Machine Learning to reconstruct the dark matter density fields from galaxy survey data.

J. Suárez-Pérez

Advisor

Jaime Forero-Romero

Universidad de los Andes

Astroandes

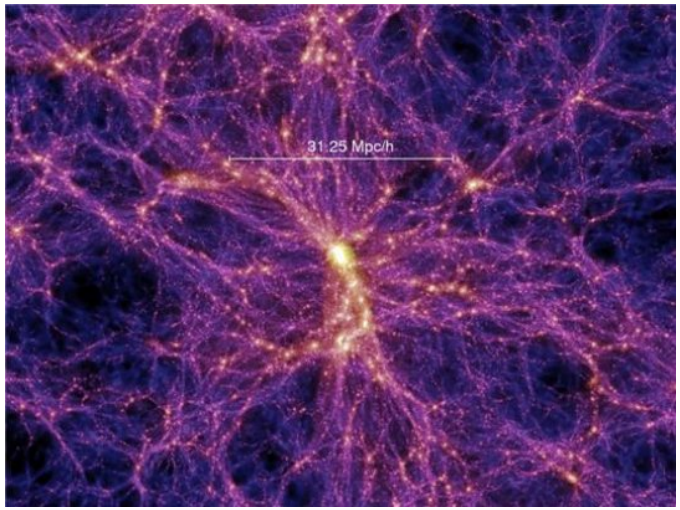
November 11th, 2019

Table of Contents

One of the main goals in cosmology is understand the distribution of dark matter in the local Universe.

The problem:The distribution of Dark Matter (DM) is not possible to observe directly.

A solution:Make an inference of the DM distribution using observational measurements of galaxies distributions as SDSS or DESI (2020).

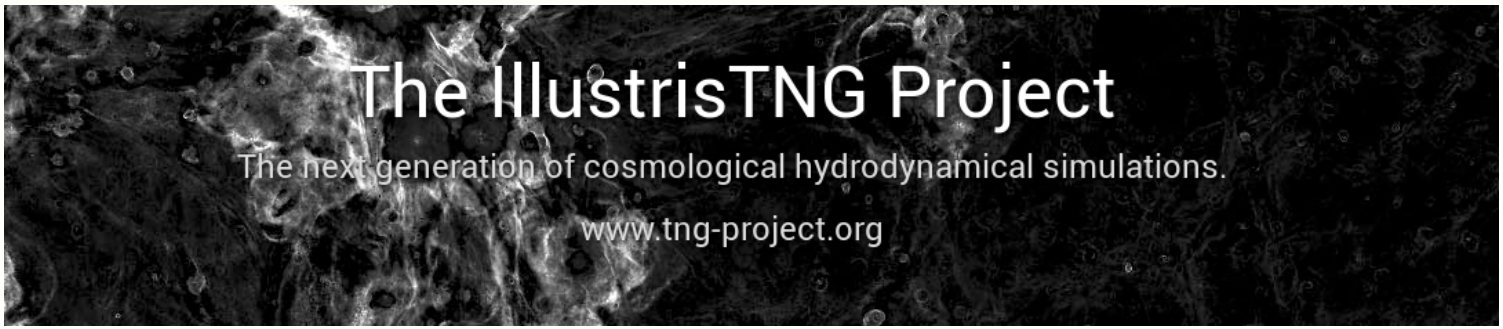


[1] Credits: V.Springel, Max-Planck Institut für Astrophysik, Garching bei München

Fisrt Step: Simulation

What is the Illustris-TNG project?

"It is a great set of simulations magneto-hydro-dynamics of galaxies formation, completed in 2019... it uses numerical algorithms and physical models. The simulation represents a combination of high resolution and high physical fidelity"[2]

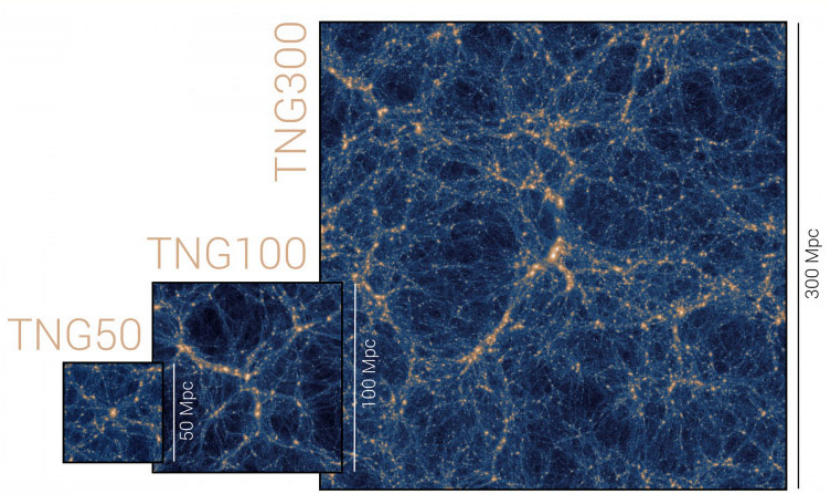


It's based in the cosmological paradigm Λ-Cold Dark Matter (Λ-CDM).

[2] http://www.tng-project.org

Description of the Illustris-TNG Project

- The simulation data includes 100 snapshots.
- Each simulation have a volume of $(302.6Mpc)^3$.
- It includes five different elements (dark matter particles, galaxies, gas cells, stars, wind stellar particles, super massive black holes, diffuse gas), in a redshift from $z = 127$ to the present $z = 0$.

