

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*

ABSTRACT

We are living the golden age of observational cosmology. There is a consolidated standard cosmological model (Λ CDM) that explains the observed Large Scale Structure (LSS) of galaxies by introducing dark matter and dark energy as the dominant Universe components along with baryonic matter. Furthermore, we are able to do precise observational measurements of the cosmological parameters in that model. Most of this success is due to computational cosmology that is now an established tool to probe theoretical models and compare them with observations. The main features of the LSS can be reproduced in large cosmological N-body simulations.

One of the most striking features in the LSS are voids; irregular volumes on the order of tens of Mpc scales, where the matter density is below the Universe average density. Statistics about voids such as its volume, shape and orientation also encode cosmological information. For this reason there is a great interest in algorithms that find and characterize voids both in simulations and observations.

The objective of this work is to develop a new void finder based on the β -Skeleton method. The β -Skeleton has been widely used on image processing, recognition and machine learning applications, it has been introduced recently in LSS analysis. It is a fast tool identifying LSS filamentary structure, and promises to be a robust tool to make cosmological tests. After developing the void finder we will characterize the β -skeleton voids in simulations and observations. We will also make prediction for the upcoming Dark Energy Spectroscopic Instrument about the void population that could be detected with that experiment.

Keywords: Large Scale Structure, cosmology, voids, computational astrophysics

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 filamentarias 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

Hace poco más de un siglo se concebía el Universo estacionario tan grande como nuestra Vía Láctea, la cosmología había sido construida sobre especulaciones y apenas empezaban las primeras observaciones con métodos y técnicas lo suficientemente sistemáticas y rigurosas para apoyar modelos cosmológicos. Las contribuciones teóricas de Albert Einstein y las observacionales de Edwin Hubble cambiaron por completo el paradigma del Universo en 1929. El Universo era ahora un ente dinámico en constante cambio, que tuvo un origen en algún momento. Lemaître propuso en 1931 un modelo del Universo basado en los modelos de Friedman, donde el universo tendría su origen en un “átomo primigenio”, una singularidad donde toda la energía y materia del Universo estaría concentrada, a una temperatura y una densidad muy altas. Afuera de esta densidad no existía el espacio ni el tiempo. Esta singularidad se expande, a medida que esto sucede la densidad de energía y materia se reducen. En intentos de desprestigiar este modelo, Hoyle lo llamó “Big Bang”, neologismo que ganó popularidad y ha perdurado hasta la fecha sin el lastre peyorativo.

En 1933, Zwicky medía la velocidad de dispersión de galaxias en el clúster Coma para calcular la masa total por el teorema del virial. Luego comparó este resultado con la masa estimada a partir de la relación Masa-Luminosidad, pero el resultado de los dos cálculos discrepaba. En el clúster había mucha más materia que la que daba cuenta de la materia luminosa ordinaria asociada a las galaxias. Esta fue la primera evidencia de la existencia de la materia oscura.

En 1946, George Gamow desarrolló un modelo para la nucleosíntesis basado en el modelo del Universo de Lemaître, con la mecánica cuántica desarrollada hasta la fecha se podía predecir temperaturas y densidades altas donde un gas de partículas elementales (protones, neutrones y electrones) podría dar origen a los núcleos de los elementos más abundantes del Universo: hidrógeno y helio. Chandrasekar y Henrick habían sentado las bases para el desarrollo de la nucleosíntesis unos años atrás. En 1948, Alpher y Herman encontraron que en cierto punto la interacción entre la materia y la radiación dejarían un remanente. En 1965, Wilson y Penzias detectaron en microondas por primera

vez ese remanente, la radiación cósmica de fondo (CMB). Hasta este entonces, el valor calculado de la constante de Hubble había oscilado entre $100 - 650 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

En 1960, von Hoerner realizó las primeras simulaciones en computador resolviendo problemas de N-cuerpos (con 16 cuerpos inicialmente) ligados gravitacionalmente. El valor medido de la constante de Hubble según distintos experimentos ahora oscilaba en el rango $50 - 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. A mediados de los 70's la capacidad de los computadores permitía realizar simulaciones cosmológicas con 300 - 1000 partículas. Las condiciones iniciales venían determinadas por la CMB y desarrollo analítico liderado por Zel'dovich, Sunyaev y Doroshkevich que describían las anisotropías iniciales descritas estadísticamente por espectros de potencias. Poco tiempo después se empezaron a resolver problemas de clusters de galaxias y cúmulos globulares de estrellas con los trabajos individuales de Aarseth, Wilem, Lecar y Almada entre otros pioneros. A mediados de los 80's se realizaban simulaciones con $N = 5.000$ cuerpos. (Aarseth 2003; Press & Schechter 1974). En simulaciones cosmológicas de formación de clusters de galaxias en los 70's se destacan los trabajos de Press & Schechter, Peebles, Efstathiou, Turner; Gott & Aarseth, trabajando con catálogos de galaxias *Catalog of Galaxies and Clusters of Galaxies (CGCG)* (Zwicky et al. 1961-1968), y en los 80's Efstathiou; Davis; Frenk & White, basados en el catálogo *CfA*. En estos trabajos se empieza a hablar de modelos con la constante cosmológica Λ y cierta densidad de materia oscura fría (CDM) para reproducir la Estructura de Gran Escala del Universo (*Large-Scale Structure LSS*); las galaxias no están puestas aleatoriamente en el espacio, obedecen a una estructura de red cósmica con filamentos y paredes donde se observa una gran densidad de galaxias y vacíos en la red donde hay pocas galaxias.

