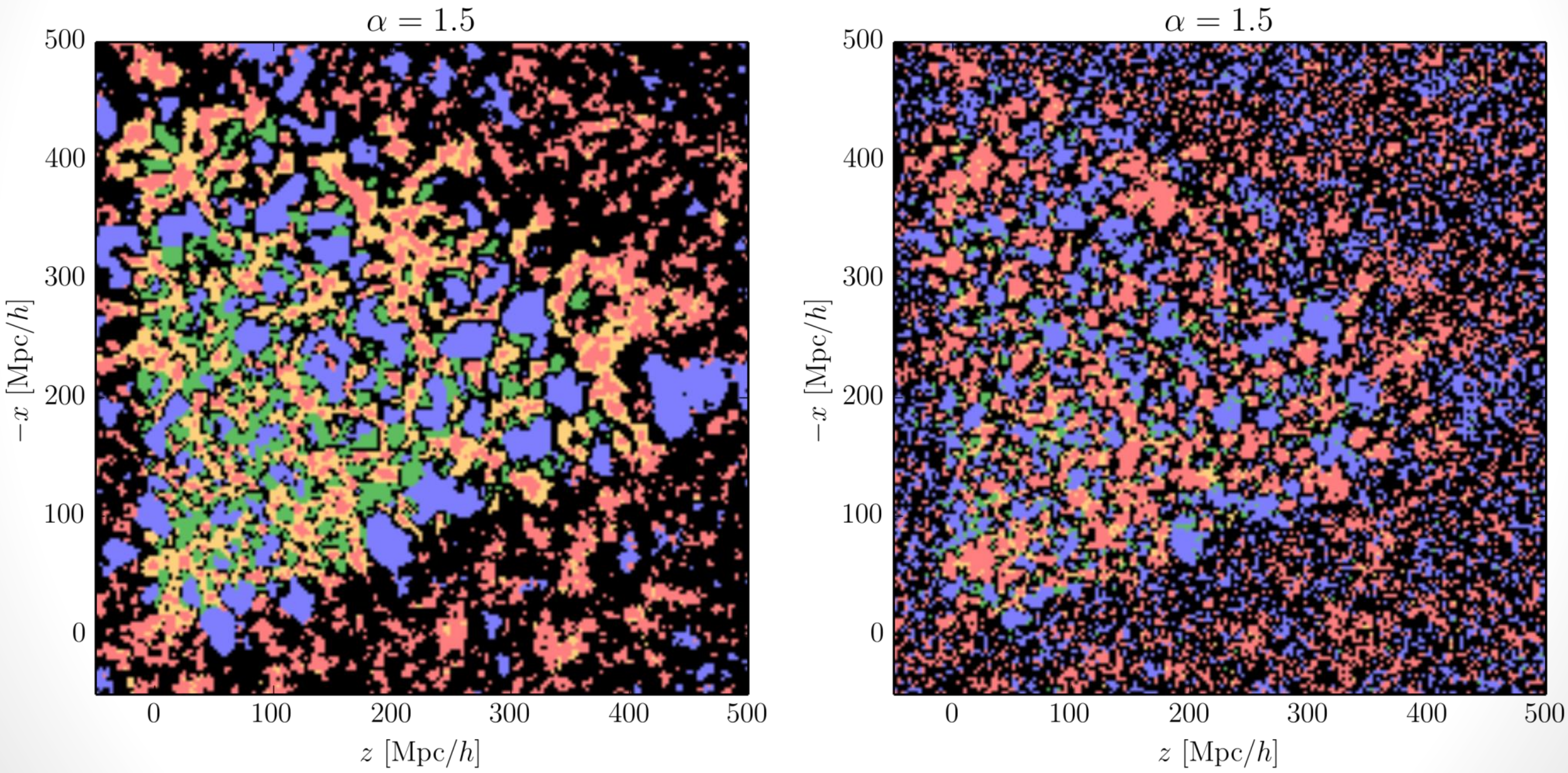
"Not playing the game"

- 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…

🟦 voids
🟨 filaments
⬛ undecided
🟩 sheets
🟥 clusters



Final conditions

Initial conditions

FL, Jasche & Wandelt 2015, arXiv:1503.00730

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** :

