

Universidad Nacional de Córdoba

FACULTAD DE MATEMÁTICA, ASTRONOMÍA, FÍSICA Y
COMPUTACIÓN

TEORÍA DE GRAFOS APLICADO A LA
ESTRUCTURA A GRAN ESCALA DEL
UNIVERSO

Futuro Astrónomo

Patricio Ezequiel Colazo

Director: Dra. Heliana Estefanía Luparello

Profesor a cargo: Dr. Diego García Lambas

25 de Marzo 2022



Teoría de grafos aplicado a la estructura a gran escala del universo por
Patricio E. Colazo se distribuye bajo una [Licencia Creative Commons
Atribución-CompartirIgual 4.0 Internacional](https://creativecommons.org/licenses/by-sa/4.0/).

Agradecimientos

Existen muchos esfuerzos invisibles. El título de licenciado no es solo mío, pertenece también a aquellas personas que se esforzaron día a día para permitirme estudiar, reír, aliviar, jugar, escuchar, aconsejar, avanzar ,etc. Esas pequeñas e imperceptibles manos son las que uno siente en la espalda antes de dar un paso atrás o de rendirse o caer. Ojo, no necesariamente esas personas son las cercanas a uno, desde el que te sonríe brevemente, hasta el que se enoja y te invita a replantearse las cosas pertenecen a este grupo. Sin embargo destaco el valioso papel que tomaron las personas en mi vida sobretodo con este logro. Gracias mamá y papá por haberme criado con la mayor libertad y amor posible. Gracias hermanos por haber allanado el camino delante para su hermanito más chico y enseñarme con el ejemplo.

Quiero dar las gracias a una persona especial, por acompañarme, en mi carrera y en mi vida, este logro es una gran parte tuya, después de todo que sería un pato sin su pata, con seguridad sería rengo pero seguro no un licenciado. Estoy donde estoy por tu apoyo, por tu amor y cariño. Por tus buenas palabras cuando necesitaba motivación, por tus abrazos cuando necesitaba contención, por tus retos cuando me hacia falta una corrección. Muy agradecido de encontrarte, siempre das más de lo que es humanamente posible, gracias por ser mi compañera/amiga/novia y permitirme aprender de vos, sos mi ejemplo, mi cariño, mi refugio. Todos Estos Años Me Obnubile, Sobre Estar Libre Escuche, pero tenerte a mi lado es sin lugar a dudas, un magnifico placer. Gracias Sele. Te amo.

Gracias a mis suegros y mi cuñado por tomarme como parte de su familia y ayudarme a avanzar en cada paso. Gracias tía por mandarme batallones de ángeles para cada examen. Gracias tíos y tías por su acompañamiento incondicional. Gracias Salva por ser mi mejor amigo y esforzarte por estar en cada paso de mi vida. Gracias amigos del secundario y de la vida por darme ese espacio de singularidad donde olvidar el mundo. Gracias a mis maestra y maestros/as de Pakua por darme la seguridad de formar la persona que soy hoy. Gracias amigos/as de Pakua quienes acompañan con mucha luz y amor cada paso que doy. Gracias a la dupla Heli y diego por ayudarme a cerrar esta etapa, con una excelente direccion en la tesis y un trato súper humano, me hicieron sentir astrónomo antes de serlo. Gracias a mis profes de la facu quienes tuvieron una ardua tarea en tenerme de alumno. Gracias al tribunal por mejorar esta tesis con sus aportes y aguantarse una lectura muy pesada. Gracias a mis amigos/as y compañeros/as de la facu, quienes realmente merecen un monumento a la paciencia y al compañerismo. Gracias a mis profesores del secundario, primaria y jardín por plantar la semilla que solo el tiempo dirá que tipo de fruto dará el árbol. Gracias a todos los no-docentes, administrativos/as, policías y docentes que no me sufrieron, por siempre tener la mejor de las voluntades en todo lo que les pedí (que fue un montón) y ser muy humanos en su trato. Gracias Martín Leiva por imprimirme la tesis y a Abril Sahade por ir a anillarla y entregarla! Gracias a toda esa gente que seguro me olvide de mencionar pero que fueron parte importante en este proceso! Además muchas gracias a mi por haberme permitido conseguido este logro, esto parece raro pero nunca solemos agradecer a la persona más importante en nuestra vida que somos nosotros mismos, que somos al fin y al cabo los que nos permitimos hacer y conseguir lo que realmente deseamos.

Pd: Si estas leyendo esto manda un mail a patricio-c@hotmail.com para ganarte un premio sorpresa por haber abierto mi tesis.

Muchas muchas muchas gracias a todos!

Patricio E. Colazo, Futuro Doctor en Astronomía)

Un día prenderán la tele y me verán y no sabrán como (ni yo ni ustedes), pero llegue a donde quería llegar.

Autor: anónimo (claramente no voy a decir quien soy.)

Resumen

En este trabajo se utiliza una muestra de 21434 BCGs en el rango $0,18 < z_{foto} < 0,42$, extraída del catálogo de galaxias centrales de cúmulos *GMBCG Cluster Catalog* y del SDSS-DR16. Se utilizan los Árboles Minimales de Expansión (MST) en la muestra principal y en una muestra de control para obtener distribuciones de los cocientes de parámetros del MST. Los resultados obtenidos son consistentes con la función de auto-correlación de cúmulos ricos y concuerdan con un Universo inhomogéneo en escalas menores a $80 Mpc \cdot h^{-1}$. El estudio de los parámetros del MST concluye que la estructura del Universo en gran escala está dominada por cúmulos ricos, cuyas galaxias centrales son masivas y rojas, encontrándose en acuerdo con los trabajos preexistentes de otros autores. Además se encuentran alineaciones relativas entre las BCGs que correlacionan con los parámetros del árbol minimal.

Palabras claves: Estructura en gran escala - BCG - Galaxias Centrales de Cúmulos - MST - Arboles minimales de Expansión - Cúmulos

Abstract

In this work we study a sample of 21434 BCGs in the range $0,18 < z_{foto} < 0,42$, obtained from the GMBCG Cluster Catalog and the SDSS-DR16. We employed a Minimal Spanning Tree (MST) for the main sample and a control sample to obtain a distribution of ratios of the parameters of the MST. The results are consistent with the auto-correlation function of rich clusters and are in agreement with an inhomogeneous Universe up to $80 Mpc \cdot h^{-1}$. The study of the MST parameters shows that the large scale structure of the Universe is dominated by rich clusters, whose central galaxies are massive and red. These results are in agreement with previous works. In addition, we found that the relative alignment between BCGs correlates with the MST parameters.

Keywords: Large scale structure - BCG - Central Cluster Galaxies - MST - Minimal Spanning Tree - Clusters

Índice general

1. Introducción	7
1.1. Marco Teórico y Herramientas	7
1.1.1. Modelo Cosmológico	7
Métrica Friedmann-Robertson-Walker	8
Mediciones actuales de los parámetros cosmológicos	9
El redshift cosmológico	9
Determinación de distancias	10
Formación de estructuras	11
Nuestro Universo	11
2. Catálogos de Galaxias	15
2.0.1. Sloan Digital Sky Survey	15
2.0.2. Magnitudes de SDSS	16
Magnitudes Model	16
2.0.3. Ángulos de posición	16
2.0.4. Cúmulos de Galaxias	17
Galaxias centrales de cúmulos de galaxias	17
Riqueza y riqueza ponderada	18
Dispersión de velocidades	19
3. Árboles Minimales de Expansión	21
Definición	21
3.0.1. Estadística de N puntos	21
Función de Correlación	22
3.0.2. Fundamento	23
Antecedentes	24
3.0.3. Algoritmo de construcción de árboles	25
Parámetros	26
4. Conjunto de Datos	29
4.1. Conjunto de Datos	29
4.1.1. Catálogo GMBCG	29
Técnica de identificación	29
4.1.2. Máscara Tridimensional	31
4.1.3. Análisis preliminar	32
Redshift fotométrico	33
Redshift espectroscópico	33
Color g-r	35
Cálculo de magnitudes absolutas	37
Riqueza y riqueza ponderada	38
4.1.4. Correlación con parámetros del SDSS-DR16	38
Ángulos de posición	39
Dispersión de velocidades	39

Cociente de semiejes	40
4.1.5. Muestra principal	41
5. Metodología	43
5.1. Metodología	43
5.1.1. Objetivo	43
5.1.2. Fundamento	43
5.1.3. Técnica de muestreo aleatorio	43
5.1.4. Cociente de estadísticos	44
5.1.5. Segregación de propiedades	45
6. Análisis de resultados	47
6.1. Análisis de resultados	47
6.1.1. Interpretación de los estadísticos l y b	47
6.1.2. Propiedades del árbol del conjunto completo	48
6.1.3. Segregación por riqueza escalada	51
Separación	51
Análisis de parámetros	51
6.1.4. Segregación por riqueza ponderada	53
Separación	53
Análisis de parámetros	53
6.1.5. Segregación en g-r	55
Separación	55
Análisis de parámetros	55
6.1.6. Segregación por dispersión de velocidades	57
Separación	57
Análisis de parámetros	57
6.1.7. Segregación por cociente de semiejes	59
Separación	59
Análisis de parámetros	59
6.1.8. Orientación relativa entre las BCGs	61
7. Conclusiones	65
8. Trabajo a futuro	71

Capítulo 1

Introducción

Para tratar de explicar los fenómenos del cosmos y los cuerpos celestes que se observan, la astronomía al igual que sus ciencias hermanas, desarrollan modelos lógicos. Estos se basan en ciertos principios establecidos previamente y van evolucionando buscando obtener altas similitudes con la realidad. Al conocer los principios que los rigen, y lograr resultados similares, los modelos buscan inferir las reglas que gobiernan a la naturaleza. Además es importante que el mismo permita predecir acontecimientos posibles de ser medidos, ya que esta es una manera de testear su validez.

1.1. Marco Teórico y Herramientas

El principio de equivalencia y las leyes de la relatividad especial sientan las bases de la teoría de Relatividad General, de la cual surge un conjunto de ecuaciones tensoriales conocidas como Ecuaciones de Campo de Einstein (1.1). Estas requieren adicionalmente una descripción geométrica del espacio y tiempo, introducida por la métrica ($g_{\mu\nu}$), el tensor de Ricci ($R_{\mu,\nu}$) (Goodstein, 1977) y el escalar de curvatura (R). También se precisa una caracterización del fluido que compone el Universo, dada por el tensor energía-momento ($T_{\mu\nu}$).

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = -8\pi GT_{\mu\nu} - \Lambda g_{\mu\nu} \quad (1.1)$$

Resolver este sistema permite explicar y predecir todo evento que ocurre en un Universo 4-dimensional y si bien su resolución no es simple, existen ciertas hipótesis que logran simplificar los cálculos.

1.1.1. Modelo Cosmológico

El modelo cosmológico nace a partir del Principio Cosmológico, el cual postula que *"el Universo en gran escala es espacialmente homogéneo e isótropo"*. Esto significa que no hay lugares ni direcciones privilegiadas donde observar las propiedades del Universo. Actualmente el modelo estándar es el conocido como Λ CDM, donde Λ es la constante cosmológica de Einstein (ecuación 1.1, haciendo referencia a la presencia de energía del vacío o energía oscura, y CDM (*Cold Dark Matter*) la materia no interactuante con la luz, más conocida como Materia Oscura. Estas componentes y sus proporciones generan un Universo con expansión acelerada, cuyo origen se puede explicar con la teoría del Big Bang seguido por una etapa inflacionaria. Existen otras componentes relevantes tales como materia bariónica, fotones, neutrinos, entre otros que componen el Universo. Sus proporciones se obtienen usando como hipótesis el principio cosmológico y ajustando las fluctuaciones de temperatura del fondo de microondas con distintas simulaciones que produzcan un fondo con las mismas propiedades estadísticas.

El modelo estándar y las hipótesis implicadas en él son de gran utilidad para estudiar las propiedades del Universo y sus componentes tal como se realiza en este trabajo. La información

puede condensarse en una métrica diagonal conocida como *Métrica de Friedmann-Robertson-Walker* o *FRW* en adelante.

Métrica Friedmann-Robertson-Walker

La métrica se puede expresar como un elemento de línea

$$ds^2 = dt^2 - a^2(t) \cdot \left(\frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \text{sen}^2(\theta)d\phi) \right) \quad (1.2)$$

en unidades donde la velocidad de la luz $c = 1$. Las variables r, θ y ϕ son coordenadas esféricas comóviles, esto significa que cada punto se mueve debido a la expansión del Universo. El factor de escala $a(t)$ es el responsable de la evolución temporal. La constante de curvatura K tiene 3 posibles valores dando origen a universos con distintas topologías espaciales (cerrada, plana o abierta) y que numéricamente adquieren los valores $(1, 0, -1)$ respectivamente.

El tensor energía-momento $T_{\mu\nu}$ que se involucra en las relaciones de Einstein describe la materia y energía que está contenida en el Universo. Depende de las propiedades intrínsecas al fluido tales como la presión p y densidad ρ , pero también tiene en cuenta la geometría del espacio (métrica $g_{\mu\nu}$).

$$T_{\mu\nu} = pg_{\mu\nu} + (p + \rho)u_\mu u_\nu \quad (1.3)$$

donde u_j son las componentes de la 4-velocidad del fluido.

Tomando la hipótesis de un fluido perfecto en la ecuación 1.3 y la métrica de FRW, las ecuaciones de campo se reducen a las ecuaciones de Friedmann.

$$H^2 = \left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2} \quad (1.4)$$

$$\left(\frac{\ddot{a}}{a} \right)^2 = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \quad (1.5)$$

La ecuación de Friedmann 1.4 puede reescribirse en función de los parámetros cosmológicos Ω_j , que son cocientes entre la densidad actual para cada fluido sobre la densidad crítica del universo ρ_c

$$H = \left(\frac{\dot{a}}{a} \right) = H_0^2 [\Omega_{rad}(1+z)^4 + \Omega_m(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda] \quad (1.6)$$

Se define H como la constante de Hubble donde H_0 es el valor actual. Sin embargo las cantidades en cosmología se dan en función de un parámetro adimensional conocido como h cuya valor es $h = \frac{H}{100 \frac{\text{km}}{\text{Mpc}}}$. En la ecuación de 1.4 el z es el corrimiento al rojo o *Redshift* cosmológico y ρ_c es

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} \quad (1.7)$$

La componente de la materia Ω_m , de la radiación Ω_{rad} y de la energía oscura Ω_Λ se definen como

$$\Omega_j := \frac{8\pi G}{3H_0^2} \rho_{j,0} \quad (1.8)$$

donde $\rho_{j,0}$ es la densidad actual de la materia, radiación o energía oscura. Para la curvatura Ω_K

$$\Omega_K := -\frac{K}{H_0^2} \quad (1.9)$$

Estas constantes permiten ajustar el modelo de Universo a partir de las observaciones. Su determinación es fundamental para entender la evolución del mismo, ya que las interacciones y proporciones de estas cantidades señalan diferentes desenlaces.

Mediciones actuales de los parámetros cosmológicos

En este trabajo, para calcular las distancias de las galaxias centrales de cúmulos de la muestra se utilizaron los parámetros actuales definidos en [Planck Collaboration et al. \(2020\)](#).

La cantidad de materia presente en el universo.

$$\Omega_m = 0,315 \pm 0,007 \quad (1.10)$$

Densidad adimensional de bariones que componen a la materia, el resto corresponde a la materia oscura

$$\Omega_b = 0,0224 * h^2 \pm 0,7 \quad (1.11)$$

Se estima la curvatura del universo plano ($K = 0$)

$$\Omega_K = 0,001 \pm 0,002 \quad (1.12)$$

La densidad adimensional de energía oscura es la componente más importante en el Universo actual.

$$\Omega_\Lambda = 0,685 \pm 0,007 \quad (1.13)$$

El redshift cosmológico

El corrimiento al rojo o redshift, denotado comúnmente como z , es la medición del efecto producido por el desplazamiento en las líneas espectrales de una fuente luminosa hacia longitudes de onda más largas. Se define en la ecuación 1.14 donde λ_{Obs} es la longitud de onda que se observa y λ_{Emit} es la longitud de onda que se emite desde la fuente luminosa.

$$z := \frac{\lambda_{Obs} - \lambda_{Emit}}{\lambda_{Emit}} \quad (1.14)$$

Para calcular el corrimiento al rojo, se utiliza el elemento de línea de la ecuación 1.2 para los fotones, es decir $ds = 0$

$$0 = dt^2 - a(t) \frac{dr^2}{1 - kr^2}$$

$$\frac{dt^2}{a(t)} = \frac{dr^2}{1 - kr^2}$$

donde el fotón recorre una distancia desde las coordenadas $(t_e, r_e, \theta_e, \phi_e)$ hasta $(t_o, r_o, \theta_e, \phi_e)$

$$\int_{t_o}^{t_e} \frac{dt^2}{a(t)} = f(r_1) \quad (1.15)$$

la distancia entre dos crestas contiguas se conoce como longitud de onda y cumplirá lo siguiente:

$$\frac{\lambda_{Emit}}{\lambda_{Obs}} = \frac{a(t_e)}{a(t_o)} \quad (1.16)$$

como el Universo esta en expansión, $a(t_o) > a(t_e)$ el cociente será menor a 1 y por lo tanto $\lambda_{Obs} > \lambda_{Emit}$. Las líneas espectrales tienden a enrojecerse.

Si se toma al factor de expansión $a(t)$ en el tiempo actual t_0 como $a(t_0) = 1$, se obtiene una relación simple con el redshift.

$$a(t) = \frac{1}{1+z} \quad (1.17)$$

La medición de redshift puede obtenerse observacionalmente a partir de un estudio espectroscópico y/o fotométrico del cielo. Cabe mencionar que su determinación se ve afectada por las velocidades propias de las fuentes, agregando una contribución debido al efecto Doppler que genera un desplazamiento extra de líneas espectrales. Estos movimientos propios en el espacio de redshift generan distorsiones o RSD por sus siglas en inglés (*Redshift Space Distortions*).

Determinar el redshift es relevante ya que permite obtener una noción de distancias y del estadio evolutivo de las galaxias.

Determinación de distancias

En los Universos de FRW existen diferentes nociones de distancia, la elección de cual emplear depende fuertemente de lo que se desea medir.

Distancia comóvil

La distancia entre dos objetos fijos en la hiper superficie que se expande por la aceleración del Universo se computa utilizando la parte espacial de la métrica FRW (ecuación 1.2). Para simplificar, se toma el origen de coordenadas y se busca la distancia a un punto de la esfera con coordenadas (r, θ, ϕ) . Se escoge un camino tal que $d\theta = 0$ y $d\phi = 0$.

$$D_c(t) = a(t) \int_0^r \frac{dr'}{\sqrt{1 - Kr'^2}} \quad (1.18)$$

Esta expresión también se conoce como distancia propia.

Distancia angular

Usando la distancia comóvil (D_c) y el ángulo que subtende el objeto θ_A , se puede definir la distancia angular D_A .

$$D_A(t) = \frac{D_c}{\theta_A} \quad (1.19)$$

Distancia de luminosidad A partir de la definición de flujo bolométrico de una fuente se puede definir la distancia de luminosidad D_L .

$$F = \frac{L}{4\pi D_L^2}$$

$$D_L = \sqrt{\frac{L}{4\pi F}} \quad (1.20)$$

donde L es la luminosidad de la fuente y F su flujo. Esta distancia permite definir el modulo de distancia MD como

$$MD := m - M = 5 \log \left(\frac{D_L}{10pc} \right) \quad (1.21)$$

La magnitud aparente se denota como m y la magnitud absoluta como M . Esta última es una medida directa de la luminosidad del objeto. Es importante destacar que ambas deben referirse al mismo filtro, el cual debe de estar desafectado por la corrección K. Esta se puede definir como la modificación sobre la magnitud absoluta de un objeto debido a que el universo se encuentra en expansión por lo tanto objetos similares a diferentes *redshift* tenderán a verse distintos. Realizar la corrección es equivalente a pensar a todos los objetos en un mismo *redshift* conocido como *restframe*.

Si la longitud de onda es λ_0 a un dado z_0 , en $z = 0$ se percibirá $\lambda_1 = \frac{\lambda_0}{1+z_0}$ haciéndolo parecer más débil y por lo tanto más lejos (Véase fórmulas en Hogg et al. (2000)).

Formación de estructuras

Al realizar observaciones se visualiza una distribución no uniforme de galaxias, que al estar ligadas gravitacionalmente forman aglomeraciones jerárquicas a distintas escalas; sobredensidades locales (grupos) y globales (cúmulos, supercúmulos) y zonas globalmente subdensas (vacíos). Sin embargo, el principio cosmológico enunciado anteriormente sigue teniendo validez para escalas mayores a 100 Mpc/h. Estadísticamente, se habla de un Universo homogéneo a grandes escalas, pero existen regiones de hasta cientos de Mpc/h que no lo son. La idea que debe mantenerse es que los apartamientos de la homogeneidad se deben a pequeñas fluctuaciones iniciales y a partir de estas se forman las estructuras del universo. El origen de las mismas yace en la evolución de pequeñas perturbaciones surgidas durante el Big-Bang. Estas pueden ser estudiadas y observadas en el fondo de microondas del cual, debido a la fuerza gravitacional, se amplifican generando estructuras tales como las galaxias, grupos, cúmulos, supercúmulos y regiones filamentosas que se pueden identificar hoy en día. Dentro del paradigma del modelo cosmológico, la formación se produce de manera jerárquica, es decir, de las estructuras más pequeñas a las más grandes. El mecanismo que lo permite son las fusiones de estructuras menores y las inestabilidades gravitacionales.

Además, las aglomeraciones en una región generan zonas subdensas, las cuales pueden ser pensadas en forma de burbujas elipsoidales llamadas vacíos cósmicos o *voids*. Esta compleja red de estructuras se conoce como red cósmica.

Ya que continuamente se habla de regiones subdensas y sobredensas, se requiere definir el concepto de densidad media. A partir de esta se podrá entender con precisión donde hay un decaimiento en la densidad y donde un exceso. La densidad media del Universo se denomina $\bar{\rho}(t)$ mientras que las perturbaciones se pueden denotar como $\delta(\vec{r}, t)$. El campo de densidad queda compuesto por dos términos.

$$\rho(\vec{r}, t) = \bar{\rho}(t)(1 + \delta(\vec{r}, t)) \quad (1.22)$$

$$\delta(\vec{r}, t) = \frac{\rho(\vec{r}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \quad (1.23)$$

donde $\rho(t) = \rho_c \Omega_m$ es la densidad media del Universo para un dado t y $\rho(\vec{r}, t)$ es la densidad en un punto específico. Para sobredensidades se obtiene $\delta > 0$ mientras que $\delta < 0$ para subdensidades.

Si se desarrolla al contraste de densidad como una serie de Fourier, se encuentra $\delta(\vec{k}, t)$, donde \vec{k} es el vector número de onda $|\vec{k}| = \frac{2\pi}{\lambda}$. Cada modo es estadísticamente independiente, acompañado por una fase aleatoria. Asumir esta independencia permite resumir toda la información estadística de las fluctuaciones primordiales en el espectro de potencias $P(k)$

$$P(k) = \langle |\delta(\vec{k})|^2 \rangle \quad (1.24)$$

donde $\delta(k) = \int_{-\infty}^{\infty} \delta(x) \cdot e^{ik \cdot x} d^3x$. Esto es promediar el espectro para cualquier dirección aplicando fuertemente el argumento de isotropía. En el modelo cosmológico estándar $P(\vec{k})$ depende de H_0, Ω_m y Λ (ver ecuación 1.10). Este espectro, al evolucionar en el tiempo, da origen a todas las estructuras actuales que componen el Universo.

Nuestro Universo

El Universo actualmente se entiende en función del modelo cosmológico estándar, con una expansión acelerada, con una geometría plana en las coordenadas espaciales ($k=0$), con presencia de materia oscura y energía oscura. La dinámica del Universo esta dominada por esta última. La constante de Hubble se encuentra en pleno foco de tensión debido a que su determinación se realiza a través de dos métodos muy aceptados que no concuerdan entre sí. Las

mediciones con el fondo de microondas arrojan valores cercano a los $H_0 = 67,4 \pm 0,5 \frac{km}{s \cdot Mpc}$ (Planck Collaboration et al., 2020), en contraste con las predicciones por supernovas arrojando valores de $H_0 = 74 \pm 1 \frac{km}{s \cdot Mpc}$. Para este trabajo, y considerando el modelo de evolución de fluctuaciones a las grandes estructuras se escoge el valor determinado a partir del fondo de microondas para la constante de Hubble.

Para determinar el resto de componentes del Universo se pueden utilizar 3 experimentos entre otros con distintas regiones de degeneración para poder estimar los parámetros cosmológicos más probables. Estas observaciones son: El fondo de microondas (CMB), la expansión medida mediante SNe Ia y el test AP usando la señal de la oscilación acústica de bariones (BAO). Esto puede verse en la figura 1.1 extraído de (Suzuki et al., 2012).

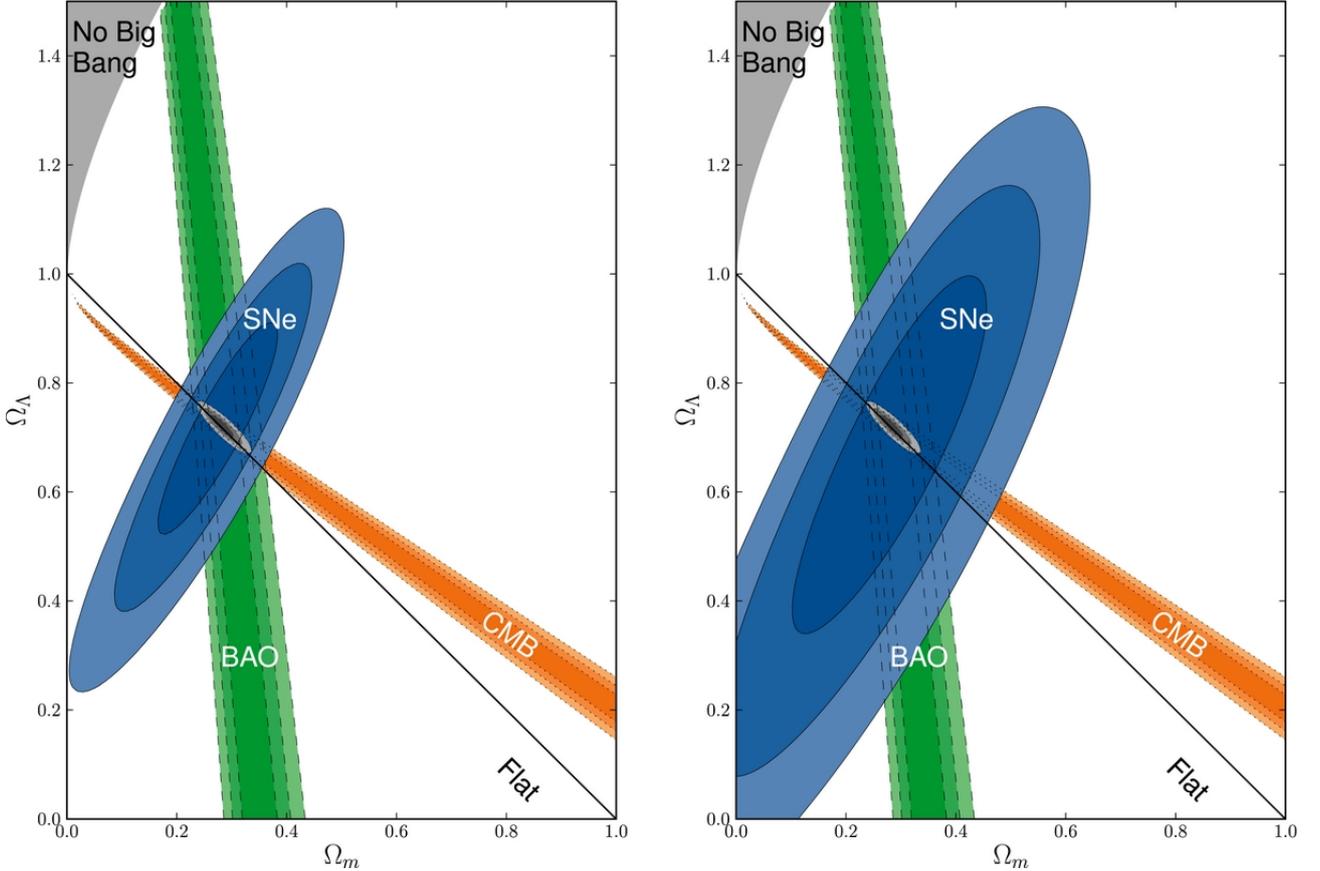


Figura 1.1: Espacio de densidades adimensionales para el modelo Λ CDM, donde el código de tonos corresponde a distintas regiones de confianza con 68,3 %, 95,4 % y 99,7 % del plano de SNe Ia combinado con las restricciones de BAO y CMB. El panel izquierdo muestra la región de confianza de SN Ia sólo incluyendo los errores estadísticos, mientras que el panel de la derecha lo hace tomando errores tanto estadísticos como sistemáticos. Extraído de (Suzuki et al., 2012).

La contribución de la radiación en el universo actual es tan baja que se suele estudiar el plano Ω_m vs Ω_Λ . Esto es así ya que las componentes que rigen la dinámica del universo van cambiando según la densidad que poseen. Si se recuerda la expresión 1.6, a alto z la contribución más relevante se debe a la radiación, luego pasando por la materia y por último a la energía oscura. La razón por la cual el termino de la densidad de curvatura no domina es porque el universo es casi plano y por lo tanto $\Omega_K \approx 0$.

