

A Large Scale Structure Void Identifier for Galaxy Surveys Based on the β -Skeleton Graph Method

FELIPE LEONARDO GÓMEZ-CORTÉS¹
MASTER STUDENT
CODE 201324084

—
JAIME E. FORERO-ROMERO¹
ADVISOR

¹*Physics Department, Universidad de Los Andes*

RESUMEN

Estamos viviendo en la era dorada de la cosmología observacional. Existe un modelo estándar cosmológico (Λ -CDM) consolidado que explica las observaciones de la Estructura de Gran Escala (LSS) de galaxias mediante la introducción de materia oscura y energía oscura como las componentes dominantes del Universo junto con la materia bariónica. Más aún, somos capaces de realizar mediciones precisas de los parámetros cosmológicos de este modelo a partir de observaciones. Gran parte de estos alcances es debido a la cosmología computacional que es ahora una herramienta fuertemente establecida para probar modelos teóricos y compararlos con las observaciones. Las características principales de la LSS pueden ser reproducidas en grandes simulaciones cosmológicas de N-cuerpos.

Una de las características más prominentes en la LSS son los vacíos: volúmenes irregulares de escalas del orden de decenas de Mpc, donde la densidad de materia está por debajo de la densidad media en el Universo. El análisis estadístico de propiedades de los vacíos, como su volumen, forma y orientación también nos puede dar información cosmológica. Por esta razón existe un gran interés en algoritmos que encuentren y caractericen vacíos tanto en simulaciones como en observaciones.

El objetivo de este trabajo es desarrollar un nuevo buscador de vacíos basado en el método β -Skeleton. El método β -Skeleton ha sido ampliamente utilizado en reconocimiento, procesamiento de imágenes y aplicaciones de *machine learning*, recientemente ha sido introducido en el análisis de LSS. Esta es una herramienta rápida para identificar estructuras filamentosarias en la LSS, y promete ser una herramienta robusta para realizar análisis cosmológicos. Luego de desarrollar el buscador de vacíos caracterizaremos los vacíos del β -Skeleton en simulaciones y observaciones. También realizaremos predicciones para el experimento en desarrollo Dark Energy Spectroscopic Instrument (DESI) acerca de la población que podrá detectar.

Palabras clave: estructura de gran escala, cosmología, vacíos, astrofísica computacional.

1. INTRODUCCIÓN

La cosmología física actual considera al Universo como un ente dinámico. Hay dos aspectos principales en esta evolución: la expansión que hace que la densidad global de materia disminuye con

el tiempo cosmológico y, el segundo, la formación de estructuras localmente cada vez más densas debido al colapso gravitacional.

Las observaciones del fondo de radiación cósmica de microondas (**referencia**) y de la distribución de galaxias a gran escala (**referencias**) apuntan a que esta evolución puede ser descrita por un puñado de parámetros cosmológicos, donde los más importantes son la densidad de materia y la densidad de energía oscura.

El reto de la cosmología actual es aumentar la precisión de las mediciones de estos parámetros cosmológicos. Esto no sólo se logra con mediciones más precisas sino con métodos independientes para acotar los parámetros cosmológicos. Aunque un método independiente pueda tener una incertidumbre grande, considerar las cotas impuestas por varios métodos simultáneamente reduce la incertidumbre sobre los parámetros cosmológicos.

Una de las pruebas cosmológicas que ha emergido en la última década es la caracterización de los vacíos cosmológicos; grandes regiones del espacio que cuentan con una baja densidad de galaxias.

2. ESTADO DEL ARTE

Los vacíos cosmológicos son evidentes en mapas de la distribución tridimensional de galaxias hechos en las últimas dos décadas ([SDSS Collaboration 2017](#)). El proyecto de mapeo más emblemático de la primera década del siglo XXI fue el *Sloan Digital Sky Survey* (*SDSS*) el cuál con el experimento BOSS ha estimado la posición en el espacio de más de 1.5 millones de galaxias de redshift $0 < z < 1.0$.