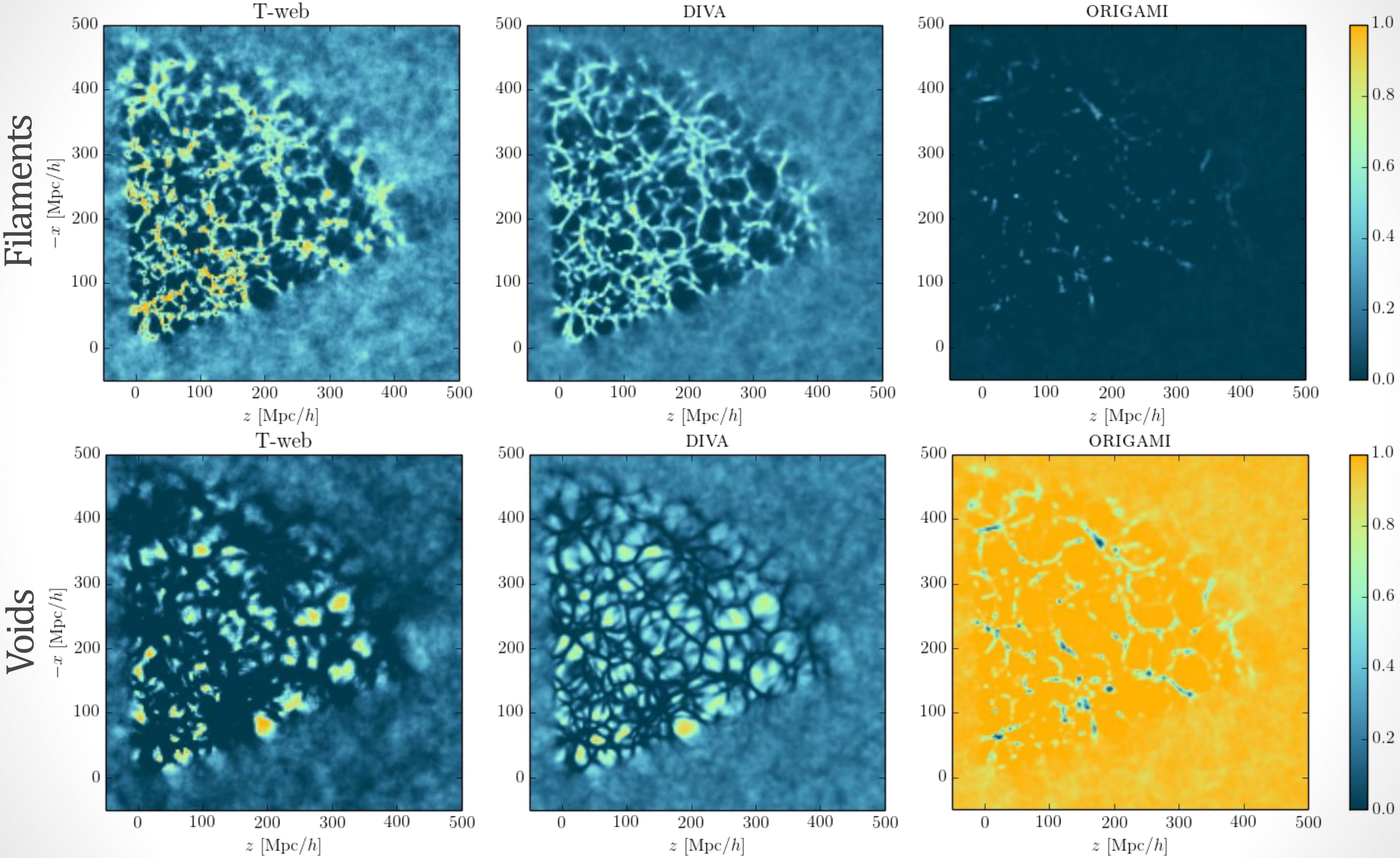
uses the dark matter "phase-space sheet" (number of orthogonal axes along which there is shell-crossing)

Falck, Neyrinck & Szalay 2012, arXiv:1201.2353

Lagrangian classifiers

now usable in real data!