En cuanto a la edad del universo, se estima en unos $13,80 \pm 0,02$ Gyr (Planck Collaboration et al., 2020), desde el Big-bang. La evolución posterior se puede entender como la propagación de fluctuaciones en la densidad. Mediante el mecanismo de la segregación jerárquica se da lugar a la formación de halos, en los cuales se forman galaxias, que a su vez se aglomeran conformando sistemas mayores. Las observaciones, respaldadas por las simulaciones numéricas

cosmológicas, han mostrado que la distribución de materia en el universo responde a un patrón espacial intrincado de estructuras con diversas morfologías, que se interrelacionan en un rango muy amplio de escalas y densidades.

Capítulo 2

Catálogos de Galaxias

Poder acceder a grandes relevamientos de galaxias es fundamental para el estudio de la astronomía extragaláctica, ya que las herramientas estadísticas requieren una gran cantidad de datos. A pesar de este hecho, la confección de catálogos de galaxias grandes es algo reciente. El primer gran relevamiento lo realizó el astrónomo George Ogden Abell en 1958 (Abell, 1958) inspeccionando de manera visual las placas fotográficas del observatorio de Palomar, California. Este contenía galaxias distribuidas en el hemisferio norte, identificando 2712 cúmulos, divididos en 6 grupos de riqueza, pero solo considerando aquellos con 50 o más miembros de galaxias. El rango de redshift que abarcaba era entre $0,02 \leq z \leq 0,2$. Con el tiempo se extendió al hemisferio sur agregando unos 1361 cúmulos en 1989 y confeccionando el hoy conocido *Catálogo Abell de cúmulos de galaxias ricos*.

Por otro lado, en 1961-1968 Zwicky y colaboradores lanzaron sus seis volúmenes con un total de 31350 galaxias y 9700 cúmulos utilizando también las placas fotográficas del Palomar. La diferencia entre ambos fue los criterios para realizar la identificación de los objetos. Mientras que Abell fue más restrictivo con la definición de cúmulos, Zwicky pidió que al menos 50 galaxias en un rango de $m < m_1 + 3$ estén dentro de un dado isocontorno. La diferencia en las definiciones generó que algunos de los objetos del catálogo de Zwicky fuesen en realidad centros de supercúmulos. Cabe mencionar que ninguno de ambos muestreos está libre de contaminación y tampoco son catálogos completos en volumen o magnitud. Sin embargo, era lo mejor que había hasta el momento y se prefirió históricamente los criterios propuestos por Abell como así también el catálogo compilado por él para realizar numerosos trabajos.

Posteriormente existieron otros trabajos que cartografiaban la distribución proyectada de galaxias en el cielo, y con el crecer de mediciones espectroscópicas, se pudo estimar la distribución tridimensional de las galaxias. Ejemplos de esto son los catálogos de Southern Sky Redshift Survey (da Costa et al., 1998), el catálogo de galaxias APM (Maddox et al., 1990), Las Campanas Redshift Survey (Shectman et al., 1996), entre otros.

El catálogo Sloan Digital Sky Survey (*SDSS* en adelante), (York et al., 2000) posee imágenes e información espectroscópica y fotométrica multibanda para una gran cantidad de galaxias. Esta es empleada para el cálculo de las propiedades de estos objetos y puede utilizarse para entender como se correlacionan efectos locales y globales en el Universo. Es por ello que en la actualidad se emplea para realizar estudios extragalácticos y cosmológicos.

2.0.1. Sloan Digital Sky Survey

El Sloan Digital Sky Survey (SDSS) (York et al., 2000) es un relevamiento fotométrico y espectroscópico construido con el telescopio de 2.5 metros del Observatorio Apache Point en Nuevo México. Su importancia radica en el muestreo del hemisferio norte celeste en 5 filtros en el óptico. El filtro *u* tiene como longitud de onda promedio ($\bar{\lambda}$ en adelante) unos $3551[\text{Å}]$ siendo el más cercano al ultravioleta, luego le sigue el filtro *g*, encontrándose entre la zona del azul y el verde del espectro electromagnético en $\bar{\lambda} = 4686[\text{Å}]$. En tercer lugar, dentro del rango de

los naranjas-rojos, se halla el filtro r en $\bar{\lambda} = 6165[\text{Å}]$ sobre el cual se realizan las calibraciones de SDSS ya que es el filtro con mayor señal/ruido, luego le siguen el filtro i entre los tonos de rojos oscuros $\bar{\lambda} = 7481[\text{Å}]$ y por último el filtro z con $\bar{\lambda} = 8931[\text{Å}]$ que pertenece al límite entre el rango óptico y el infrarrojo cercano.

El SDSS cuenta con una base de datos que con el tiempo se fue extendiendo al realizar nuevas observaciones. Para cada actualización (*realese*) se introdujeron nuevos conjuntos de información y actualizaron algunos de los preexistentes, brindando medidas de las propiedades de numerosos objetos utilizando distintos algoritmos.

En el presente trabajo, se utilizó el DR7 (*Data Release 7*), cuyo relevamiento fotométrico abarca unos 11663 grados cuadrados con una PSF de ancho $1,4''$ mediana en r cuya magnitud petrosiana posee un límite en $22,70$ ¹. Por otro lado, el relevamiento espectroscópico cubre unos 9380 grados cuadrados, y un límite en magnitud petrosiana para galaxias en $r < 17,77$. Los espectros se obtienen usando un espectrógrafo de fibras, siendo la separación entre dos fibras sobre una placa de $55''$ al menos. La ubicación de cada fibra esta dada por un algoritmo que maximiza el numero de objetos que pueden ser observados. Debido a la colisión de fibras, las magnitudes aparentes que se deben de tener en cuenta son de $14 < r$.

2.0.2. Magnitudes de SDSS

Cuando se trata de medir magnitudes de objetos extendidos como lo son las galaxias, se pueden emplear diferentes criterios para definir una apertura en la cual medir el flujo del objeto. En este trabajo se utilizaran las magnitudes model para estudiar las propiedades de las galaxias.

Magnitudes Model

Las magnitudes Model miden el flujo empleando aberturas equivalentes en todas las bandas. El algoritmo que se utiliza ajusta dos modelos a la imagen bidimensional de cada objeto en cada banda, ya sea un perfil de deVaucouleurs o un perfil exponencial. Los modelos están convolucionados con un ajuste de doble gaussiana a la función de desparramamiento de un punto (PSF de aquí en adelante) mejorando los parámetros estimados para cada modelo. Esta función modela como se observa una fuente puntual debido a todos los efectos físicos y observacionales obteniendo una distribución de luz más desparramada que un único punto. Posteriormente se elije el que mayor probabilidad tenga de ajustar en el filtro r , y se aplica en las otras bandas después de convolucionar con el PSF apropiado para cada una de ellas. Esto permite que solo varíe la amplitud. Estas magnitudes son útiles para medir los colores de manera no sesgada, ya que se calculan midiendo su flujo a través de aberturas equivalentes en todas las bandas.²

2.0.3. Ángulos de posición

El ángulo de posición de una galaxia es una medida de la orientación de su eje mayor proyectado en el plano del cielo respecto a la dirección del polo norte celeste. Para determinar este ángulo se ajustan isofotas elípticas (niveles de conteo de igual cantidad de fotones) y se mide en relación con la fila y la columna de los CCD orientados al este del norte³.

En el trabajo de [Lupton et al. \(2001\)](#) se describe que dicho parámetro se encuentra muy sesgado especialmente en el caso de las galaxias casi frontales y muy inclinadas. Este parámetro es útil debido a la información que aporta sobre la orientación de las galaxias y permite estudiar si existe o no una alineación en la estructuras en gran escala.

¹<http://classic.sdss.org/dr7/>.

²http://skyserver.sdss.org/dr7/en/help/docs/algorithm.asp?search=mag_model&submit1=Search

³<https://www.sdss.org/dr16/algorithms/classify/>

2.0.4. Cúmulos de Galaxias

Muchas de las galaxias del Universo actual se encuentran en grupos y cúmulos en los que la densidad numérica de galaxias es de decenas a cientos de veces mayor a la media. Las agrupaciones más densas y pobladas se denominan cúmulos de galaxias, que suelen contener más de 50 galaxias relativamente brillantes en un volumen de sólo unos pocos megaparsecs. Las agrupaciones más pequeñas y menos pobladas se denominan "grupos", aunque no existe una distinción bien definida. Los grupos y cúmulos son los objetos más masivos y virilizados del Universo, y son importantes laboratorios para estudiar la evolución de la población de galaxias (Mo et al., 2010). Debido a sus altas densidades superficiales y al gran número de galaxias muy luminosas que los componen, pueden ser identificados hasta distancias muy grandes.

Galaxias centrales de cúmulos de galaxias

En el centro de un cúmulo de galaxias suele encontrarse una galaxia elíptica con luminosidad y masa mayor a todas las galaxias del cúmulo. Esto se corresponde con los modelos de formación jerárquica, ya que los mecanismos dentro de estas estructuras permiten que se produzcan las condiciones óptimas para el crecimiento de las llamadas galaxias centrales de cúmulos (*Brightest Cluster Galaxy* o BCG en adelante). Ejemplos de posibles razones que permiten tal crecimiento son:

i) *Canibalismo galáctico*, que se produce cuando las galaxias caen hacia el centro del cúmulo debido a la fricción dinámica perdiendo tanto gas como estrellas que son acretadas por la galaxia central, contribuyendo al aumento de su masa (Ostriker & Tremaine, 1975).

ii) *Fusiones entre galaxias*, las interacciones en ambientes con sobredensidades producen un incremento en la probabilidad de interacción entre galaxias. Estos fenómenos pueden producir combinaciones entre dos o más elementos del cúmulo, haciendo que se incrementen la masa y la luminosidad de uno de los miembros de la fusión (Merritt, 1985).

iii) *Formación estelar a partir del enfriamiento del gas*, en los cúmulos la presencia de gas caliente representa casi un 15% de la masa del mismo. Para poder ser usado para la formación de estrellas se necesita enfriar lo suficiente (a un ritmo de $10^1 - 10^3 M_* \cdot \text{yr}^{-1}$) debido a la pérdida de energía por la emisión de rayos X. Este proceso se denomina *cooling* y es eficiente ya que en los cúmulos hay grandes reservorios de gas frío (Véase Silk (1976) y Fabian (1994)).

Sin embargo, no todos los cúmulos disponen una galaxia central muy brillante. Estos objetos se encuentran en su mayoría en cúmulos regulares los cuales poseen un centro y una estructura esférica bien definida. La galaxia central dentro del cúmulo tiene una posición espacial que coincide con el pico más bajo del pozo de potencial, por ello suele decirse que están en el centro del cúmulo. Además su ubicación suele estar relacionada con la región de mayor actividad de rayos X.

Una correlación muy aceptada es la del diagrama color magnitud de las galaxias. En el se puede identificar una relación conocida como secuencia roja y es integrada por objetos más elípticos, con poca formación estelar y con ausencia de gas. Por otra parte, puede observarse un conjunto de galaxias presentes en el diagrama que se distribuyen en una zona difusa llamada "nube azul". Las propiedades de estos objetos son una formación estelar muy activa, con morfología tipo espiral o irregular y con alto contenido en gas. Este diagrama fue el resultado del estudio realizado para entender la bimodalidad presentada en la distribución de color de las galaxias.

Riqueza y riqueza ponderada

El parámetro que normalmente se emplea para referirse a cuan poblado es un cúmulo es el de riqueza. Definir la extensión de los cúmulos no es trivial por lo que se requiere determinar un criterio para el borde de la estructura. Para abordar este problema, se suele emplear una apertura a partir de su centro dentro de la cual se contabilizan la cantidad de objetos.

Por otro lado los objetos superpuestos a lo largo de la línea de la visual dificultan la determinación de la riqueza. Si se dispone de mediciones multibanda se pueden emplear dos técnicas para descontaminar la medición, tal como se menciona en (Hao et al., 2010):

1. Usar los colores, con ellos se pueden estimar los desplazamientos al rojo fotométricos y a partir de los mismos desproyectar.
2. Usar la secuencia roja para detectar la agrupación directamente en el espacio de color.

Se escoge el segundo camino tomando en consideración el método de detección, las estimaciones de cuan sesgadas están ambas opciones (a partir de simulaciones) y la presencia de la BCG que agrega características únicas a la secuencia roja.

Al cuantificar los miembros se suele preferir una apertura fija con la idea de mantener una estimación de riqueza consistente. Lo ideal sería examinar y elegir una serie de radios de apertura que maximice la S/N pero debido a su costo computacional se implementa un enfoque de dos pasos similar a Koester et al. (2007b). En primer lugar, se mide la riqueza del cúmulo usando una apertura fija; luego se escala el radio en función de la riqueza medida con esta apertura y se obtiene un radio escalado. Posteriormente se realiza una segunda vez todo el procedimiento pero cambiando la apertura fija por la nueva apertura que tiene como radio el radio escalado y al contabilizar los miembros dentro de esta nueva apertura se obtiene la riqueza escalada.

El procedimiento de medición se detalla en el trabajo de (Hao et al., 2010):

1. Se toma como apertura fija un círculo de radio de 0,5 Mpc alrededor del candidato BCG en su redshift fotométrico y se cuentan como miembros todas las galaxias más débiles que la BCG, pero más brillantes que el corte de $0,4L^*$ en el z_{foto} mencionado anteriormente.
2. se pretende ajustar en el espacio de color dos gaussianas, una correspondiente a las galaxias de fondo y otra a la de secuencia roja, puede pasar que el candidato BCG tenga más probabilidad de pertenecer a la gaussiana ancha, en cuyo caso es eliminada de la lista de candidatos por que no pertenece a la secuencia roja. Otra posibilidad es que solo se pueda ajustar una sola gaussiana en la ventana de 0.5 Mpc con centro en la BCG, por lo tanto en este caso se toma como ancho máximo de esta gaussiana un ancho de color de 0,16 (aproximadamente el doble del ancho intrínseco) como un umbral apropiado en los colores g-r y r-i.⁴ En el caso que la única gaussiana encontrada exceda este umbral, se procede a descartar la BCG por no pertenecer a una secuencia roja. El criterio para escoger cuando es posible ajustar una o más gaussianas se determina a partir del criterio de información de Akaike (AIC) (Akaike, 1974). Una vez ajustada la gaussiana correspondiente a la secuencia roja y la confirmación de que la candidata a BCG pertenezca a esta, todas las galaxias cuyos colores están dentro de ± 2 desviaciones estándar de la media de la componente gaussiana correspondiente, se marcan como miembros.
3. A continuación, luego de obtener una estima de riqueza se procede a reescalar el círculo de 0,5 Mpc. Para ello se asume que hay una relación entre la apertura y la riqueza medida (Vease (Hansen et al., 2009)).

$$R_{scale} = N \left(N_{gals}^{0,5Mpc} \right)^P \quad (2.1)$$

⁴Las pruebas exhaustivas en grupos ricos en el SDSS establecen este valor

donde N es una normalización y P la potencia tal que R_{scale} sea aproximadamente cercano al R_{200} . Para ajustarlos se emplean datos provenientes del catálogo maxBCG clusters (Koester et al., 2007a) con lo que estiman $N_{gals}^{0,5Mpc}$ y con análisis de lentes débiles (Johnston et al., 2007) (Hansen et al., 2009) obtienen R_{200}^{lens} . Se consigue la siguiente relación.

$$R_{scale} = 0,133 \left(N_{gals}^{0,5Mpc} \right)^{0,539} \quad (2.2)$$

4. Finalmente se repite el proceso de medición de miembros con este nuevo radio. Aquí queda mencionar que el conteo de miembros se puede realizar con un modelo de mezcla de gaussianas (GMM) o uno más general que incorpore corrección de errores de las mediciones (ECGMM). Cuando se aplica GMM (ECGMM) para ajustar la distribución de color, cada componente gaussiano tiene un peso. Este cuantifica cuánto de la población total proviene del componente gaussiano corresponde a la secuencia roja. Al multiplicar el peso relativo del componente del cúmulo por el número total de galaxias en el campo se mide la riqueza ponderada. Mediante un proceso de Montecarlo se analiza cual algoritmo responde para estimar la riqueza o la riqueza ponderada. Como resultado, el autor encuentra que GMM ajusta mejor la riqueza mientras que ECGMM la riqueza ponderada. Es por esta razón que se escoge una manera híbrida de determinar estas cantidades. Primero se detecta la secuencia roja usando ECGMM y luego se mide la riqueza ponderada. A continuación se emplea GMM con un número fijo de mezclas (según los resultados de ECGMM) y se seleccionan los miembros de secuencia roja. El número de miembros estimado a partir de GMM se denota como riqueza y el obtenido de ECGMM, riqueza ponderada.

Dispersión de velocidades

La dispersión de velocidades de las galaxias (σ) es una medida de la varianza de la distribución de las velocidad de las estrellas de una galaxia determinado a partir del ensanchamiento de las líneas espectrales. Esta dispersión esta fuertemente correlacionada con la distribución de masa ya que esta genera el movimiento de las estrellas.

Para estimar la dispersión de velocidades se requieren datos espectroscópicos con calidad suficiente que permitan estimar los anchos de las líneas de absorción. El procedimiento puede encontrarse detallado en el sitio web de SDSS ⁵. El procedimiento incluye la comparación con plantillas estándares (*templates*) generados con poblaciones estelares a σ constante.

⁵<http://www.sdss3.org/dr8/algorithms/redshifts.php>

Capítulo 3

Árboles Minimales de Expansión

Desde los conocimientos provenientes de la teoría de grafos, surge la técnica conocida como Árbol minimal de expansión o MST por sus sigla en inglés. Como se menciona en [Graham & Hell \(1985\)](#), el MST se suele enunciar como un problema de comunicaciones donde cada punto representa una ciudad y cada línea del grafo representa el costo de hacer la comunicación. La idea era encontrar la red de menor costo que conectara todas las ciudades.

Definición

El MST es un algoritmo de agrupamiento o *clustering* del cual se puede extraer información intrínseca de una distribución de puntos. Para una mayor comprensión es más simple vislumbrar como se construye, pero para ello hace falta introducir ciertas ideas previas. Un grafo está compuesto por una colección de nodos, bordes o aristas (líneas rectas que unen nodos) y longitudes de bordes. Para el presente trabajo, los nodos serán las posiciones tridimensionales de las galaxias del catálogo y las longitudes de borde serán las distancias tridimensionales entre cada par de nodos. Una secuencia de aristas conectadas entre sí es un camino; un camino cerrado se llama circuito y un grafo se llamará conectado si hay un camino entre cualquier par de nodos ([Barrow et al., 1985](#)). Un grafo conectado que no contiene circuitos es un árbol. Si el árbol de un grafo conectado, denotado como F , contiene todos los nodos del conjunto de datos, se denomina árbol de expansión. La longitud de un árbol se define como la suma lineal de las longitudes de los bordes de sus componentes. El árbol de expansión mínima (MST) es el árbol de expansión de longitud mínima. Si no hay dos longitudes de borde iguales, el MST será único.

Algunas definiciones útiles son:

Grado: el número de aristas que salen desde un nodo. **Peso:** el valor asociado a cada arista calculado a partir de una métrica dada. **Rama:** cadenas de aristas conectadas desde un nodo grado 3 a un nodo grado 1 donde todos los nodos intermedios tienen grado 2. **Nodo hoja:** nodo de grado 1.

Poda: Un MST se poda al nivel p cuando se han eliminado todas las k -ramas con $k \leq p$.

Separación: eliminar del MST cualquier borde cuya longitud exceda algún corte.

3.0.1. Estadística de N puntos

Los estudios de las distribuciones de puntos suelen emplear estadísticas de dos puntos (entre pares), pero sólo en los casos donde la distribución es gaussiana, esta estadística es suficiente para describir toda la distribución. Para cualquier otro caso, ya no es suficiente y se requiere calcular la estadística de N puntos de la distribución, que en principio, posee toda la información que puede describir una distribución de puntos ([Szapudi & Szalay, 1998](#)). Para obtener la estadística de 2 puntos se calcula la distribución de distancias entre pares, para la estadística de 3 puntos se agrupan de a triplete y se observan la distribución de ellos. Generalizando, al tomar las distancias de N -tuplas y estudiar su distribución se obtiene la estadística de N puntos. Al realizar estas correlaciones para todos los ordenes, se obtiene la información total de

la distribución. Sin embargo, el problema recae en que esta técnica requiere una gran cantidad de computo haciendo inviable su calculo.

Es por ello que se buscan métodos alternativos como el MST, que si bien no contiene toda la información presente en los estadísticos de N puntos, permite capturar parte de esta información e identificar patrones esqueléticos (Naidoo, 2019). Antes de abordar los fundamentos del MST, se describe una de las herramientas más empleadas para estudiar una distribución de puntos, la función de correlación de pares. Esta constituye una base de resultados que pueden compararse con los obtenidos a partir del MST.

Función de Correlación

La función de correlación $\xi(r)$ es una herramienta estadística simple y poderosa que tiene como objetivo cuantificar el exceso de probabilidad de encontrar N objetos a una dada distancia respecto de una distribución de Poisson. Teóricamente, existen las funciones de correlación de orden superior, las cuales son una generalización de la función usualmente empleada para dos puntos. Si bien estas herramientas logran capturar mayor información de una distribución son relativamente difíciles medir debido a su gran costo computacional.

Debido a esto, la función de correlación de dos puntos es la más utilizada. Se define como la probabilidad de encontrar un par de objetos ($dP_{1,2}$) a una dada distancia. Si la \bar{n} es la densidad media de galaxias, la probabilidad de encontrar un objeto en el elemento de volumen dV_1 y correlacionarlo con la probabilidad de encontrar otro en el volumen dV_2 dentro de una distancia $\vec{r}_{1,2}$ está dado por

$$dP_{1,2} = \bar{n}^2 \cdot [1 + \xi(\vec{r}_{1,2})] \cdot dV_1 dV_2 \quad (3.1)$$

Su importancia es la de medir el grado de agrupación o *clustering* presente en el conjunto de datos. Debido a las hipótesis de isotropía global, la correlación en gran escala solo puede depender de la separación relativa entre los dos objetos. Si se tiene un campo de densidad continuo, la expresión 3.1 se transforma en

$$\langle \rho(\vec{r}_1) \rho(\vec{r}_2) \rangle = \bar{\rho}^2 \cdot [1 + \xi(\vec{r}_1, \vec{r}_2)] \quad (3.2)$$

El conteo de pares de galaxias separadas por distancias entre r y $r + dr$ puede denotarse como $DD(r)$ y se expresa

$$DD(r) = \frac{N_{pares}(r)}{N_{tot}} \quad (3.3)$$

donde $N_{pares}(r)$ es la cantidad de pares que hay dentro del cascarón esférico y N_{tot} es la cantidad total de pares. Por otro lado, el conteo de pares random se denota como $RR(r)$ y puede computarse de manera análoga a partir de un catálogo de partículas aleatorias. Ambas probabilidades están relacionadas por la función de correlación de dos puntos que se puede aproximarse mediante el estimador natural:

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1 \quad (3.4)$$

Existen otros estimadores de la función de correlación de 2 puntos. Una manera alternativa es la dada por Davis & Peebles (1983). La diferencia con la expresión 3.5 es que emplean la variable $DR(r)$ en vez de $RR(r)$. Con ella se cuantifica la probabilidad de encontrar pares con un elemento en el catálogo de galaxias y el otro en un catálogo de partículas aleatorias.

$$\xi(r) = \frac{DD(r)}{DR(r)} - 1 \quad (3.5)$$

Estos estimadores son importantes para el presente trabajo por dos motivos. En primer lugar existen muchos estudios basados en la implementación de esta herramienta y por lo tanto se

dispone de valores de comparación. En segundo punto, la idea de cotejar los pares encontrados en los datos con una muestra aleatoria. Esta idea se detalla en la construcción del método que se explica en las secciones subsiguientes (Véase capítulo 5).

A través de la función de correlación se puede cuantificar el grado de estructura del Universo. Esta puede ser empleada no sólo para correlacionar galaxias, sino que también para correlacionar cualquier tipo de objetos (cúmulos, grupos, etc.), e incluso para realizar correlaciones cruzadas (galaxia-cúmulo, void-galaxia, etc).

Por la naturaleza del presente trabajo, es relevante entender el comportamiento de la función de correlación cúmulo-cúmulo. Para tener una referencia, en [Hong et al. \(2012\)](#) calculan la correlación para 13904 cúmulos de galaxias en un rango $z \leq 0,4$. Estos datos son seleccionados a partir de un catálogo de cúmulos ([Wen et al., 2009](#)) y los resultados se cotejan con otras muestras de cúmulos. Se extrae del trabajo la Figura 3.1 y se presenta la función de correlación respecto de la distancia de correlación para dos muestras de cúmulos. Para diferenciar los conjuntos de puntos se escoge separar los valores en $2h^{-1}Mpc$ hacia la derecha para los puntos en negro respecto de los cuadrados blancos. Se observa en línea sólida empleando el modelo Λ CDM que mejor ajusta la muestra completa. El cuadro realizado dentro de la figura corresponde a $\xi(r) \cdot r^2$ para mostrar con mayor claridad la correlación detectada en $r \approx 110 h^{-1} \cdot Mpc$. Este máximo en el conteo de a pares se corresponde con la evolución de las fluctuaciones iniciales en el presente. Esto concuerda con el modelo de formación jerárquica que establece la formación de grandes estructuras a partir de otras más pequeñas. Cabe mencionar que se observan correlaciones bien definidas en escalas menores a $80Mpch^{-1}$.

[Hong et al. \(2012\)](#) encuentran que la función de correlación depende del parámetro de riqueza (cuantas galaxias hay en un cúmulo en un cierto radio), es decir, la longitud de la correlación aumenta a medida que se incrementa la riqueza de los cúmulos. Además encuentran que los cúmulos más ricos tienen una mayor longitud de correlación. Estas conclusiones fueron confirmadas por otros autores utilizando varias muestras de cúmulos ([Croft et al. \(1997\)](#); [Collins et al. \(2000\)](#); [Lee & Park \(2002\)](#); [Bahcall et al. \(2003\)](#); [Estrada et al. \(2009\)](#)).

3.0.2. Fundamento

En el trabajo de [Barrow et al. \(1985\)](#), se utilizó esta técnica para evaluar patrones intrínsecos en conjuntos de datos puntuales (vértices). Esta idea es importante debido a que, por construcción, al asignarle a cada arista una cierta función de peso, puede mostrar algún tipo de patrón no intuitivamente visible. Además esta técnica permite la inclusión de información estadística de orden mayor sobre patrones en la red cósmica, que puede por ejemplo proporcionar información adicional para mejorar las restricciones de los parámetros cosmológicos ([Naidoo, 2019](#)).

En el trabajo de [Graham & Hell \(1985\)](#) se hace una recopilación de los tres métodos más utilizados para encontrar un árbol de expansión minimal a partir de un grafo (resumido a continuación). Si bien al pensar en esta técnica se citan los trabajos de [Kruskal \(1956\)](#) y [Prim \(1957\)](#), el pionero en el tema fue [Borvka \(1926\)](#), quien sentó las bases del problema intentando solucionar la minimización de costos para una red eléctrica. Los tres Algoritmos más importantes que solucionaron eficientemente este problema son:

i) El algoritmo 1 (Kruskal): Primero se clasifica los bordes por peso; luego se toman las aristas en el orden de aumento de peso teniendo en cuenta el siguiente criterio: si el borde examinado no crea un ciclo con los bordes del subárbol actual se escoge y de lo contrario, se descarta. Este algoritmo necesita ordenar al comienzo (*sort*) que sea eficiente.

ii) El algoritmo 2 (Prim): Para comenzar se toma un vértice cualquiera y se ordena las aristas de ese vértice, que luego se une utilizando la de menor peso. Posteriormente, se toma un vértice aleatorio, y se escoge la arista de menor peso, en este punto pueden darse dos situaciones. La primer alternativa es que el borde examinado conecte con un vértice que no esta en el subárbol,

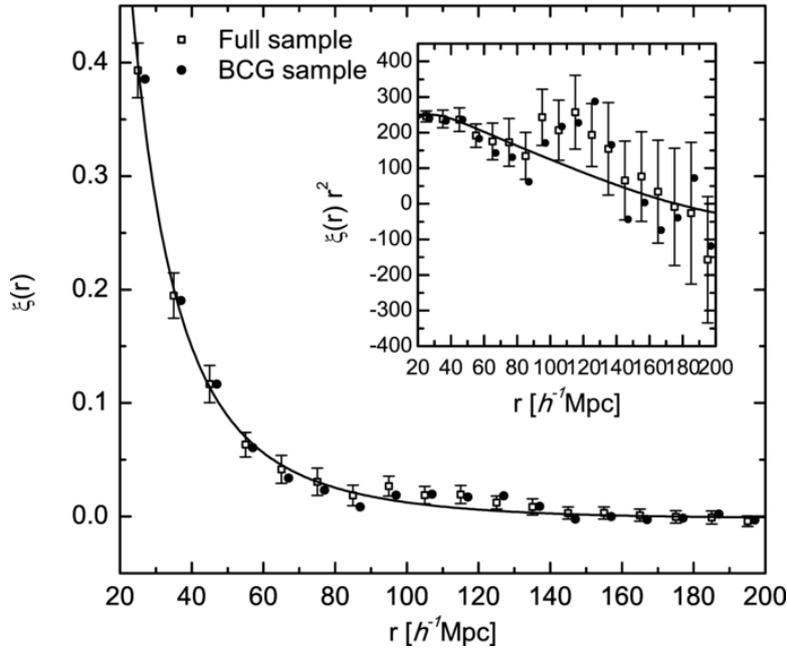


Figura 3.1: La función de correlación para una muestra de 13904 cúmulos (puntos cuadrados) y otra muestra de 11103 cúmulos con desplazamientos al rojo conocidos de BCGs (puntos en negro, desplazados a la derecha en $2h^{-1}Mpc$ para mayor claridad). La línea sólida es el modelo Λ CDM que mejor ajusta la muestra completa. El cuadro realizado dentro de la figura corresponde a un ajuste $\xi(r) \cdot r^2$ para mostrar con mayor claridad la correlación detectada en $r \approx 110 h^{-1} \cdot Mpc$. Este máximo en el conteo de a pares se corresponde con la evolución de las fluctuaciones iniciales en el presente. Esto concuerda con el modelo de formación jerárquica que establece la formación de grandes estructuras a partir de otras más pequeñas.

en cuyo caso, se guarda la información momentáneamente, pero no se realiza la conexión. La segunda posibilidad es que la arista conecte con un nodo que este en el subárbol, entonces se realiza la conexión y se busca en la información guardada momentáneamente si algún vértice conectaba con este nuevo vértice del subárbol. Cabe destacar que si ambos vértices pertenecen al árbol, la arista en cuestión no se escoge ya que daría lugar a un ciclo.

iii) El algoritmo 3 (Boruvka) se hacen pequeños fragmentos hasta que cada vértice pertenezca a un subárbol, luego se unen los subárboles.