El trabajo de interpretación de estos mapas en términos de parámetros cosmológicos requiere la realización de simulaciones que sigan la formación de estructuras en un Universo en expansión. Una simulación más emblemática en este aspecto es el *Millenium Run* hecha en el Instituto Max Planck de Astrofísica hace 15 años.

La Figura 1 compara los mapas de la distribución de galaxias obtenidos por simulaciones (rojo) y por observaciones (azul). Cada punto representa una galaxia. En estas imágenes los grandes filamentos donde se aglomeran la galaxia se complementan con las regiones donde hay menos galaxias que conforman los vacíos que nos interesan en esta tesis.

El *Dark Energy Survey Instrument* (DESI) es uno de los catálogos de galaxias de la nueva generación. DESI es un experimento que actualmente se encuentra en etapa de construcción. Usa el telescopio Mayall de 4m en Kitt Peak, California. Tiene dispositivos de ubicación automática brazos robóticos con fibra óptica para tomar en simultánea el espectro de hasta 5,000 objetos por exposición. En la operación inicial se observará el cielo por cinco años, tomando el espectro de más de 30 millones de galaxias previamente seleccionadas. Puede ser considerado una versión potenciada del SDSS (640 espectros simultáneos con las fibras ubicadas manualmente, en un telescopio de 2.5 m).

El desarrollo de DESI ha involucrado varias partes: Por un lado el equipo de construcción del *hardware* de posicionamiento de las fibras, construcción del espectrógrafo y montaje sobre el telescopio. Por otro lado se tiene el equipo de programación de la ubicación de las fibras y el equipo de simulación, donde se corren simulaciones cosmológicas para generar catálogos *MOCKS* del cielo. Este equipo simula también las condiciones observacionales y limitaciones que perturban cada observación, estos datos alimentan el sistema de observación, reducción y análisis de datos para comparar el cielo “verdadero” de los *MOCKS* contra el mapa que construye DESI.

Al estudiar el cielo en esta escala nunca antes vista, DESI generará catálogos de galaxias que permitirán realizar una gran cantidad de pruebas para acotación de parámetros cosmológicos, en

