

Applying Density-Based Algorithms to Galaxy Cluster Catalogs 2dFGRS and SDSS.

Unveiling Galaxy Structure with Unsupervised Clustering and 2PCF.

Autor: Carlos Toro Peñas.

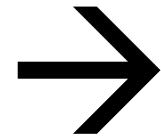
Tutor: Laura Ruiz Dern.

Profesor: David Masip Rodo.

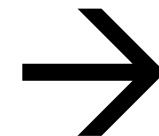
Índice

1. [Cartografiados: Redshift-Space.](#)
2. [Estructura a gran escala en el Universo.](#)
3. [Grupos de Galaxias.](#)
4. [Análisis mediante algoritmos basados en densidad.](#)
 - [Distorsiones en el Redshift-Space.](#)
 - [Mitigación 1: sAlgoritmos.](#)
 - [Mitigación 2: Real-Space.](#)
5. [Densidad a nivel estadístico: 2PCF.](#)
 - [Huellas de las Oscilaciones Bariónicas Acústicas \(BAO\).](#)
6. [Futuros proyectos.](#)
7. [Sources.](#)
8. [Bibliografía.](#)

Cartografiados: Redshift-Space.

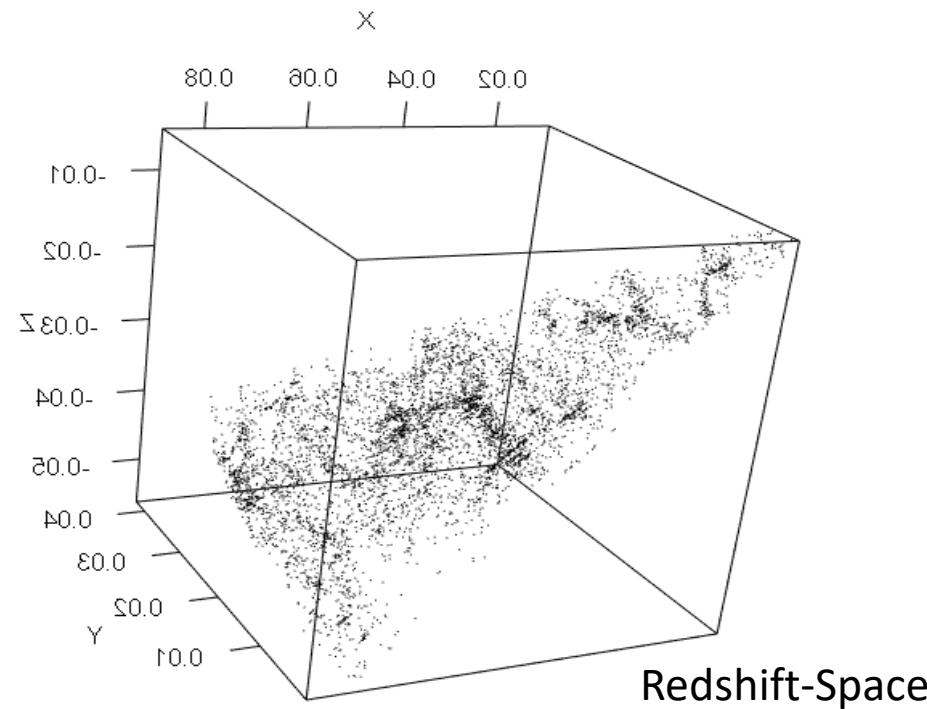


Modelos
Cosmológicos
 λCDM



Observaciones

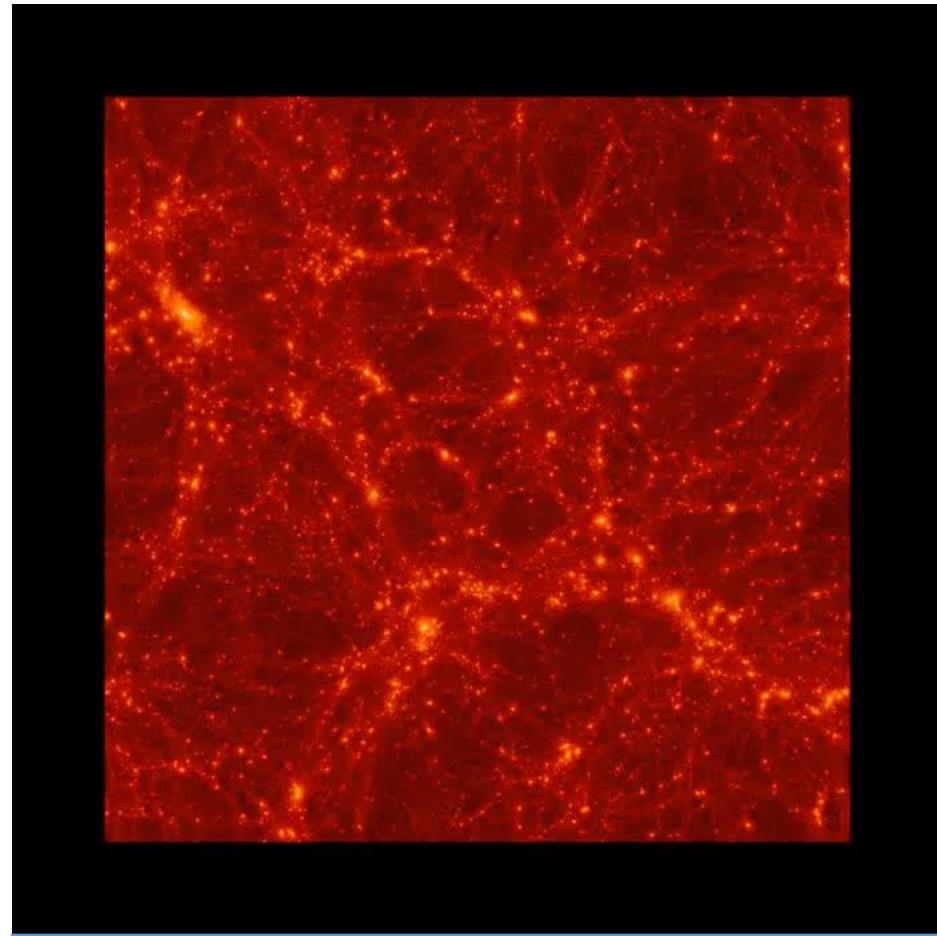
Cartografiados.
Permiten tener
representaciones en
coordenadas x, y, z.



Redshift-Space

Estructura a gran escala en el Universo.

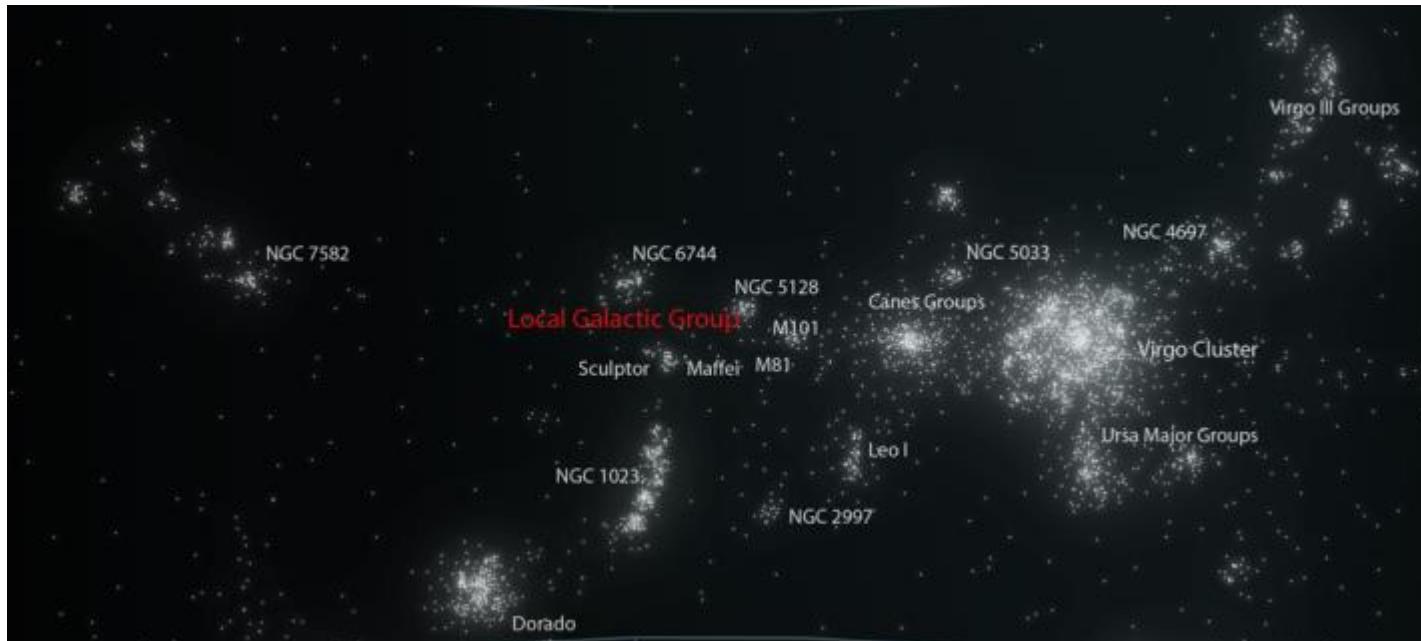
- A muy gran escala se observa la materia agrupada en una estructura de filamentos y vacíos.
- En los filamentos y sobre todo en las intersecciones se encuentran los mayores cúmulos de galaxias conteniendo centenares o miles de ellas.



Fuente: Instituto de Astronomía Teórica y Experimental.

Grupos de Galaxias.