Para elegir que método aplicar, [Graham & Hell \(1985\)](#) hace comparaciones entre ellos concluyendo lo siguiente: En las implementaciones modernas, se prefiere el algoritmo 1 cuando los bordes están preordenados o pueden ordenarse en tiempo lineal. En tal caso, el algoritmo 1 tiene un límite de tiempo de $O(e)$, donde e hace referencia a la cantidad de aristas del grafo. En todos los demás casos, el algoritmo 2, con el uso de montones, es al menos tan bueno como el algoritmo 1, y para gráficos densos, es preferible utilizar la implementación Prim-Dijkstra original sin montones.

El algoritmo 3 tiene un límite de tiempo de $O(e \cdot \log(\log(v)))$ donde e son la cantidad de aristas y v la cantidad de vértices. De hecho, existen algoritmos prácticos estrechamente relacionados con él con los mismos límites de tiempo del peor caso y con límites de tiempo del caso promedio de $O(e)$.

Antecedentes

No es la primera vez que esta técnica se emplea en el área de astronomía, ya que varios trabajos en la literatura la utilizan de manera muy distinta. [Barrow et al. \(1985\)](#) usa la técnica para evaluar patrones intrínsecos en conjuntos de datos puntuales mientras que [Campana et al.](#)

(2008) lo emplea sobre imágenes bidimensionales de rayos γ donde se introducen algunos filtros para seleccionar clusters de datos y reducir las detecciones espurias. Otros trabajos se centran en estudios de forma (Véase [Cartwright & Whitworth \(2009\)](#)) cuyo objetivo es caracterizar la geometría de un cúmulo de estrellas, en especial su elongación.

Por supuesto se pueden encontrar trabajos más cercanos al área de gran escala (Véase [Krzewina & Saslaw \(1996\)](#) o [Ruiz et al. \(2015\)](#)). En este último por ejemplo se indaga sobre la existencia de correlaciones entre la dinámica del interior de los vacíos y su entorno. Empleando el MST [Pereyra et al. \(2020\)](#) analiza las propiedades de los filamentos cúmulo-cúmulo dentro de una simulación cosmológica.

Si bien el método sirve como identificador para muestras cosmológicas, también a sido empleado de formas versátiles como puede verse en el trabajo de [Beuret et al. \(2017\)](#) en el área de astronomía estelar. En el se investigan la agrupación de las fuentes en el infrarrojo lejano del catálogo Hi-GAL. Motivado por los antecedentes y su potencialidad es que se propuso emplear esta técnica en la realización del presente trabajo

3.0.3. Algoritmo de construcción de árboles

El trabajo de [Naidoo \(2019\)](#) se basa en la implementación del algoritmo MST optimizado para aplicaciones cosmológicas. El paquete denominado *MisTree* permite construir rápidamente el MST creando inicialmente un grafo de k-vecinos más cercanos (kNN, un grafo que vincula cada punto con los k-vecinos más cercanos, utilizando la función `kneighbours_graph` de scikit-learn) que mejora la velocidad al limitar el número de aristas consideradas en la construcción, pasando de n^2 a kn (donde $k \ll n$). A continuación implementa el algoritmo de Kruskal ([Kruskal, 1956](#)) (mediante la función `minimum_spanning_tree` de scipy ¹) obteniendo como resultado final un árbol minimal de expansión. La figura 3.2 extraída del trabajo de [Naidoo \(2019\)](#) muestra un ejemplo claro de este procedimiento. A partir de una distribución de puntos (panel izquierdo) se construye un grafo kNN que vincula todos los puntos con sus k-vecinos más cercanos, 20 para este trabajo, (panel central) y luego ejecuta el algoritmo de Kruskal para construir el MST (panel derecho).

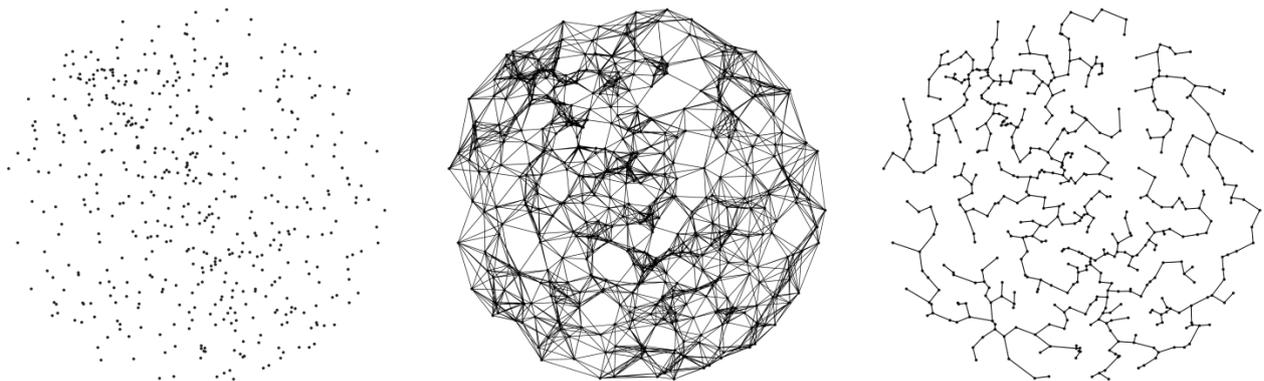


Figura 3.2: A partir de una distribución de puntos (panel izquierdo) se construyen un gráfico kNN que vincula todos los puntos con sus k-vecinos más cercanos (panel central) y luego el paquete *MisTree* ejecuta el algoritmo de Kruskal para construir el MST (panel derecho). Esta figura fue extraída del trabajo de [Naidoo \(2019\)](#)

Además *MisTree* permite al usuario medir las estadísticas propuestas para el árbol como lo son el grado, el peso de las aristas, el peso de las ramas y un parámetro de forma del árbol. Si bien el paquete fue diseñado para ser utilizado en cosmología, podría utilizarse en cualquier campo que requiera extraer información no gaussiana de las distribuciones de puntos. Esta es

¹Véase <https://scipy.org/>

la herramienta que se utilizará para este trabajo. Los pesos que se emplearán para las aristas son las distancias entre los puntos; es decir, la distancia euclidiana para las coordenadas de 2/3 dimensiones, polares esféricas, o las distancias angulares.

Parámetros

Los parámetros que se pueden extraer de un árbol son variados, pero los estándar se describen en [Naidoo \(2019\)](#). Estos parámetros son:

- Grado (d) - el número de aristas que tiene cada nodo. (ver Fig. 3.3)
- Longitud de las aristas (l) - la longitud de las aristas en el MST. (ver Fig. 3.4)
- Longitud de rama (b) - la suma de las longitudes de las aristas miembros. (ver Fig. 3.5)
- Forma de la rama (s) - la distancia en línea recta entre las puntas de las ramas dividida por la longitud de la rama. (ver Fig. 3.6)

Este último estadístico no se analizará a lo largo del trabajo ya que su interpretación tiene sentido solo al cortar el árbol en subgrupos e identificar características geométricas de los mismos. Como el MST no se pretende emplear como un método de identificación de estructuras sólo se emplearán los estadísticos d , l y b de aquí en adelante.

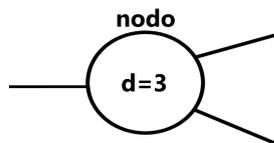


Figura 3.3: Parámetro d . El grado, el número de aristas que conectan un nodo.

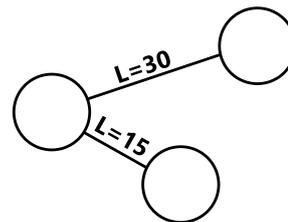


Figura 3.4: Parámetro l . El peso de la arista o longitud de la misma.

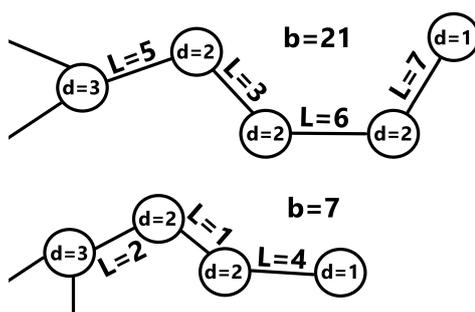


Figura 3.5: Parámetro b . La rama, la suma de los pesos de las aristas desde un nodo de grado 3 a uno de grado 1.

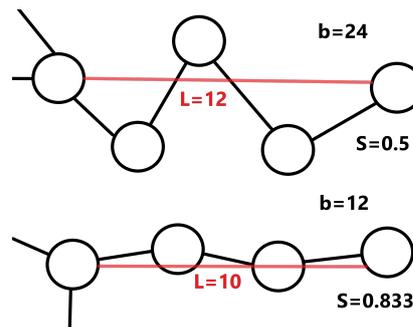


Figura 3.6: Parámetro s . Que tanto dista una rama en ser lineal.

En la sección 3.0.2 se mencionaron aplicaciones del algoritmo MST para la determinación de mediciones astronómicas. Sin embargo se hace una especial mención al trabajo de [Naidoo et al. \(2020\)](#) debido a la utilidad de sus resultados para el presente estudio. En el se analiza la sensibilidad de los estadísticos antes mencionados para restringir los parámetros cosmológicos. Con este fin, el autor emplea las simulaciones Illustris ([Nelson et al., 2015](#)) y confecciona una muestra de control que sea muy similar a la simulación de tal manera que no se encuentren

diferencias al comparar las funciones de correlación de a pares. Como las distribuciones son ligeramente distintas, se requiere alguna herramienta estadística que puedan extraer información en ordenes más altos que considerar la distancia de a pares. Empleando la técnica del MST a ambas muestras, logra encontrar diferencias en las propiedades de los conjuntos de puntos. Se concluye entonces que los estadísticos d , l , b y s son especialmente sensibles para detectar variaciones de alto orden en la geometría de la red cósmica. Posteriormente realiza estudios empleando muestras de control afectadas por distintos sesgos tales como una geometría particular (mediante el uso de máscaras angulares), efectos en el espacio de redshift o cambios en la densidad media del Universo. Mediante estas comparaciones se cuantifica la manera en que los estadísticos se ven afectados por cada uno de estos sesgos.

Por último, emplea todos los conocimientos obtenidos en estos análisis para estudiar las posibles nuevas restricciones de los parámetros cosmológicos. Como conclusión, entre otros resultados, logra mejorar la restricción en 1σ para Ω_m respecto lo obtenido a partir de la función de correlación de dos puntos.

Capítulo 4

Conjunto de Datos

4.1. Conjunto de Datos

4.1.1. Catálogo GMBCG

El catálogo con el que se realizó el presente trabajo consta de una muestra de galaxias centrales de cúmulos (*GMBCG Cluster Catalog*) [Hao et al. \(2010\)](#), la cual abarca unos 8240 grados cuadrados del relevamiento SDSS-DR7. Este dispone de mediciones de corrimiento al rojo estimados fotométricamente para todos los datos $0,1 \leq z \leq 0,55$ y espectroscópicamente para una submuestra. Además, se dispone de las coordenadas ecuatoriales (α, δ) junto con otros tres parámetros: la riqueza del cúmulo, las magnitudes aparentes petrosianas y las magnitudes aparentes model.

Técnica de identificación

Para la confección de este catálogo [Hao et al. \(2010\)](#) utilizó la identificación de la secuencia roja de un cúmulo tal como se explico en la sección 2.0.4. Dicha secuencia se encuentra al realizar un diagrama color-magnitud sobre el centro de un cúmulo rico poblado por muchas galaxias elípticas. Al colocar cada uno de estos objetos en el diagrama se obtiene una pendiente de ajuste. Además, si la estructura posee una BCG no hay duda que lo observado pertenece a un cúmulo. El problema que tiene esta técnica para la detección de estructuras es que solo identifica cúmulos ricos con centros poblados de galaxias tempranas con una BCG, dada cualquier otra estructura menos rica o con más asimetrías, no sería detectada.

Por otro lado, la distribución de los colores de las galaxias se puede aproximar adecuadamente mediante una mixtura de dos distribuciones gaussianas ([Hao et al., 2009](#)). Una población es más roja y estrecha (secuencia roja), mientras que la otra es más azul y ancha (nube azul). En un cúmulo, esta última incluye galaxias tanto en primer y segundo plano junto con sus miembros.

En este marco, los algoritmos de mezcla de gaussianas son fundamentales para ajustar la distribución de color de estas observaciones. La modelización de los errores debe ser satisfactoria ya que no es posible despreciarlos. El algoritmo cuenta con varios pasos sucesivos. Primero se emplean métodos paramétricos o no paramétricos para analizar la intensidad de la señal de agrupación. Luego se toma un rango de corrimiento al rojo, se identifica la BCG y se adopta el redshift fotométrico de la BCG como un buen estimador de la distancia del cúmulo. Al tomar un gran rango en z , existen tres complicaciones:

- i) El corrimiento de color de la secuencia roja a lo largo del rango de redshift
- ii) La superposición de cúmulos en la línea de la visual
- iii) Complejidad de realizar una definición de una medida consistente de la riqueza a través

de más de un color.

Para solucionar el problema los autores proponen las siguientes soluciones. i) como el rango de redshift se extiende hasta $z_{foto} \leq 0,55$ y la secuencia roja hasta $z_{foto} \leq 0,43$, se puede identificar utilizando el g-r y entre $z \in [0,43 - 0,55]$ se emplea el color r-i. Un cuestionamiento válido es pensar que ocurre con los cúmulos cerca de $z_{foto} \approx 0,43$. Cerca de la transición de los filtros, el patrón de secuencia BCG más rojo es evidente en ambos colores adyacentes. En el caso de ii) se toma la precaución de adoptar un corte en luminosidad a partir de una ventana de $z_{foto} \pm 0,25$. Para escoger el corte para las potenciales galaxias miembros se emplea un criterio según su brillo; se pide que sean más brillantes que $0,4L^*$, donde L^* es la luminosidad característica correspondiente a la función de luminosidad de Schechter (Schechter, 1976). Mientras que para iii) como el efecto para la confección del catálogo es pequeño solo se ajustan las definiciones de riqueza para dar como resultado una transición suave entre filtros.

Como las detecciones se centran alrededor de la BCG, aumenta la eficiencia del algoritmo ya que omite las búsquedas alrededor de galaxias intrínsecamente débiles. Para asegurar que la BCG pertenece a la secuencia roja identificada y que no se trate de una superposición a lo largo de la visual, se toma como criterio que el color de la misma se encuentre dentro de 3 desviaciones estándar del pico de la Gaussiana de secuencia roja identificada. Luego conociendo los posibles miembros y el objeto central (BCG) se calcula la fuerza del agrupamiento espacial en el plano ascensión recta y declinación mediante una convolución de los miembros seleccionados con un perfil NFW¹ proyectado (Véase Bartelmann (1996); Navarro et al. (1997); Koester et al. (2007b)) empleando un kernel radial.

La fuerza de agrupamiento que se menciona se define como

$$S_{cluster} = \sum_{k=1}^{N_g} \frac{2\rho_s r_s}{x_k^2 - 1} f(x_k) \quad (4.1)$$

donde N_g es el número total de galaxias miembros y $r_s = r_{200}/c^2$ es el radio de escala, ρ_s es la densidad crítica proyectada, $x = r/r_s$, y

$$f(x) = \begin{cases} 1 - \frac{2}{\sqrt{1-x^2}} \tanh^{-1} \sqrt{\frac{1-x}{x+1}} & \text{si } x < 1 \\ 0 & \text{si } 1 = x \\ 1 - \frac{2}{\sqrt{x^2-1}} \tan^{-1} \sqrt{\frac{x-1}{x+1}} & \text{si } 1 < x \\ 0 & \text{si } 20 < x \end{cases} \quad (4.2)$$

Similar a Koester et al. (2007b), se emplea el valor de $r_s = 150$ kpc, independientemente de la riqueza del cúmulo. El parámetro de fuerza de agrupamiento $S_{cluster}$ es esencialmente la altura del pico del campo de densidad de secuencia roja suavizada en la posición de la BCG como se menciona en Hao et al. (2010).

El procedimiento consta de tres pasos:

1. Para cada galaxia del catálogo, se calcula la fuerza de agrupamiento $S_{cluster}$ dentro de una apertura de búsqueda de 0,5 Mpc. Utilizando las galaxias más débiles que el candidato BCG y perteneciente a la secuencia roja identificada.
2. Mediante un procedimiento de percolación, se clasifican las galaxias candidatas a BCG por su fuerza de agrupamiento y se eliminan las que se identifican como "miembros" de otro cúmulo con mayor fuerza de agrupamiento.
3. Repetir el proceso anterior y obtener una lista de BCG y sus miembros del grupo. Luego a partir de la riqueza medida en 0,5 Mpc, se calcula un escalado R_{scaled} para cada

¹El perfil NFW es un modelo del perfil de densidad para un halo de materia oscura que suele ajustar en la distribución de masas en las simulaciones

² r_{200} es el radio en el cual la densidad de materia es 200 veces la densidad crítica del Universo.

aglomeración. Posteriormente se repiten los procesos 1) y 2) cambiando la apertura de búsqueda a R_{scaled} desde 0,5 Mpc. Así se completa la lista final de miembros de cúmulos con riqueza escalada N_{gals}^{scale} .

4.1.2. Máscara Tridimensional

Una máscara angular es una sección delimitada en el cielo que cumple la función de especificar que no se trabaja con toda la esfera celeste si no con un área en particular. Se suele emplear cuando un relevamiento posee datos medidos en una región en particular del cielo. Para este trabajo se requiere construirla ya que se pretende emplear datos con posiciones aleatorias en el cielo como muestra de control pero en el mismo volumen y con la misma geometría angular que poseen los datos. Esto da lugar a una máscara tridimensional la cual delimita la zona angular de la esfera celeste y el rango de redshift en la dirección radial. Un punto importante a destacar es que al utilizar una máscara para el análisis de los resultados de la muestra aleatoria se evitan posibles sesgos debido a la geometría del catálogo en los resultados.

La herramienta (MST, capítulo 3) que se emplea en este trabajo requiere un área conexa. En la figura 4.1 se visualiza distribución de galaxias del catálogo GMBCG en coordenadas ecuatoriales (α, δ) , donde en azul esta representada cada galaxia de la muestra y puede verse que existen zonas angulares disconexas con forma de tiras alargadas o de una pequeña isla cerca de la zona central. Se escoge conservar la zona angular con mayor cantidad de datos y se corta la máscara tal como se muestra en línea de trazos roja en las figuras 4.1 y posteriormente 4.2. A partir de este preprocesamiento se obtiene un área adecuada para la posterior aplicación del algoritmo de identificación MST tal como se puede ver en la figura 4.3. En ella se presenta un mapa angular sobre la esfera y sus respectivas coordenadas angulares α y δ . La máscara angular se muestra en verde, y sobre ella se visualiza la distribución de galaxias de la muestra (puntos azules).

Para generar la máscara se hace uso de la herramienta Helpix (Górski et al., 2005) que es un acrónimo de *Hierarchical Equal Area isoLatitude Pixelization of a sphere*. Este algoritmo permite reconstruir la geometría angular del catálogo, recreando la misma a partir de píxeles de áreas iguales entre sí proyectados en la esfera. El tamaño del pixel es configurable así como también la cantidad utilizada a través de los parámetros N_{side} y N_{pix} respectivamente. El primero de ellos debe ser potencia de 2 (2^n con $1 \leq n \leq 8192$), mientras que el segundo es la cantidad de píxeles que se empleará. Como se mencionó anteriormente la idea es tener una superficie conexa. Para esto se pretende mapear el área del cielo que contenga los datos del catálogo en las zonas limitadas por los cortes (Véase figura 4.3). Con este fin se elijen píxeles grandes ($N_{side} = 32$) evitando huecos en la máscara. Fijado el N_{side} , el N_{pix} queda entonces $N_{pix} = 12 \cdot N_{side}^2$. A continuación se procede con el siguiente algoritmo:

1. Se confecciona la máscara angular. Para ello se mapea la posición angular de cada galaxia del catálogo asignándole un píxel, creando así un arreglo de píxeles de tamaño $N_{side} = 32$ con dos posibles valores, 0 si ningún valor cae dentro del píxel o 1 si una o más galaxias caen dentro.
2. Se fijan los límites tridimensionales. Tomando las distancias mínimas y máximas en la dirección radial (calculado a partir de la distancia de luminosidad; expresión 1.20).

La máscara angular visualizada en la figura 4.3 muestra un arreglo de píxeles que junto con las restricciones en redshift definen un volumen análogo al cubierto por el conjunto de datos formando un área conexa que será empleada en la confección de una muestra de control aleatoria.

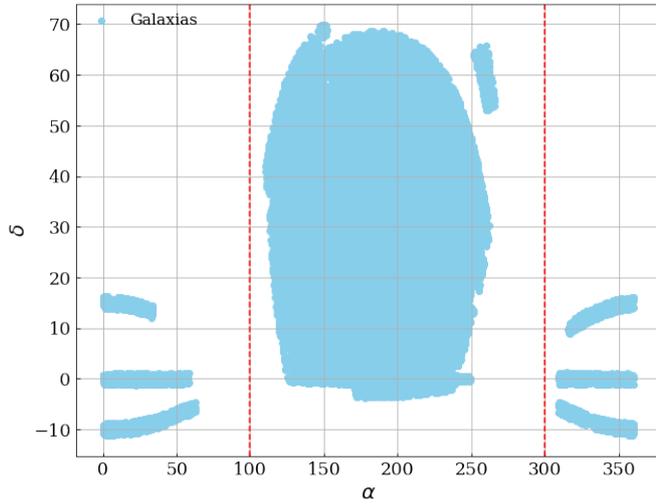


Figura 4.1: Distribución de galaxias en la sección angular del cielo. En puntos azules se presentan las galaxias de la muestra de datos. Pueden observarse un conjunto de tres áreas similares a tiras alargadas en ambas direcciones respecto del centro. La línea punteada en color rojo representa los cortes propuestos que se realizan para lograr obtener un área lo más conexa posible.

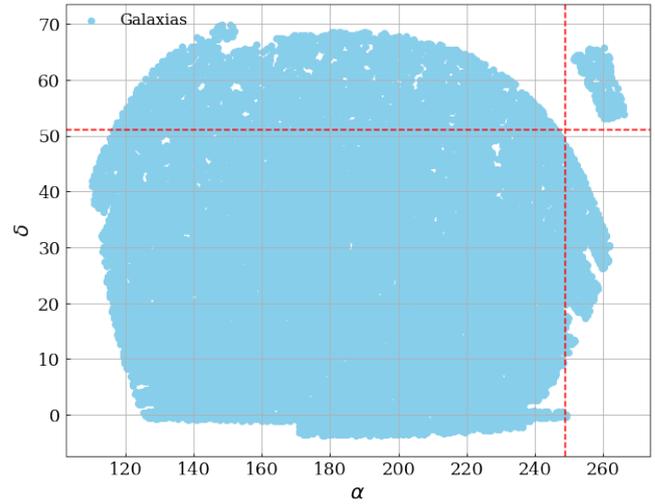


Figura 4.2: Distribución de galaxias en la zona angular del cielo luego de la primera restricción. En puntos azules se visualizan las galaxias de la muestra. En la esquina superior derecha existe un área disconexa de la mayor cantidad de datos. El área superior derecha exhibe las restricciones propuestas para lograr obtener una sección lo más conexa posible con la mayor cantidad de objetos.

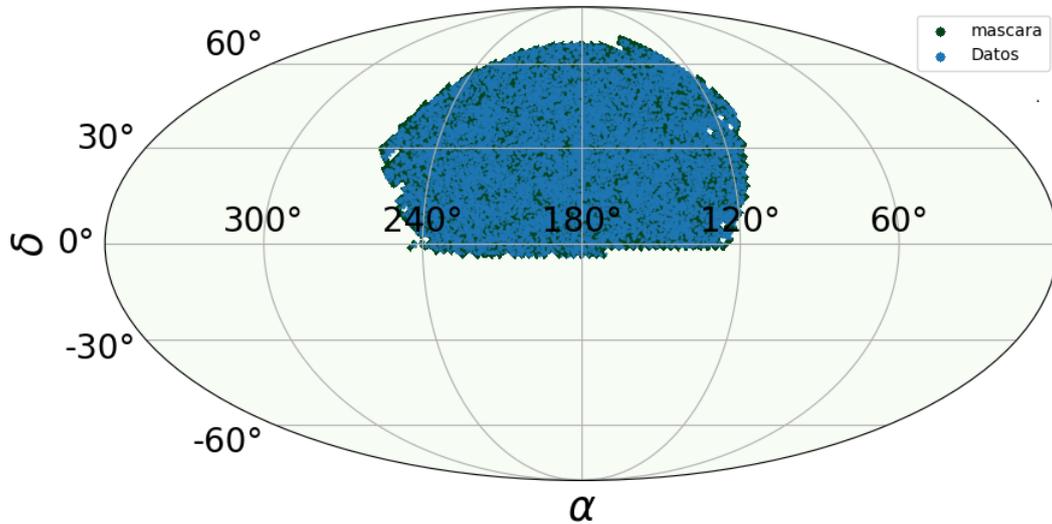


Figura 4.3: Máscara y Distribución de galaxias luego de la aplicación de las restricciones. Se presenta un mapa angular sobre la esfera con sus respectivas coordenadas angulares α y δ . La máscara angular se presenta con un área en verde sobre la cual se visualiza la distribución de galaxias de la muestra (puntos azules).

4.1.3. Análisis preliminar

Antes de comenzar a utilizar el catálogo, siempre es útil analizar los datos disponibles. Como se mencionó anteriormente, el catálogo GMBCG posee un rango de redshift $0,1 \leq z \leq 0,55$, α , δ , magnitudes model aparentes de las 5 bandas de SDSS (sección 2.0.2), colores g-r y r-i, riqueza (sección 2.0.4), riqueza ponderada (sección 2.0.4), la fuerza de clustering ($S_{cluster}$, sección 4.1.1) y un parámetro que indica cuando la riqueza ponderada es recomendada en vez de la riqueza.

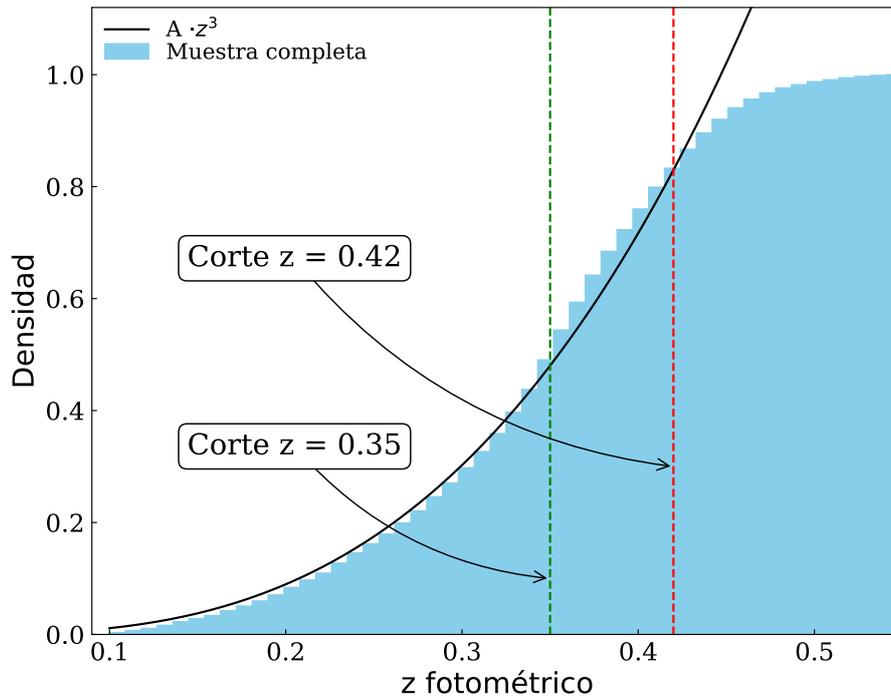


Figura 4.4: Distribución acumulada de z_{foto} para todas las galaxias del catálogo. Se muestra un ajuste de $A \cdot z^3$ (línea negra) el cual es utilizado para definir muestras con volumen completo a bajo redshift (se lo extiende hasta 0.4). Se realizan dos cortes en $z=0.35$ (muestra de referencia, línea verde) y en $z=0.42$ (muestra principal, línea roja). La muestra de referencia se empleará para comparar la incerteza entre utilizar los redshifts fotométricos o los espectroscópicos, mientras que la muestra principal será la empleada para realizar el presente trabajo.