Comparing classifiers



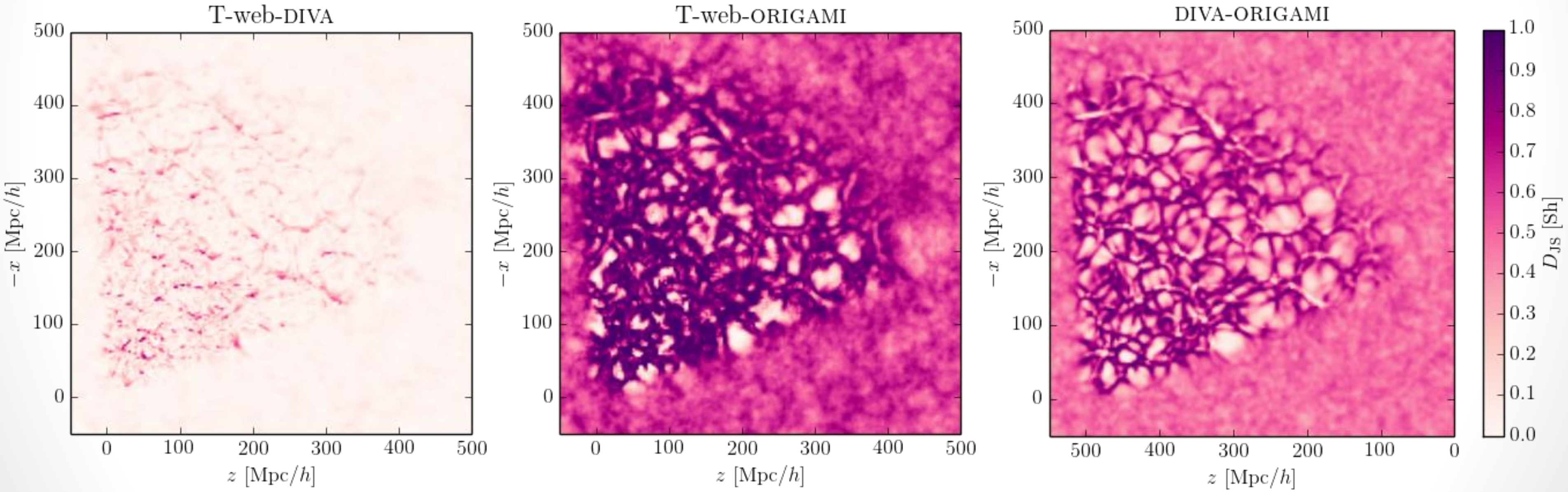
FL, Jasche & Wandelt 2015, arXiv:1502.02690

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

How similar are different classifications?

Jensen-Shannon divergence

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



(more about the Jensen-Shannon divergence later)

FL, Lavaux, Jasche & Wandelt, 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

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

- 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.

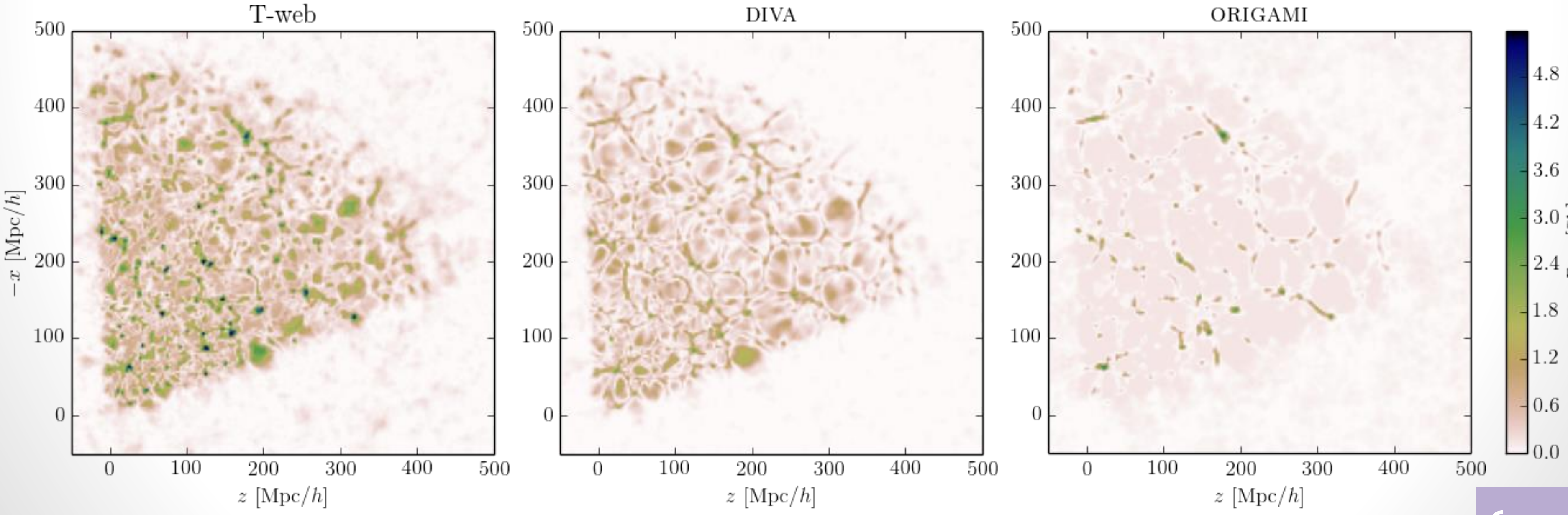
FL, Lavaux, Jasche & Wandelt, in prep.

1. Utility for parameter inference:

cosmic web analysis

- In analogy with the formalism of **Bayesian experimental design**: maximize the **expected information gain** for cosmic web maps $U_1(d, \xi)(\vec{x}_k) = D_{\mathrm{KL}}\left[\mathcal{P}(\mathrm{T}(\vec{x}_k)|d, \xi)||\mathcal{P}(\mathrm{T}|\xi)\right]$

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



FL, Lavaux, Jasche & Wandelt, in prep.