A principios de 1980, Guth propuso un modelo de universo inflacionario para resolver problemas de cómo perturbaciones cuánticas podrían interactuar con regiones más allá del horizonte de eventos, problema señalado dos años antes por Zel'dovich trabajando en el problema de los monopolos magnéticos. Según el modelo inflacionario, el Universo tuvo una expansión tipo exponencial en los primeros 10^{-32} s, después de esto se expande según el modelo de Friedmann. Este modelo de inflación permite conectar causalmente regiones del Universo que hoy en día se encuentran a distancias mayores a lo que la luz podría viajar en la edad del Universo, de este modo se puede explicar la homogeneidad del Universo a gran escala.

En esta misma década se lanzó la misión COBE para estudiar con mayor detalle la CMB, y a la par se realizaban grandes proyectos para generar catálogos tridimensionales de galaxias, estos mapeos del cielo como el *Center for Astrophysics (CfA) Survey* (1978-1982) aprovecharon la reciente tecnología de cámaras CCD. Se encontraron grandes estructuras de galaxias ligadas gravitacionalmente, del orden de $\sim 100h^{-1}\text{Mpc}$. Posteriormente, en los 90's, vienen grandes catálogos como *Las Campanas Redshift Survey (LCRS)*, se tiene en órbita en plenas condiciones de observación el Telescopio Espacial Hubble, se detectan lentes gravitacionales, empiezan a medirse grandes distancias cosmológicas usando Supernovas Ia y con esto se descubre que el universo tiene una expansión acelerada, definida por la misma constante cosmológica Λ que Einstein había propuesto en su modelo del universo. Mientras tanto, el desarrollo de microprocesadores ya permite que las simulaciones cosmológicas manejen cerca de 256^3 partículas (~ 17 millones). El valor de la Constante de Hubble medido por distintos experimentos se restringe al rango $60 - 85 \text{ km s}^{-1} \text{ Mpc}^{-1}$. (Schneider 2014)

2. ESTADO DEL ARTE

El proyecto de mapeo más grande de la primera década del siglo XXI fue el *Sloan Digital Sky Survey (SDSS)*, utiliza un telescopio refractor de 2.5 m de diámetro, dedicado exclusivamente a este

proyecto. Se encuentra actualmente en operación (2018). Proyectos como BOSS del SDSS han tomado espectros de más de 1.5 millones de galaxias de redshift $z < 0.7$, cubriendo $10,000 \text{ deg}^2$, se han estudiado espectros de 160,000 cuásares $2.2 < z < 3$. La última entrega de datos se hizo en 2017, la DR17 es la segunda entrega de la cuarta fase del catálogo (SDSS-IV), incluye datos de BAO *extended Baryon Oscillation Spectroscopic Survey (eBOSS)*, datos de espectroscopía de campo integral de galaxias cercanas *Mapping Nearby Galaxies at APO (MaNGA)*, espectros en infrarojo de estrellas *Apache Point Observatory Galaxy Evolution Experiment 2 (APOGEE-2)*, así como datos reducidos y recalibrados de fases anteriores del proyecto. (SDSS Collaboration 2017)

Wilkinson Microwave Anisotropy Probe (WMAP) es un satélite que operó entre 2001 y 2010. Midió la CMB con mejor resolución espacial y espectral que su predecesora COBE. Permitió calcular la constante de Hubble, las densidades de energía y materia (bariónica y oscura) en el Universo con precisión del $\sim 1.5\%$. La amplitud del espectro primordial con precisión del 3% , restricción del valor de curvatura del espacio $\Omega_k = -0.0027_{-0.0038}^{0.0039}$, la suma de la masa de los neutrinos a $\sum m_\nu < 0.44 \text{ eV}$ con un intervalo de confianza del 95% , y polarización de la CMB. Es de resaltar el trabajo conjunto con el SDSS utilizando los resultados de BOSS, WiggleZ (Australia - UK) y 6dFGS (Australia - UK) para mejorar la precisión en los cálculos de los parámetros cosmológicos. Esto ha convertido a la cosmología en una ciencia de precisión. La última entrega de datos del WMAP fue tras 9 años de operación, dando espacio para una misión con mejor resolución. (Hinshaw et al. 2013)