Redshift fotométrico

La distribución acumulada de *redshift* fotométrico que se observa en la figura 4.4 permite ajustar una curva $A \cdot z^3$ (línea negra) para estimar las restricciones necesarias que deben tomarse para obtener una muestra con volumen completo. Este ajuste es una aproximación adecuada para bajo redshift (aunque se extiende). En función de esto, se decide tomar dos intervalos, uno en $0,1 \leq z_{foto} \leq 0,35$ (muestra de referencia en adelante) y otro en $0,1 \leq z_{foto} \leq 0,42$ (muestra principal). La primera muestra se toma con el objetivo de estudiar la relación entre el redshift fotométrico y espectroscópico en un intervalo en donde la comparación resulte adecuada por la cantidad de datos espectroscópicos disponibles. Esta se realiza con la intención de estimar el error introducido al usar los redshifts fotométricos en lugar de los redshifts espectroscópicos (z_{spect}). Usualmente se prefiere el z_{spect} sobre el z_{foto} ya que su incerteza es menor, pero al querer analizar datos a distancias más grandes se necesita disponer de mediciones para todos los objetos. Esta última condición es cumplida por los z_{foto} . Posteriormente se emplea la muestra principal para realizar el análisis de agrupamientos.

Redshift espectroscópico

La determinación de la distancia es más precisa cuando se emplean los z_{spect} , pero al no disponer de mediciones para todos los objetos se complica su utilización. Es por ello que se analizó la proporción de objetos con medidas espectroscópicas en los distintos rangos de z_{foto} . En la figura 4.5 se presentan intervalos de z_{foto} sobre los que se calculó la fracción de valores medidos de z_{spect} que se disponían. Puede verse que para todos los bins para $z_{foto} > 0,22$, los porcentajes de mediciones disponibles son muy bajos (menores al 50%). Este resultado no permite generar una muestra espectroscópica para realizar un estudio estadístico como el que se pretende.

A continuación se busca estimar la diferencia entre los z_{foto} y z_{spect} . Analizando galaxias que cuenten con ambos valores medidos, se espera cuantificar en qué grado las mediciones fotométricas son consistentes con las espectroscópicas. Con este fin, en la figura 4.6 se expone la diferencia $z_{foto} - z_{spect}$ en función del z_{spect} empleando todos los datos de la muestra principal con ambas estimaciones de redshift. Se separa el comportamiento en los cuatro cuartiles de la distribución de z_{spect} , donde cada uno tiene estimada su media. Los rectángulos poseen un ancho $2 \cdot \sigma$ y están centrados en los valores medios de la diferencia. La media de la resta entre ambos parámetros no supera los 0,01 en ninguno de los casos. Sin embargo, al examinar la dispersión, el primer, segundo y tercer intervalo poseen desviaciones que son del orden de 0,02. Como estas 3 secciones abarcan el rango $z_{spect} \in [0,1 - 0,25]$ se concluye que para cada z_{spect} , el valor de z_{foto} representa una buena aproximación de esta cantidad en el rango mencionado. En el caso del último rectángulo, se puede ver que tiene una dispersión mayor que en los intervalos anteriores. Esto se corresponde a la gran incerteza en las mediciones a *redshift* intermedio.

Se determina emplear a lo largo del trabajo el redshift fotométrico en lugar del espectroscópico. Debido a la baja dispersión que se presenta en la figura 4.6 y que el z_{foto} de las BCG suelen estar entre ($z_{foto} - z_{real} \sim 0,015$ para SDSS DR7) (Hao et al., 2010) se realiza la siguiente estimación para cuantificar el apartamiento en distancia

$$z \cdot c = D \cdot H \quad (4.3)$$

Tomando $h = H/100$

$$\frac{z \cdot c}{100} = D \cdot h$$

y $c = 3 \times 10^5 \frac{Km}{s}$ y $z = 0,015$ se obtiene

$$45 \frac{Mpc}{h} = D$$

se concluye que la distancia si estará afectada al emplear el redshift fotométrico en vez del espectroscópico. Sin embargo se podrán obtener resultados cualitativos en el presente trabajo.

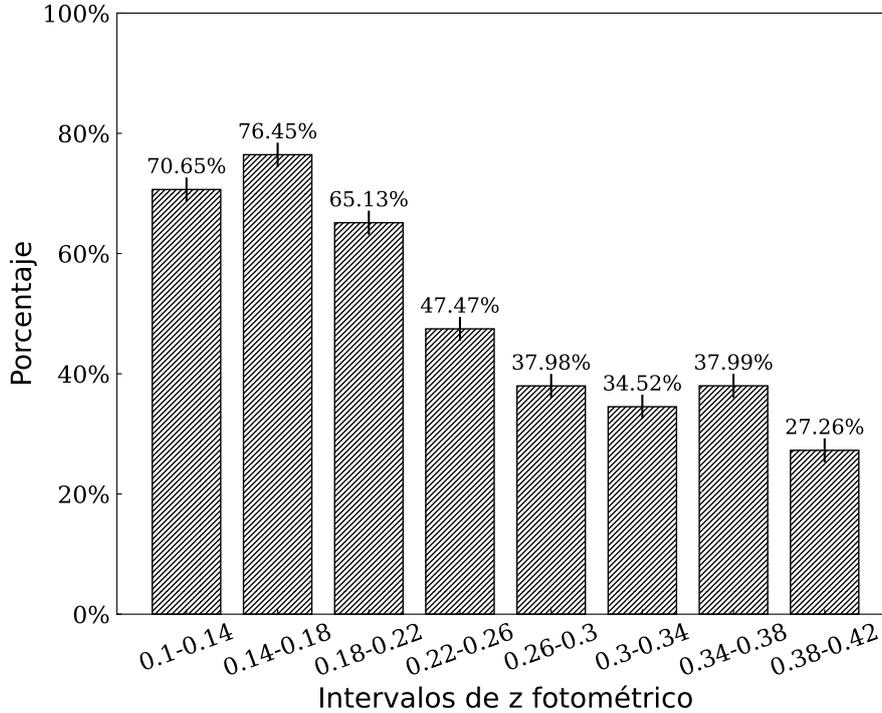


Figura 4.5: Fracción de datos disponibles de z_{spect} para cada intervalo de z_{foto} . Puede verse que en los primeros bins ($z_{foto} < 0,22$) hay un porcentaje de datos mayor al 60% mientras que para todos los intervalos con $z_{foto} > 0,22$ no alcanzan el 50%.

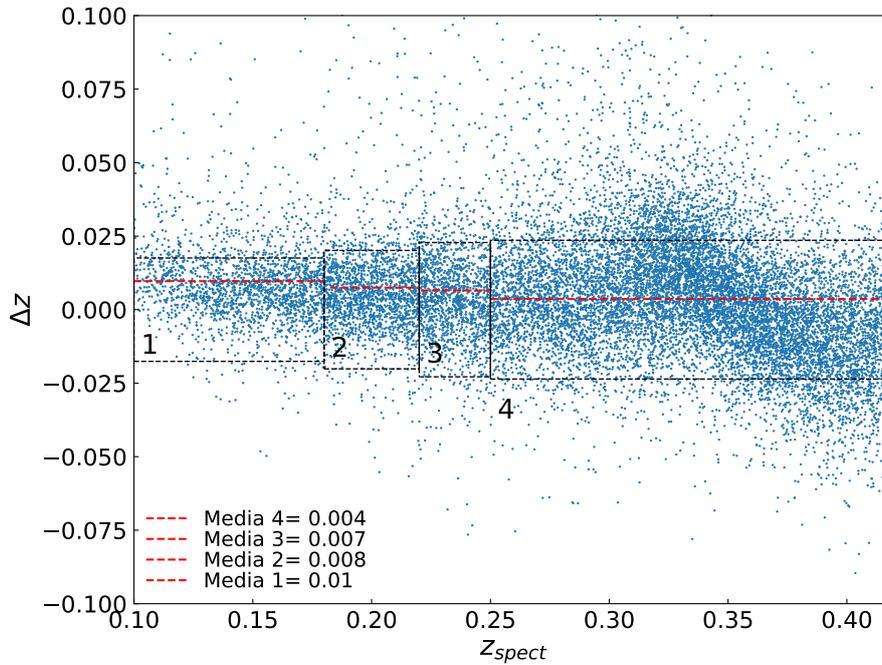


Figura 4.6: Diferencia entre z_{spect} y z_{foto} denotado como Δz . En esta figura se emplean todos los datos de la muestra principal con ambas estimaciones de redshift. Cada rectángulo corresponde a los intervalos intercuartílicos de la distribución de z_{spect} con una estimación de la media (denotada en recta roja) y de la dispersión (el ancho de los rectángulos es de $2 \cdot \sigma$). Se observa que para todos los casos, la media1 no supera los 0,01 y la dispersión para el primer, segundo y tercer cuadrilátero es del orden de 0,02. En el caso del último rectángulo, se puede ver una dispersión mayor que en los intervalos anteriores.

Color g-r

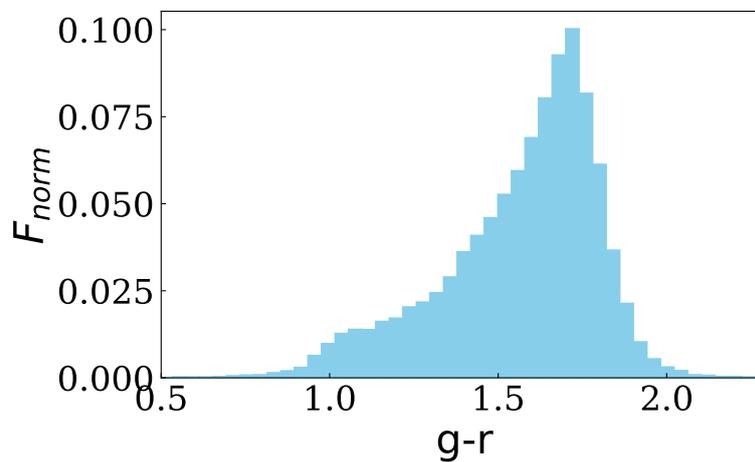


Figura 4.7: Frecuencia normalizada de la distribución del parámetro $g-r$ para la muestra principal. Al analizar la figura pueden identificarse dos comportamientos distintos con máximos en $g-r \approx 1,12$ y $g-r \approx 1,72$.

Al analizar la distribución del parámetro $g-r$ observado de la muestra principal se distingue una bimodalidad (figura 4.7). Al visualizar la figura pueden identificarse dos comportamientos distintos con máximos en $g-r \approx 1,12$ y $g-r \approx 1,74$. Tal como se describió en la sección 2.0.4, las galaxias esperadas en la muestra principal poseen colores rojos. Debido a lo observado, se propone descartar aquellos objetos más azules.

Para ello se realiza una regresión de dos gaussianas que ajusten este comportamiento bimodal en $g-r$. Se obtiene la figura 4.8 donde la suma de las dos poblaciones correlaciona con

la distribución del parámetro. Además se logra identificar dos distribuciones de colores, una centrada en $g - r = 1,48$ (más azul) y otra en $g - r = 1,71$ (más roja). La existencia de la cola azul en la figura 4.7 y 4.8 no es fácil de descontar de la muestra. Se propone entonces dividir la distribución empleando los cuartiles para detectar si el comportamiento bimodal se extiende a toda la muestra o solo para un cierto rango de redshift. A cada intervalo se le intenta ajustar dos gaussianas y en caso en la cual el ajuste no se logre realizar se optara por ajustar una sola.

En la figura 4.9 se visualizan los resultados obtenidos. Se presentan cuatro paneles con los ajustes realizados para cada intervalo intercuartilico en redshift fotométricos. En cada caso, se estiman los valores de la media y su desviación. Solo para el primer cuartil, se logra ajustar una distribución bimodal con componentes comparables. Para los otros tres casos se observa una única gaussiana con un desplazamiento del máximo hacia colores más rojos.

Estas observaciones son consistente con los resultados que se describen en Hao et al. (2010). En ese trabajo hay una sección haciendo alusión a este problema, y se recomienda emplear un corte en $z_{foto} = 0,16$.

Se estudia la restricción propuesta en la figura 4.10. En ella se comparan la distribución de $g-r$ para dos intervalos acotados en z_{foto} . En el panel de la izquierda se muestra el rango $0,16 \leq z_{foto} \leq 0,22$ con la restricción propuesto por el autor y en el de la derecha el intervalo $0,18 \leq z_{foto} \leq 0,22$ con el valor mínimo de z_{foto} propuesto en este trabajo. Para determinar que restricción utilizar, se emplea como criterio el mejor ajuste de una única gaussiana. Se espera que si la distribución solo tuviese galaxias rojas, la distribución se podría representar adecuadamente mediante una única gaussiana. Al realizar el ajuste, se escoge restringir a la muestra en $0,18 \leq z_{foto}$ por ajustar mejor el valor maximo de la distribución.

La distribución analizada es de $g-r$ y no la de $r-i$ debido a que la secuencia roja con la cual se identificaron los objetos en el catálogo se visualiza en este color hasta $z_{foto} \leq 0,43$ (Véase sección 4.1.1). Como la muestra principal abarca las galaxias hasta $z_{foto} \leq 0,42$ no se analizaran los otros colores.

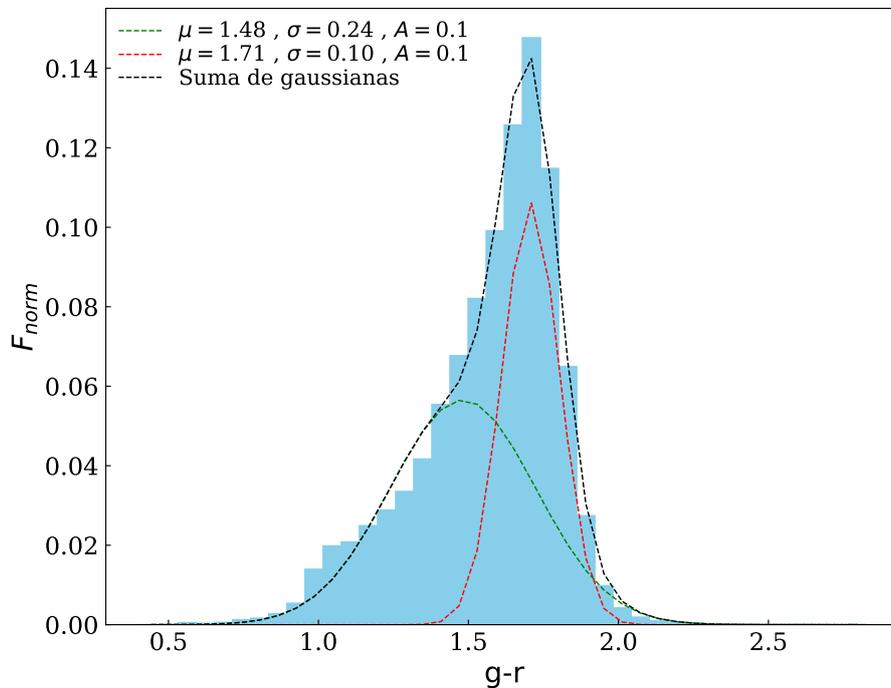


Figura 4.8: Distribución de $g-r$ con el ajuste de dos gaussianas. El comportamiento bimodal en $g-r$ se ajusta con la envolvente de las dos poblaciones obtenidas. La distribución de los objetos mas azules esta centrada en $g - r = 1,48$ mientras que $g - r = 1,71$ corresponde a objetos más roja.

Cálculo de magnitudes absolutas

A partir de las magnitudes model en la banda r (sección 2.0.2) y el redshift fotométrico, se estima la distancia de luminosidad fotométrica expresión 1.20 asumiendo la cosmología presentada en Planck Collaboration et al. (2020) y utilizando una corrección k con un *restframe* en $z = 0,3$ elegido en función del valor medio de la distribución de redshift.

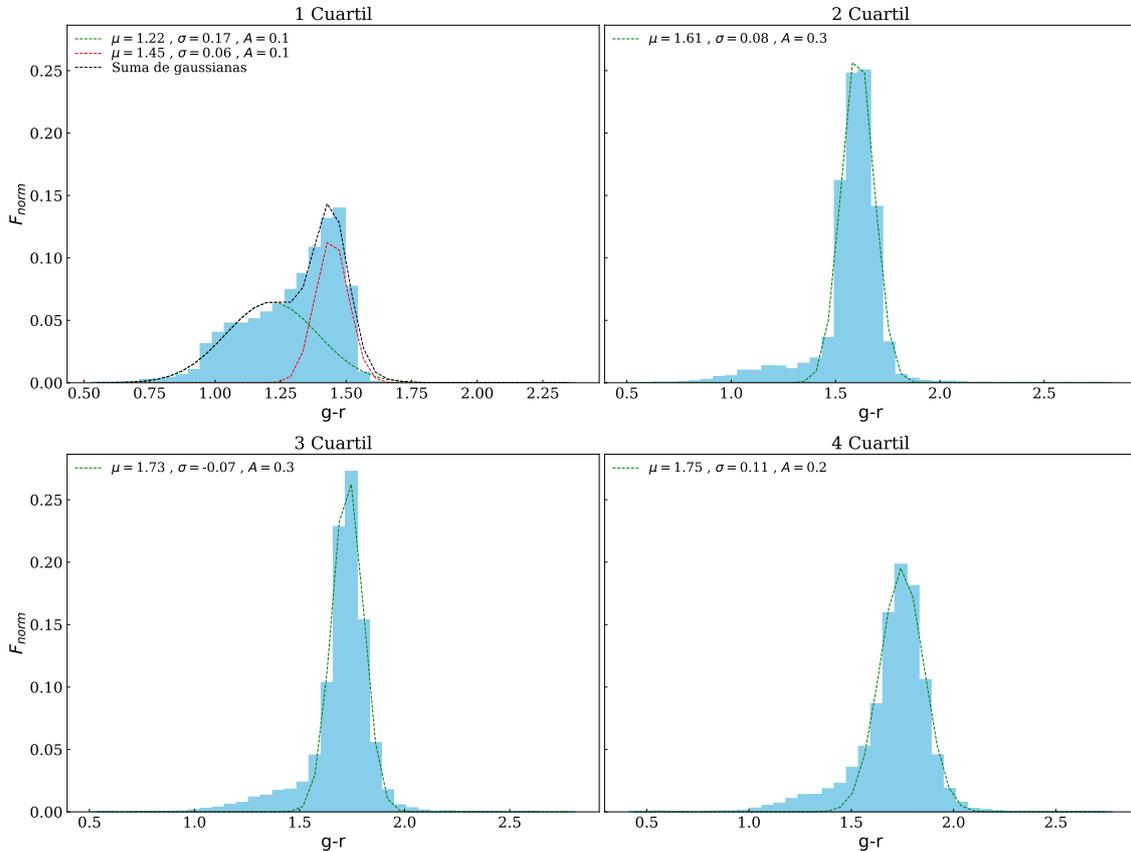


Figura 4.9: Intervalos intercuartílicos de redshift fotométrico para la distribución de g-r. Para cada panel se intenta ajustar dos gaussianas y en caso contrario se ajusta una sola componente. Solo en el primero intervalo se presenta una bimodalidad, mientras que para los otros tres casos se observa una única gaussiana con un desplazamiento del máximo hacia colores más rojos.

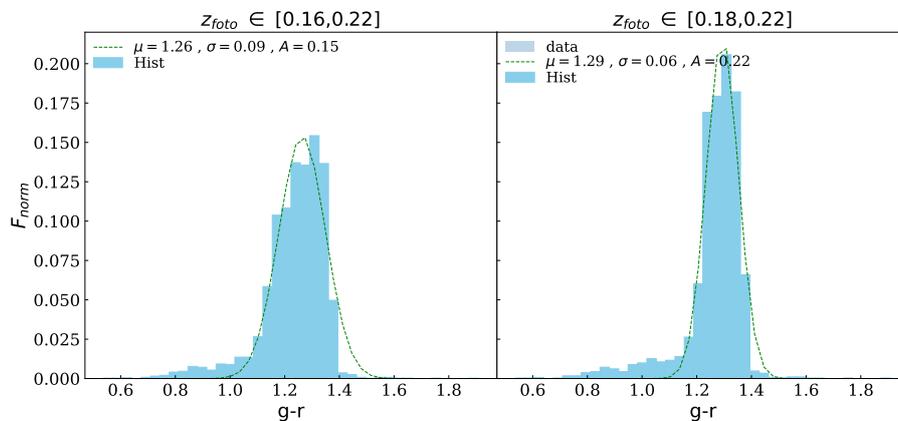


Figura 4.10: Comparación entre las distribuciones de g-r para dos intervalos acotados en z_{foto} . En el panel de la izquierda se muestra el rango $0,16 \leq z_{foto} \leq 0,22$ con la restricción propuesta por el autor y en el de la derecha el intervalo $0,18 \leq z_{foto} \leq 0,22$ con el valor mínimo de z_{foto} propuesto en este trabajo. En ambos casos se ajusta una única componente gaussiana. Se observa un mejor ajuste a favor de la figura de la izquierda.

Riqueza y riqueza ponderada

La distribución de los parámetros de riqueza (o riqueza escalada) y riqueza ponderada se definen con precisión en la sección 2.0.4. La distribución de valores en la muestra principal se analizan en las figuras 4.11 y 4.12 respectivamente. En ambas figuras se descartan aquellos valores muy extremos ($Riq > 60$).

El valor máximo para la riqueza escalada esta en el primer bin (≈ 10) mientras que para la riqueza ponderada es más cercano a ≈ 13 . En cuanto a la forma de la distribución, en la figura 4.11 se visualiza una fuerte caída hacia altas riquezas. La riqueza ponderada por otra parte, tiene una distribución más simétrica respecto del valor máximo respetando la rápida disminución para valores grandes.

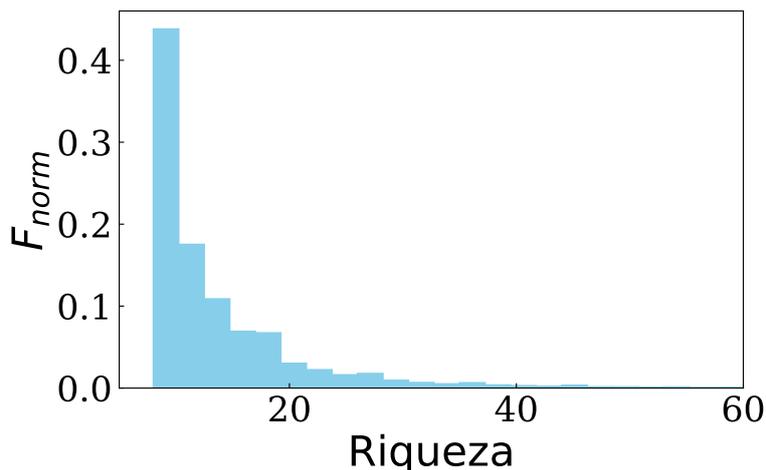


Figura 4.11: Distribución del parámetro de riqueza escalada para la muestra principal de los datos. El valor máximo se encuentra en el primer bin (≈ 10) y se destaca una fuerte caída hacia altas riquezas.

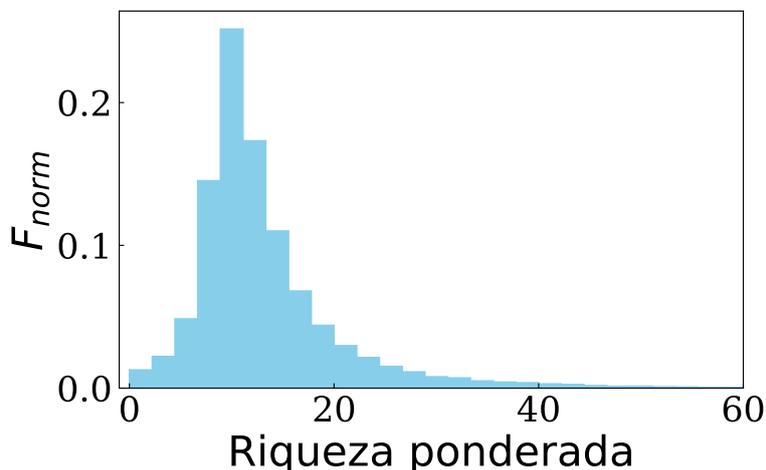


Figura 4.12: Distribución del parámetro de riqueza ponderada para la muestra principal. El valor máximo se encuentra cercano a ≈ 13 y presenta una distribución casi simétrica respecto del valor máximo. Posee un exceso de valores grandes que decrecen para valores más altos.

4.1.4. Correlación con parámetros del SDSS-DR16

Se busca obtener propiedades adicionales de los objetos de la muestra. Para ello se correlacionaron los datos del catálogo GMBCG con la información del DR16 de SDSS empleando el identificador único de cada objeto (ObjID). Se buscó obtener las dispersión de velocidades,

los ángulos de posición y los cocientes de semiejes para cada una de las galaxias del catálogo GMBCG. La idea detrás de expandir el espacio de parámetros es poder obtener algunas características para luego poder construir distintos tipos de árboles en función de las propiedades de los nodos.

Ángulos de posición

Los ángulos de posición (definidos en la sección 2.0.3) pueden ser utilizados para analizar la orientación de los objetos respecto de la estructura en gran escala. La distribución de valores de los ángulos de posición para las BCG, se presenta en la figura 4.13. En la misma se observa un rango de valores entre 0° y 180° . Cabe señalar que la forma de la distribución presenta una homogeneidad en todo el rango de valores con un pequeño aumento en el intervalo $[60^\circ, 140^\circ]$.

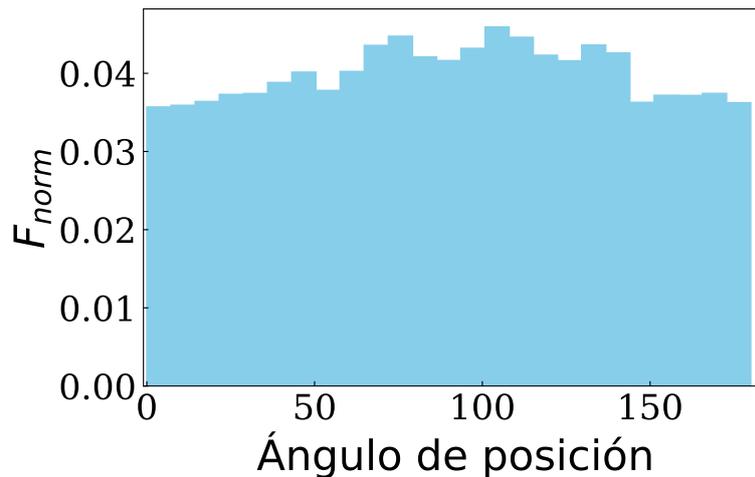


Figura 4.13: Distribución del parámetro ángulo de posición para la muestra principal. Se observa una distribución casi uniforme con valores en el rango $[0^\circ, 180^\circ]$, con un leve aumento en el intervalo $[60^\circ, 140^\circ]$.

Dispersión de velocidades

La distribución de dispersión de velocidades (Véase sección 2.0.4) permite obtener una estima de la masa de la BCG. En la figura 4.14 se presenta la distribución de dicho parámetro en la cual puede observarse un valor máximo alrededor de $250 \frac{km}{s}$. Al igual que lo mencionado para la riqueza, los valores altos más allá de $600 \frac{km}{s}$ no se consideran por el argumento de ser valores muy extremos de la distribución.

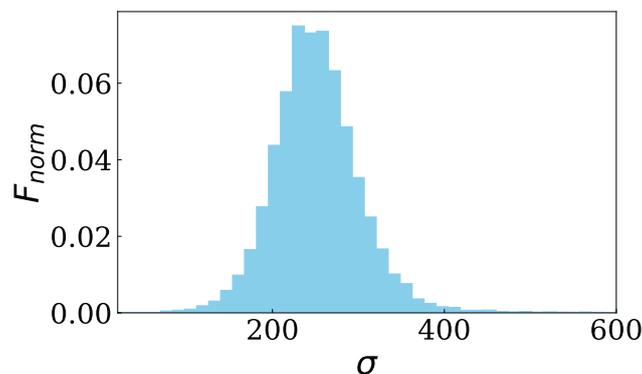


Figura 4.14: Distribución del parámetro dispersión de velocidades para la muestra principal. En la misma se observa un valor máximo alrededor de $250 \frac{km}{s}$.

Cociente de semiejes

Por último, el cociente de semiejes corresponde a la razón entre el semieje menor proyectado (b_p) y el semieje mayor proyectado (a_p) de la galaxia ($\frac{b_p}{a_p}$) en el filtro r. Para calcular este parámetro se emplea el uso de ajustes por isofotas similar al utilizado para calcular los ángulos de posición (Véase sección 2.0.3).

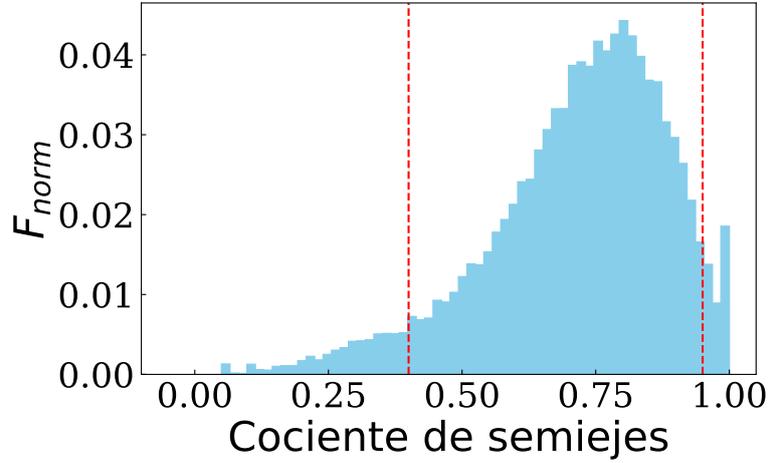


Figura 4.15: Distribución del parámetro cociente de semiejes para la muestra principal de datos. La forma de la distribución posee su punto máximo cercano a $\frac{b_p}{a_p} \approx 0,77$ donde b_p corresponde al semieje menor proyectado y a_p al semieje mayor proyectado. Se observa un pico elevado en 1 y una cola de valores cercanas a 0.1. Se restringe el rango al interior del recuadro limitado por las líneas discontinuas rojas.