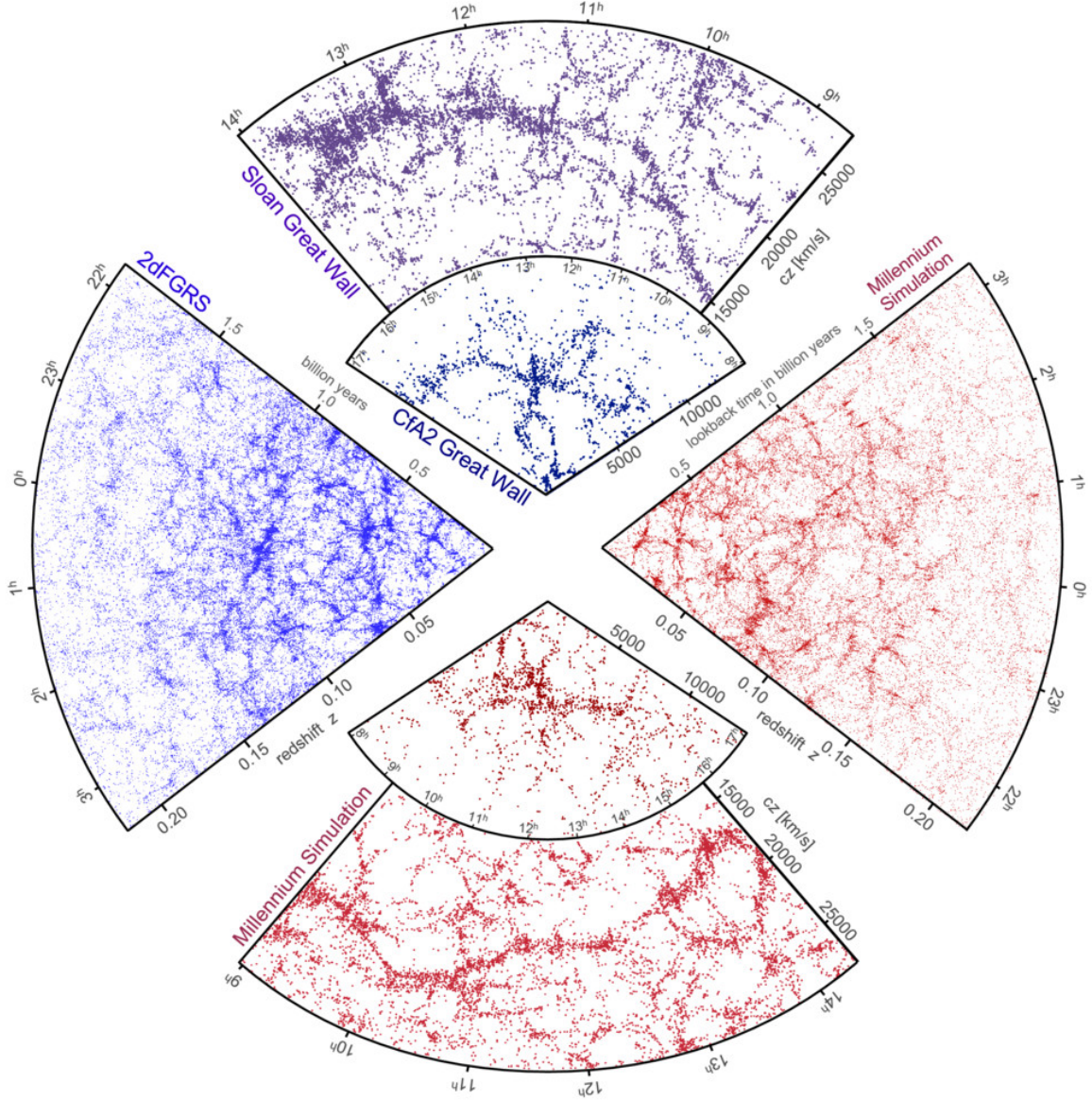


Figure 1. Distribución espacial de galaxias observada en mapeos como el SDSS y el 2dFGRS (en azul) comparadas con resultados de la simulación Millenium (rojo). Instituto de Astrofísica Max Planck Institute. Cada punto representa una galaxia. El punto de observación se encuentra en el vértice central de los mapas. Los mapas se encuentran a escalas distintas apuntando a regiones distintas del cielo. En el segmento de la izquierda vemos la estructura a gran escala compuesta por filamentos donde se agrupan las galaxias y grandes vacíos donde es baja la densidad de galaxias. Este segmento es un extracto del catálogo 2dFGRS, es difícil observar las galáxias lejanas poco brillantes, por eso hacia el exterior se reduce la cantidad de galaxias observadas. A la derecha vemos en un segmento de la simulación Millenium, se tiene en cuenta el mismo efecto para galaxias lejanas poco brillantes. En los dos segmentos superiores vemos estructuras de varias decenas de Mpc encontradas en dos catálogos (SDSS y 2dFGRS). En los dos segmentos inferiores, vemos estructuras con tamaños similares generadas por la misma simulación Millenium.

particular densidad de materia oscura. Este parámetro puede estudiarse mediante análisis de vacíos de la red cósmica.

3. MARCO TEÓRICO

Al estudiar la LSS se encuentra que la mayor parte del volumen del Universo está dominado por los vacíos de la red cósmica donde la densidad de materia es inferior al 20% de la densidad media del universo. Es interesante ver el contraste de la dinámica no lineal de evolución de las regiones con sobredensidades donde se forman estructuras ligadas gravitacionalmente cuyos marcadores son las galaxias (agrupadas en filamentos, paredes, clústers y súperclusters), contra las regiones de baja densidad que obedecen dinámicas lineales, están más cercanas a las condiciones iniciales del Universo. En estas regiones de baja densidad la métrica está dominada por la energía oscura. Aquí yace el interés por estudiar los vacíos en la red cósmica para explorar parámetros cosmológicos.