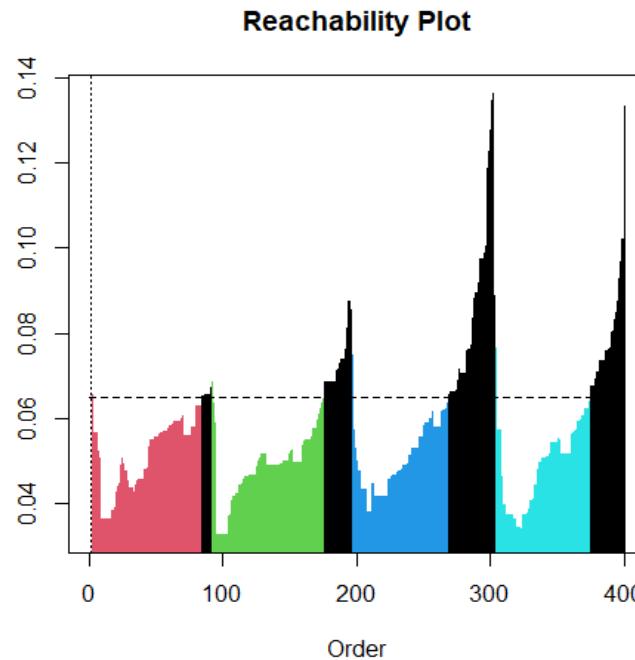
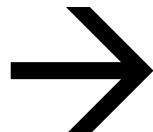
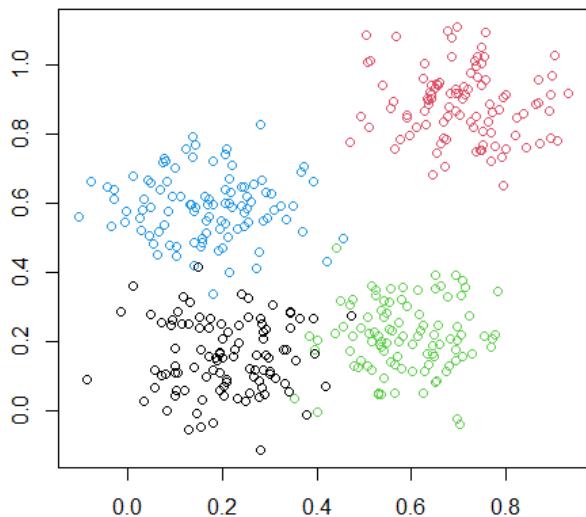
A escalas menores encontramos los grupos de galaxias, que contienen desde 5 hasta 50 miembros.



Los halos de materia oscura son el andamiaje para grupos, cúmulos y supercúmulos de galaxias.

Algoritmos basados en Densidad.

DBSCAN, OPTICS , HBSCAN, DPC ...



¿Cómo funcionan?
Permiten obtener modelos de datos según su densidad bajo la suposición de que hay agrupaciones de objetos separados por áreas más dispersas

¿Pueden los algoritmos basados en densidad modelizar la estructura en grupos observada en los cartografiados el redshift-space?