La distribución de los cocientes de semiejes se visualiza en la figura 4.15. El intervalo observado de valores está comprendido entre $\frac{b}{a} \in [0, 1]$. En cuanto a la forma de la distribución posee su punto máximo cercano a $\frac{b_p}{a_p} \approx 0,77$. Se observa una frecuencia elevada en $\frac{b_p}{a_p} = 1$ y una cola de valores cercanas a $\frac{b_p}{a_p} = 0,1$. Se advierte que las cantidades mayores a 0.95 y menores a 0.4 son un problema. Para entender la razón detrás de este intervalo se plantean los siguientes argumentos:

i) Los valores mayores a 0,95 no son confiables. Debido a la esfericidad que poseen, y teniendo en consideración que si bien algunos de los cocientes realmente están bien medidos, en su mayoría podrían ser proyecciones visuales, se decidió no incluirlos en el análisis.

ii) En cuanto a la cota inferior en 0,4, al estar trabajando con BCGs el cociente tiene que ser alto, lo que se espera al tener morfologías elípticas. Por lo tanto se decide tomar como restricción a las galaxias con valores más bajos que $\frac{b_p}{a_p} \leq 0,4$.

Se concluye realizar ambas restricciones en este parámetro y mantener los valores intermedios por los criterios antes mencionados.

4.1.5. Muestra principal

Una vez realizadas las restricciones a los diferentes parámetros de las muestras, para la muestra de referencia resultan 13993 objetos, mientras que para la muestra principal 21434 de un total de 55423 del catálogo original. En la figura 4.16 se muestran las distribuciones finales para la muestra principal. El primer histograma correspondiente a la distribución de z_{foto} se encuentra restringido en el rango entre $[0,18 - 0,42]$.

En el caso de los colores, el índice g-r no muestra bimodalidad luego de la restricción, donde el máximo de la distribución se encuentra en $g - r = 1,74$ debido a una superposición de gaussianas con distintos valores medios dependiente del redshift de los objetos.

En el caso de la distribución de las riquezas, se observa una rápida disminución de la misma hacia cúmulos con alto número de miembros, mientras que la riqueza ponderada posee un máximo en ≈ 12 . En cuanto a los ángulos de posición, de acuerdo a lo esperado se mantiene una distribución homogénea.

La distribución de dispersión de velocidades tiene un máximo en $250 \frac{Km}{s}$ comprendida en el rango $(0, 600) \frac{Km}{s}$ consistente con la dispersión de velocidades de las estrellas en galaxias luminosas. Por último, la distribución de cocientes de semieje con máximo cercano a $\frac{b}{a} \approx 0,77$ en acuerdo con galaxias de tipo esferoidal.

En adelante solo se utilizará la muestra principal obtenida a partir del catálogo ya que maximiza la cantidad de objetos con las restricciones impuestas.

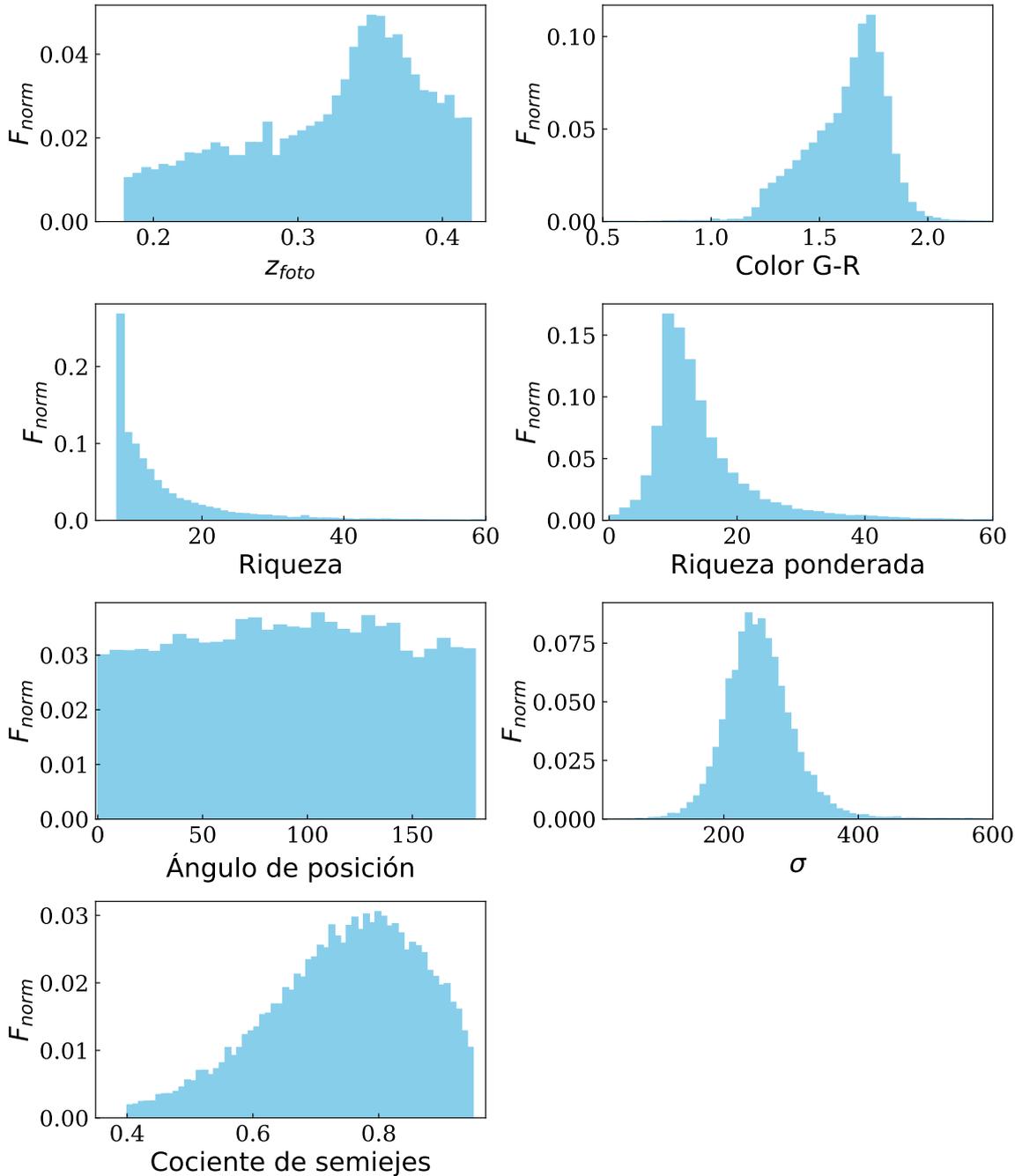


Figura 4.16: Distribuciones finales de los parámetros para la muestra principal. En ella se muestran 8 paneles correspondientes a las distribuciones explicadas en las secciones 4.1.3 y 4.1.4. Para el panel referido al z_{foto} se visualizan las restricciones colocadas sobre su rango en $[0,18 - 0,42]$. En el caso del color g-r, la distribución posee un máximo en los colores rojos $g - r = 1,74$ y ya no presenta una bimodalidad. Sin embargo debido a una superposición de gaussianas con distintos valores medios se visualiza una distribución asimétrica con una base amplia. La distribución de riquezas presenta una rápida disminución de la misma hacia cúmulos con alto numero de miembros, con un valor máximo establecido en el primer bin. Por su parte el histograma de riqueza ponderada posee un máximo en ≈ 13 y una distribución mas simétrica respecto del panel anterior. En cuanto a los ángulos de posición, de acuerdo a lo esperado se mantiene una distribución homogénea a lo largo de todo su rango. El panel referido a la dispersión de velocidades tiene un máximo en $250 \frac{Km}{s}$ y comprende un rango entre $[0, 600] \frac{Km}{s}$ consistente con la dispersión de velocidades de las estrellas en galaxias luminosas. Por último, la distribución de cocientes de semieje alcanza su valor más alto cercano a $\frac{b}{a} \approx 0,77$ en acuerdo con galaxias de tipo esferoidal.

Capítulo 5

Metodología

5.1. Metodología

Para realizar este trabajo, se busca comparar los resultados obtenidos para las BCGs con una muestra de control apropiada (muestra estocástica) la cual está afectada con los mismos sesgos que los datos. Al comparar los estadísticos de ambas muestras empleando cocientes se consigue eliminar el sesgo geométrico de la presencia de la máscara y observar efectos cualitativos en los resultados.

5.1.1. Objetivo

La teoría de grafos ofrece herramientas poderosas y de aplicación relativamente sencilla, que permiten analizar las formas de agrupar los nodos y así explorar un espacio definido a partir de un conjunto de puntos con distintas características. El objetivo general de este trabajo consiste en la aplicación de dicha teoría sobre la distribución de galaxias en gran escala. Se pretende confeccionar árboles minimales de expansión (MST) seleccionando los nodos (entendidos como galaxias que son centros de cúmulos) de acuerdo a sus propiedades astrofísicas, y luego contextualizar los resultados obtenidos a partir de diferentes elecciones de estos conjuntos. Para realizarlo, se estudia como varían los parámetros escogidos del MST a partir de modificar los nodos seleccionados al construir el árbol. Al modificar los valores de entrada, se obtendrán diferentes árboles. El objetivo final es entender el impacto de como la elección en las propiedades conllevan a predicciones físicas de la estructura en gran escala.

5.1.2. Fundamento

La fortaleza de los árboles minimales de expansión es justamente su habilidad para extraer patrones subyacentes en conjuntos de puntos dados, y es precisamente por esto que han sido usados para identificar rasgos particulares de la red cósmica de estructuras. Cada parámetro del catálogo puede entregar una cierta cantidad de información al explorarlo. En principio, si se pudiese calcular de forma eficiente la estadística de N puntos, se obtendría toda la información disponible (Szapudi & Szalay, 1998). El MST tiene en cuenta antes de realizar la conexión de un nodo con otro, un cierto entorno logrando capturar más información que las distancias entre pares la cual es empleada en el cálculo de la función de correlación como puede verse en la sección 3.0.1.

5.1.3. Técnica de muestreo aleatorio

A partir de los datos del catálogo se confecciona la máscara descrita en la sección 4.1.2 y se lleva a cabo una realización montecarlo con el mismo número de objetos de la muestra de acuerdo al siguiente algoritmo:

1. Empleando números pseudo-aleatorios, se sortea un α y un δ para la posición angular de un objeto y se evalúa si el objeto cae dentro de la máscara angular, de lo contrario se vuelve a sortear una nueva posición. Los números pseudo-aleatorios se escogen entre $-180^\circ \leq \alpha_{ran} \leq 180^\circ$ y $-1 \leq \cos \delta_{ran} \leq 1$ donde para obtener la distribución aleatoria de δ se calcula la inversa consiguiendo valores entre $-90^\circ \leq \delta_{ran} \leq 90^\circ$
2. Por otro lado, se elije al azar un valor de redshift, siempre respetando la distribución de z de la muestra principal. Para ello se hace uso del método de la función inversa. Primero se realiza la construcción de la distribución z acumulada de la muestra (ver figura 5.1). Luego se sortea un valor entre 0 y 1. A partir del mismo se traza una recta paralela al eje X (recta roja en la figura). Una vez que la recta intersecte la distribución acumulada, se hace una interpolación lineal entre el valor inmediatamente mayor y menor al cruce. Luego se calcula la función inversa de la recta obtenida y se evalúa en el valor aleatorio (línea negra de la figura) logrando así la distribución deseada.

Este procedimiento se realiza hasta obtener una distribución estocástica con el mismo número de objetos y misma distribución de z que la muestra de datos.

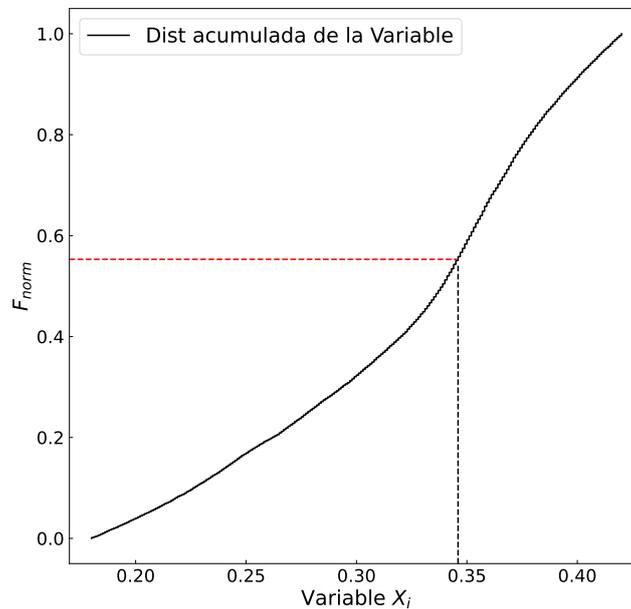


Figura 5.1: Método de la función inversa ejemplificado con una variable X_i (donde X_i representa el parámetro z_{foto} para cada submuestra). La distribución acumulada de la variable se visualiza como la curva negra, donde en el eje de las ordenadas se sortea un valor aleatorio (línea roja) y al producirse la intersección con la curva se obtiene el valor de la variable del eje X (línea negra).

5.1.4. Cociente de estadísticos

La distribución de los estadísticos del MST en este trabajo están sesgados por la presencia de la máscara tridimensional tal como se describe en [Naidoo et al. \(2020\)](#). Usando esta muestra como normalización, se puede desafectar los estadísticos de los sesgos geométricos al calcular el cociente entre las distribuciones de los valores de los parámetros en los intervalos correspondientes. A partir de las distribuciones de d , b y l para la muestra de los datos y la muestra de control se define la distribución de los cocientes ($C_{estadísticos}$). Su importancia radica en que permiten observar las diferencias intrínsecas entre las muestras y obtener propiedades respecto de una distribución homogénea. Si los resultados se comparan con los obtenidos para una muestra que cumpla homogeneidad en toda escala, se puede obtener el impacto de las

fluctuaciones sobre la estructura presente en los datos. Además, al usar una muestra aleatoria para la normalización, se pueden comparar entre sí los resultados provenientes de árboles con distinto número de nodos.

Para realizar la comparación, la distribución del estadístico de ambas muestras debe cumplir una serie de condiciones. En primer lugar debe poseer una cantidad de objetos mínima (mayor a cero) para que la señal pueda identificarse. Es relevante señalar que al utilizar el cociente, se intenta evitar cualquier división por el elemento nulo. La segunda condición es pedir una relación S/N mínima, ya que de lo contrario se obtendrán fluctuaciones aleatorias propias del ruido Poisson y no de las fluctuaciones cósmicas. Para ello se generan bins de distinto espesor para ambas distribuciones buscando el tamaño de bin necesario para que simultáneamente ambas muestras superen un cierto número de objetos.

En la figura 5.2 se ejemplifica lo detallado con anterioridad. Se visualiza una distribución de un estadístico x (donde x representa alguno de los tres estadísticos d , b o l) para la muestra de datos (muestra azul) y su respectiva muestra de control (muestra roja). Como se puede ver, los primeros bins no superen una frecuencia mínima (línea a trazos). Debido a esto, los primeros intervalos se suman en cada distribución hasta que la frecuencia en el bin de ancho variable para ambos casos supere el valor umbral. Realizando iterativamente este proceso para todo el rango del estadístico, se definen bins de ancho variable. Estos se emplean posteriormente para calcular el cociente entre los puntos rojos y azules obteniendo la distribución del cociente de un dado estadístico. Se puede ver que en el caso del último bin, la frecuencia no consigue superar la línea de corte, por lo tanto, el último recuadro usualmente no se considera en la evaluación de los resultados.

Para calcular el cociente se realiza $\frac{F_{datos}}{F_{randoms}}$ donde F_{datos} es la frecuencia del bin para el histograma de datos y $F_{randoms}$ para los randoms. Para computar los errores se aplica un modelo tipo Poisson $\Delta_{F_{Datos}} = \sqrt{F_{Datos}}$ y utilizando propagación de errores del cálculo de un cociente.

$$\Delta_{Cociente} = \sqrt{\left(\frac{1}{F_{Randoms}} \cdot \sqrt{F_{Datos}}\right)^2 + \left(\frac{F_{Datos}}{F_{Randoms}^2} \cdot \sqrt{F_{Randoms}}\right)^2} \quad (5.1)$$

5.1.5. Segregación de propiedades

Como el objetivo del presente trabajo es el entender el efecto de las propiedades de la BCGs (nodos) sobre la geometría de la estructura resultante, se realiza el siguiente procedimiento.

A partir de la distribución de una dada propiedad, por ejemplo, la riqueza, se subdivide la muestra en partes iguales empleando como separador al valor de la mediana. Con esto se consiguen dos muestras con la misma cantidad de objetos y por lo tanto, comparables entre sí. Para cada submuestra se construye su muestra de control (sección 5.1.3). Para este fin se emplea la máscara original del catálogo y la distribución de redshift de la submuestra. De esta manera, aplicando el algoritmo MST se obtienen los parámetros normalizados para dos árboles, donde uno presenta nodos ricos, mientras que el otro presenta nodos con menor cantidad de miembros que la mediana de la muestra total.

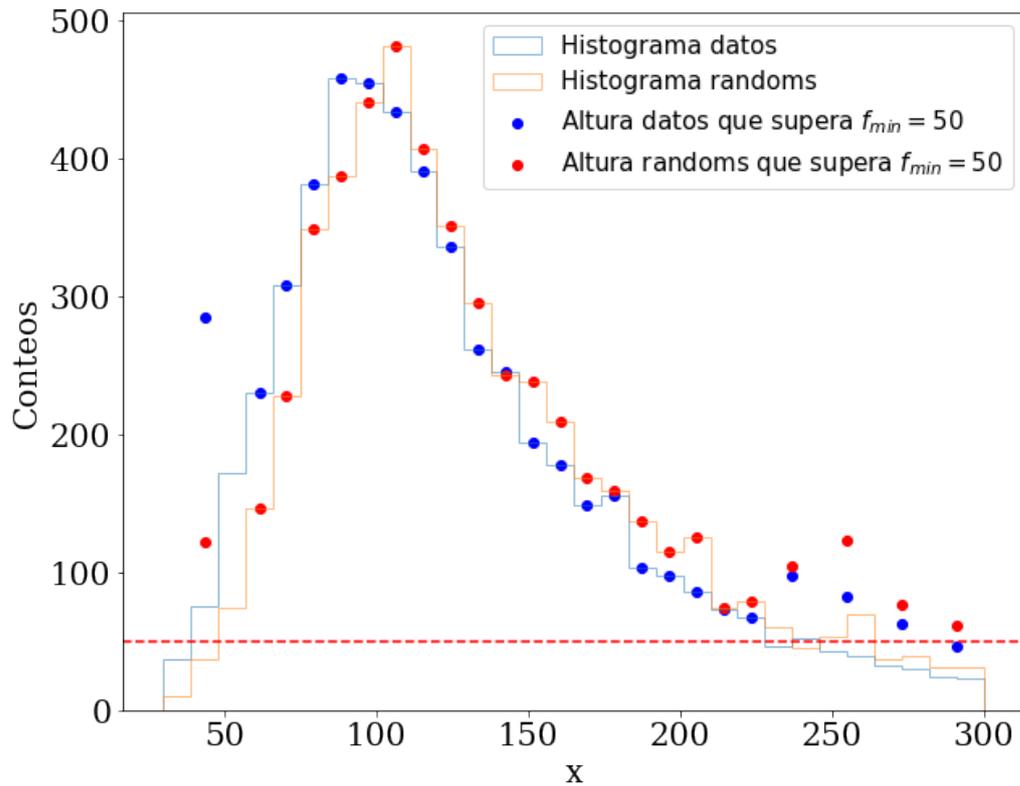


Figura 5.2: Distribuciones de un estadístico x (donde x representa alguno de los tres estadísticos d , b o l) para las muestras de datos (azul) y de números aleatorios (rojo). Los puntos con el mismo código de color significa que proviene de su muestra correspondiente y su valor de frecuencia viene dado por la suma de los bins anteriores que no lograron superar el umbral propuesto (línea a trazos).

Capítulo 6

Análisis de resultados

6.1. Análisis de resultados

En esta sección se analizan las distribuciones de los cocientes de estadísticos (Véase sección 5.1.4) para cada propiedad de las galaxias BCG de la muestra principal. En el presente capítulo los resultados se obtuvieron mediante el siguiente procedimiento:

i) A partir de la muestra principal se realizó una muestra de control apropiada tal como se detalla en la sección 5.1.3 (Técnica de muestreo aleatorio). A continuación se aplicó el algoritmo MST a ambos conjuntos de datos obteniendo para cada uno de ellos un árbol minimal de expansión. Comparando los resultados mediante el método de cociente de estadísticos se obtiene una distribución no sesgada por la geometría de la muestra principal. En la siguiente sección 6.1.2 se estudian los resultados y sus implicaciones.

ii) Las secciones desde 6.1.3 hasta 6.1.7 subdividen a la muestra principal en dos grupos según el valor de la mediana para cada parámetro (véase los parámetros disponibles en las secciones 4.1.3 y 4.1.4). De manera análoga al paso anterior, de cada submuestra se obtiene una muestra de control apropiada. En ambos conjuntos se emplea el algoritmo MST y se comparan los estadísticos obteniendo la distribución de los cocientes. En cada sección se analizan los resultados obtenidos.

Por otro lado, la importancia del análisis de la distribución de los cocientes de estadísticos permite distinguir diferentes mecanismos físicos que inhiben o promueven una mayor cantidad de distancias cortas/largas o nodos con más o menos conexiones.

6.1.1. Interpretación de los estadísticos l y b

En primer lugar se analiza el l . Su comportamiento es similar al de la función de correlación de dos puntos (véase sección 3.0.1) exceptuando que posee mayor información que la distancia entre pares de puntos (Naidoo, 2019). Mediante el análisis de un cierto entorno antes de realizar una conexión entre nodos, el parámetro l captura parte de la estadística de N puntos (Sección 3.0.3). En la figura 6.1 se presenta un ejemplo de interpretación de este parámetro. El panel superior muestra un conjunto de puntos provenientes de una muestra con distribución homogénea (muestra random) mientras que los puntos del panel inferior corresponden a posiciones en una muestra con aglomeraciones (muestra datos). Para ambos casos se realizó un MST obteniendo aristas de distinto peso en función de la distancia entre los puntos. Puede verse que el exceso de pesos menores a $l \leq 3 [Mpc]$ se obtiene en la muestra de datos (panel inferior) mientras que en la muestra de los randoms la distribución de pesos es más diversa. Como ambas muestras poseen la misma cantidad de aristas, el cociente entre la frecuencias del estadístico l entre ambas sera mayor a 1 para las distancias $l \leq 3 [Mpc]$ y menor a 1 para los valores $l \geq 3 [Mpc]$. Es por esta razón que se interpreta un exceso en el cociente de l como el fenómeno de agrupación (*clustering*). Cuando existen excesos en distancias cortas en comparación con una muestra homogénea, en la muestra de datos existe un mecanismo que produce un

acercamiento entre puntos.

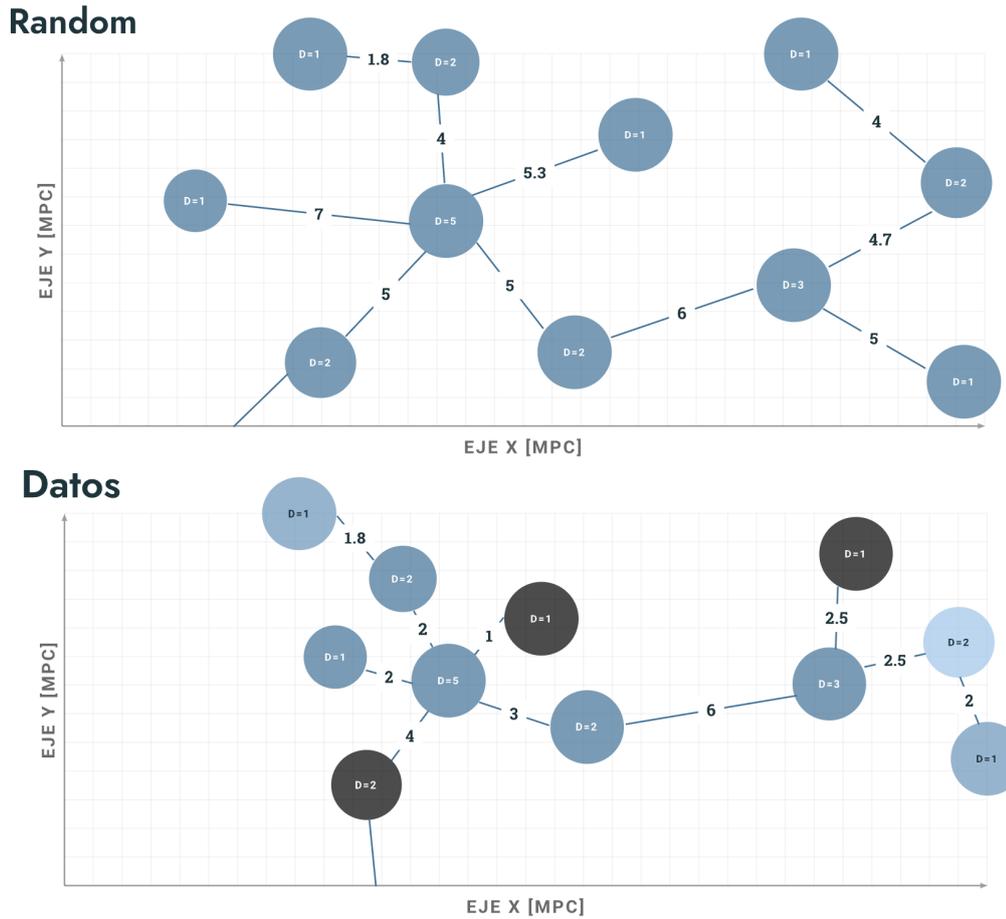


Figura 6.1: Interpretación sobre la distribución del cociente del estadístico l en unidades de Mpc. El panel superior muestra un conjunto de puntos provenientes de una muestra con distribución homogénea (muestra random) mientras que los puntos del panel inferior corresponden a la posición esperada para una distribución de galaxias en el universo a gran escala (muestra datos). Para ambos casos se realizó un MST obteniendo aristas de distinto peso en función de la distancia en el plano de los nodos. Puede verse que el exceso de pesos menores a $l \leq 3$ [Mpc] se obtiene en la muestra de datos (panel inferior) mientras que en la muestra de los randoms la distribución de pesos es más diversa.

El estadístico b se corresponde a la suma de los pesos de las aristas que conforman las ramas en un árbol minimal de expansión (Véase secciones 3.0.3 y el capítulo 3). En la figura 6.2 se presentan dos paneles, el superior consta de nodos distribuidos homogéneamente unidos mediante aristas formando 2 ramas con valores $b = 5$ [Mpc] y $b = 12$ [Mpc] y el inferior posee vértices que corresponden a la muestra de datos formando dos ramas de $b = 12$ [Mpc] en ambos casos. La diferencia entre los paneles consta de una orientación en gran escala hasta una cierta distancia. En este estadístico no sólo se tiene en cuenta la separación entre nodos, si no la concatenación entre ellos. Estos caminos o ramas pueden interpretarse como indicadores de la presencia de estructura.

6.1.2. Propiedades del árbol del conjunto completo

En la figura 6.3 se presenta la distribución de los cocientes de los estadísticos d , b y l (Véase su definición en la sección 3.0.3) obtenidos entre los conjuntos de datos (C_i son los cociente del estadístico $i \in (d, b, l)$ entre la muestra de control y la muestra de datos). La línea roja corresponde a $C_i = 1$ significando la ausencia de señal respecto de la muestra de control. La

$C_i \geq 1$ corresponde a un exceso de objetos en los datos respecto de la muestra de aleatoria mientras que la $C_i \leq 1$ a una menor cantidad de objetos en la muestra de datos respecto de la muestra de control, para cada intervalo del estadístico.

Las barras azules denotan al intervalo de error o $1-\sigma$. En cuanto a los puntos rojos en los último bins de datos son un indicativo de una baja S/N. Si un bin no posee una frecuencia mayor a la frecuencia mínima (se tomo $f_{min} = 50$) el valor no debe considerarse.

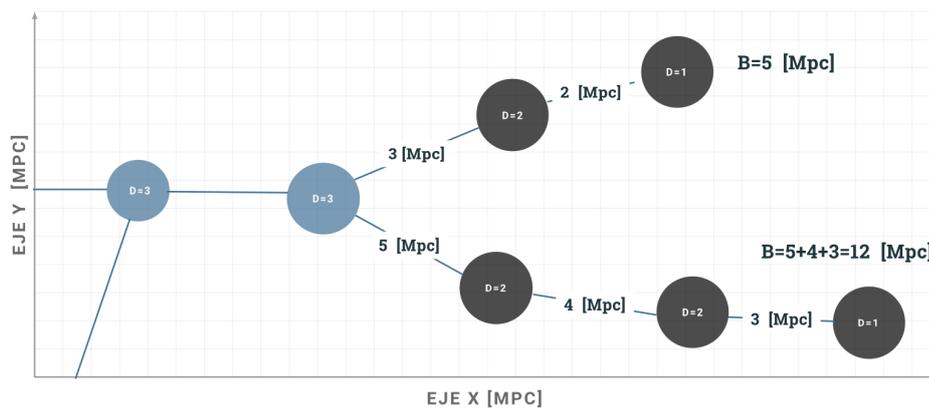
Por la naturaleza del cálculo se nota que muchos estadísticos compensan excesos en algunas escalas con decaimientos en otras.