En la parte computacional, se construye la simulación cosmológica más grande hasta la fecha, (Springel et al. 2005). La simulación Millenium corrió en el clúster de cómputo del Max Plank Institute. Cuenta con $N = 2,160^3 \sim 10^{10}$ partículas, inicia en redshift $z = 127$ y evoluciona hasta el presente, emulando una región de $500h^{-1} \text{ Mpc}^{-1}$. Tiene suficiente resolución como para reproducir galaxias más brillantes que la Pequeña Nube de Magallanes. Las condiciones iniciales vienen dadas por los primeros resultados de la CMB obtenidos por la WMAP. Esta simulación tiene diez veces más resolución que otras de su clase. Tres años después se lanza una versión similar de la simulación en un volumen reducido para mejorar su resolución. Es un gran logro, reproduce la estructura observada a gran escala.

La misión Planck entró en operación en 2009 y tomó datos hasta 2012 cuando agotó su reserva de helio líquido. Este satélite contaba con equipos para la medición de microondas con mejor resolución y mejor resolución espacial. Se encuentra pronto a salir el último artículo de la colaboración Planck.

Parámetro	Símbolo	WMAP2 (2013)	PLANCK(2015)
Edad del Universo (Ga)	t_0	13.74 ± 0.11	13.799 ± 0.021
Densidad de materia Bariónica	$\Omega_b h^2$	0.022 ± 0.0009	0.02230 ± 0.00014
Contenido de materia Oscura	$\Omega_c h^2$	0.1138 ± 0.0045	0.1186 ± 0.0020
Constante de Hubble ($\text{km s}^{-1} \text{ Mpc}^{-1}$)	H_0	69.32 ± 0.80	67.74 ± 0.46
Fluctuaciones de Densidad @ $8h^{-1} \text{ Mpc}$	σ_8	$0.820_{-0.014}^{+0.013}$	0.8159 ± 0.086
Profundidad óptica de reionización	τ	0.081 ± 0.012	0.066 ± 0.016
Índice Espetral	n_s	0.9608 ± 0.0080	0.968 ± 0.006

Bolshoi-Planck-MD (PENDIENTE)

DESI (PENDIENTE)