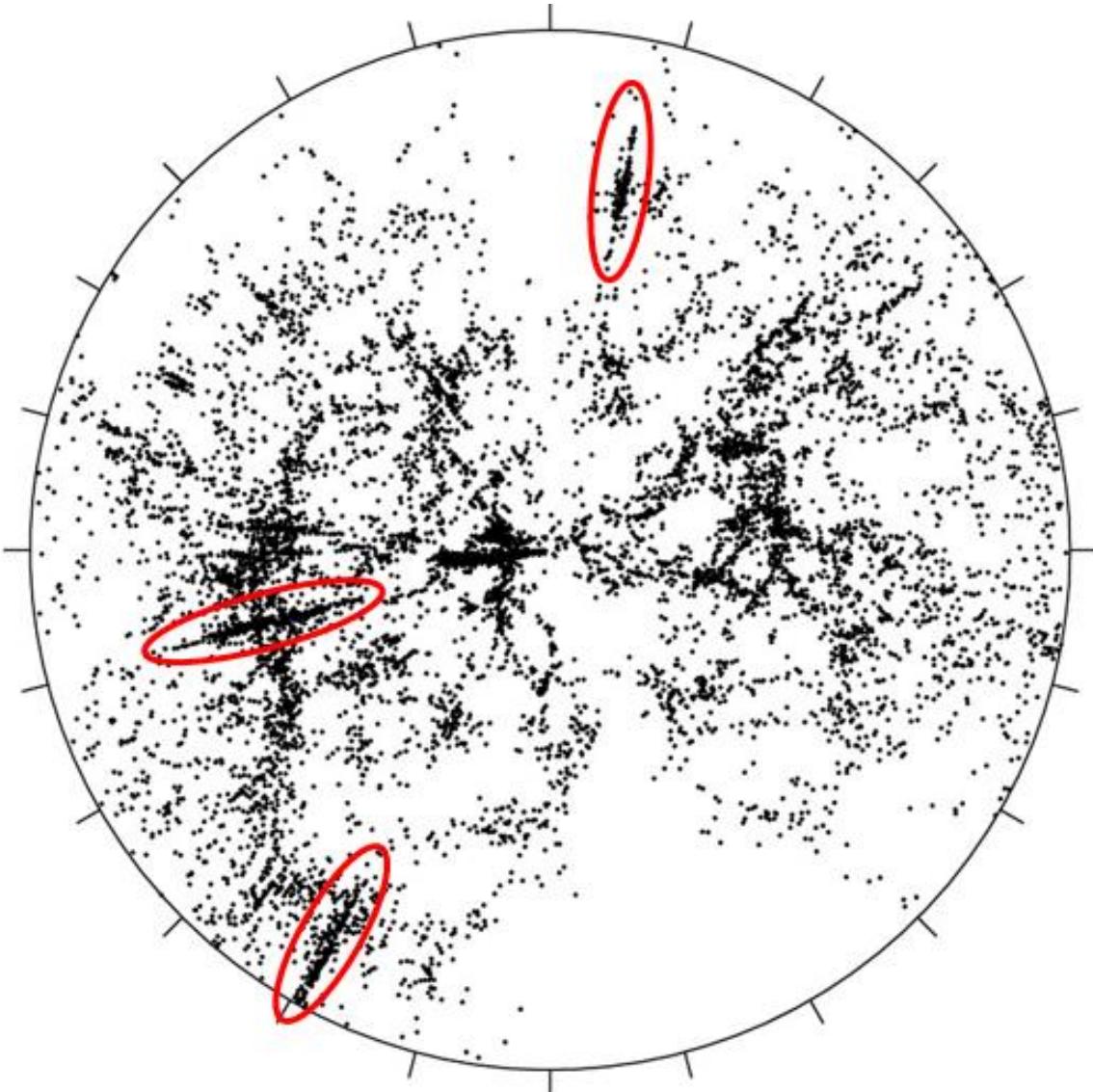
NO*.

Cada algoritmo tiene unas particularidades, pero en general, los datos presentan unas distorsiones en la línea de visión del observador.

En el mejor de los casos se recupera un 50-55% de los grupos originales.

*: Coincide con la literatura: Ma et al: **sOPTICS: A Modified Density-Based Algorithm for Identifying Galaxy Groups/Clusters and Brightest Cluster Galaxies**. arXiv:2405.09855v2 [astro-ph.CO] 18 Jan 2025.

El Redshift-Space.



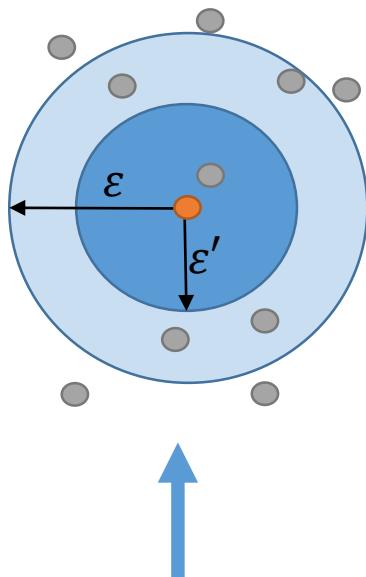
Distorsiones.

Los grupos y clústers aparecen alargados hacia el observador, causado por los efectos Fingers Of God (FOG) y, en menor escala, efecto Kaiser.

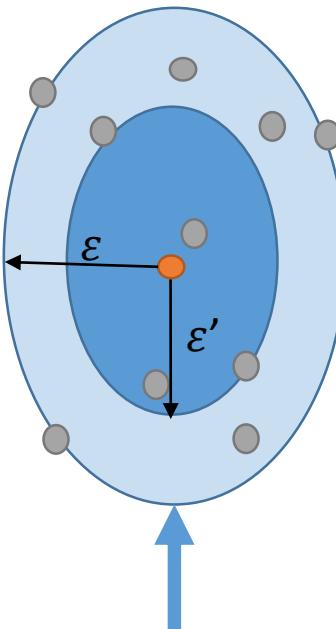
¿Hay forma de mitigarlos?

Mitigación 1: Algoritmos sOPTICS y sDBSCAN.

Escala normal



Escala de s-distance



¿Cómo funciona?

Compensa el efecto FOG sustituyendo la distancia usual (Euclídea), por una distancia alargada en sentido de la línea de visión.