Colberg et al. (2008) realizaron una comparación de distintos algoritmos para encontrar vacíos en la red cósmica. Estos pueden clasificarse en dos: Métodos que utilizan distribución de materia oscura y métodos que utilizan galaxias/halos.

Los métodos que buscan vacíos en la red cósmica usando materia oscura operan directamente sobre resultados de simulaciones. De un *snapshot* de la simulación extraen el catálogo de todas las partículas en la simulación. Algunos algoritmos generan un campo de densidad suavizado de materia a partir de la distribución discreta de partículas para identificar las regiones de densidad inferior al 20% de la densidad media del Universo(?), este campo puede dividirse en una grilla 3D y se detectan los vacíos como conjuntos de celdas interconectadas que cumplan el criterio de baja densidad (?), otros métodos definen los vacíos de la red como regiones del espacio encerradas por superficies donde la densidad de materia es constante (?). Los vacíos en la red cósmica pueden verse también desde el espacio de fase como las regiones donde las partículas tienden siempre a escapar, por medio de tensores de marea son identificados como regiones de inestabilidad (?). También pueden estudiarse los vacíos utilizando la Transformada de *Watershed* para identificar contornos y delimitar regiones a partir de un campo escalar de densidad (?).

Otra aproximación consiste en buscar vacíos a partir de marcadores de la densidad de materia en el Universo, estos marcadores pueden ser galaxias (directamente de *surveys* experimentales como el SDSS o el 2dFGRS, o bien de simulaciones utilizando métodos semianalíticos para generación de galaxias) o halos de materia oscura (desde catálogos producidos en simulaciones).

Una primera aproximación consiste en buscar regiones esféricas donde no haya galaxias encerradas (?). Otra busca diferenciar galaxias que hacen parte efectiva del contorno de los vacíos según número de vecinos y distancia a los primeros vecinos y repetir el proceso de búsqueda de esferas huecas vacías o con escasas galaxias clasificadas como no pertenecientes a los contornos (?). Otros algoritmos permiten a los vacíos tener formas irregulares. Puede dividirse el espacio en celdas e identificar las celdas vacías o que se encuentren a una cierta distancia mínima de las celadas con galaxias. (?).

3.1. Test AP

Al construir los mapas tridimensionales del universo se convierten los observables (posición angular y corrimiento al rojo) en coordenadas cartesianas en el marco de referencia comovil. Para hacer esta conversión es necesario conocer la métrica de nuestro Universo; cómo se expande el espacio en función del tiempo. El modelo cosmológico estándar Λ CDM describe esta rata de expansión en función de algunos parámetros como la densidad de materia y la densidad de energía oscura en el Universo.