2. Utility for model selection:

dark energy equation of state

- For example, consider three dark energy models with

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

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

$$\boxed{U_2(\xi) = I\left[\underset{\substack{\nearrow\\ \text{model classifier}}}{\mathcal{M}} : \underset{\substack{\nwarrow\\ \text{mixture distribution}}}{\mathcal{R}(d)} \middle| \xi\right]}$$

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

FL, Lavaux, Jasche & Wandelt, in prep.

3. Utility for prediction of new data:

galaxy colors

- Maximize the **expected information gain** for some new quantity $U_3(d, \mathrm{T}, \xi) = D_{\mathrm{KL}}\left[\mathcal{P}(c|d, \mathrm{T}, \xi) || \mathcal{P}(c|\xi)\right]$

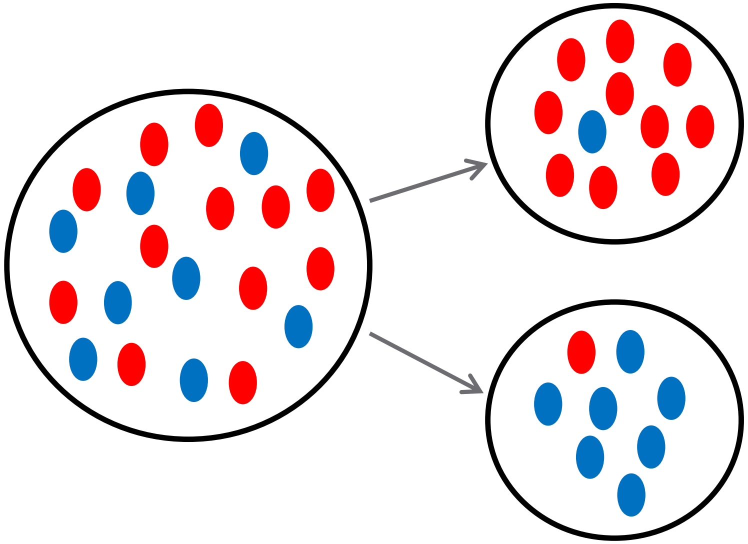
$$\boxed{U_3(\xi) = I[\underset{\substack{\nearrow \\ \text{predicted data}}}{c} : \underset{\substack{\nwarrow \\ \text{classification}}}{\mathrm{T}} | \xi]}$$

FL, Lavaux, Jasche & Wandelt, in prep.

3. Utility for prediction of new data:

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:

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

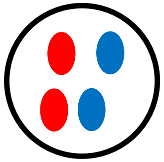
information gain for this split: $0.9709 - 0.4682 = 0.5027 \text{ Sh}$

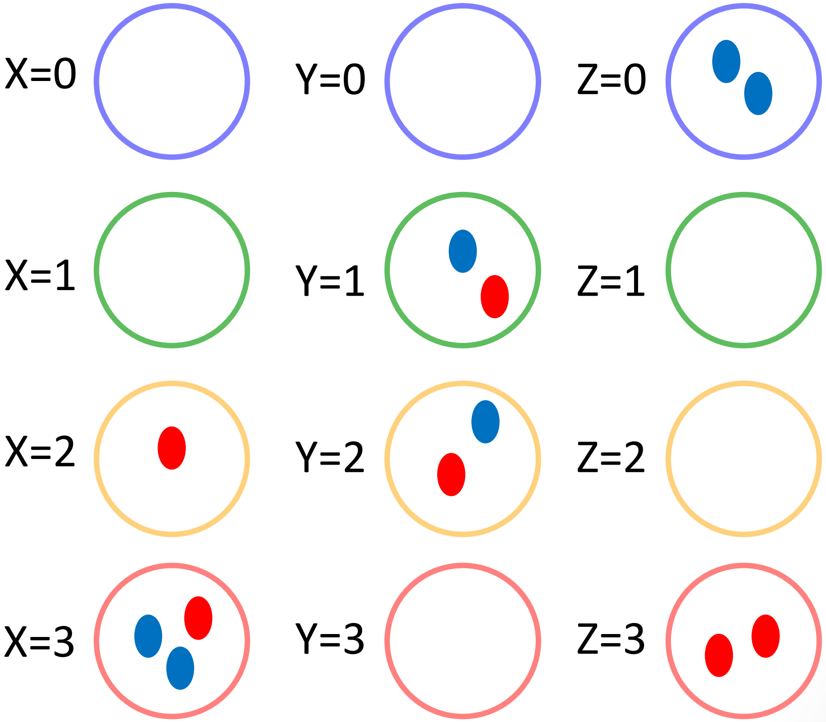
3. Utility for prediction of new data:

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, 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

FL, Jasche & Wandelt 2015, arXiv:1503.00730

- It is now possible to **use Lagrangian classifiers with real data**!

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

- 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)

FL, Lavaux, Jasche & Wandelt, in prep.