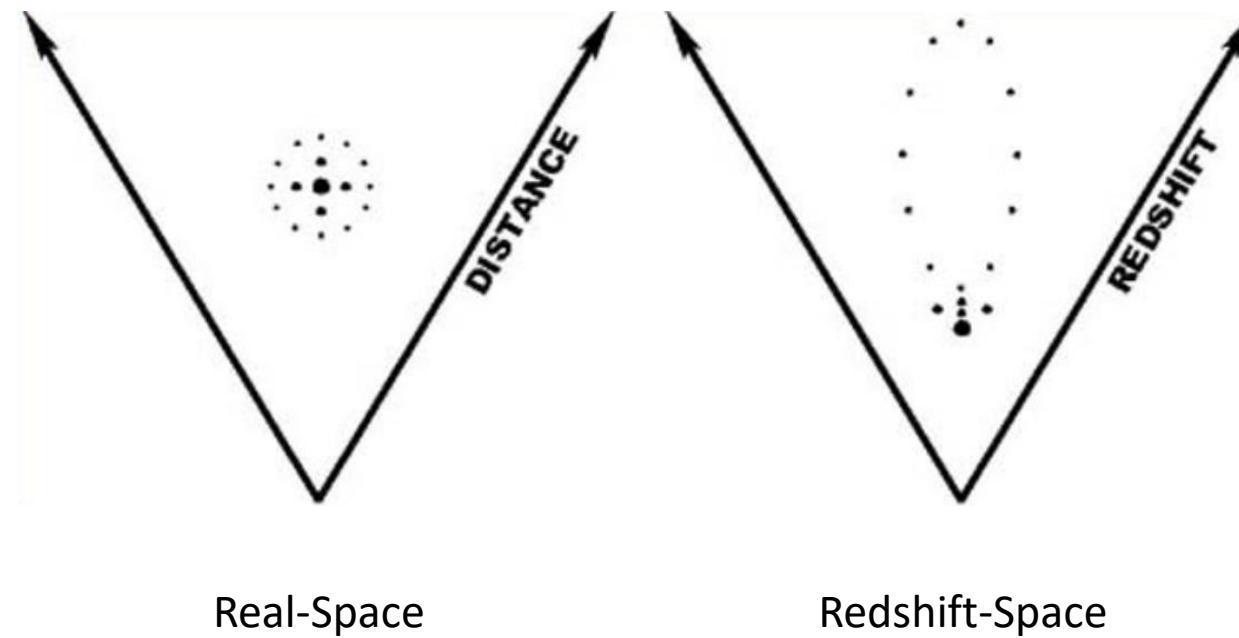
Mitigación 1 (sOPTICS): Resultados

| Alg. | Hyperparam. | Data Sample | Outcomes | $N_{det}(\%) - N_{match}/Total$ | Conclusion |
|----------------|--|-------------|--|---------------------------------|---|
| DBSCAN | $\epsilon = 6 \times 10^{-4}$ $minPts = 5$ | Non-scaled | $P = 0.65$ $C = 0.87$ $R = 0.42$ $U = 21$ | $73(67\%) - 55/92$ | Reasonable cluster detection. Low recovery-rate. |
| HDBSCAN | - | - | - | - | Not good in cluster detection. |
| DPC | $\rho = 8.3 \times 10^{-4}$ $\delta = 0.9985$ | Non-scaled | - | - | Cluster-center detection did not match. |
| <i>sOPTICS</i> | $\epsilon = 11 \times 10^{-5}$ | Non-scaled | $P = 0.84$ $C = 0.84$ $R = 0.86$ $U = 12$ | $85(78\%) - 83/92$ | Best results |
| <i>sDBSCAN</i> | $minPts = 5$ | | | | |
| DBSCAN | $\epsilon = 6 \times 10^{-4}$ $minPts = 5$ | Scaled | $P = 0.72$ $C = 0.81$ $R = 0.41$ $U = 22$ | $65(60\%) - 47/92$ | Acceptable cluster detection. Low recovery-rate. |
| OPTICS | - | - | - | Good in reachability plot. | |

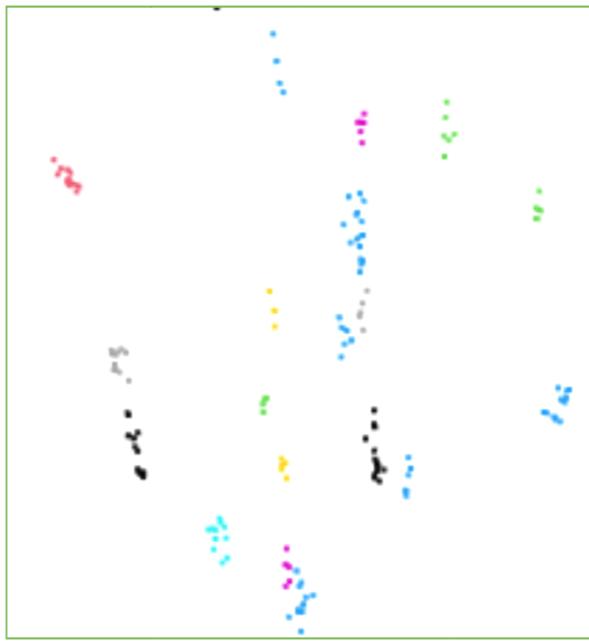
sOPTICS/ sDBSCAN
Entre un 70 y un 75% de los grupos originales.