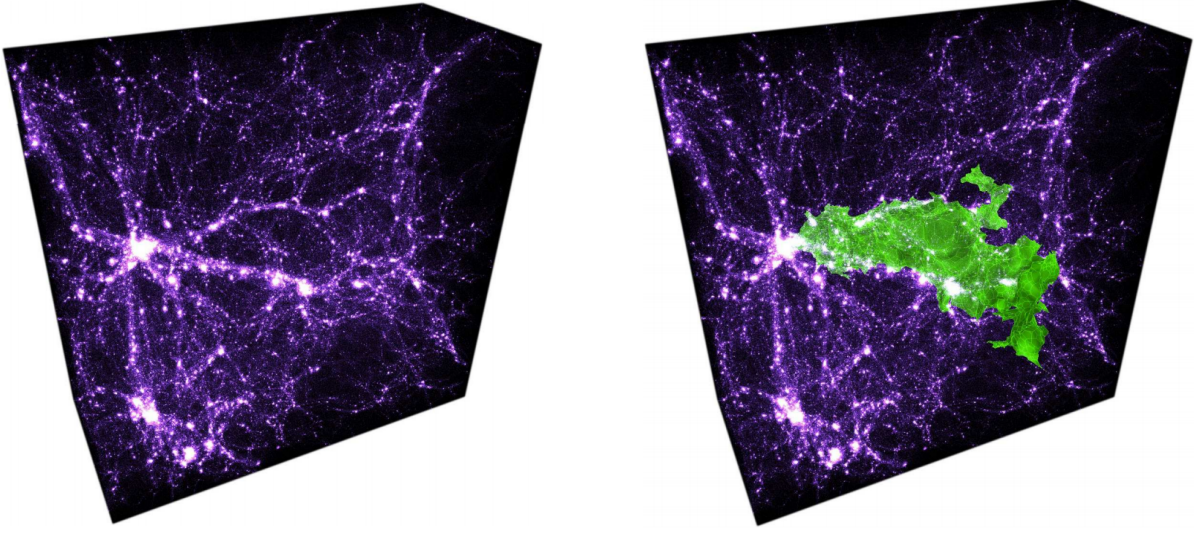


Figure 2. Izquierda: Distribución de materia oscura en una caja de $50 \times 50 \times 20 \text{ h}^{-1} \text{ Mpc}$ obtenida de una simulación de 512^3 partículas en un cubo de $100 \text{ h}^{-1} \text{ Mpc}$ en cosmología ΛCDM . Derecha: Un vacío de la red identificado con el algoritmo DisPerSE. Imagen de ?.

Estos datos se pueden obtener con precisión de sondas como Planck y otro tipo de experimentos (BAO por ejemplo). Si se construye un mapa del universo con los parámetros cosmológicos erróneos, se observarán distorsiones en la distribución de los objetos. Este es el principio del test de [Alcock-Paczynski \(1979\)](#).

Se puede aplicar este test AP al convertir observaciones de vacíos (a partir de las observaciones de galaxias como marcadores de sus límites) a coordenadas comóviles. Una descripción detallada de esta metodología se brinda en [Hamaus et al. \(2015\)](#):

Sea x el vector que denota las coordenadas de un objeto en el espacio. La distancia comóvil en la dirección de la línea de visión se calcula como:

$$x_{\parallel} = \int_0^z \frac{c}{H(z')} dz' \quad (1)$$

Donde c es la velocidad de la luz y $H(z)$ describe la expansión de Hubble como función del corrimiento al rojo z . Si el objeto tiene velocidad radial, esta induce un corrimiento adicional hacia el rojo o hacia el azul. Al tener esto en cuenta queda:

$$\tilde{x}_{\parallel} = \int_0^{z + \frac{v_{\parallel}}{c}} (1+z) \frac{c}{H(z')} dz' \simeq x_{\parallel} + \frac{v_{\parallel}}{H(z)} (1+z) \quad (2)$$

En tanto el ángulo θ entre un par de objetos se relaciona con la distancia comóvil según

$$x_{\perp} = D_A(z) \theta \quad (3)$$

En una métrica de Friedmann-Lemaître-Robertson-Walker de curvatura k , la distancia comóvil angular viene dada por:

$$D_A(z) = \frac{c}{H_0 \sqrt{-\Omega_k}} \sin \left(H_0 \sqrt{-\Omega_k} \int_0^z \frac{1}{H(z')} dz' \right). \quad (4)$$

Este término es bastante sensible al parámetro de curvatura $\Omega_k = 1 - \Omega_m - \Omega_\Lambda$, en tanto la rata de expansión de Hubble depende en sí misma del contenido de energía y materia en el Universo como:

$$H(z) = H_0 \sqrt{\Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda} \quad (5)$$

De este modo se puede ver cómo es necesario asumir los valores actuales correctos de los parámetros cosmológicos para generar mapas tridimensionales de los objetos en el cielo. Si estos valores no coinciden con los parámetros cosmológicos reales, las distancias calculadas en la línea de visión y perpendicular a ella serán incorrectas. Una de las herramientas fundamentales para realizar tests AP son las funciones de correlación. Estas pueden ser aplicadas entre pares de galaxias-vacíos o funciones de autocorrelación vacío-vacío.