La distribución de conteos relativos respecto del estadístico d (C_d) de la figura 6.3, no presenta un comportamiento distinto respecto de la muestra de control. Por otro lado, en la distribución del cociente l (C_l) se observa un exceso en distancias cortas que se mantiene hasta unos $35 \text{ Mpc} \cdot h^{-1}$, luego el cociente se estabiliza alrededor de uno y posteriormente se genera el efecto de compensación obteniendo una escasez de distancias medias entre 55 y $75 \text{ Mpc} \cdot h^{-1}$. Estos resultados pueden ser comparados con la función de la autocorrelación de cúmulos ricos, entre 20 y $40 \text{ Mpc} \cdot h^{-1}$ (Véase sección 3.0.1).

Similarmente el cociente relacionado al estadístico b (C_b) presenta excesos en ramas cortas en el rango $[0, 80] \text{ Mpc} \cdot h^{-1}$ que es consistente con un Universo inhomogeneo en escalas menores a $80 \text{ Mpc} \cdot h^{-1}$. Posteriormente se vuelve a observar la compensación entre 220 y $240 \text{ Mpc} \cdot h^{-1}$.

Los parámetros del algoritmo son especialmente sensibles a los patrones de la red cósmica (Naidoo, 2019). Los resultados obtenidos a partir de la figura 6.3 utilizando la muestra principal

Random



Datos

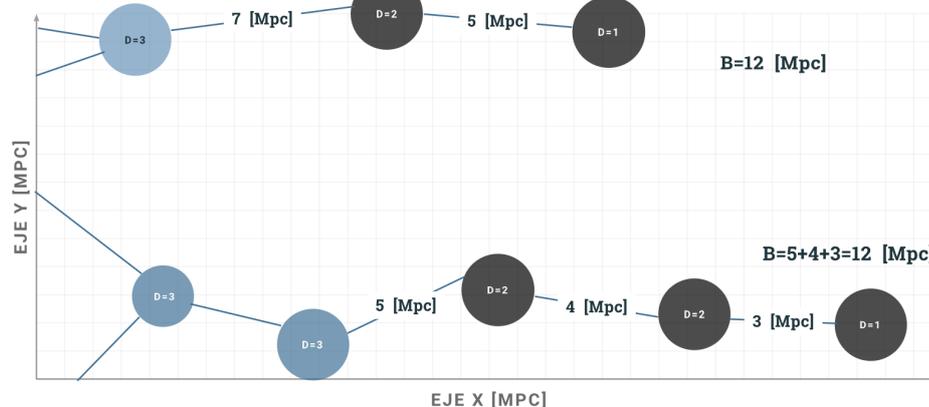


Figura 6.2: Interpretación sobre la distribución del cociente del estadístico b en unidades de Mpc . El panel superior presenta dos ramas construidas a partir de una distribución de puntos homogénea con valores $b = 5 \text{ [Mpc]}$ y $b = 12 \text{ [Mpc]}$ (suma de los pesos de las aristas). El panel inferior posee vértices distribuidos de manera no uniforme formando dos ramas de $b = 12 \text{ [Mpc]}$ en ambos casos. La diferencia entre los paneles sugiere una orientación hasta una cierta distancia.

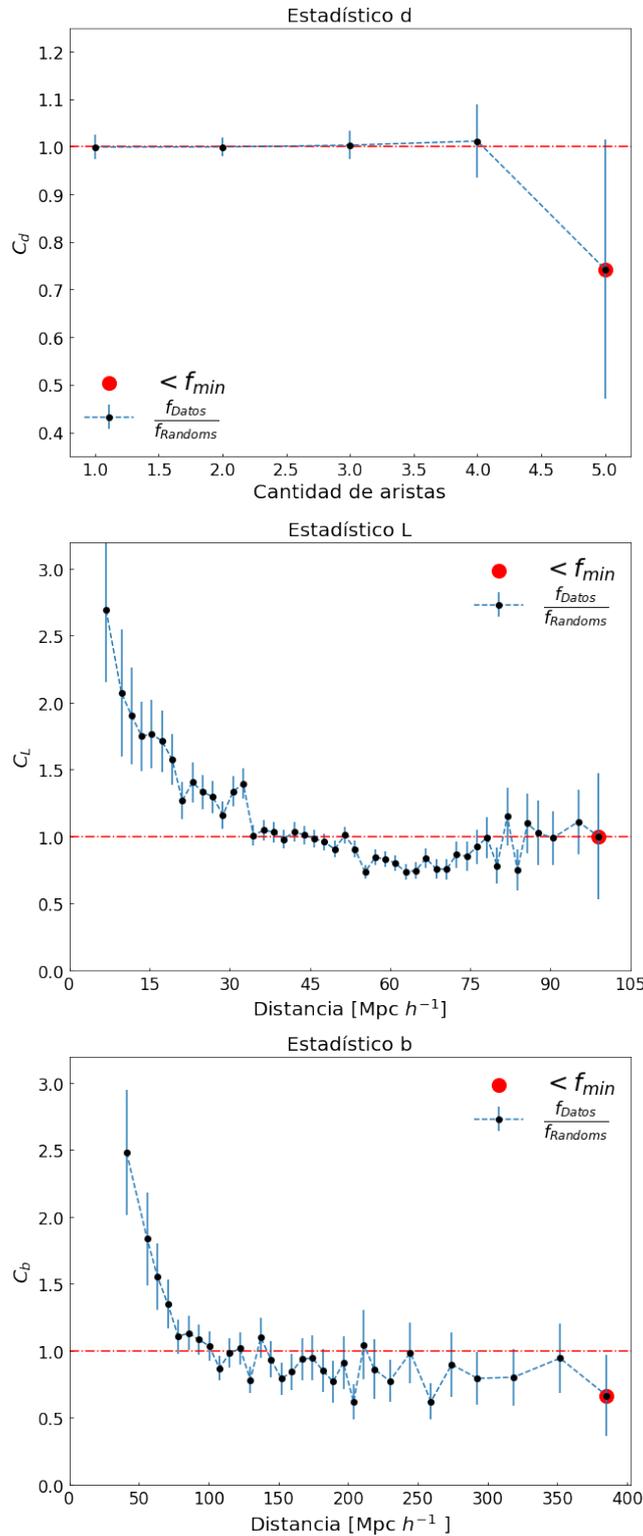


Figura 6.3: Distribución de los tres cocientes de estadísticos (d, l y b) para la muestra principal. En el panel d (C_d) no se visualiza un comportamiento distinto respecto de la muestra de control. En contrapartida, la distribución del cociente l (C_l) posee un exceso en distancias cortas que se mantiene hasta unos $35 Mpc \cdot h^{-1}$, luego el cociente se estabiliza alrededor de uno y posteriormente se genera el efecto de compensación obteniendo una escasez de distancias medias entre 55 y $75 Mpc \cdot h^{-1}$. Similarmente el cociente relacionado al estadístico b (C_b) presenta excesos en ramas cortas en el rango $[0, 80] Mpc \cdot h^{-1}$. En este cociente se observa una compensación entre 220 y $240 Mpc \cdot h^{-1}$ debido al exceso en los primeros bins.

demuestra la potencialidad de esta herramienta.

6.1.3. Segregación por riqueza escalada

Separación

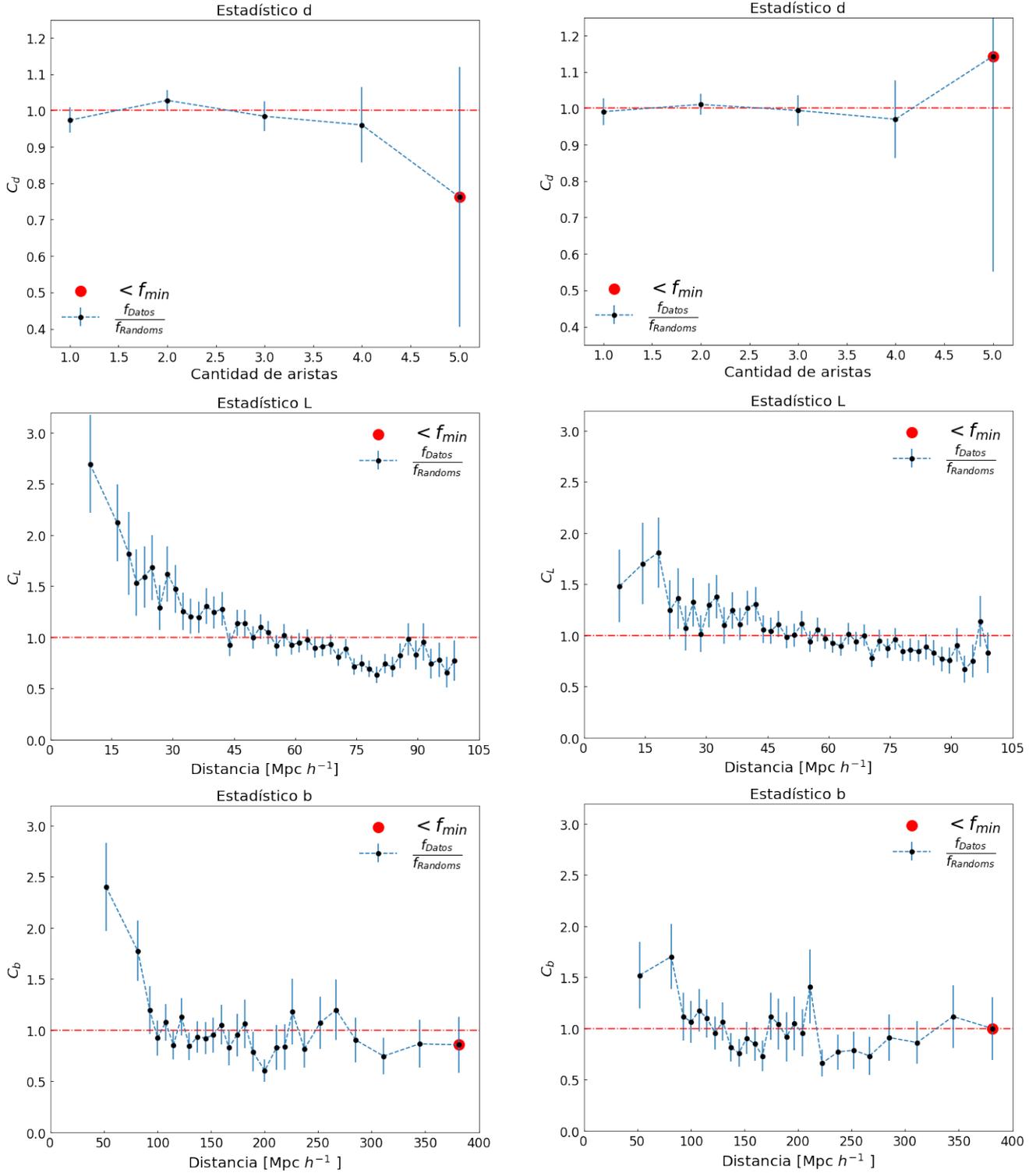
El parámetro riqueza escalada tal como se define en la sección 2.0.4, se emplea para referirse a cuan poblado es un cúmulo. Este parámetro correlaciona con cuan masivo son estas estructuras, es decir, a mayor riqueza se espera que también el cúmulo sea más masivo.

En este trabajo se pretende estudiar el impacto que tiene la riqueza de los cúmulos en la estructura en gran escala del universo. Para ello se confeccionan dos submuestras a partir de la muestra principal que sean estadísticamente comparables entre si. Se escoge separar los nodos por el valor de la mediana de la distribución de riqueza ($Riq_{Med} = 12$ véase figura 4.11), obteniendo dos subconjuntos de datos con la misma cantidad de puntos. Se espera que la primera submuestra conformada por cúmulos ricos (mayor a la mediana) posea correlaciones en escalas más grandes que la submuestra de cúmulos con baja riqueza. En terminos de los estadísticos, se espera encontrar C_l y C_b ambos > 1 para distancias cortas.

Análisis de parámetros

Para cada submuestra se realiza el procedimiento descrito al comienzo del capítulo, obteniendo las figuras 6.4a y 6.4b. En ambas se presentan los cocientes de los 3 estadísticos a analizar. El C_d no presenta señal al igual que lo ocurrido cuando se analizó la muestra total de datos pero si se aprecian señales para los C_l y C_b . Comparando la distribución del cociente de l en las dos figuras se aprecian excesos en distancias menores a $20 \text{ Mpc} \cdot h^{-1}$ pero se observa una mayor potencia para la submuestra de cúmulos ricos. Esta mayor amplitud se corresponde con un mayor agrupamiento de cúmulos en escalas chicas/moderadas. El rango de exceso también difiere para ambas distribuciones, mientras que los cúmulos con menor riqueza poseen $C_l > 1$ en el intervalo $[0, 20] \text{ Mpc} \cdot h^{-1}$, los cúmulos ricos alcanzan distancias mayores $[0, 45] \text{ Mpc} \cdot h^{-1}$. Este último resultado concuerda con lo esperado tal como se mencionó en la sección 6.1.3.

Por otro lado, el cociente del estadístico b en las dos submuestras revelan abundancias en las ramas entre $[0, 90] \text{ Mpc} \cdot h^{-1}$ respecto de su respectiva muestra de control. En este caso la mayor diferencia entre ambas figuras se encuentra en las distintas amplitudes para los primeros bins. Debido a la mayor presencia de agrupamiento los cúmulos ricos tienen una mayor influencia en la estructura en gran escala comparado con la submuestra de menor riqueza. Al emplear la correlación riqueza-masa del cúmulo, se puede concluir que el comportamiento en gran escala esta dominado por las estructuras más masivas del universo.



(a) Figura 6.4a. Distribución de los tres cocientes de estadísticos (d, l y b) para la submuestra de riqueza mayor a 12 miembros. En esta figura puede verse que el panel izquierdo C_d no presenta señal. El panel central correspondiente a C_l posee excesos en longitudes de aristas hasta $45 \text{ Mpc} \cdot h^{-1}$. La potencia para los primeros bins se encuentra dos veces por encima respecto de la distribución homogénea. Por ultimo el panel de la derecha muestra el comportamiento del cociente de ramas C_b . En este caso, el comportamiento revela abundancias en las ramas entre $[0, 90] \text{ Mpc} \cdot h^{-1}$ respecto de su respectiva muestra de control.

(b) Figura 6.4b. Distribución de los tres cocientes de estadísticos (d, l y b) para la submuestra de riqueza menor a 12 miembros. En esta figura pueden verse que el panel izquierdo C_d no presenta señal. El panel central C_l posee excesos en longitudes de aristas hasta $20 \text{ Mpc} \cdot h^{-1}$ cuya potencia no supera en ningún caso $C_l > 2$. El cociente de ramas C_b revela un comportamiento similar al la figura 6.4a, con abundancias en ramas entre $[0, 90] \text{ Mpc} \cdot h^{-1}$ pero con una amplitud menor.

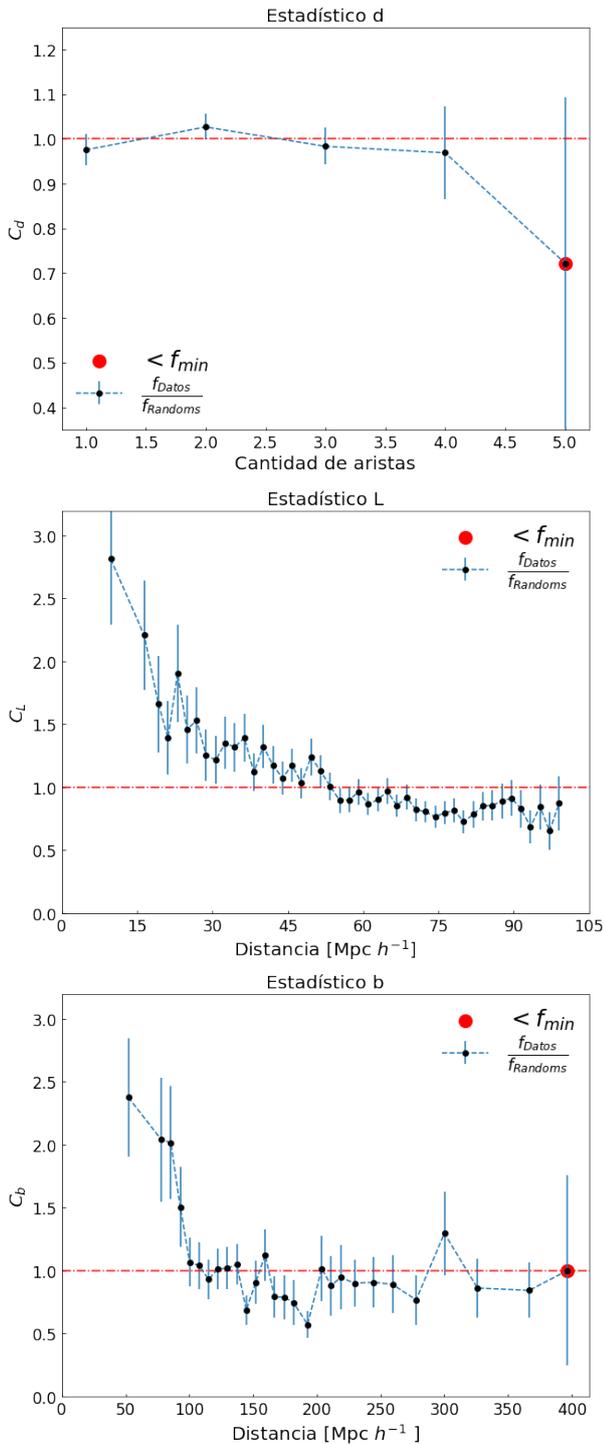
6.1.4. Segregación por riqueza ponderada

Separación

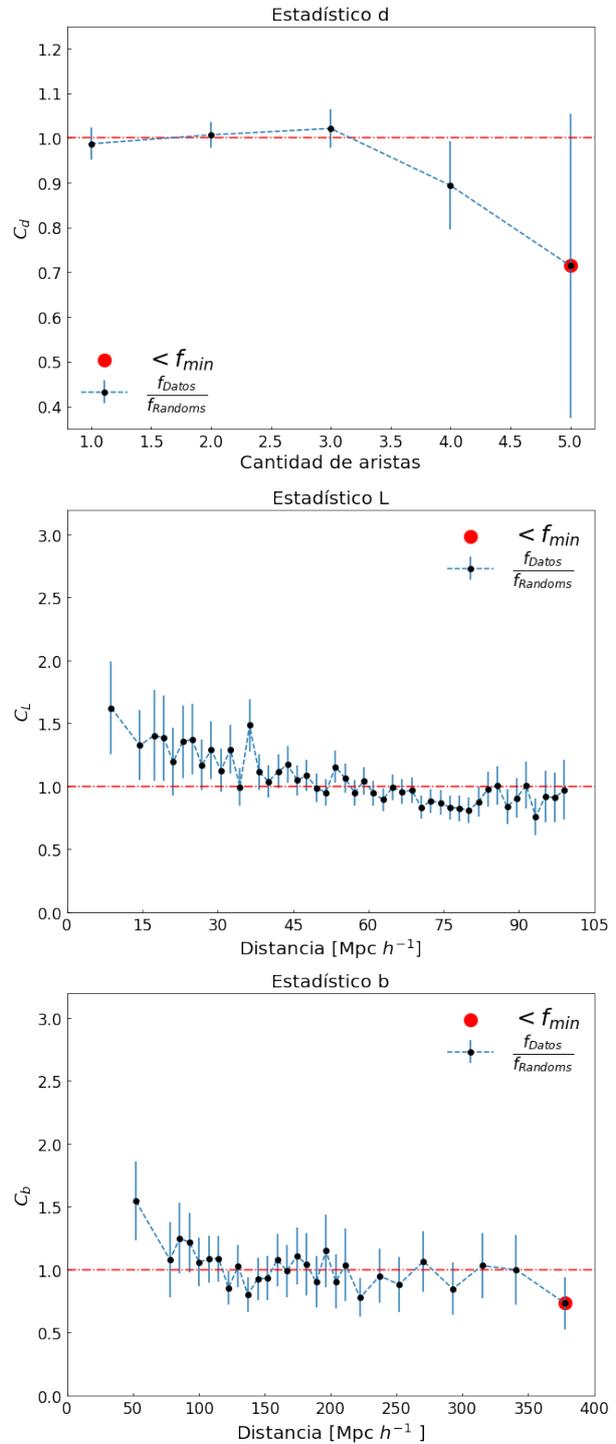
La riqueza ponderada es un parámetro similar a la riqueza escalada como se explica en la sección 2.0.4, sin embargo para algunos cúmulos es una mejor estimación de su riqueza real (Véase Hao et al. (2010)). Es por esta razón que se espera que los resultados estén en concordancia con lo encontrado en el análisis de los cocientes para la segregación de riqueza escalada. Se procede de manera análoga a la sección anterior, buscando separar en dos muestras comparables por encima y por debajo de la mediana y posteriormente cotejar los resultados.

Análisis de parámetros

Se presentan las figuras 6.5a y 6.5b con características similares a los resultados anteriores. Para los cúmulos más ricos, el C_l es mayor en amplitud y en rango que para los cúmulos menor ricos, respetando los valores mencionados en la sección 6.1.3. Algo similar ocurre en el C_b , con la diferencia de una disminución para el rango y amplitud de los nodos con riquezas bajas. Este resultado respalda la conclusión que el comportamiento en gran escala esta dominado por las estructuras más masivas del universo.



(a) Figura 6.5a. Distribución de los cocientes de estadísticos d, l y b para la submuestra de riqueza ponderada mayores a 12,048 miembros. El cociente de grado (C_d) no presenta apartamientos respecto de la muestra de control. Para el C_l se visualiza abundancias en distancias menores a $40 \text{ Mpc} \cdot h^{-1}$. Cabe mencionar que la potencia en los primeros bins se encuentra dos veces por encima de lo esperado al igual que lo ocurrido en los resultados de riqueza escalda. Al analizar el cociente C_b se encuentra una gran cantidad de ramas entre $[0, 90] \text{ Mpc} \cdot h^{-1}$ en el conjunto de datos respecto de su respectiva muestra de control. Además se menciona que su amplitud en los primeros valores se encuentra por encima de 2.



(b) Figura 6.5b. Distribución de los cocientes de estadísticos d, l y b para la submuestra de riqueza ponderada menores a 12,048 miembros. El cociente de grado (C_d) no presenta apartamientos respecto de la muestra de control. Para el cociente C_l se logran apreciar apartamientos respecto de la homogeneidad con aristas de pesos por debajo de $18 \text{ Mpc} \cdot h^{-1}$. Al analizar el cociente C_b el exceso de ramas se identifica en el rango $[0, 60] \text{ Mpc} \cdot h^{-1}$. Sin embargo el panel central como el panel de la derecha poseen una separación muy baja respecto de lo observado en la submuestra antagónica.

6.1.5. Segregación en g-r

Separación

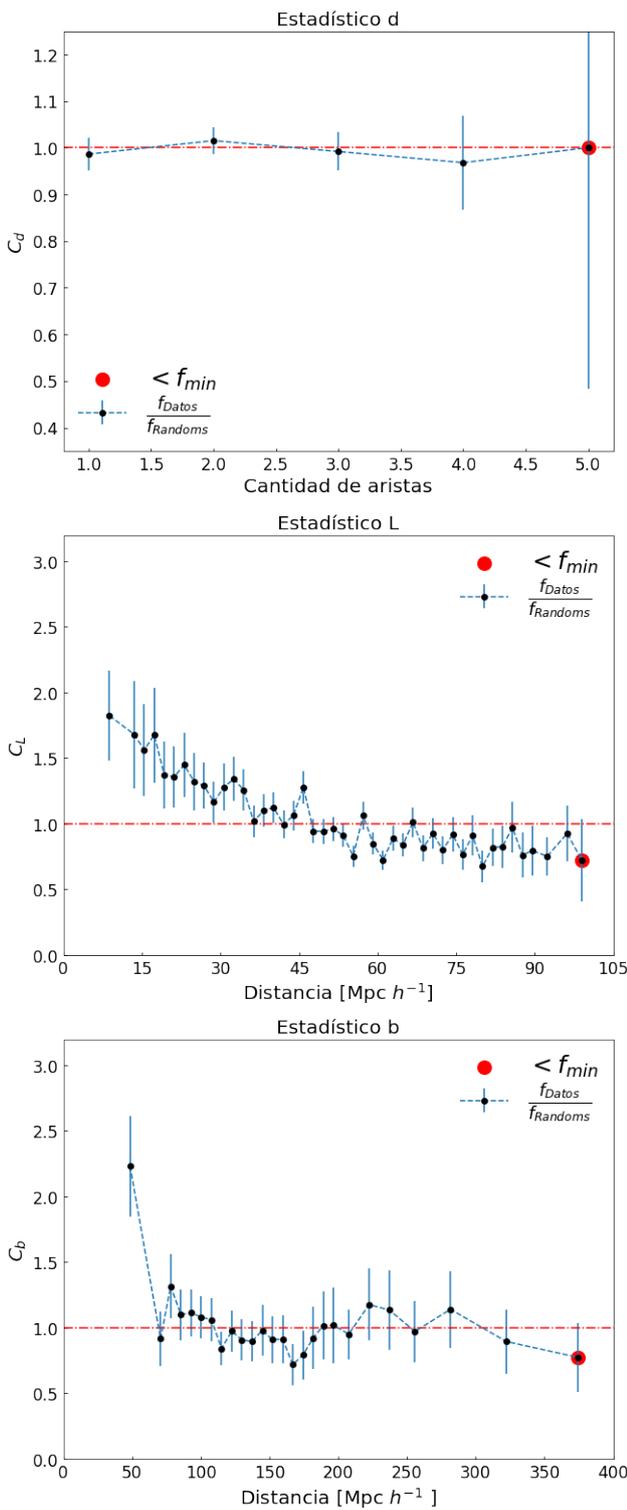
La técnica de identificación con la que se confeccionó el catálogo emplea la secuencia roja de los cúmulos, la cual es detectable hasta $z_{foto} = 0,43$ en el espacio g-r (Sección 4.1.1). La muestra principal tiene como redshift fotométrico máximo $z_{foto} = 0,42$, es por ello que al realizar el presente estudio se escoge trabajar con dos submuestras separadas por la distribución de g-r en vez del color r-i. La confección de las submuestras se realiza sobre un conjunto de datos cuyos nodos son de por sí rojos ya que esas son las propiedades de la muestra principal (Véase sección 4.1.3). Al separar por la mediana ($g - r_{Med} = 1,67$ ver figura 4.7), se obtienen dos conjuntos de datos, donde probablemente aquellas BCG con un índice de color muy rojo se correspondan con un gran contenido de polvo. Las submuestras obtenidas se denotan como galaxias muy rojas y galaxias rojas.

Análisis de parámetros

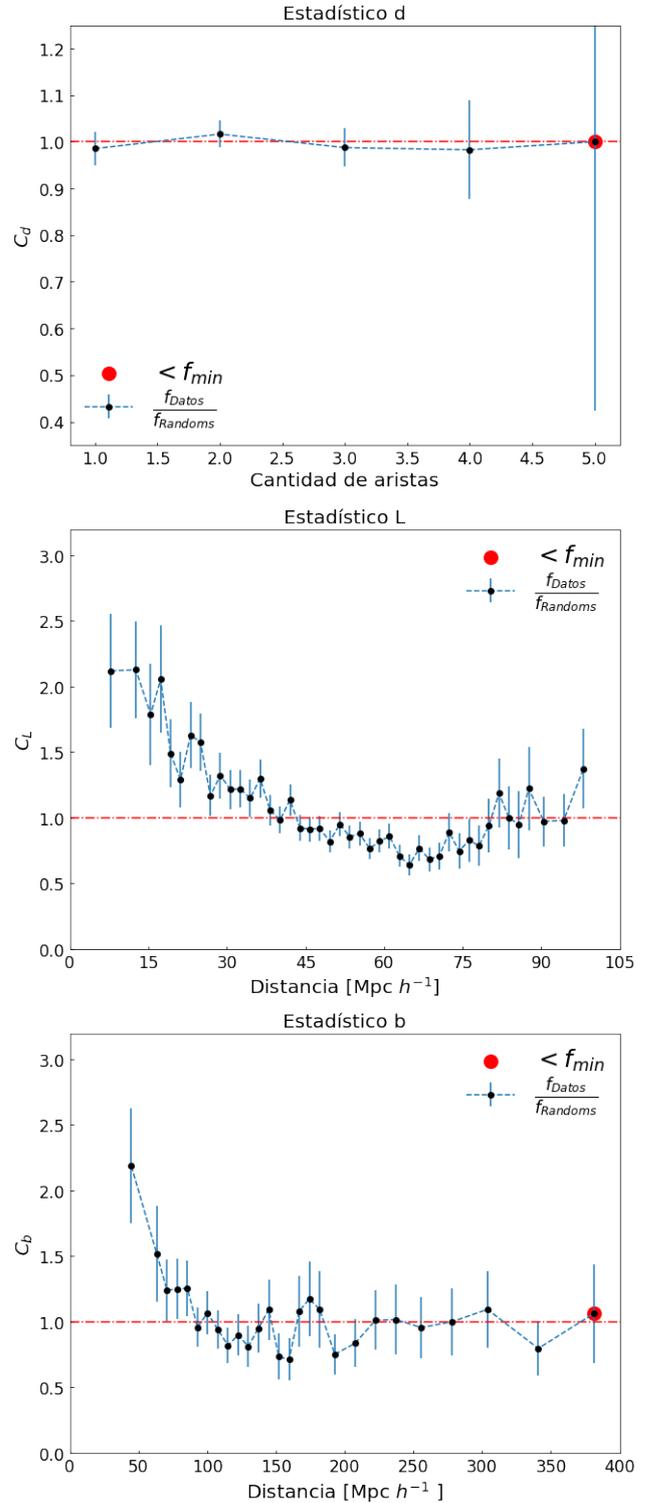
Para la submuestra de las galaxias muy rojas se obtiene la figura 6.6a y para las galaxias rojas la figura 6.6b donde se presenta la distribución de los cocientes de estadísticos d, l y b. La distribución del cociente del estadístico l para ambas muestras posee diferencia en la amplitud de los primeros bins de distancia (escalas menores a $20 \text{ Mpc} \cdot h^{-1}$) teniendo mayor potencia para las galaxias menos enrojecidas.