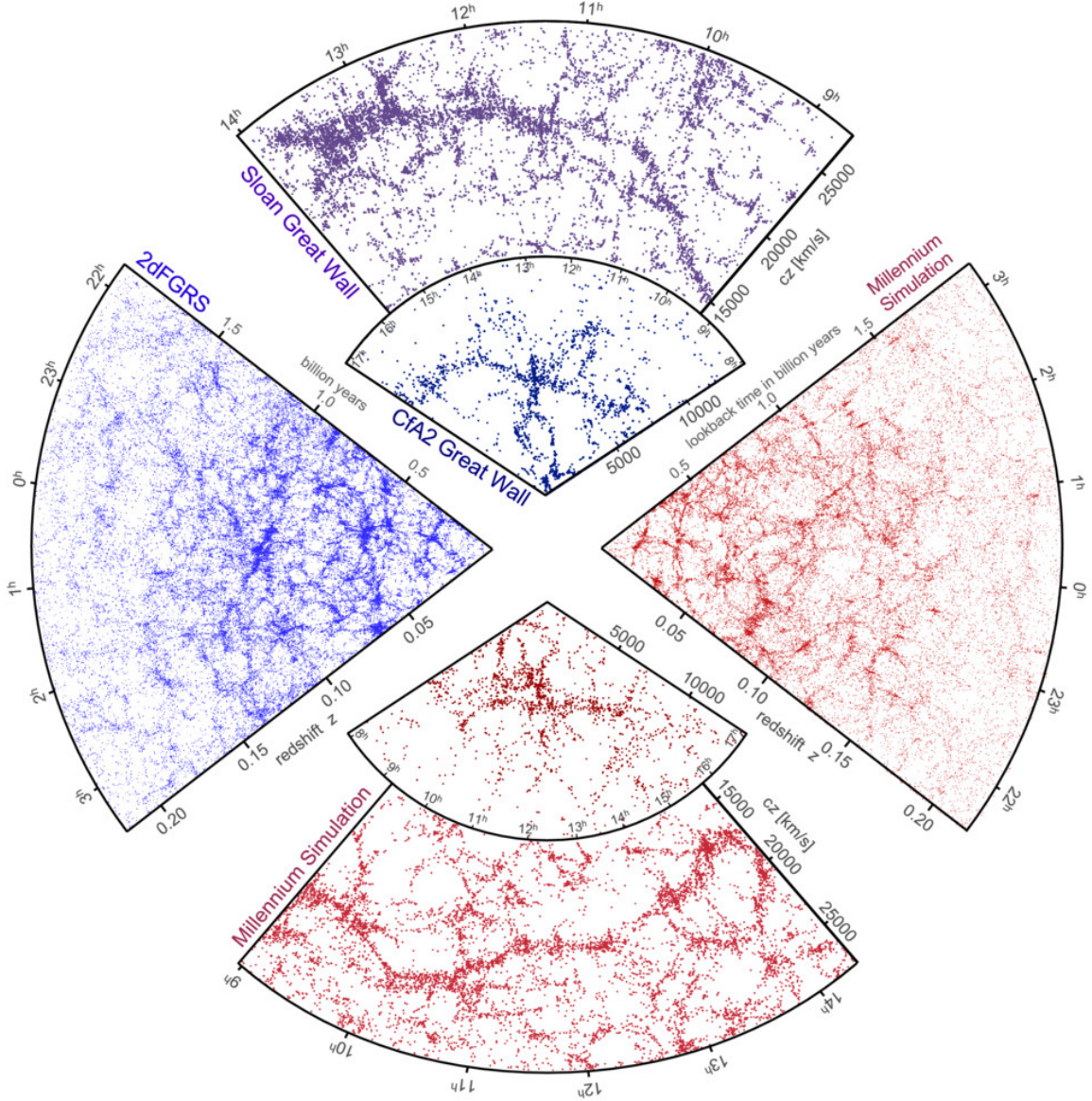


Figure 1. Estructura a gran escala (LSS) observada en mapeos como el SDSS y el 2dFGRS (en azul) comparadas con resultados de la simulación Millenium (rojo). Max Planck Institute for Astrophysics.

3. MARCO TEÓRICO

Actualmente se dispone tanto de catálogos de halos de materia oscura de simulaciones como de catálogos de galaxias de observaciones.

Se dispone 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.

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.

Análisis estadístico por medio de funciones de correlación de dos puntos

PENDIENTE

4. OBJETIVOS

4.1. *Objetivo Principal*

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

4.2. *Objetivos Específicos*

- Identificar y catalogar vacíos de la red cósmica en catálogos de halos de materia oscura de simulaciones y en catálogos de galaxias desde observaciones.
- Calcular el parámetro cosmológico asociado a la energía oscura a partir del catálogo de vacíos de la red cósmica
- Estimar los posibles resultados observacionales que pueda medir el experimento DESI.

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. En la segunda parte se utilizará el código para obtener catálogos de vacíos de la red a partir de catálogos de halos de materia oscura de simulaciones o catálogos de galaxias de observaciones, mediante análisis estadístico de los catálogos de vacíos se compararán simulaciones y observaciones. En la tercera parte se hará un análisis estadístico para determinar parámetros cosmológicos. Finalmente se podrá dar un estimado de las observaciones que podrá llegar a medir el experimento DESI.

5.1. *Desarrollo del código*

El código será escrito en Python3 o C++, podrá ejecutarse en el Colaboratorio o en el 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.

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.

5.2. *Análisis Estadístico de Catálogos de Vacíos de la Red Cósmica*

Se calculan propiedades de las regiones como tamaño, forma, orientación. (Tensor de momento de inercia) (El-Ad & Piran 1997)

Se estudian propiedades de distorsión de esferas hacia elipses por el efecto Alcock-Paczynski (1979).

PENDIENTE TBD

5.3. *Análisis de Parámetros Cosmológicos*

http://cosmo17.in2p3.fr/talks/parallel/Achitouv_COSMO17.pdf

PENDIENTE TBD

5.4. *Posibles Resultados del Experimento DESI*

TBD PENDIENTE

6. CRONOGRAMA

1. Desarrollo del código del Buscador de Vacíos Basado en β -Skeleton.
2. Calibración de los parámetros del código y comparación con otros buscadores.
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

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

7. RESULTADOS ESPERADOS

REFERENCES

- Aarseth, S. J. 2003, “Gravitational N-Body Simulations”, Cambridge University Press.
- Alcock, C. & Paczyński, B. 1979, *Nature*, 281, 358
- Correa, Carlos & Lindstrom, Peter. 2011, *IEEE TVCG* 17,12 (Dec 2011), 1852-1861
- El-Ad, H. & Piran, T. 1997, *ApJ*, 491, 2, 421
- Fang, F.; Forero-Romero, J.; Rossi, G.; Li, X. & Feng, L 2018, arXiv, 1809.00438 astro-ph
- Leclercq, F.; Jasche, J. et al. 2015, *JCAP*, 03, 047
- 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.