3.2. *Análisis de elipticidad*

A partir del análisis estadístico de la morfología y la dinámica de los vacíos se pueden estudiar los parámetros cosmológicos. Las condiciones iniciales de distribución de materia en el Universo se toman como el modelo analítico de Zel'dovich.

? demuestran cómo la elipticidad de los vacíos es sensible a los parámetros cosmológicos, como un balance entre efectos de marea producidos por la componente de materia del Universo y la expansión misma del Universo: las perturbaciones iniciales tienden a crecer esféricamente por la expansión del Universo, aquí entra el parámetro Ω_Λ , en tanto las fuerzas de marea (la contribución de Ω_m) tienden a deformar los vacíos, alejándolos de la forma esférica. Desarrollan modelos analíticos sobre la distribución de elipticidad para la población de vacíos según los parámetros cosmológicos.

? clasifica los vacíos según sus propiedades dinámicas como verdaderos vacíos, filamentosales y tipo pancake. La proporción de estos cambia con los parámetros cosmológicos que se utilicen de fondo. En ? se comparan incluso cosmologías distintas al modelo estándar Λ CDM. Esta referencia muestra un desarrollo completo del principio de análisis de la elipticidad de los vacíos en la red cósmica.

Una vez se han definido los vacíos se calcula su tensor de inercia, donde $a \geq b \geq c$ son los valores propios del tensor análogos a los semiejes en un elipsoide. Se define la esfericidad en términos de la elipticidad $s = 1 - \epsilon = c/a$ y el achatamiento como $p = b/a$. Con esto el volumen del elipsoide equivalente será $V = \frac{4}{3}\pi abc$. Se puede calcular el tensor de marea T_{ij} del potencial gravitacional ϕ como:

$$T_{ij} = \frac{\partial^2 \phi}{\partial x_i \partial x_j} - \frac{1}{3} \nabla^2 \phi \delta_{ij}, \quad (6)$$

los vectores propios de este tensor serán $\lambda_1 > \lambda_2 > \lambda_3$, cuya suma es $\delta_\nu = \sum \lambda_i$, y están relacionados con la esfericidad y el achatamiento según:

$$\lambda_1 = \frac{1 + (\delta_\nu - 2)s^2 + p^2}{p^2 + s^2 + 1} \quad (7)$$

y

$$\lambda_2 = \frac{1 + (\delta_\nu - 2)p^2 + s^2}{p^2 + s^2 + 1}. \quad (8)$$

La función de densidad de probabilidad $f(s)$ para una esfericidad dada viene dada por:

$$\begin{aligned}
f(1 - \epsilon; z) &= f(s; z, R_L) = \int_0^1 \mathcal{P}[p, s | \delta = \delta_\nu; \sigma(z, R_L)] dp \\
&= \int_s^1 dp \frac{3375\sqrt{2}}{\sqrt{10\pi}\sigma^5(z, R_L)} \times \exp\left[\frac{-5\delta_\nu^2 + 15\delta_\nu(\lambda_1 + \lambda_2)}{2\sigma^2(z, R_L)}\right] \\
&\quad \times \exp\left[-\frac{15(\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2)}{2\sigma^2(z, R_L)}\right] (2\lambda_1 + \lambda_2 - \delta_\nu) \\
&\quad \times (\lambda_1 - \lambda_2)(\lambda_1 + 2\lambda_2 - \delta_\nu) \frac{4(\delta_\nu - 3)^3 ps}{(p^2 + s^2 + 1)^3}
\end{aligned} \tag{9}$$

$\sigma(z, R_L)$ es la fluctuación RMS del campo de materia suavizado en una escala R_L a redshift z :

$$\sigma^2(z, R_L) = D^2(z) \int_0^\infty \frac{k^2 dk}{2\pi^2} P(k) W^2(k R_L) d \ln k \tag{10}$$