Mitigación 2: Real-Space.

Se trata de una transformación en las posiciones de las galaxias que corrige el efecto de alargamiento en la línea de visión.



Mitigación 2: Real-Space.



Línea de visión



Línea de visión

¿Cómo funciona?

Se realiza a nivel estadístico.
Dependiendo de la masa del grupo y lo lejos que la galaxia esté del centro la corrección es mayor

Mitigación 2 (Real-Space): Resultados

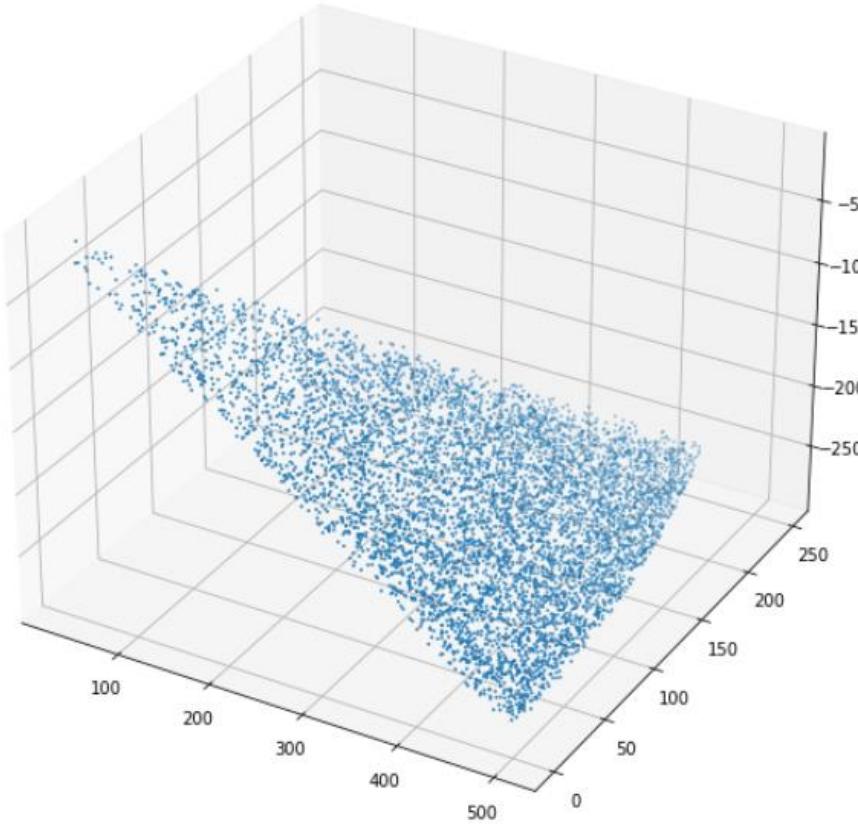
| Alg. | Hyperparam. | Data Sample | Outcomes | $N_{det}(\%) - N_{match}/\text{Total}$ | Conclusion |
|---------|--|-------------|---|--|--|
| DBSCAN | $\epsilon = 3 \times 10^{-4}$ $minPts = 5$ | Non-scaled | $\mathcal{P} = 0.83$ $\mathcal{C} = 0.92$ $\mathcal{R} = 0.99$ $\mathcal{U} = 6$ | 93(88.3%) - 93/95 | Worked in cluster detection and recovery-rate. |
| HDBSCAN | - | - | - | - | Not good in cluster detection. |
| DPC | $\rho = 8.5 \times 10^{-4}$ $\delta = 0.9986$ | Non-scaled | - | - | Cluster-center detection did not match. |
| DBSCAN | $\epsilon = 2.6 \times 10^{-2}$ $minPts = 5$ | Scaled | $\mathcal{P} = 0.88$ $\mathcal{C} = 0.88$ $\mathcal{R} = 0.96$ $\mathcal{U} = 7$ | 91 (86%) 91/95 | Good in cluster detection |
| OPTICS | - | Scaled | - | Good in reachability plot. | |

DBSCAN con Real-Space.