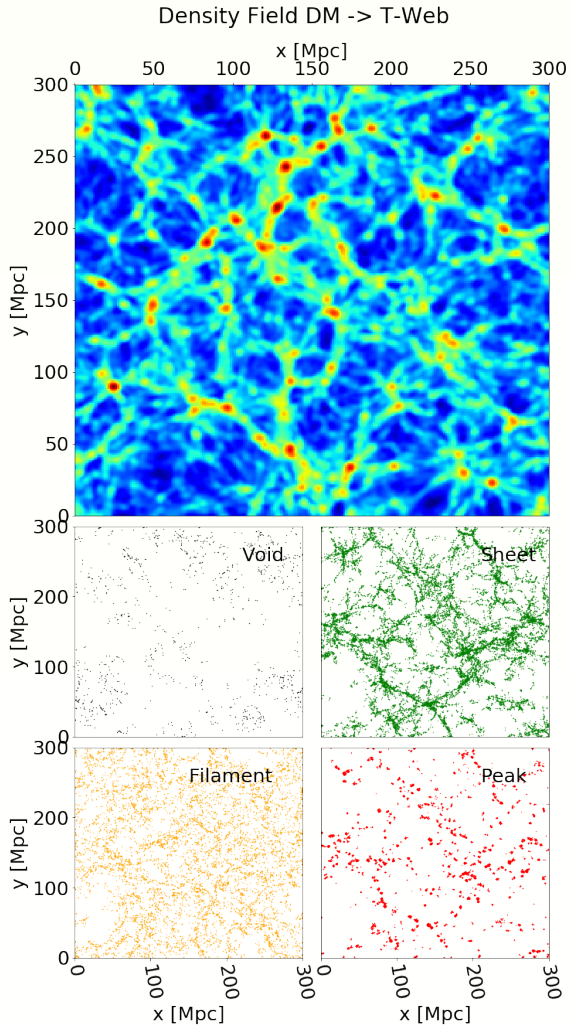


Classification of the Cosmic Web: T-Web

"It is to possible make a classification of the cosmic web as a function of the local density, for make this classification is used the gravitational potential"

$$T_{\alpha\beta} = \frac{\partial^2 \phi}{\partial r_\alpha \partial r_\beta}$$

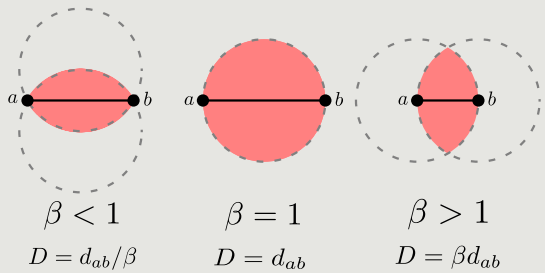
It depends on the value of the eigenvalues respect to a cut value λ_{th}, it is possible to make a classification by environments between peaks, sheets, filaments and voids. This classification is called the T-Web [5]



[5]A dynamical classification of the cosmic web. Forero–Romero J. et al. MNRAS. 2009

Characterization of the galaxies distribution using the β-skeleton.

The characterization is obtained using the β-skeleton algorithm, this algorithm allow us identify graph.



From the graph is possible to compute the number of connections by galaxy (node), the average length of connections, the eigenvalues of the inertia matrix, a relative volume and density and the anisotropy fractional.

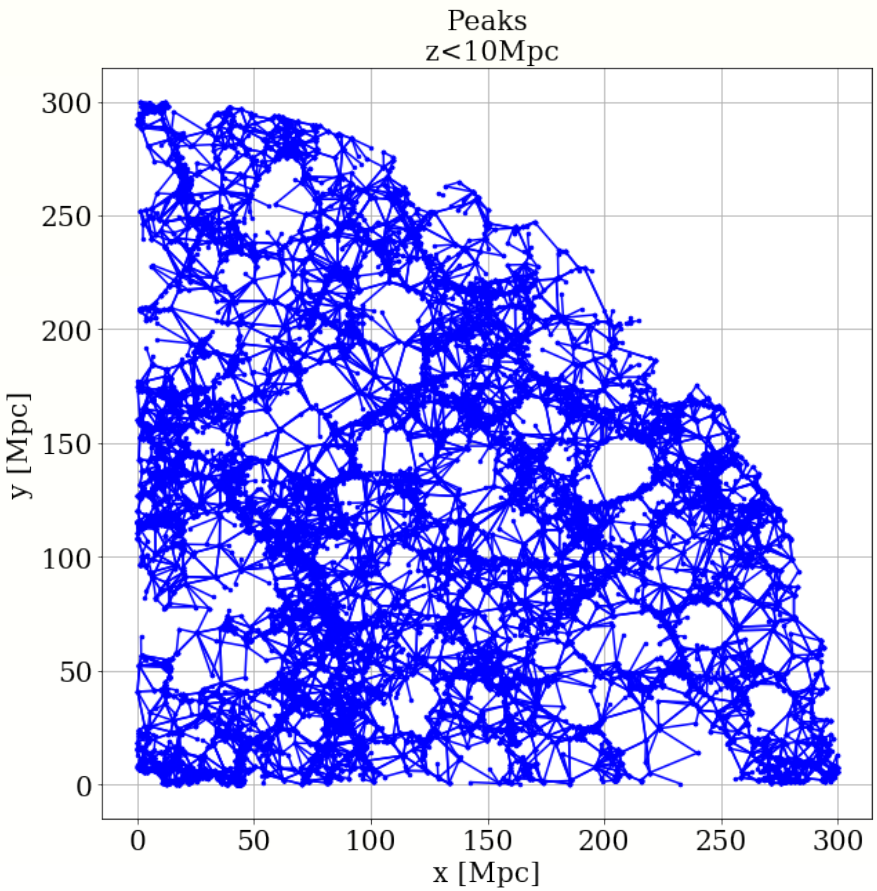


Fig. 1: Graph for the galaxies distribution of TNG with for a region $z < 10Mpc$ and $\beta = 1$.

First Step: Reconstruction of the T-Web

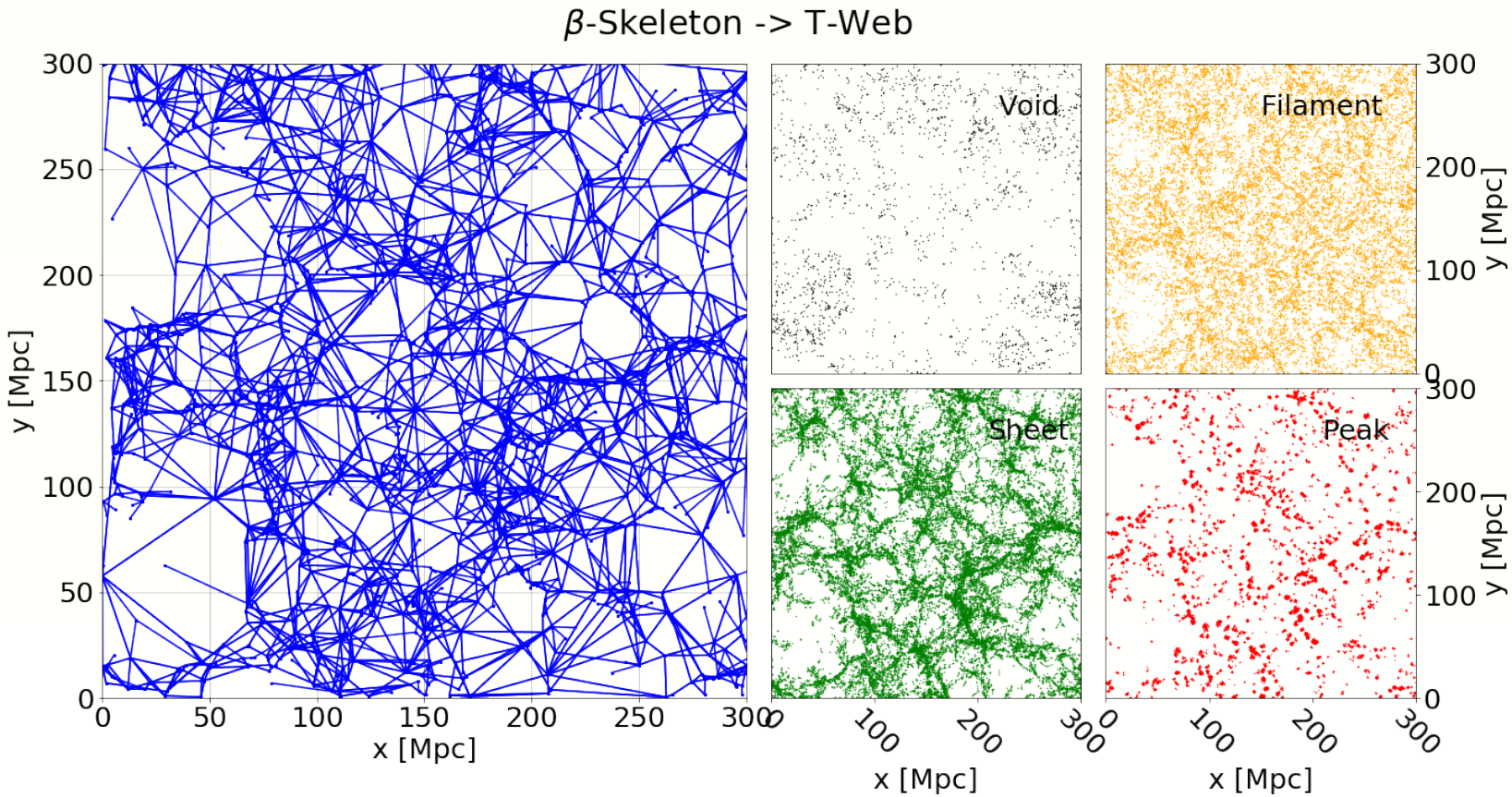


Fig. 2: Reconstruction of the T-Web from the β-skeleton

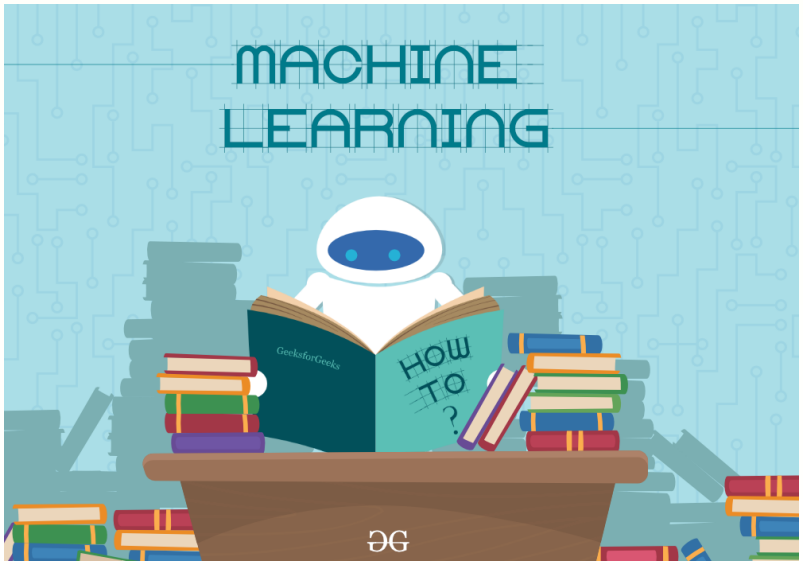
¿Machine Learning?

¿Machine Learning?

¿Why?

- It is not possible to make direct observations of the DM.
- We can to make an inference from information that can be measuring.

«Training with simulations predicting with observations».



[6]From https://www.geeksforgeeks.org/machine-learning/

Feature Space. Galaxies.

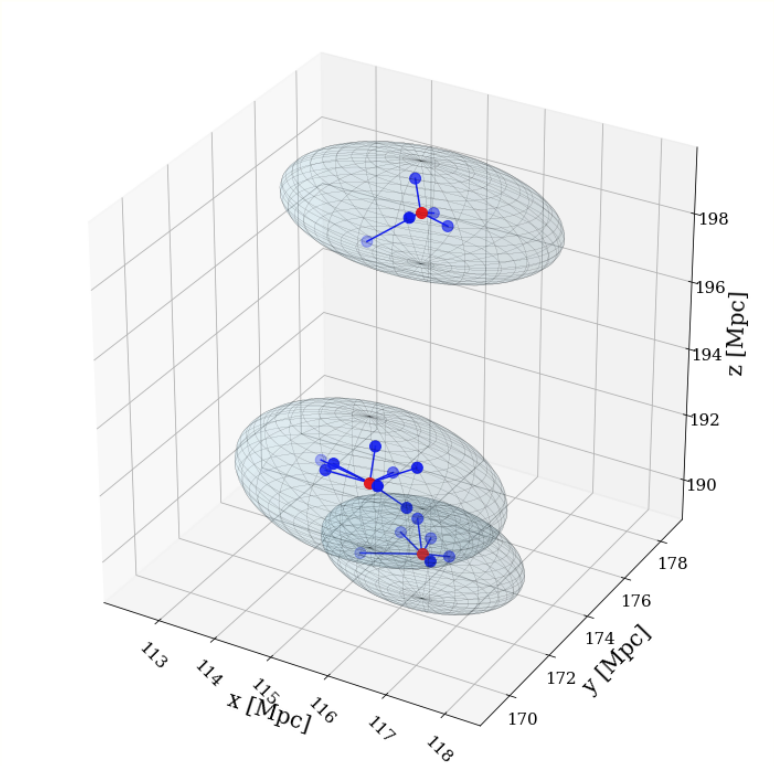


Fig. 3: Pseudo-Volumen using the parameters **a**,**b** y **c**.

- For $\beta = 1.0$
- By structure is possible define a inertia matrix and compute its eigenvalues.(σ_1, σ_2 y σ_3).

Defined:

- $a = \sqrt{\sigma_1}$, $b = \sqrt{\sigma_2}$ y $c = \sqrt{\sigma_3}$.
- The pseudo-volume $V = abc$ and pseudo-density as $\rho = \frac{1}{abc}$.
- The fractional anisotropy as $fa = \sqrt{\frac{(a-c)^2+(c-b)^2+(b-a)^2}{a^2+b^2+c^2}}$.

Feature Space. Galaxies.

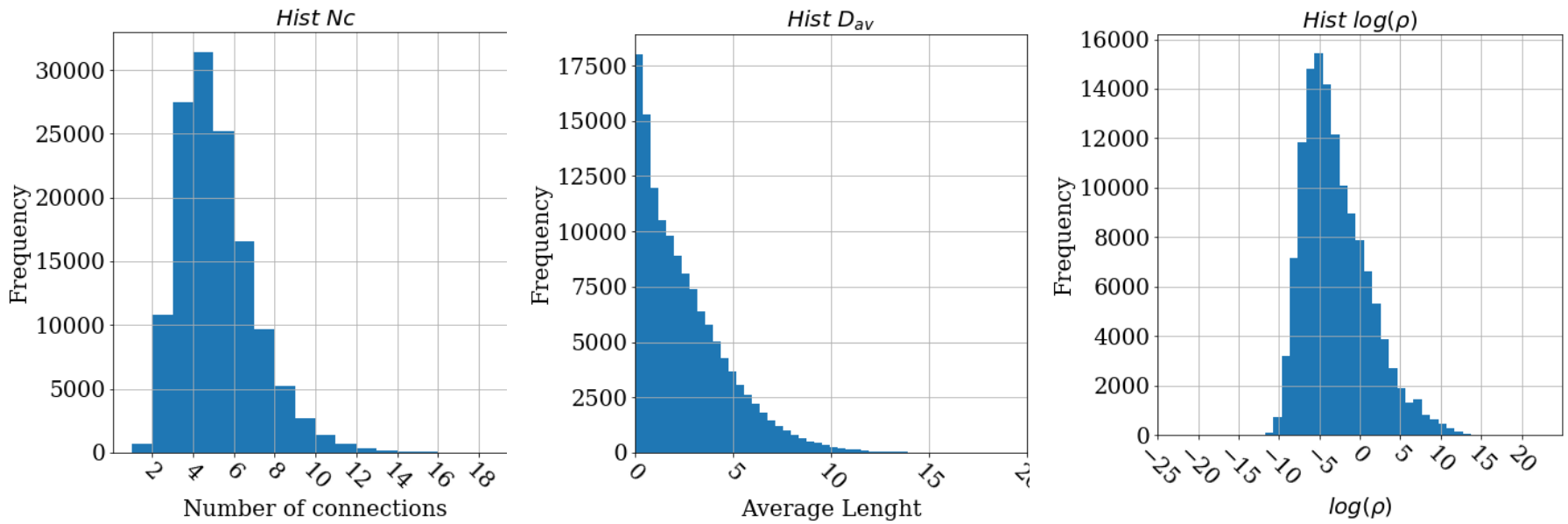


Fig. 4: a) Number of connections, **b)** average distance, **c)** pseudo-density computed from the β-skeleton.

Feature Space. Galaxies. Local Parameters.

Also was computed a set of local parameters that include the information of the first neighbors. This information is define as $\Delta f = \bar{f} - f$.

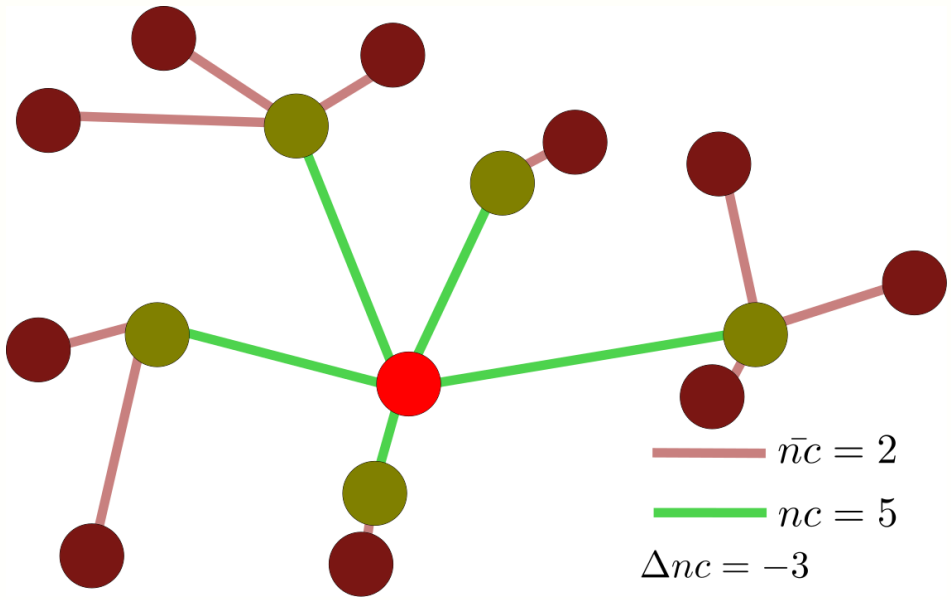


Fig. 5: Representation for the local parameters.

Feature Space. Density field of DM.

- The smoothing (sm) is a tuning parameter over the density field.

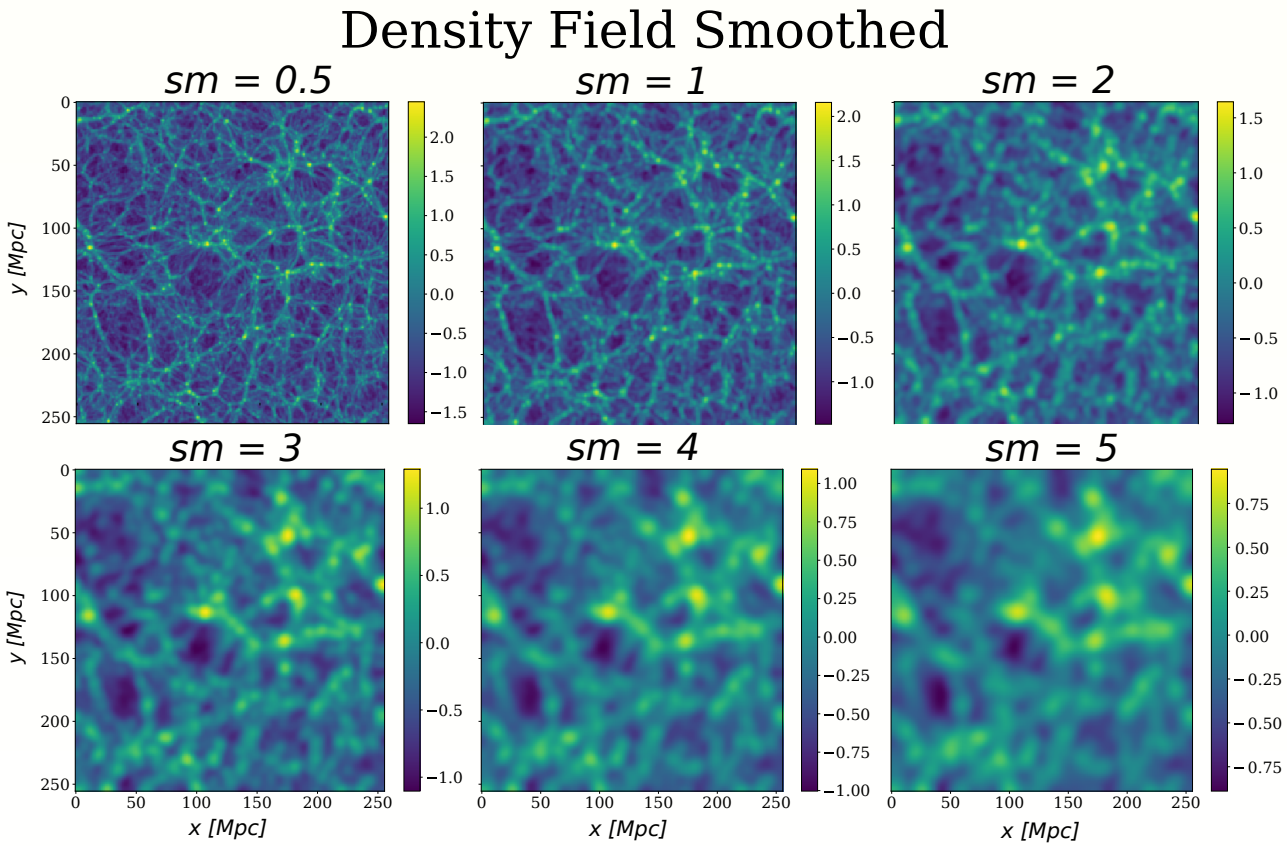


Fig. 6: Density field for different smoothing ***sm***.

Feature Space. Density field of DM.

The eigenvalues computed over the Hessian $T_{\alpha\beta}$, it is possible tuning a cut value λ_{th} that allow us make a classification by environments between peaks, sheets, filaments and voids.

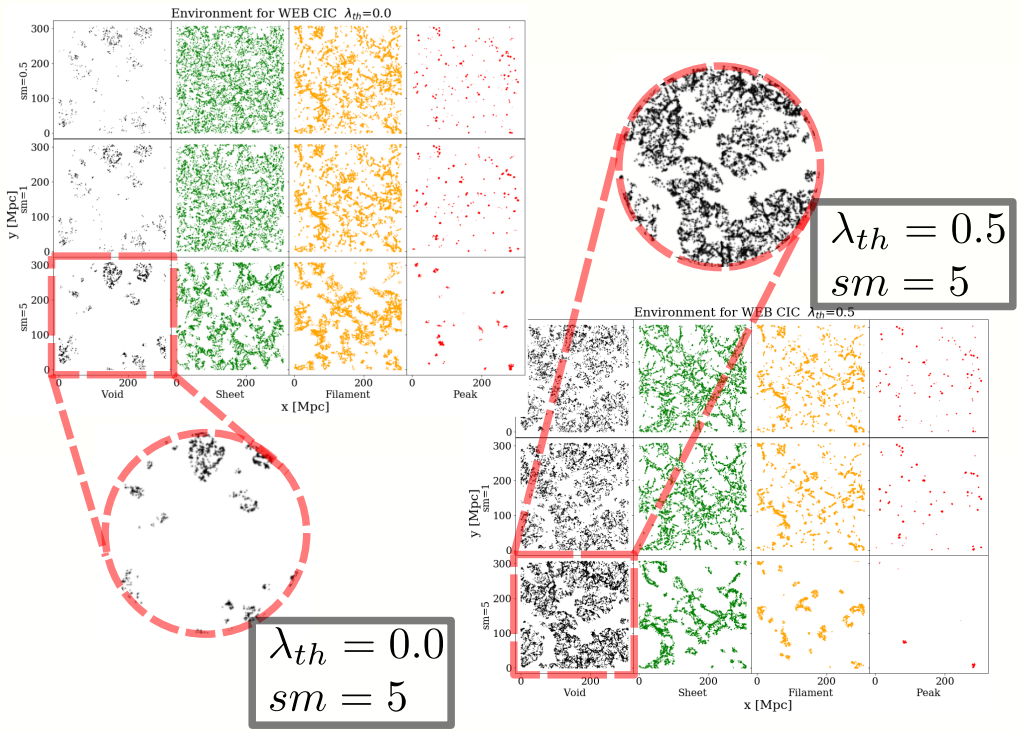


Fig. 7: Classification by environments for different cuts in λ_{th} y sm.

Feature Space. Density field of DM.

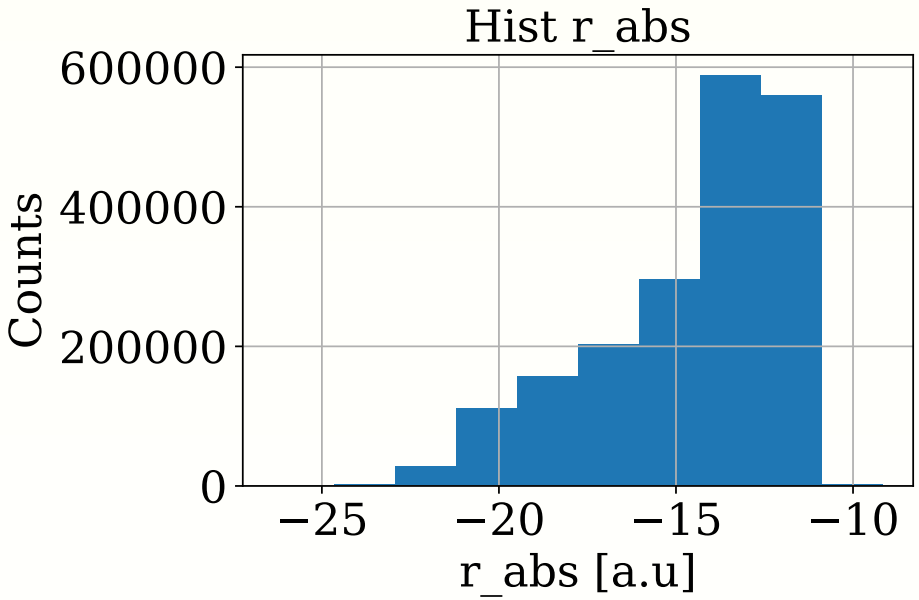
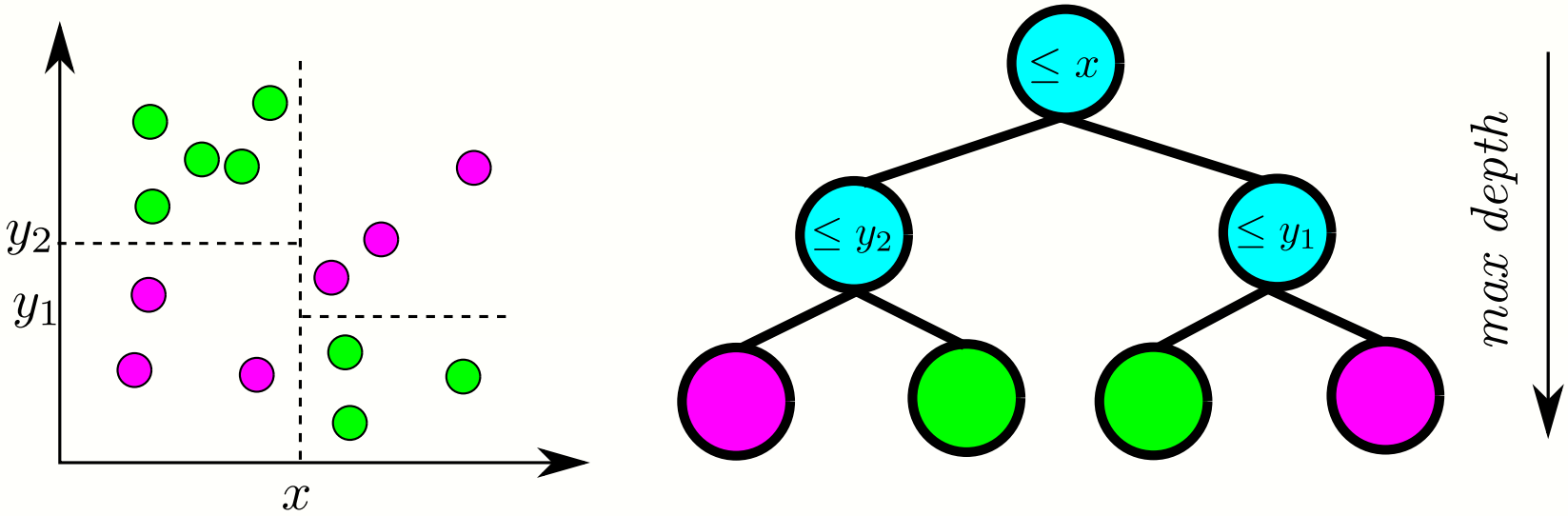


Fig. 8: Distribution of the luminosity in the galaxies distribuion

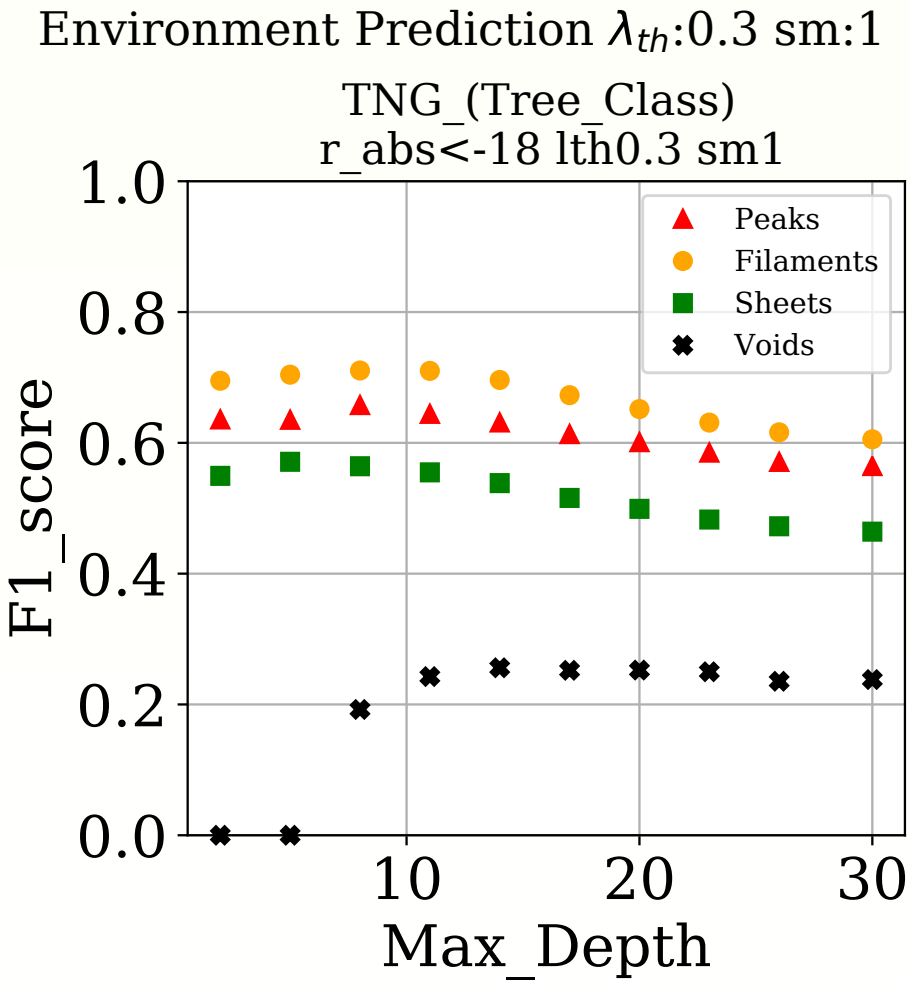
- Stelar Luminosity in the r filter. $\sim 500nm$

Classification Trees



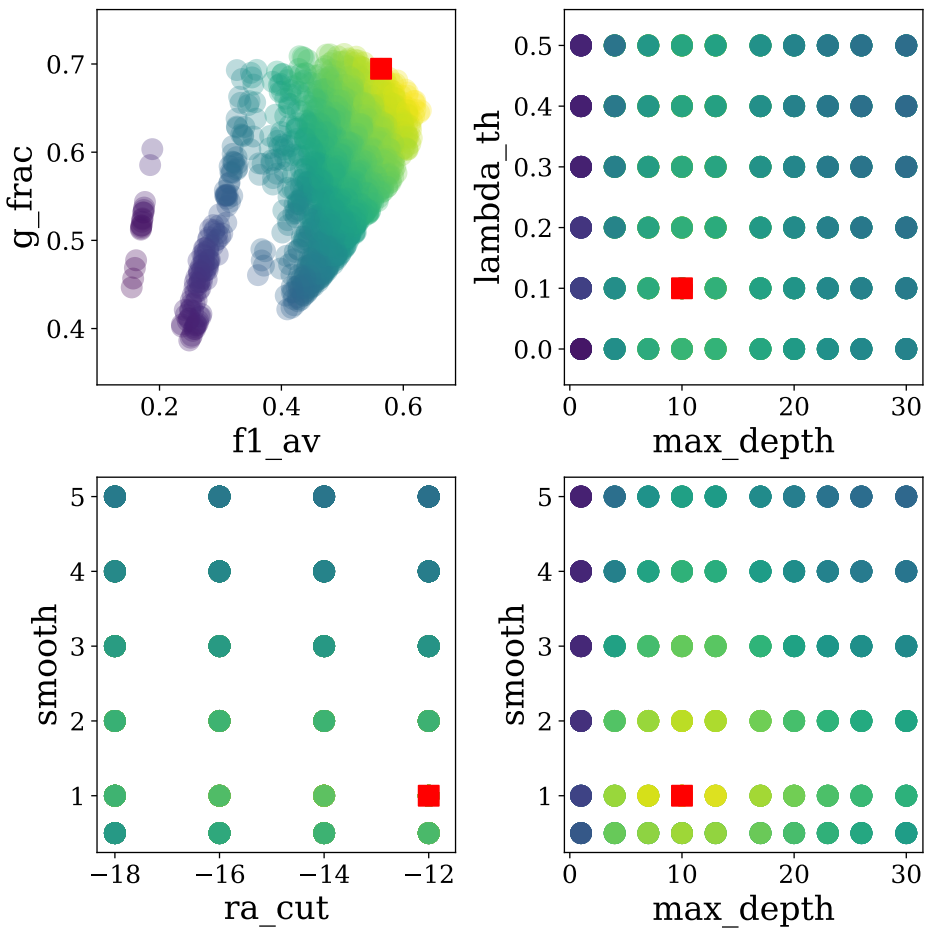
Train: 50 % **Test:** 30 %

Classification trees



The prediction of the classification trees is evaluated with the F1 score as a function of the *max_depth*.

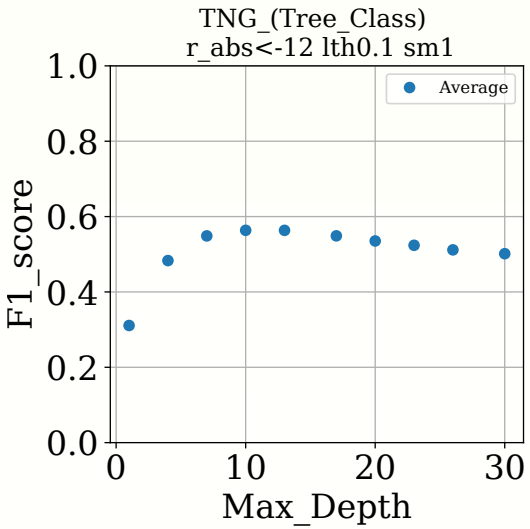
Evaluation in the Feature Space

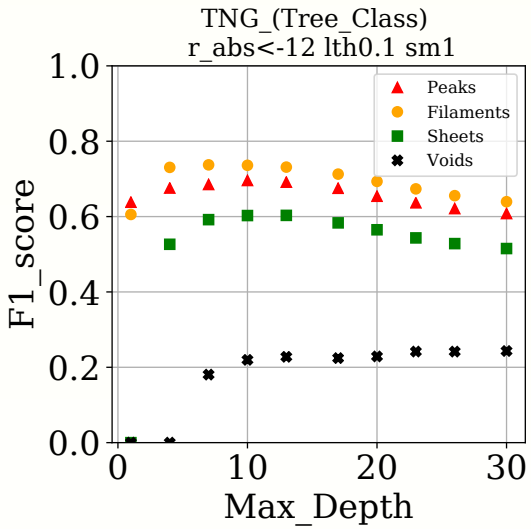


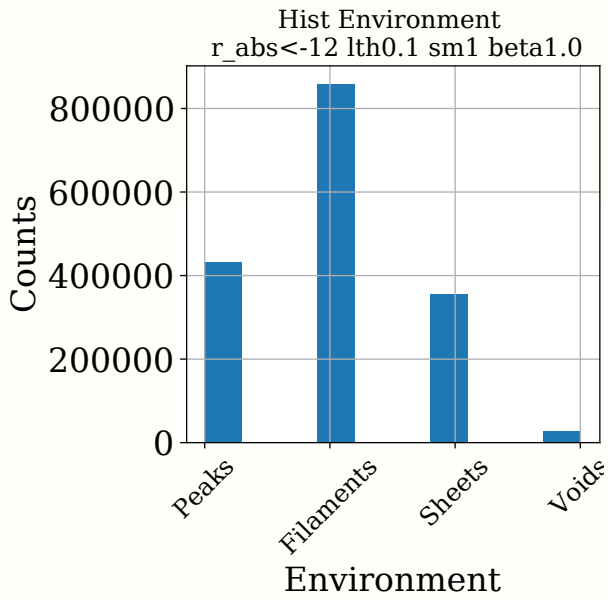
Selection of the optimal parameters in the feature space that maximize the F1 score and the *g_frac* factor.

Best Prediction

Environment Prediction λ_{th}:0.1 sm:1

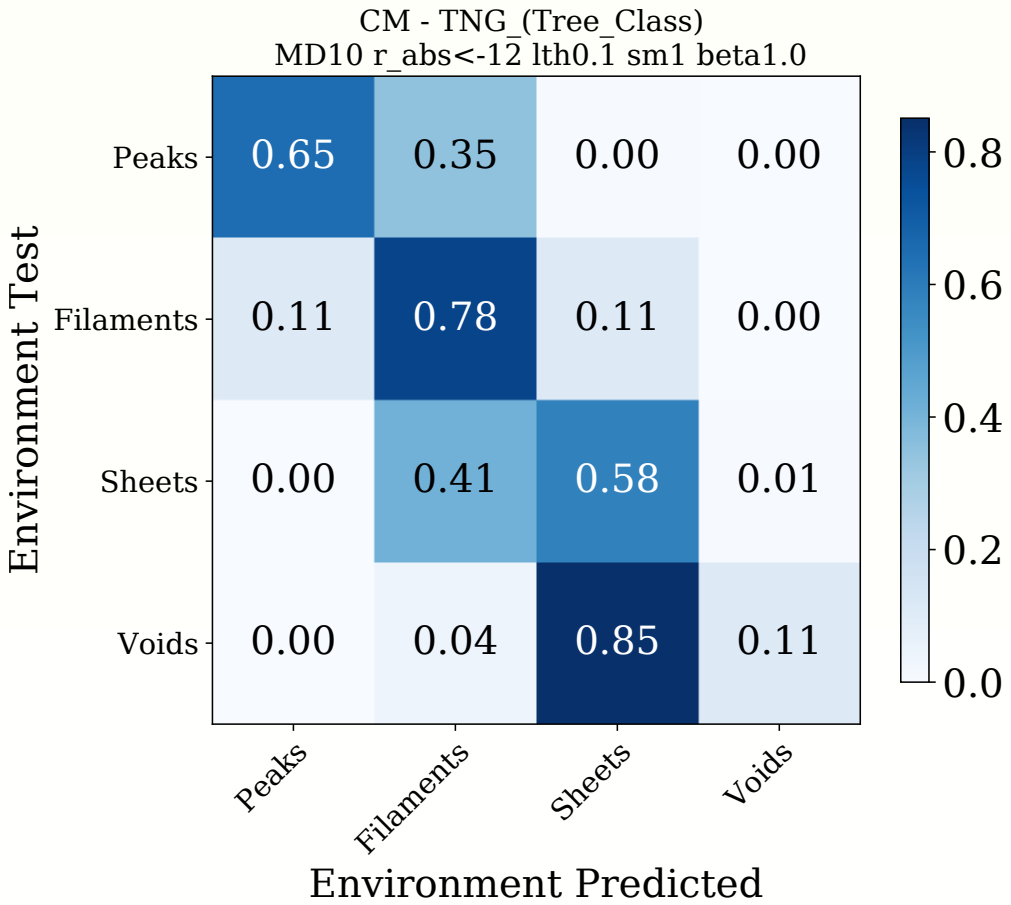






- The maximum value for the average of F1 is around of 0.6.
- The best prediction was for the filaments.
- The maximum F1 is around of 0.7 for the filaments.

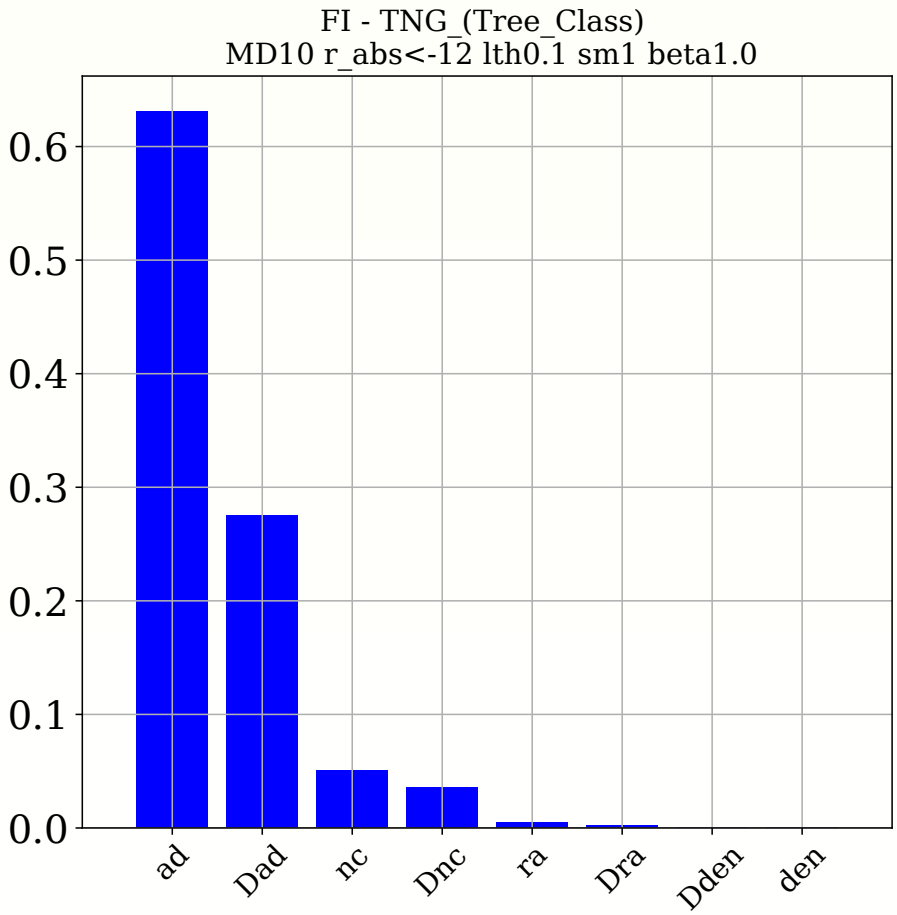
Confusion Matrix



- 65/100 peaks was correctly predicted.
- 11/100 voids was correctly predicted.

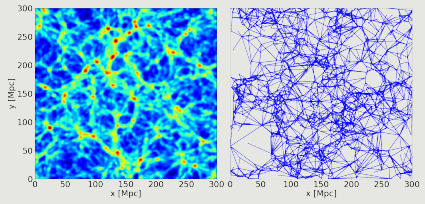
Importance of the features

- nc: Number of connections
- ad: Average distance
- ra: Luminosity
- den: Pseudo-density
- Dx: Local parameters

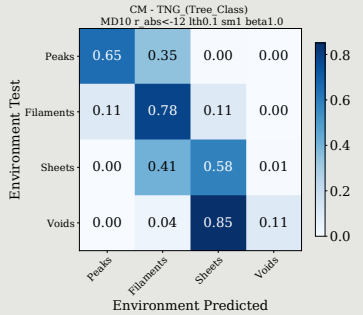


Conclusions

- It is possible to make a characterization of the T-Web through the implementation of the β-skeleton algorithm.



- The classification three in according with the confusion matrix is good to predict filaments, however it is not good when try to predict voids.



Future

- Consider other values for the β parameter.
- Consider other physical parameters.

Thanks!!

jf.suarez@uniandes.edu.co