- el Beta skeleton como un nuevo metodo para describir la red cosmica
- Uso de las simulaciones presentadas en este paper <https://arxiv.org/abs/1809.07355> para encontrar los vacios que se esperan en DESI

El método β -Skeleton ha sido introducido en astrofísica recientemente (Fang et al. 2018) para encontrar el grafo de la LSS en catálogos de halos de materia oscura de distintas simulaciones. Ha probado ser una herramienta útil para detectar estructuras subyacentes diferentes en simulaciones cosmológicas generadas por métodos distintos.

4. OBJETIVOS

4.1. *Objetivo Principal*

Desarrollar un nuevo buscador de vacíos de la red cósmica en catálogos de galaxias basado en el método β -Skeleton.

4.2. *Objetivos Específicos*

- Identificar y catalogar vacíos de la red cósmica en simulaciones y observaciones.
- Acotar los parámetros cosmológicos asociado a la densidad de materia y de energía oscura a partir del catálogo de vacíos de la red cósmica
- Crear un catálogo de vacíos con simulaciones del Dark Energy Spectroscopic Instrument.

5. METODOLOGÍA

La primera parte de este trabajo consiste en desarrollar el código del buscador de vacíos en la red cósmica. Aquí se incluye una etapa de calibración de parámetros libres del buscador para obtener catálogos de vacíos de la red a partir de simulaciones y observaciones.

Luego se hará un análisis estadístico para determinar parámetros cosmológicos. Se espera encontrar estadísticamente que no hayan direcciones privilegiadas para la orientación de los ejes (asumiendo formas elipsoidales en los halos), funciones de correlación en distancias transversales y longitudinales a la línea de observación para hacer pruebas de [Alcock-Paczynski \(1979\)](#). La estadística de la morfología de los halos puede ser comparada con modelos analíticos para revisar proporciones entre poblaciones e identificar si los parámetros cosmológicos con los que se han realizado los cálculos son los parámetros adecuados.

Finalmente se crearemos un catálogo de vacíos en simulaciones que predican el tipo de mapa que se producirá con el Dark Energy Spectroscopic Instrument a finales del 2024.

5.1. Desarrollo del código

El código será escrito en Python3 o C++, podrá ejecutarse en el cluster HPC de la facultad de ciencias de la Universidad de los Andes.

Para desarrollar el código se parte de la lectura de archivos estándar de catálogos de halos de materia oscura o distribución de galaxias. Estos puntos se ubican en un espacio tridimensional. Disponemos de la librería “NGL” ([Correa & Lindstrom 2012](#)) (Neighborhood Graph Library) para calcular la estructura β -Skeleton. Esta librería es de uso libre, está escrita en C++. Inicialmente fue desarrollada para estudiar topología de conjuntos de datos con un número de muestra pequeño. Puede encontrar vecinos en los conjuntos de puntos usando distintos métodos como los Grafos de Gabriel y el β -Skeleton, estos métodos resultan computacionalmente más económicos y rápidos que otros métodos más robustos como la triangulación de Delaunay.

Por definición los vacíos cósmicos son regiones con baja densidad de materia, así que se estudiarán regiones de un campo escalar. Para esto es conveniente dividir el espacio en celdas discretas. El tamaño de las celdas será el primer parámetro a calibrar en el código.

Se aplica el método β -Skeleton para conectar los puntos en el espacio y trazar los filamentos de la red cósmica. En este punto se utilizará la librería NGL ([Correa & Lindstrom 2012](#)). El parámetro β de este método es otro parámetro a calibrar en el código. Se puede utilizar como guía el resultado obtenido por [Fang et al. \(2018\)](#). Se transforman estos puntos y filamentos en un campo de densidad de materia. Este primer campo escalar de materia es bastante discreto y discontinuo. Será suavizado mediante un kernel Gaussiano, el número de veces que sea suavizado el campo de densidad y el tamaño de las celdas del kernel serán parámetros a calibrar en el código.