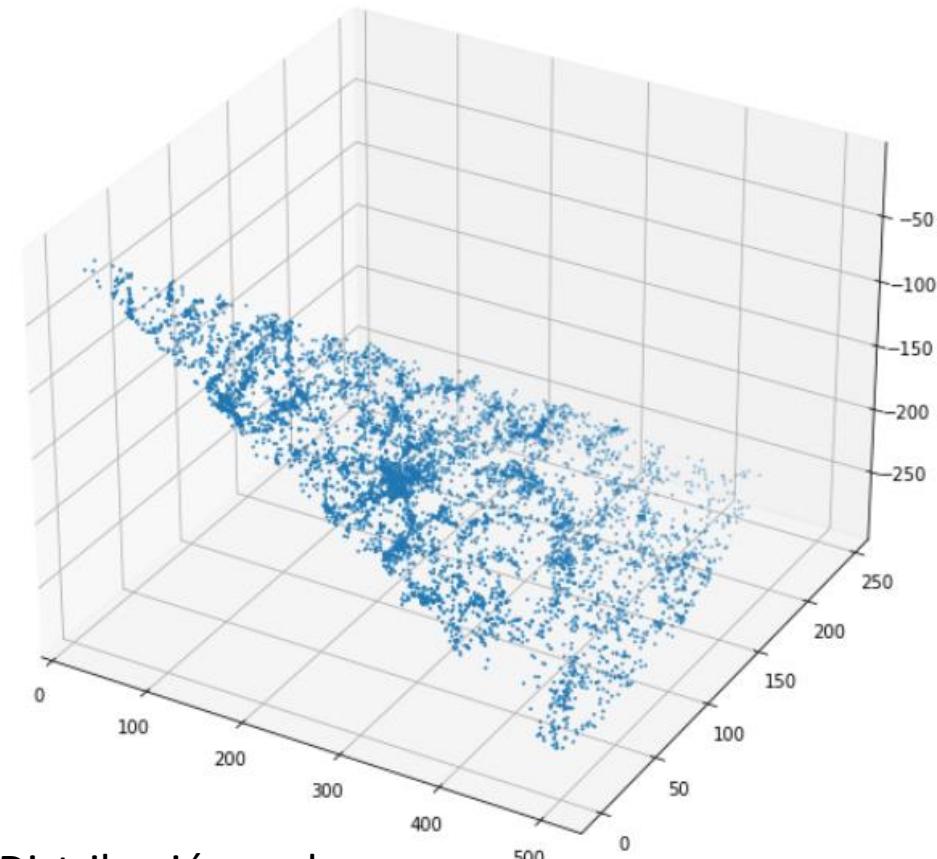
Consigue recuperar un 86% de grupos originales, todos puros y completos.

2PCF: Densidad a nivel estadístico.

¿Cómo están agrupados los datos de una muestra frente a una muestra aleatoria?



Distribución aleatoria



Distribución real

2PCF: Densidad a nivel estadístico.

La función de correlación a dos puntos: $\xi(r)$. Analiza la probabilidad de encontrar pares de galaxias diferentes escalas de distancias frente a un catálogo simulado.

Estimadores de $\xi(r)$

$$\hat{\xi}_N = \frac{DD}{RR} - 1$$
$$\hat{\xi}_{DP} = \frac{DD}{DR} - 1$$
$$\hat{\xi}_{Ha} = \frac{DD\ RR}{DR^2}$$
$$\hat{\xi}_{LS} = \frac{DD - 2DR + RR}{RR}$$

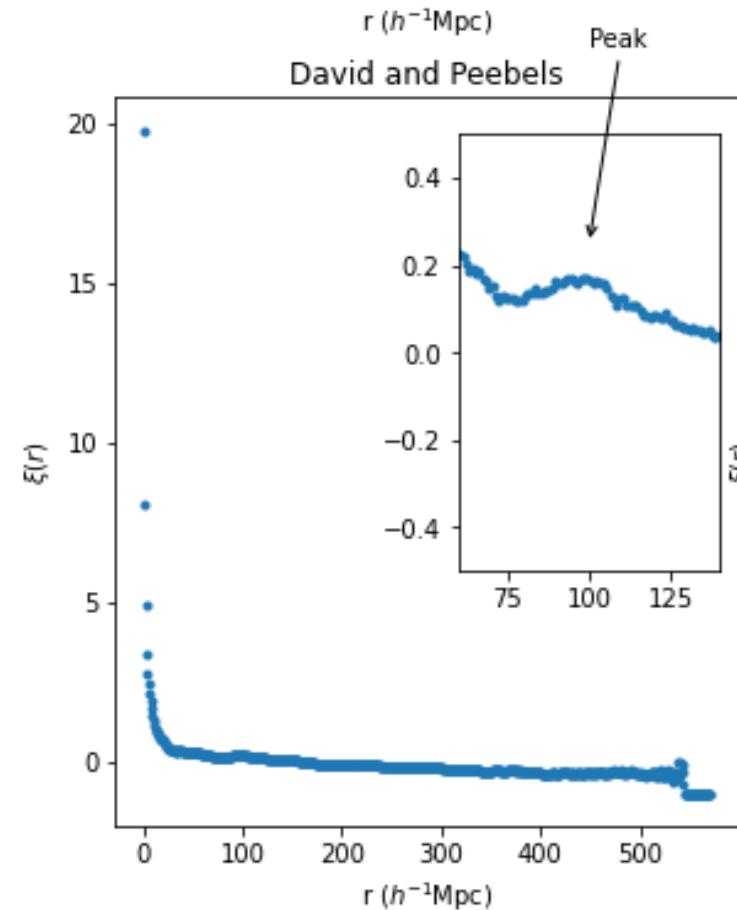
Pares de Distancias:

- DD: Catálogo real.
- RR: Catalogo simulado.
- DR: Simulado-Real

2PCF: Densidad a nivel estadístico.