Visualizando los paneles correspondientes al cociente del estadístico b, las dos distribuciones coinciden en valor para el primer bin, pero luego se diferencian. En el caso de las muy rojas (figura 6.6a) se estabiliza en valores cercanos a $C_b \approx 1$ mientras que las galaxias rojas continúan con excesos de ramas respecto de la muestra de control hasta $80 \text{ Mpc} \cdot h^{-1}$.

Se concluye que las galaxias rojas marcan la misma tendencia que se observa para la muestra total y para los cúmulos ricos mientras que las galaxias muy rojas presentan una menor señal en comparación. Probablemente un índice de color elevado corresponda a la presencia de enrojecimiento interestelar más que intergaláctico por lo tanto al ser una propiedad más local de las BCG no afecte en gran medida a la estructura en gran escala.



(a) Figura 6.6a. Distribución de los cocientes de estadísticos para la submuestra de g-r mayor a 1,67. Esta submuestra corresponde a galaxias muy rojas. El análisis de C_d no presenta señal respecto de la muestra de control. En oposición, el C_l exhibe excesos de distancias para el conjunto de datos con escalas menores a $30 Mpc \cdot h^{-1}$ y una amplitud por debajo del valor 2. El cociente C_b por otro lado, posee un solo bin que se separa del comportamiento esperado en una distribución homogénea.



(b) Figura 6.6b. Distribución de los cocientes de estadísticos para la submuestra de g-r menor a 1,67. Esta submuestra corresponde a galaxias rojas. El análisis de C_d no presenta señal respecto de la muestra de control. En contrapartida, el C_l exhibe excesos de distancias para el conjunto de datos con escalas menores a $30 Mpc \cdot h^{-1}$ pero con una amplitud considerable por encima del valor 2 respecto de la muestra de control. Se observa para el cociente C_b un apartamiento en los primeros cinco bins hasta escalas menores de $80 Mpc \cdot h^{-1}$. Este resultado es similar a lo encontrado en la figura 6.3.

6.1.6. Segregación por dispersión de velocidades

Separación

La dispersión de velocidades de la muestra principal es una medida de la varianza de la distribución de velocidad de las estrellas de la BCG (sección 2.0.4). Las velocidades de las estrellas están determinadas por la masa gravitatoria en el interior de su órbita y la distancia a la que se encuentren del centro de masa de la galaxia. Es por esta razón que el σ permite estimar la cantidad total de masa que genera la dispersión observada. Una vez más se emplea la mediana como corte para generar dos submuestras (corte $\sigma_{Med} = 250,15 \frac{km}{s}$ como se puede ver en la figura 4.14). Una aclaración importante es que la dispersión de velocidades habla de la masa de la BCG mientras que el parámetro de la riqueza de la masa del cúmulo.

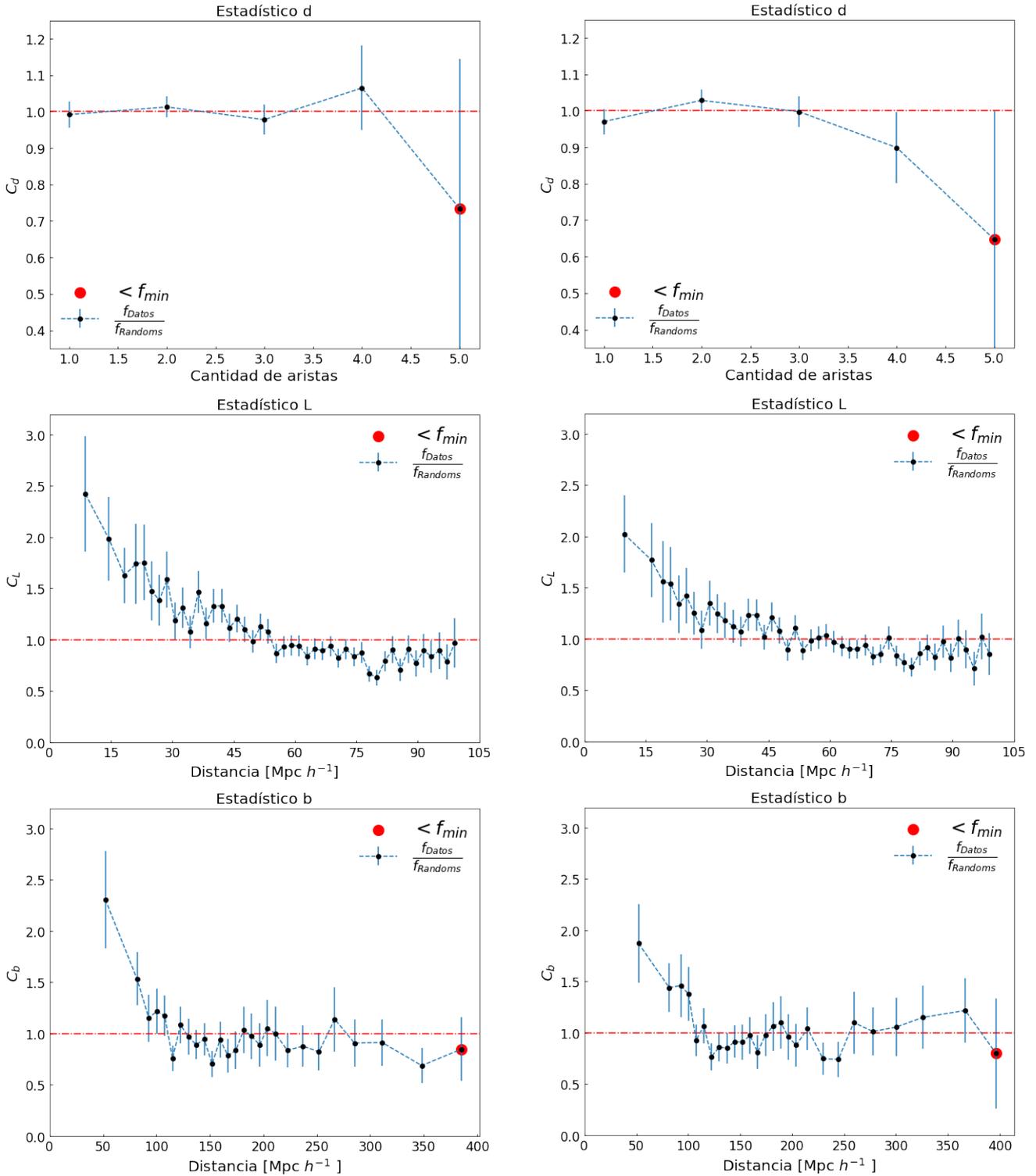
Análisis de parámetros

Siguiendo el mismo procedimiento utilizado en las secciones anteriores, se obtienen dos figuras 6.7a y 6.7b para las submuestra de alta dispersión y la submuestra de baja dispersión, respectivamente. Se espera que las galaxias con dispersión más alta posean una escala de correlación más grande que las galaxias con menos dispersión. Esto se debe a la correlación que puede existir con el parámetro de riqueza. En primer lugar la masa total de la BCG y la dispersión de velocidades están relacionadas. En segundo término se espera que sea más probable encontrar en un cúmulo masivo una galaxia central masiva. Es entonces que debido a los resultados de la sección 6.1.3 es natural pensar que las BCG masivas al igual que los cúmulos ricos mantienen correlaciones a escalas mayores.

Las distribuciones del C_d son muy similares para ambas submuestras sin mostrar ninguna señal para los 5 bins. En cuanto a la distribución de C_l , para la submuestra de alta dispersión se encuentra una potencia mayor para el primer bin respecto de la submuestra de baja dispersión. Además si se estudia el rango de inhomogeneidad, en ambos conjuntos el rango es practicamente el mismo (teniendo en cuenta que las distancias no estan determinadas con z_{spect}). Combinando ambas observaciones, se puede concluir que la correlación entre galaxias centrales más masivas es mayor que para las galaxias menos masivas (Véase sección 3.0.1)

La distribución de C_b de la submuestra de alta dispersión presenta excesos en las ramas de longitudes menores a $80 Mpc \cdot h^{-1}$. Al comparar en este intervalo con la figura 6.7b se observa que posee una mayor amplitud. Este resultado indica una relevancia más importante de las galaxias con altas dispersiones en la estructura en gran escala. Sin embargo, para ramas con distancias mayores a $80 Mpc \cdot h^{-1}$, las galaxias de alta dispersión disminuyen su señal y la mayor contribución en la identificación de estructuras en gran escala se deben a las galaxias con baja dispersión (excesos de ramas con valores menores a $110 Mpc \cdot h^{-1}$). Cabe mencionar que al utilizar la muestra principal, tal como se estudio en las sección 6.1.2, no se visualiza excesos en ramas mayores a $80 Mpc \cdot h^{-1}$, por lo tanto este efecto de inhomogeneidad respecto de la muestra de control queda compensado al incluir todos los nodos.

Como conclusión y teniendo en cuenta la correlación masa y dispersión de velocidades, las galaxias centrales muy masivas están distribuidas más inhomogeneamente en el Universo que las no tan masivas.



(a) Figura 6.7a. Distribución de los cocientes de estadísticos para la submuestra de dispersión de velocidades mayor a $\sigma_{Med} = 250,15 \frac{km}{s}$. El cociente C_d no percibe apartamientos significativos. Con respecto al panel central, el cociente de pesos de las aristas exhibe excesos hasta escalas de $[0 - 35] Mpc \cdot h^{-1}$. Para los dos primeros valores de C_l se logra detectar una amplitud mayor o igual a 2. En el caso del panel de la derecha, el C_b permite inferir abundancias en las ramas de tamaño menor a $80 Mpc \cdot h^{-1}$.

(b) Figura 6.7b. Distribución de los cocientes de estadísticos para la submuestra de dispersión de velocidades menor a $\sigma_{Med} = 250,15 \frac{km}{s}$. El cociente C_d no percibe apartamientos significativos respecto de la muestra de control. El C_l presenta una estructura similar a 6.7a con la diferencia sobre la amplitud en cada paso. En los primeros bins, la diferencia no es mayor a 0.4. El C_b posee un comportamiento similar lo ocurrido con el panel anterior. Si bien permite inferir abundancias en las ramas de tamaño menor a $80 Mpc \cdot h^{-1}$, la potencia observada es menor para esta submuestra que en la figura 6.7a

6.1.7. Segregación por cociente de semiejes

Separación

El cociente de semiejes como su nombre lo indica es la razón entre el semieje menor proyectado y el semieje mayor proyectado de la galaxia (Véase sección 4.1.4). Como la muestra se compone de galaxias BCG que tienen típicamente una estructura elíptica, se busca estudiar la correlación con la estructura en gran escala en función de la morfología proyectada de la galaxia central. Con este fin se busca separar en dos submuestras por la mediana a la muestra principal ($\frac{b}{a_{Med}} = 0,76$). Una vez realizado se obtiene dos submuestras: la primera con simetría proyectada más esférica (mayor al valor de la mediana) mientras que la segunda corresponde a simetrías proyectadas más elipsoidales. (Véase figura 4.15).

Análisis de parámetros

Se comienza analizando los resultados de haber aplicado el algoritmo en la figuras 6.8a para las BCG más esféricas y 6.8b) para las más elípticas. En primer lugar, la distribución del cociente del estadístico d permite concluir que no se detecta ninguna diferencia aparente con la muestra de control.

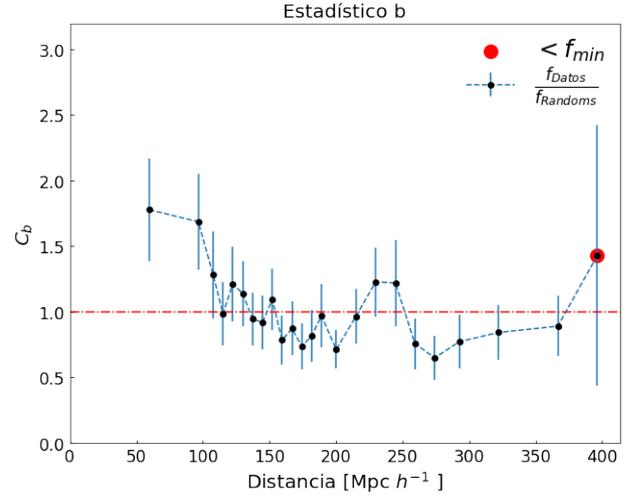
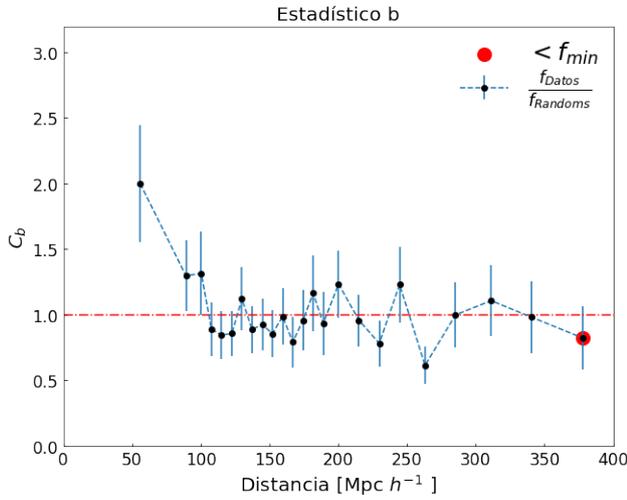
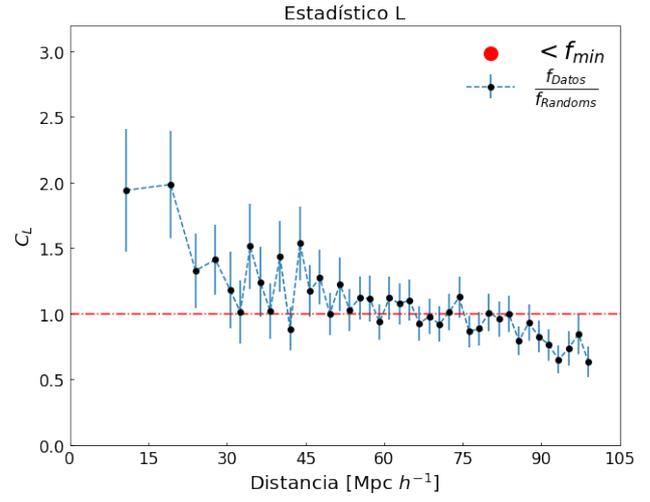
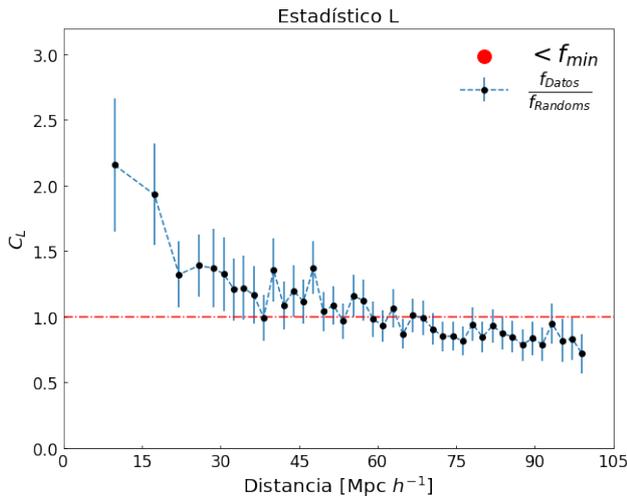
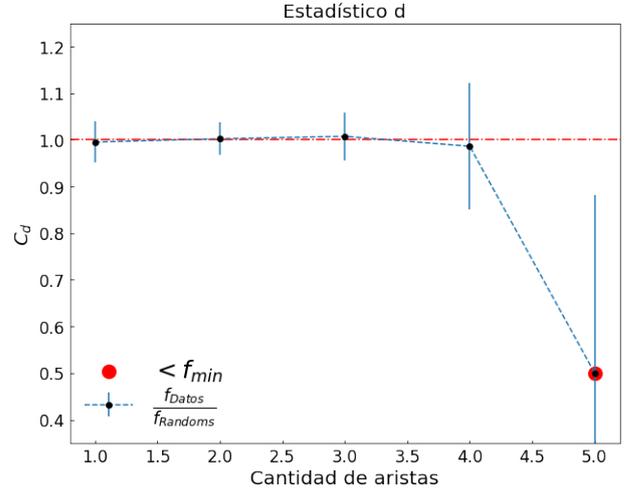
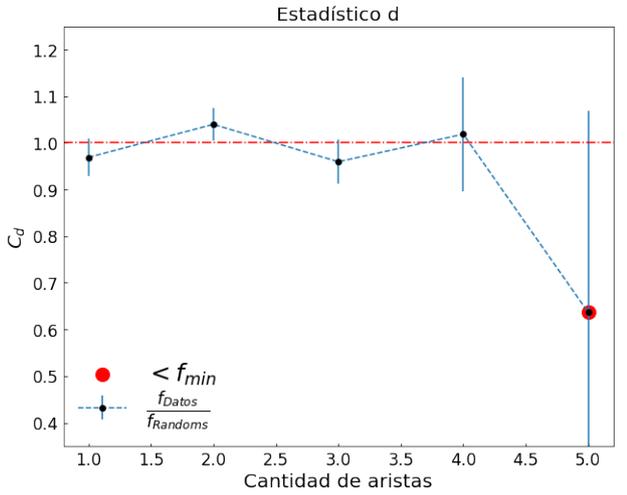
En cuanto al estadístico l , se analiza el intervalo donde $C_l > 1$ y la amplitud que poseen los bins entre ambas figuras. En este caso no se logran detectar grandes diferencias al realizar la comparación pero se encuentran que, tanto para galaxias elípticas como las esféricas, existe una correlación hasta una escala de $30 \text{ Mpc} \cdot h^{-1}$. Este resultado esta en acuerdo con lo encontrado para el árbol completo.

Similarmente al estudiar la distribución del C_b no se observan diferencias entre las submuestras. Ambos intervalos de $C_b > 1$ alcanzan valores máximos de hasta $100 \text{ Mpc} \cdot h^{-1}$, lo cual es superior a lo esperado en comparación con la muestra total (Véase figura 6.3).

A partir de este análisis se encuentra que no hay diferencias significativas entre ambas submuestras. Por lo tanto existen dos posibles escenarios:

i) La primer alternativa se debe a la mediciones de los semiejes proyectados. Este parámetro al ser una proyección puede no ser una medida fiable de una cantidad intrínseca del objeto. Debido a ello el algoritmo no esta separando dos grupos con propiedades distintas y por lo tanto no tienen un efecto considerable en la estructura en gran escala que sea diferente al analizado para la muestra completa.

ii) El segundo escenario propuesto es que el parámetro de semiejes verdaderos sea una propiedad local y por lo tanto no afecte a las distribución espacial en gran escala.



(a) Figura 6.8a. Distribución de los cocientes de estadísticos para la submuestra de cociente de semillas mayores a 0,76. En el caso del C_d , se observa un comportamiento indistinguible de la señal proporcionada por la muestra de control a diferencia de los cocientes C_l y C_b . El intervalo de exceso de pesos menores es de $30 \text{ Mpc} \cdot h^{-1}$ es del orden de 2. Por otro lado el cociente C_b logra alcanzar escalas de excesos de hasta $100 \text{ Mpc} \cdot h^{-1}$ pero no logra superar el valor de cociente $C_b = 2$.

(b) Figura 6.8b. Distribución de los cocientes de estadísticos para la submuestra de cociente de semillas menores a 0,76. En el caso del C_d , se observa un comportamiento indistinguible de la señal proporcionada por la muestra de control a diferencia de los cocientes C_l y C_b . El intervalo de exceso de pesos menores es de $30 \text{ Mpc} \cdot h^{-1}$ y los valores obtenidos en el eje de las y son del orden de 2. Por otro lado el cociente C_b logra alcanzar escalas de excesos de hasta $100 \text{ Mpc} \cdot h^{-1}$ pero no logra superar el valor de cociente $C_b = 2$.

6.1.8. Orientación relativa entre las BCGs

Ya se ha mencionado que analizar el cociente del estadístico l es similar al estudio de la función de correlación de 2 puntos, pero considerando información de más alto orden (Secciones 3.0.1 y 3.0.3). Por otro lado, el cociente del estadístico b es similar a pensar una orientación entre nodos y aristas que conforman una rama. Los resultados de los análisis presentados en esta sección, sugieren que se puede hallar información relevante estudiando alineaciones relativas entre las BCG en relación con el árbol de expansión. Como una primera aproximación, se propone estudiar la distribución de la diferencia entre ángulos de posición (sección 2.0.3) para dos nodos conectados por una arista. Es interesante realizar este análisis ya que es más frecuente pensar en alineaciones entre galaxia y estructura en gran escala que el que se pretende para este trabajo (galaxia-galaxia). Para este fin se empleará la muestra principal y el árbol generado a partir de ella. Como un nodo puede tener más de una arista conectada (grado mayor a 1) se analizan en primera instancia la orientación relativa entre nodos conectados mediante una arista. Para poder comparar los resultados encontrados, se construye una muestra de control donde las posiciones de las galaxias son las mismas que las de la muestra principal pero asignando una distribución aleatoria y uniforme de ángulos de posición. Así, ambas muestras tienen el mismo MST.

En ambas muestras, para cada par de galaxias conectadas por una arista se calcula la orientación relativa mediante la resta de sus respectivos ángulos de posición. Cabe mencionar los ángulos calculados corresponden al ángulo agudo entre la orientación de ambas galaxias, es decir, si el ángulo relativo entre los nodos (θ_{rel}) es mayor a 90° , se utiliza $180 - \theta_{rel}$.

Para la muestra principal de BCGs se espera encontrar algún indicio de orientaciones correlacionadas entre nodos conectados por el árbol. De existir se desea investigar en futuros trabajos la correlación de esta orientación en función de la escala empleando la longitud de las aristas (véase sección 8).

La figura 6.9 presenta la distribución de ángulos agudos relativos para la muestra de datos (color azul) y la muestra confeccionada a partir de ángulos de posición aleatorios (color naranja). Las distribuciones observadas pueden parecer uniformes, pero para comprobar la presencia de señales significativas se calcula cuantas desviaciones estándar difieren respecto de la media de cada distribución utilizando la aproximación de fluctuaciones Poissonianas:

$$\sigma = \frac{N_{bin} - \bar{N}_{tot}}{\sqrt{\bar{N}_{tot}}} \quad (6.1)$$

donde N_{bin} es el número de objetos para cada intervalo y \bar{N}_{tot} el número medio de objetos en la distribución.

La muestra proveniente de ángulos aleatorios funciona como muestra de control. Esta se emplea para estimar la magnitud de la variación natural de una distribución uniforme de una cierta cantidad de objetos que se encuentran binned con un dado tamaño. El criterio escogido para obtener un valor estimativo de señal es el cálculo del percentil 99 (P_{99}) de la distribución de σ para la muestra de ángulos aleatorios. Al estimar el valor para P_{99} de los σ calculados a partir de la ecuación 6.1 se obtiene un límite inferior a partir del cual se puede considerar una detección de la señal.

La muestra de control presenta un $P_{99} = 2,57 \sigma$. Los valores más altos encontrados en la muestra de datos que superan estos valores son $3,82 \sigma$ para el primer bin ($\theta_{rel} \approx 3^\circ$) y $2,99 \sigma$ para el sexto ($\theta_{rel} \approx 33^\circ$). Esto indica la presencia de una señal de alineación relativa para las galaxias de la muestra principal.

A partir de estos resultados, y teniendo en cuenta el número de aristas que conectan cada galaxia con sus vecinas, se decide analizar los nodos en función de su grado. Para ello se construyen cinco submuestras (con $d = 1, \dots, 5$) en donde se calcula la distribución de ángulos agudos relativos y se realiza el procedimiento mencionado anteriormente para detectar señales de orientación. Los resultados se exponen en la tabla 6.1. En ella se presentan los valores para

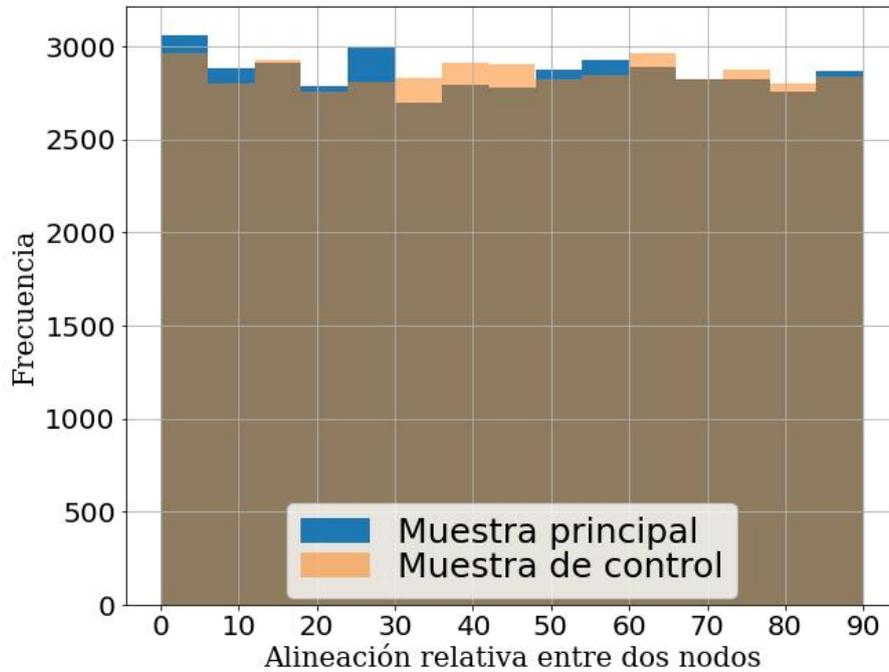


Figura 6.9: Distribuciones de las orientaciones relativas entre los nodos conectados por una arista. Los datos de la muestra principal se exhiben en color azul mientras que la muestra en naranja corresponde a los ángulos de orientación relativa, calculados a partir de una distribución uniforme de ángulos de posición. Puede observarse un máximo alrededor de 10° para el conjunto de datos y un mínimo entre 30 y 40° .

Muestra de Grado	P_{99}	Valor de sigma $>P_{99}$	θ_{rel}
$d = 1$	$2,03 \sigma$	$2,21 \sigma$	57°
$d = 2$	$2,38 \sigma$	$3,57 \sigma$	3°
$d = 3$	$2,10 \sigma$	$2,83 \sigma$	45°
$d = 4$	$2,27 \sigma$	-	-
$d = 5$	$2,39 \sigma$	-	-

Tabla 6.1: Los resultados para las 5 muestras segregadas por el grado de los nodos. En la columna P_{99} se presenta el valor estimado del percentil 99% para la muestra de control. Las columnas *Valor de sigma* $>P_{99}$ y θ_{rel} corresponden a los valores obtenidos a partir de la muestra principal y la ecuación 6.1. Para las muestras con grados altos ($4 \leq d$) no se encuentran señales de orientaciones relativas. La muestra $d = 2$ coincide con lo encontrado para la totalidad de la muestra.

las cinco muestras segregadas por el grado de los nodos. En la columna P_{99} se observan el percentil 99% estimado para la muestra de control. Las columnas *Valor de sigma* $>P_{99}$ y θ_{rel} corresponden a los valores obtenidos a partir de la muestra principal. La muestra $d = 2$ coincide con lo encontrado para la muestra principal (figura 6.9) mientras que el comportamiento de las orientaciones relativas a 57° y 45° (submuestras $d = 1$ y $d = 3$ respectivamente) se compensan en la muestra principal. Para los conjuntos de grados más altos ($4 \leq d$) no se encuentran señales de orientaciones relativas. Se concluye entonces que la aparición de una señal en la figura 6.9 proviene de los nodos con solo dos aristas. Se presenta la figura 6.10 en donde se observa la distribución de los ángulos agudos relativos entre galaxias con dos aristas, realizado para la muestra de control (naranja) y la referida a los datos (azul). En la misma se visualiza una señal de orientación relativa hasta $\theta_{rel} \leq 6^\circ$, consistente con BCGs muy alineadas.

Teniendo en consideración estos resultados, se buscó alguna dependencia de las orientaciones relativas de acuerdo al largo de las aristas que conectan las BCGs, pero no se encontró una

Tipo de proyección	P_{99}	Valor de sigma $>P_{99}$	θ_{rel}
Esféricas	2,50 σ	2,66 σ y 3,70 σ	3° y 33°
Elipsoidales	2,37 σ	2,76 σ y 4,32 σ	3° y 33°

Tabla 6.2: Resultados para la muestra principal segregada de acuerdo al cociente de semiejes de las BCGs. En la columna P_{99} se presenta el valor estimado del percentil 99 % para la muestra de control. Las columnas *Valor de sigma* $>P_{99}$ y θ_{rel} corresponden a los valores obtenidos a partir de la muestra principal y la ecuación 6.1. Para las muestras con galaxias más esféricas se puede ver una señal con mayor intensidad hacia los valores de 33° respecto de 3°. Los nodos con una forma elipsoidal tienden al mismo resultado.

dependencia clara con este parámetro.