Una vez suavizado el campo de densidad de materia se identifican los centroides de los vacíos de la red. Esto se puede realizar revisando, por ejemplo, puntos intermedios en conexiones largas que pueden aparecer en un β -skeleton con parámetro β bajo, pero no aparecen en la red obtenida con el valor de β apropiado.

Ya identificados los centroides de los vacíos, se utiliza el método *watershed* para identificar las regiones de baja densidad ([Sutter et al. 2015](#)). Con este método se planta una semilla en los centroides, todas las semillas crecen a la misma velocidad; una celda vecina a la vez. Se establecen ciertas reglas, por ejemplo las regiones no crecerán en las regiones donde el campo escalar de densidad tenga un valor alto, o se detendrán cuando se encuentren con otra región en crecimiento. Otro método a probar para identificar regiones es la construcción de esferas dentro de los vacíos de la red.

6. CRONOGRAMA

1. Desarrollo del código del Buscador de Vacíos Basado en β -Skeleton.

Tareas \ Semanas	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	X	X	X	X	X	X										
2					X	X										
3							X	X								
4							X	X	X							
5									X	X						
6											X	X	X			
7														X	X	X
8	X	X	X	X	X	X	X									
9							X	X	X							
10										X	X	X	X			
11														X	X	X
12																X

2. Calibración de los parámetros del código.
3. Obtención de catálogos de vacíos de la red a partir de simulaciones.
4. Obtención de catálogos de vacíos de la red a partir de observaciones.
5. Comparación por análisis estadístico entre simulaciones y observaciones.
6. Cálculo de la constante cosmológica a partir de catálogo de vacíos de la red en observaciones.
7. Estimación de resultados para el experimento DESI.
8. Escritura del Documento: Introducción y marco teórico.
9. Escritura del Documento: Desarrollo y calibración del Código.
10. Escritura del Documento: Análisis de catálogos de vacíos de la red cósmica.
11. Escritura del Documento: Estimación de posibles resultados del experimento DESI.
12. Escritura del Documento: Conclusiones

7. RESULTADOS ESPERADOS

(Falta esta parte)

REFERENCES

- | | |
|--|--|
| <p>Aarseth, S. J. 2003, “Gravitational N-Body Simulations”, Cambridge University Press.</p> <p>Alcock, C. & Paczyński, B. 1979, Nature, 281, 358</p> <p>Colbert, J. M. et al. 2008, MNRAS, 387, 933.</p> <p>Correa, Carlos & Lindstrom, Peter. 2011, IEEE TVCG 17,12 (Dec 2011), 1852-1861</p> <p>El-Ad, H. & Piran, T. 1997, ApJ, 491, 2, 421</p> | <p>Fang, F.; Forero-Romero, J.; Rossi, G.; Li, X. & Feng, L 2018, arXiv, 1809.00438 astro-ph</p> <p>Hamaus, N.; Sutter, P.M.; Lavaux, G. & Wandelt, B. D. 2015, JCAP, 11, 036</p> <p>Leclercq, F.; Jasche, J. et al. 2015, JCAP, 03, 047</p> |
|--|--|

- Longair, M. S. 2004, “A Brief History of Cosmology”, “Measuring and Modeling the Universe”, Carnegie Observatories Astrophysics Series, Vol 2.
- Press, W. H. & Schechter, P. 1974, ApJ, 187, 425-438
- Schneider, P. 2014, “Extragalactic Astronomy and Cosmology”, Springer
- Springel, V. et al. 2005, Nature, 435, 639
- Sutter, P. M.; Lavaux G.; Hamaus, N; et al. 2015, A&C, 9, 1-9
- van de Weygaert, Rien 2014, Proceedings of the IAU, 308, 493
- SDSS Collaboration 2017, arxiv, 1707.09322 astro-ph
- Hinshaw, G. et al. 2013, ApJS, 208, 20.