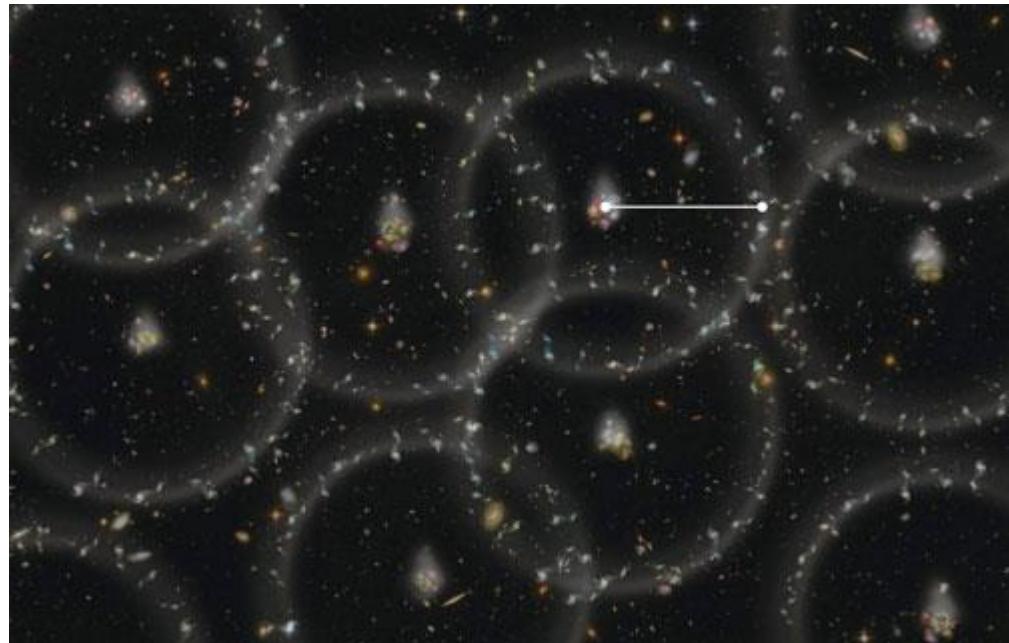
¿Qué se observa?

1. La probabilidad de encontrar pares disminuye con la distancia.
2. A $100 h^{-1}\text{Mpc}$ de distancia se observa un pico de densidad.
3. Los 4 estimadores obtienen lo confirman.
4. Se explica mediante las Oscilaciones Bariónicas Acústicas (BAOs).



Ejemplo inspirado en ejercicio de la asignatura programación en Python en el Master de Data Science de la UOC.

Huellas de las Oscilaciones Bariónicas Acústicas (BAO)



Fuente: BOSS Project.

¿Qué son?

Son el remanente de unas ondas de presión esféricas causadas por la separación de materia oscura y bariónica antes del desacoplamiento de los fotones.

Futuros Proyectos

- Otros catálogos: Dark Energy Spectroscopic Instrument (DESI) de la misión Euclid.
- Análisis de densidad a mayor escala → cúmulos y supercúmulos de galaxias.
- Machine Learning con redes Neuronales → Automatizar el proceso de corrección para el Real-Space.
- 2PCF sobre SDSS → CasJOB SQL interface.
- 3PCF. → Otro análisis estadístico.

Sources.

- <https://earthsky.org/clusters-nebulae-galaxies/what-is-the-local-group/>
- <https://iate.oac.uncor.edu/academicas-2/investigacion/estructura-en-gran-escala-del-universo/>
- <https://medium.com/starts-with-a-bang/theres-no-such-thing-as-a-supercluster-d28272ec7796>
- <https://astronomy.stackexchange.com/questions/18743/fingers-of-god-effect-for-galaxy-clusters>
- <https://www.thunderbolts.info/tpod/2004/arch/041018fingers-god.htm>
- <https://astro.ucla.edu/~wright/BAO-cosmology.html>

Todos los resultados de este estudio y el código fuente están disponibles en:

- <https://github.com/carlostp12/tfm/blob/main/README.md>

Bibliografía.

- Longair S. Malcom. (1996). Our Evolving Universe. Cambridge University press, United Kingdom, UK.
- Cepa J. (2023). Cosmología Física. Ediciones Akal, Barcelona, ES.
- Group catalog. Group Catalogues for 2dFGRS and SDSS. Available at <https://gax.sjtu.edu.cn/data/Group.html>.
- Ma et al.(2025). sOPTICS: A modified density-based algorithm for identifying galaxy/groups/clusters and brightest cluster galaxies. ArchivX.
- Kerscher M. et al. (2000). A comparison of estimators for the two-point correlation function.

Gracias UOC.