Por otra parte, la precisión en la medición de la orientación de una galaxia está relacionada con su cociente de semiejes. La forma de la isofota es utilizada para determinar ambos parámetros (véase sección 4.1.4 y 2.0.3) y por lo tanto podría tener algún efecto sobre los alineamientos. Para estudiar esta dependencia se separa a la muestra principal de acuerdo a la mediana de la distribución de los cocientes de semiejes de las BCGs, tal como en la sección 6.1.7 y se procede a medir los ángulos de orientación relativos de las BCGs que se encuentran unidas mediante una arista. A continuación se exponen los resultados de manera análoga a la segregación por grados. En la tabla 6.2 se analizan los valores obtenidos. Para las muestras con galaxias más esféricas se puede ver dos señales. La de mayor intensidad corresponde a la orientación en 33° similarmente a lo ocurrido para los nodos con una forma elipsoidal. Se puede inferir que los valores con orientaciones 33° se deben en su mayoría a la forma de las isofotas mientras que las orientaciones cercanas a 3° son regidas por los nodos con grado dos.

Se plantea la posibilidad de exigir ambas condiciones y caracterizar dichas muestras. Con este fin, se emplean los nodos de grado dos y se separan según su cociente de semiejes. los resultados se presentan en la tabla 6.3. Para las muestras con galaxias con $d = 2$ y con forma

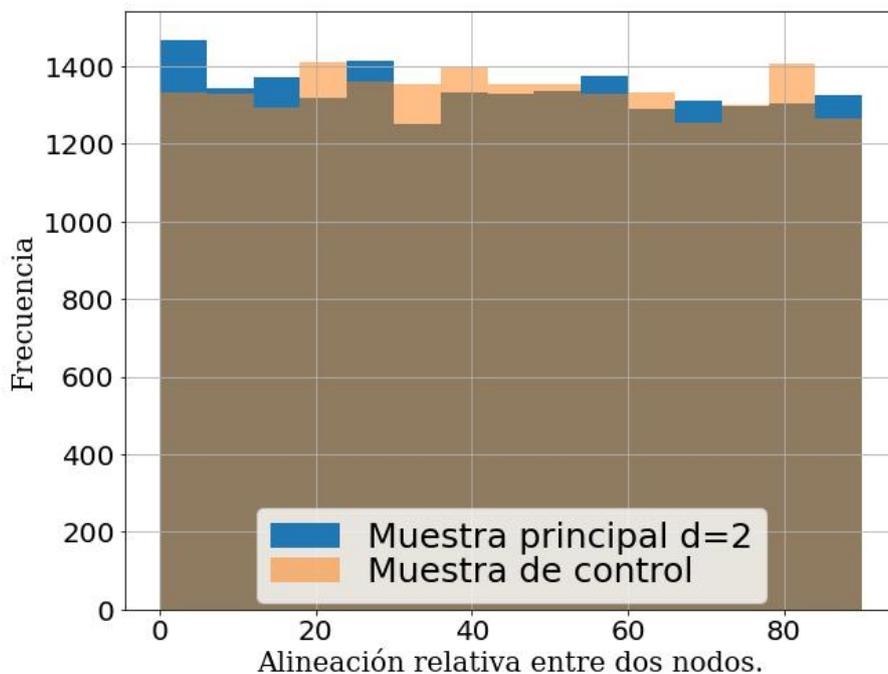


Figura 6.10: Distribución de orientaciones relativas entre galaxias con dos aristas, realizada para la muestra de control (naranja) y la referida a los datos (azul). Puede observarse un máximo en el primer bin para el conjunto de datos en 3° y un decaimiento para los orientados a 33°

Tipo de proyección	P_{99}	Valor de sigma $>P_{99}$	θ_{rel}
d=2 y Esféricas	2,29 σ	3,49 σ y 2,36 σ	27° y 75°
d=2 y Elipsoidales	2,22 σ	2,90 σ y 3,45 σ	3° y 33°

Tabla 6.3: Los resultados para las 2 muestras segregadas en grado $d=2$ y por el cociente de semiejes. En la columna P_{99} se presenta el valor estimado del percentil 99% para la muestra de control. Las columnas *Valor de sigma* $>P_{99}$ y θ_{rel} corresponde a los valores obtenidos a partir de la muestra principal y la ecuación 6.1. Para las muestras con galaxias más esféricas con nodos de grado dos se puede ver una señal con mayor intensidad hacia los valores de 27° respecto de 75°. Los nodos con $d = 2$ y con una forma elipsoidal poseen consistencia con las alineaciones encontradas.

elipsoidal se encuentran consistencias con las alineaciones reportadas para los ángulos agudos de 3° y 33°.

En cuanto a los nodos de la muestra más esférica de grado dos se puede ver una señal con mayor intensidad hacia los valores de 27° respecto de 75°.

Cabe recordar que los ángulos de posición son medidas de orientación proyectadas sobre la esfera celeste, por lo tanto para obtener información más precisa sobre las alineaciones reales de las BCGs es necesario hacer estudios en simulaciones numéricas. Sin embargo, el presente análisis da indicios de la presencia de correlación entre las orientaciones de las galaxias, y puede considerarse como un primer paso para un posterior estudio de las alineaciones de las BCGs, tanto entre ellas como en relación a las aristas del MST, consideradas como trazadoras de la geometría de la estructura en gran escala.

Capítulo 7

Conclusiones

A lo largo de este trabajo se estudió un conjunto de 55423 galaxias BCG provenientes del catálogo de galaxias centrales de cúmulos (*GMBCG Cluster Catalog*) elaborado por Hao et al. (2010). Este abarca unos 8240 grados cuadrados del relevamiento SDSS-DR7 (York et al., 2000). En él se encuentran objetos con mediciones de redshift fotométrico en un intervalo de $0,1 \leq z \leq 0,55$ y determinaciones de redshift espectroscópico para una submuestra. Además para cada cúmulo se dispone de coordenadas ecuatoriales (α, δ), estimaciones de riquezas (escalada o ponderada), la fuerza de agrupamiento, la magnitud aparente petrosiana, la magnitud aparente models y los colores g-r y r-i. Los objetos del catálogo se identificaron a través de la secuencia roja de los cúmulos empleando los algoritmos de mezcla de gaussianas (véase sección 4.1.1). Posteriormente se expande el espacio de parámetros correlacionando el catálogo con los datos del DR16 de SDSS. Se obtienen como parámetros adicionales la dispersión de velocidades, los ángulos de posición y los cocientes de semiejes.

Como los datos poseen mediciones angulares en una región en particular del cielo se requiere la construcción de una máscara angular. Esta cumple la función de especificar que no se trabaja con toda la esfera celeste sino con una sección en particular. Para este fin se emplea la herramienta Helpix (Górski et al., 2005) que es un acrónimo de *Hierarchical Equal Area isoLatitude Pixelization of a sphere* (sección 4.1.2).

A partir del conjunto de datos obtenidos se construyen dos muestras. La primera es una muestra de referencia empleada para determinar la relación entre el redshift fotométrico y espectroscópico en un intervalo comparable. Se emplean los redshifts fotométricos para la muestra de BCG ya que no todos los objetos tienen determinado el redshifts espectroscópicos. Esto impide determinar con precisión las distancias y se utilizan la estimación del z_{foto} y se interpretan resultados de manera cualitativa considerando un error en $45 \frac{Mpc}{h}$.

La segunda muestra es el conjunto de datos principal que se empleó a lo largo del trabajo. En este caso el intervalo de redshift fotométrico se escoge en $0,1 \leq z_{foto} \leq 0,42$ para obtener una muestra completa en volumen. Luego, a partir del análisis de las propiedades de las BCGs, se descartan los objetos con valores muy extremos para cada distribución y se restringen los parámetros de cocientes de semiejes y de g-r para garantizar que los objetos de la muestra sean galaxias centrales de cúmulos, con morfología elíptica y con colores rojos. Finalmente queda determinada la muestra principal con un total de 21434 objetos, en un rango de $0,18 < z_{foto} < 0,42$ (Véase sección 4.1.3 para una explicación detallada del valor mínimo en el rango de redshift).

En cuanto a la herramienta que se emplea para el análisis de la muestra, proviene de la teoría de grafos y es conocida como Árbol Minimal de Expansión o MST por sus siglas en inglés. El MST se puede definir como un algoritmo de agrupamiento o clustering del cual se puede extraer información intrínseca de una distribución de puntos. Para el presente trabajo, los nodos son las posiciones tridimensionales de las galaxias del catálogo y las longitudes de borde las distancias tridimensionales entre cada par de nodos. Como el árbol es de expansión mínima se escoge el árbol cuya suma lineal de las longitudes de los bordes de sus componentes sea la mínima. Si no

hay dos longitudes de borde iguales, el MST será único. Para construirlo se emplea el paquete denominado *MisTree* creado por Naidoo (2019). El mismo se basa en la implementación del algoritmo de Kruskal (1956) optimizado para aplicaciones cosmológicas (sección 3.0.3). Para caracterizar un árbol se escogen los siguientes 3 parámetros (sección 3.0.3):

- Grado (d) - el número de aristas que tiene cada nodo. (ver Fig. 3.3)
- Longitud de las aristas (l) - la longitud de las aristas en el MST. (ver Fig. 3.4)
- Longitud de rama (b) - la suma de las longitudes de las aristas miembros. (ver Fig. 3.5)

El principal objetivo del trabajo es la utilización del análisis de árboles minimales de expansión, para cuantificar el impacto de las propiedades locales y globales de las galaxias centrales de cúmulos en la estructura presente en la red cósmica del universo. Teniendo en cuenta el efecto geométrico que introduce la presencia de una sección angular y un intervalo de redshift limitados sobre los estadísticos (véase Naidoo et al. (2020)) se emplea una muestra de control estadísticamente comparable (misma cantidad de puntos en igual volumen) con distribución homogénea (sección 5.1.3). Con ella se pretende evitar los sesgos geométricos en los resultados. Además, al usar una muestra aleatoria para la normalización, se pueden comparar entre sí los resultados provenientes de árboles con distinto número de nodos.

Con este fin se calcula la distribución de los cocientes para cada bin, de los estadísticos de la frecuencia de los datos respecto de la frecuencia en la muestra de control ($C_{Estadisticos} = \frac{F_{datos}}{F_{Randoms}}$). Una vez obtenido el cociente se analizan la señales de cada bin para valores distintos de $C_{Estadisticos} = 1$. El análisis de la distribución de los cocientes de estadísticos permite distinguir diferentes mecanismos físicos que inhiben o promueven una mayor cantidad de distancias cortas/largas o nodos con más o menos conexiones.

Para entender los resultados primero se realiza una interpretación sobre la distribución del cociente del estadístico l, siendo similar al análisis de la función de correlación de dos puntos (Véase sección 3.0.1) exceptuando que posee mayor información que la distancia entre pares por la naturaleza de su construcción (Véase Naidoo (2019) y Sección 3.0.3). En la figura 6.1 se presenta un ejemplo de interpretación de este parámetro. En ambos casos se presenta un árbol minimal de expansión para cada conjunto de datos. El exceso de pesos menores a $l \leq 3[Mpc]$ que se obtiene en la muestra de datos (panel inferior) corresponde al fenómeno de agrupación o (*clustering*) (el C_l será mayor a 1 para las distancias $l \leq 3[Mpc]$).

Por otra parte el estadístico b se corresponde a la suma de los pesos de las aristas que conforman las ramas en un árbol minimal de expansión (Véase secciones 3.0.3 y el capítulo 3). En la figura 6.2 se presentan dos paneles, con un análisis similar a lo explicado anteriormente. La distribución homogénea forma 2 ramas con valores $b = 5 [Mpc]$ y $b = 12 [Mpc]$ mientras que la muestra de datos también posee la misma cantidad de ramas pero con valores de $b = 12 [Mpc]$ en ambos casos. Se interpreta la diferencia entre los paneles como una orientación en gran escala hasta una cierta distancia. En este estadístico no sólo se tiene en cuenta la separación entre nodos, sino la concatenación entre ellos. Estos caminos o ramas pueden interpretarse como indicadores de la presencia de alineamientos en la estructura en gran escala.

En cuanto al cociente del estadístico d, en ningún caso se logra obtener alguna señal significativa. Una posible razón es la insensibilidad de este parámetro a la estructura presente en un único árbol o la imposibilidad de dejar desechar las aristas que cumplen la función de mantener al árbol conexo (véase capítulo 3)

Luego de realizar estas interpretaciones, se procede a realizar el estudio con la muestra principal. Para ello se confecciona una muestra de control adecuada, luego se aplica el algoritmo MST y posteriormente se realizan los cocientes entre las distribuciones de los estadísticos obtenidos. Los parámetros del algoritmo son especialmente sensibles a los patrones de la red cósmica (Naidoo, 2019). Los resultados obtenidos a partir de la figura 6.3 demuestran la potencialidad de esta herramienta. En ella se exhiben 3 paneles con la distribución de cada cociente de los

distintos estadísticos.

La distribución del cociente del estadístico d (C_d) de la figura 6.3, no presenta un comportamiento distinto respecto de la muestra de control. Por otro lado, en la distribución del cociente l (C_l) se observa un exceso en distancias cortas que se mantiene hasta unos $35 \text{ Mpc} \cdot h^{-1}$, luego el cociente se estabiliza alrededor de uno y posteriormente se genera el efecto de compensación obteniendo una escasez de distancias medias entre 55 y $75 \text{ Mpc} \cdot h^{-1}$. Los resultados obtenidos son consistentes con la longitud de correlación encontrada por la función de la auto-correlación de cúmulos ricos entre 20 y $40 \text{ Mpc} \cdot h^{-1}$ (Véase sección 3.0.1).

De manera análoga, el cociente relacionado al estadístico b (C_b) presenta excesos en ramas cortas en el rango $[0, 80] \text{ Mpc} \cdot h^{-1}$. Este resultado está en acuerdo con un Universo inhomogéneo en escalas menores a $80 \text{ Mpc} \cdot h^{-1}$.

A continuación se subdividen a la muestra principal en dos grupos según el valor de la mediana para cada parámetro. De manera similar a lo realizado para el conjunto principal en cada submuestra se obtiene una muestra de control apropiada y se emplea el algoritmo MST. Luego se comparan los estadísticos obteniendo la distribución de los cocientes. La primer propiedad que se analiza es la riqueza de los cúmulos definida en la sección 2.0.4. Este parámetro se emplea para referirse a cuan poblado son estas estructuras. Como la riqueza es una estimación de la masa de los cúmulos, estudiar los resultados obtenidos a partir de dos submuestras de bajas y altas riquezas es análogo a estudiar los cúmulos en función de su masa. Se espera que a mayor riqueza sea también más masivo.

A partir de las dos submuestras se obtienen la figuras 6.4a y 6.4b. El C_d no presenta señal en ninguno de los dos casos a diferencia de los C_l y C_b . Al Comparar ambas figuras, en la distribución del cociente de l se observa una mayor potencia para la submuestra de cúmulos ricos en los primeros bins (hasta $20 \text{ Mpc} \cdot h^{-1}$) que corresponde en ambos casos a excesos de aristas cortas respecto de la muestra de control. Esta mayor amplitud se interpreta como un mayor agrupamiento de cúmulos en escalas chicas/moderadas. Por otro lado, al analizar el intervalo donde se presenta el exceso en l , los cúmulos con menor riqueza poseen $C_l > 1$ en el intervalo $[0, 20] \text{ Mpc} \cdot h^{-1}$ mientras que los cúmulos ricos mantiene este efecto en una mayor distancia $[0, 40] \text{ Mpc} \cdot h^{-1}$. Los cúmulos ricos sostienen correlaciones en escalas más grandes que la submuestra de cúmulos con baja riqueza.

Por otro lado, el cociente del estadístico b en las dos submuestras revelan abundancias en las ramas entre $[0, 90] \text{ Mpc} \cdot h^{-1}$ respecto de su respectiva muestra de control. En este caso la mayor diferencia entre ambas figuras se encuentra en las distintas amplitudes para los primeros bins. Este resultado permite inferir que los cúmulos ricos tienen una mayor influencia en la estructura en gran escala comparado con la submuestra de menor riqueza. Se puede concluir que el comportamiento en gran escala observado en la muestra principal está dominado por las estructuras más masivas del Universo.

En segundo lugar, se analiza la riqueza ponderada. Se espera encontrar algo similar a la riqueza escalada (véase sección 2.0.4) ya que para algunos cúmulos es una mejor estimación de su riqueza real. Los resultados obtenidos están en concordancia con lo encontrado en el análisis de los cocientes para la segregación de riqueza escalada explicados anteriormente.

El tercer parámetro a analizar es $g-r$. Como la muestra principal consta de galaxias rojas, a partir de esta separación se obtienen dos conjunto de datos, donde probablemente aquellas con un índice muy rojo se corresponda a un gran contenido en polvo presente en la BCG respecto de las otras. Para la submuestra de las galaxias muy rojas se obtiene la figura 6.6a y para las galaxias rojas la figura 6.6b. La diferencias entre ambas puede observarse en la distribución del cociente del estadístico l . La amplitud para los primeros bins de distancia (escalas menores a $20 \text{ Mpc} \cdot h^{-1}$) es mayor en la submuestra de las galaxias menos enrojecidas. Los paneles correspondientes al cociente del estadístico b , presentan una coincidencia en valor para el primer bin, pero luego se diferencian. En el caso de las muy rojas (figura 6.6a) se estabiliza en valores cercanos a $C_b \approx 1$ mientras que las galaxias rojas continúan con excesos de ramas respecto de la muestra de control hasta $80 \text{ Mpc} \cdot h^{-1}$. Estas diferencias permiten concluir

que las galaxias rojas marcan la misma tendencia que se observa para la muestra total y para los cúmulos ricos mientras que las galaxias muy rojas presentan una menor señal en comparación. Probablemente como se menciono anteriormente, un índice de color elevado corresponde a la presencia de enrojecimiento interestelar más que intergaláctico. Al ser una propiedad local de las BCG existe la posibilidad que no afecte en gran medida a la estructura en gran escala.

Se puede resumir lo encontrado hasta el momento de la siguiente manera. La riqueza traza propiedades mas globales del cúmulo que el color de la BCG y por lo tanto es mayormente responsable de la estructura de la red cósmica.

El cuarto parámetro que se estudia es la dispersión de velocidades. Este parámetro es una medida de la varianza de la distribución de velocidad de las estrellas de la BCG (sección 2.0.4). Como las velocidades de las estrellas están determinadas por la masa gravitatoria en el interior de su orbita y la distancia a la que se encuentre desde el centro de masa de la galaxia, este parámetro correlaciona con la masa total de la BCG. Siguiendo el mismo procedimiento empleado en las secciones anteriores, se obtienen dos figuras 6.7a y 6.7b para las submuestra de alta dispersión y la submuestra de baja dispersión, respectivamente. Se espera que las galaxias con dispersión más alta posean una escala de correlación más grande que las galaxias con menos dispersión. La distribución del C_d son muy similares para ambos subconjuntos sin mostrar ninguna señal para los 5 bins. En cuanto a la distribución de C_l , para la submuestra de alta dispersión se encuentra una potencia mayor para el primer bin respecto de la submuestra de baja dispersión. El rango en el que se encuentran excesos en distancia de a pares es hasta $[0 - 35] \text{ Mpc} \cdot h^{-1}$ para ambas submuestras. A partir de estas observaciones, se puede concluir que la correlación entre galaxias centrales más masivas es mayor que para las galaxias menos masivas en función de la amplitud de los primeros binnes (Véase sección 3.0.1). La distribución de C_b de la submuestra de alta dispersión presenta excesos en las ramas de longitudes menores a $80 \text{ Mpc} \cdot h^{-1}$. Al comparar en este intervalo con la figura 6.7b se observa que posee una mayor amplitud. Este resultado indica una relevancia más importante de las galaxias con altas dispersiones en la estructura en gran escala. Sin embargo, para ramas con distancias mayores a $80 \text{ Mpc} \cdot h^{-1}$, las galaxias de alta dispersión disminuyen su señal y la mayor contribución en la identificación de estructuras en gran escala se deben a las galaxias con baja dispersión (excesos de ramas con valores menores a $110 \text{ Mpc} \cdot h^{-1}$). Cabe mencionar que al utilizar la muestra principal, no se visualiza excesos en ramas mayores a $80 \text{ Mpc} \cdot h^{-1}$, por lo tanto este efecto de inhomogeneidad respecto de la muestra de control queda compensado al incluir todos los nodos. Como conclusión y teniendo en cuenta la correlación masa y dispersión de velocidades, las galaxias centrales muy masivas están distribuidas más inhomogeneamente en el Universo que las no tan masivas. Este resultado es esperado debido a que el hecho que dos galaxias se aproximen corresponde un poco a la gravitación mutua, pero en mayor medida esta subyugado por los grandes flujos dominados por la estructura en gran escala.

El quinto parámetro de la muestra que se analiza son los semiejes proyectados de las galaxias. Como su nombre lo indica es la razón entre el semieje menor proyectado y el semieje mayor proyectado de la galaxia (Véase sección 4.1.4). Con esta propiedad se espera estudiar la correlación de la estructura en gran escala en función de la morfología proyectada de la galaxia central. para ello se separa la muestra en dos subconjuntos, uno con simetría proyectadas más esférica y otro simetrías proyectadas más elipsoidales. Se analizan los resultados en las figuras 6.8a (más esféricas) y 6.8b) (más elípticas) . En primer lugar, la distribución del cociente del estadístico d permite concluir que no se detecta ninguna diferencia aparente con la muestra de control. Tanto para el estadístico l como el estadístico b, se analiza el intervalo de excesos para los primeros bins y su amplitud. En ambos casos no se logran detectar diferencias significativas. Cabe señalar que a pesar de no existir comportamientos distintos entre ambas figuras, si presentan concordancia con lo encontrado para el árbol completo. Teniendo en cuenta estos resultados se proponen dos posibles escenarios por lo cual no se logran diferenciar efectos entre ambas submuestras:

i) La primer alternativa se debe a la mediciones de los semiejes proyectados. Este parámetro al ser una proyección puede no ser una medida fiable de una cantidad intrínseca del objeto. Debido a ello el algoritmo no esta separando dos grupos con propiedades distintas y por lo tanto no tienen un efecto considerable en la estructura en gran escala que sea diferente al analizado para la muestra completa.

ii) El segundo escenario propuesto es que el parámetro de semiejes verdaderos sea una propiedad local sin efecto sobre la distribución espacial en gran escala.

Luego de examinar el espacio de parámetros, se puede concluir que el Universo en gran escala esta dominado por cúmulos ricos, cuyas galaxias centrales son masivas y rojas. Estos resultados se obtuvieron a partir de la utilización de una herramienta simple pero sensible a la red cósmica como lo es el MST. Los resultados obtenidos se encuentran en acuerdo con los trabajos preexistentes de estructura en gran escala. Se pretende emplear el potencial de este algoritmo para estudiar más relaciones entre parámetros de las BCG y la estructura en gran escala (Véase sección 8).

Por ultimo, los resultados de los análisis presentados, sugieren que se puede hallar información relevante estudiando alineaciones relativas entre las BCG en relación con el árbol de expansión. Para ello se calcula la distribución de la diferencia entre ángulos de posición (sección 2.0.3) para dos nodos conectados por una arista empleando como conjunto de datos la muestra principal. El estudio de las alineaciones son realizadas en mayor medida entre galaxia y estructura en gran escala. Resulta de gran interés indagar este efecto entre galaxia-galaxia.

Como un nodo puede tener más de una arista conectada (grado mayor a 1) se analizan en primera instancia la orientación relativa entre nodos conectados mediante una arista. Para poder comparar los resultados encontrados, se construye una muestra de control donde las posiciones de las galaxias son las mismas que las de la muestra principal pero asignando una distribución aleatoria y uniforme de ángulos de posición. Así, ambas muestras tienen el mismo MST. Utilizando cada par de galaxias conectadas por una arista se calcula la orientación relativa mediante la resta de sus respectivos ángulos de posición. Cabe mencionar los ángulos calculados corresponden al ángulo agudo entre la orientación de ambas galaxias. Para identificar si existe o no una señal en los resultados, se propone calcular a cuantos sigmas respecto de la media de cada distribución se encuentra cada una de las frecuencias observadas empleando la ecuación 6.1. La muestra proveniente de ángulos aleatorios funciona como muestra de control. Esta se emplea para estimar la magnitud de la variación natural de una distribución uniforme de una cierta cantidad de objetos que se encuentran bineados con un dado tamaño. Se obtiene un límite inferior a partir del cual se puede considerar una detección de la señal empleando el cálculo del percentil 99 (P_{99}) de la distribución de σ para la muestra de ángulos aleatorios.

Con este análisis, se detecta la presencia de una señal de alineación relativa para las galaxias de la muestra principal en ($\theta_{rel} \approx 3^\circ$) y ($\theta_{rel} \approx 33^\circ$). A partir de estos resultados, y teniendo en cuenta el número de aristas que conectan cada galaxia con sus vecinas, se decide analizar los nodos en función de su grado. Para ello se construyen cinco submuestras (con $d = 1, \dots, 5$). Los resultados indican que la muestra $d = 2$ coincide con lo encontrado para la muestra principal concluyendo que la aparición de una señal proviene de los nodos con solo dos aristas. En la misma se visualiza una señal de orientación relativa hasta $\theta_{rel} \leq 6^\circ$, consistente con BCGs muy alineadas.

A continuación se buscó alguna dependencia de las orientaciones relativas de acuerdo al largo de las aristas que conectan las BCGs, pero no se encontró una dependencia clara con este parámetro. Por otra parte, la precisión en la medición de la orientación de una galaxia está relacionada con su cociente de semiejes. Con esta consideración se separa a la muestra principal de acuerdo a la mediana de la distribución de los cocientes de semiejes de las BCGs, y se procede a medir los ángulos de orientación relativos entre los nodos unidos mediante una arista. De los resultados se puede inferir que los valores con orientaciones 33° se deben en su

mayoría a la forma de las isofotas mientras que las orientaciones cercanas a 3° son regidas por los nodos con grado dos. El análisis presentado da indicios de la presencia de correlación entre las orientaciones de las galaxias, y puede considerarse como un primer paso para un posterior estudio de las alineaciones de las BCGs, tanto entre ellas como en relación a las aristas del MST, consideradas como trazadoras de la geometría de la estructura en gran escala.

Capítulo 8

Trabajo a futuro

A través del estudio realizado en este trabajo, se plantea la posibilidad de aplicar la misma técnica utilizando una muestra de galaxias BCG con mediciones de datos espectroscópicos. Se espera sin embargo que los resultados no difieran de los hallados.

En segundo lugar, se deja como trabajo a futuro el cotejar los resultados con otra muestra de control. Ejemplo de esto es lo realizado por [Naidoo et al. \(2020\)](#) donde efectúa comparaciones con un conjunto de datos construido de tal manera para que la función de correlación de dos puntos no genere señales distintas a la muestra de datos. En el estudio que realizó el autor, empleó el algoritmo MST y logró encontrar diferencias entre ambas distribuciones de puntos. Estos resultados le permitieron concluir que los estadísticos d , l , b y s son suficientemente sensibles como para capturar parte de la estadística de más alto orden. En el caso de este trabajo se pretende comparar los resultados observacionales con catálogos sintéticos tales como el de galaxias MDPL2-SAG [Cora et al. \(2019\)](#); [Knebe et al. \(2018\)](#) del grupo MultiDark ([Riebe et al. \(2013\)](#)); [Klypin et al. \(2016\)](#)).

Se planea utilizar dos o más propiedades (empleando las correlaciones encontradas) en conjunto para generar submuestras que permitan hallar características novedosas que aporten información a como están relacionadas las propiedades locales con la formación de la estructura en gran escala.

Actualmente, los arboles minimales de expansión siguen contribuyendo al descubrimiento del Universo tal como puede verse en el trabajo de [Naidoo et al. \(2021\)](#). En él combina el MST y el espectro de potencia para romper muchas de las degeneraciones individuales en la determinación de parámetros cosmológicos. Solo empleando los arboles de expansión se pueden encontrar restricciones más estrictas sobre la suma de las masas de neutrinos, la constante de Hubble h , la inclinación espectral n_s y la densidad de energía bariónica Ω_b .

Por ultimo en el análisis de orientaciones existen muchas posibilidades a explorar. En primer instancia como se desea encontrar alguna alineación entre nodos conectados por una arista, se pretende estudiar la correlación de esta orientación en función de la escala empleando el estadístico l (longitud de las arista). Además mediante la aplicación de triángulos esféricos se puede proyectar sobre la esfera celeste el ángulo agudo subtendido entre la galaxia y la arista que la une a otro nodo. Con esta técnica se podría investigar si los puntos conectados por el MST están preferencialmente en alguna orientación. Otra posible camino es medir los ángulos relativos entre galaxias para las distintas ramas y buscar posibles interpretaciones del estadístico b en función de estas conclusiones. Por ultimo, se proyecta realizar el análisis de orientaciones en simulaciones numéricas buscando obtener información más precisa sobre las alineaciones reales de las BCGs.

Referencias

- Abell G. O., 1958, , [3](#), [211](#)
- Akaike H., 1974, [IEEE Transactions on Automatic Control](#), 19, 716
- Bahcall N. A., Dong F., