

# Entendiendo la distribución de materia oscura a gran escala con algoritmos de machine learning.

**J. Suárez-Pérez**

Asesor

**Jaime Forero-Romero**

Universidad de Los Andes

Astroandes

Departamento de Física

Mayo 28 de 2019



# Tabla de Contenidos

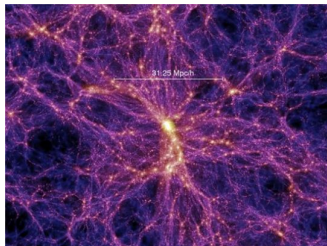
- 1 Descripción del problema
- 2 ¿Qué es Illustris-TNG?
- 3 Clasificación de la red cósmica
- 4 ¿Machine Learning?
- 5 Conclusiones



**Objetivo:** Entender la distribución de materia oscura en el universo local es uno de los principales objetivos de la cosmología.

**El problema:** Sin embargo, esta distribución no se puede observar directamente.

**Una solución:** Inferir esta distribución a partir de medidas observacionales de distribuciones de galaxias tales como SDSS o DESI (2020).



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



# Simulación



## Qué es el proyecto Illustris-TNG?

*"Es una gran serie de simulaciones magneto-hidro-dinámicas de formación de galaxias, completada en el 2019 ... utiliza algoritmos numéricos y modelos físicos. La simulación representa una combinación de alta resolución y alta fidelidad física."*[2]



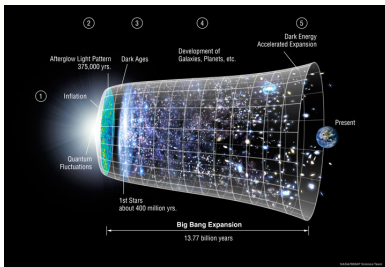
Basada en el paradigma cosmológico  $\Lambda$ -Cold Dark Matter ( $\Lambda$ -CDM).

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



## $\Lambda$ Cold Dark Matter ( $\Lambda$ -CDM)

*El modelo  $\Lambda$ -CDM representa el modelo de concordancia del Big Bang. Este modelo permite explicar las observaciones cósmicas de la radiación de fondo de microondas, la estructura a gran escala del universo y las observaciones realizadas de supernovas. Permite realizar una explicación de la expansión acelerada del Universo.*



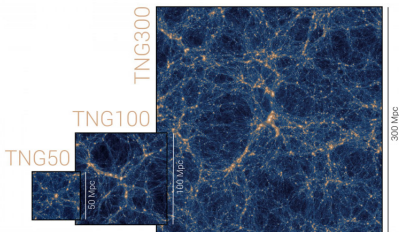
**Fig. 1:** Representación de la expansión del universo. [3]

[3] Tomada de NASA - Universe Evolution

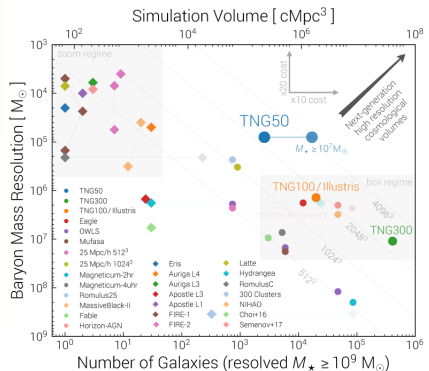


## Descripción del proyecto Illustris

- ▶ Los datos obtenidos por la simulación incluyen 100 snapshots. Estos snapshots incluyen todas las partículas/celdas en todo el volumen.
- ▶ Cada simulación tiene un volumen de  $(302.6 \text{ Mpc})^3$ . Además incluye cinco diferentes elementos (partículas de materia oscura, galaxias, celdas de gas, estrellas, partículas de viento estelar, agujeros negros supermasivos, gas difuso), en un redshift desde  $z = 127$  hasta el presente  $z = 0$ .



# Comparación entre diferentes simulaciones.



- ▶ Resolución  $\sim 10^7$
- ▶ Número de galaxias  $\sim 10^6$
- ▶ Volumen de simulación  $\sim 300 Mpc$

**Fig. 2:** Configuraciones para diferentes simulaciones. [4]

[4] Tomada de <http://www.tng-project.org/>

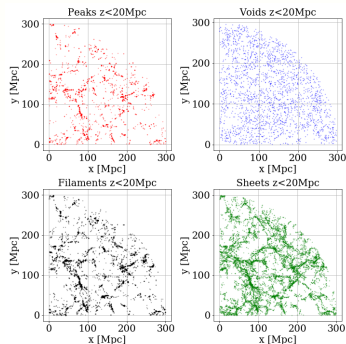
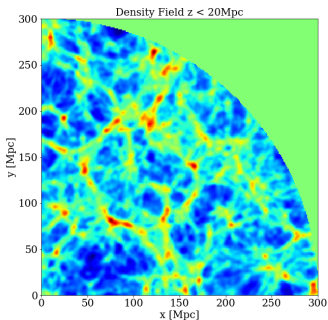




*Es posible hacer una clasificación de ambientes en función de la densidad local haciendo uso del potencial gravitacional ...*

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

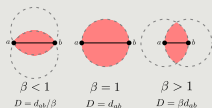
*... si todos los autovalores del tensor son negativos, se tiene un vacío, una hoja para uno positivo, un filamento para dos positivos y un pico para tres positivos.[5]*



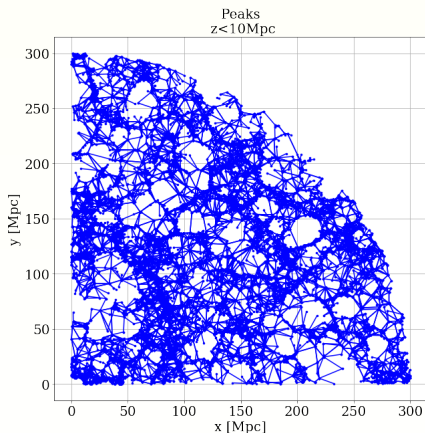
[5] A dynamical classification of the cosmic web. Forero-Romero, J. et al. MNRAS. 2009

# Caracterización de la distribución de galaxias usando el $\beta$ -skeleton.

Esta caracterización se hace utilizando el algoritmo de  $\beta$ -skeleton, un algoritmo que nos permite identificar grafos.

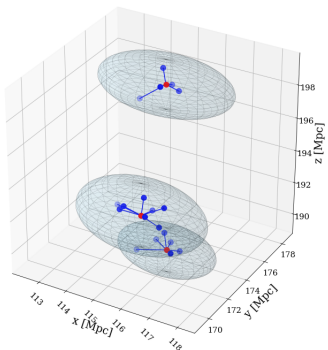


- Es posible extraer información relacionada con la distribución de galaxias tal como el número conexiones por galaxia, la distancia promedio, además de algunos parámetros locales.



**Fig. 3:** Grafo para una muestra de distribución de galaxias de TNG con  $\beta = 1$  para  $z < 10 Mpc$ .

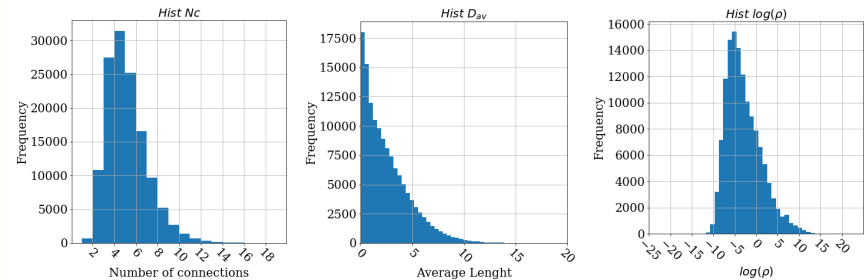
# Espacio de configuraciones.



**Fig. 4:** Volumen relativo utilizando los parámetros **a, b y c**.

- ▶ Por estructura es posible construir un tensor de inercia y calcular sus autovalores ( $\sigma_1$ ,  $\sigma_2$  y  $\sigma_3$ ).
- ▶ Definimos  $a = \sqrt{\sigma_1}$ ,  $b = \sqrt{\sigma_2}$  y  $c = \sqrt{\sigma_3}$ .
- ▶ Definimos el volumen  $V = abc$  y la pseudo-densidad como el inverso del volumen  $\rho = \frac{1}{abc}$ .
- ▶ Definimos la anisotropía fraccional como  $f_a = \sqrt{\frac{(a-c)^2 + (c-b)^2 + (b-a)^2}{a^2 + b^2 + c^2}}$ .



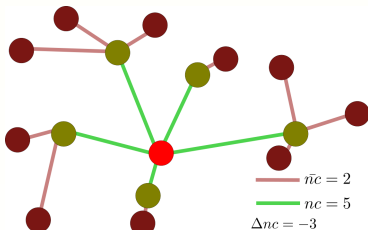


**Fig. 5: a)** Número de conexiones, **b)** distancia promedio, **c)** pseudo densidad calculadas usando el  $\beta$ -skeleton.



## Parámetros locales

Además de extraer estos parámetros consideramos además un conjunto de parámetros locales que incluyen la información de las primeras conexiones definidas como  $\Delta f = \bar{f} - f$ .



**Fig. 6:** Esquema para la definición de parámetros a primeras conexiones.



# ¿Machine Learning?



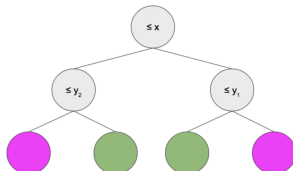
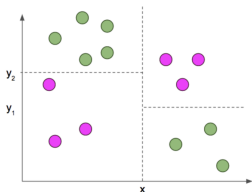
# ¿Machine Learning?

## ¿Por qué?

- ▶ Dado que no tenemos medidas directas de materia oscura, una posibilidad de entender su naturaleza es a través de simulaciones. Inferencia a partir de información que si puede ser medida.

«Entrenar con simulaciones, predecir con observaciones».

## Árboles de Clasificación

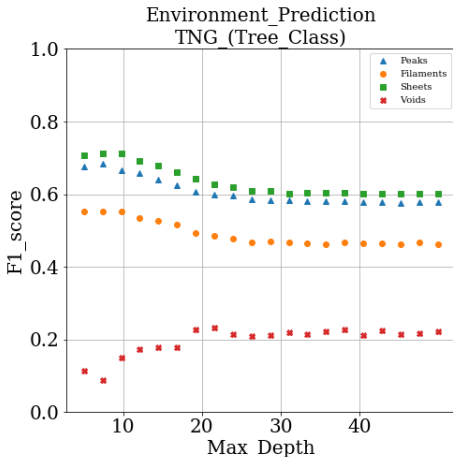


**Train: 50 % Test: 30 %**



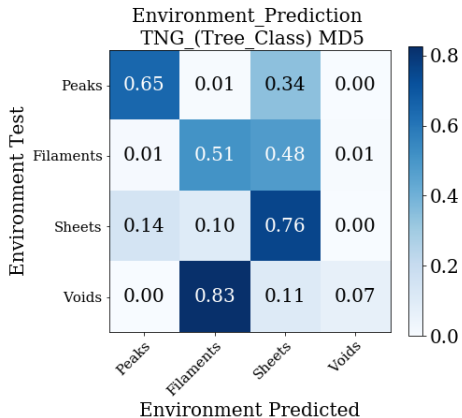
# Árboles de Clasificación

- La clasificación se hace en función del parámetro *max\_depth*.





# Matriz de Confusión



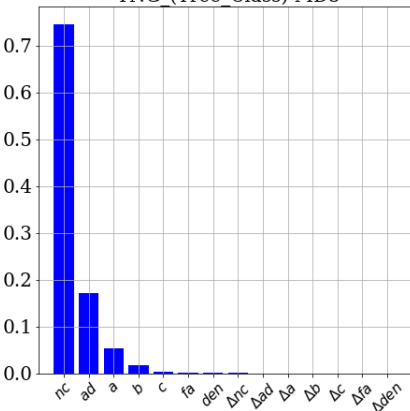
- ▶ De 100 peaks 65 fueron predichos correctamente.
- ▶ De 100 voids 83 fueron predichos como filamentos.



# Importancia de Parámetros

- ▶  $nc$ : Número de conexiones
- ▶  $ad$ : Distancia promedio
- ▶  $a, b, c$ : Momentos principales
- ▶  $\rho$ : Pseudo-densidad
- ▶  $\Delta x$ : Parámetros locales

Environment Prediction Feature Importances  
TNG (Tree Class) MD5





# Gracias!!

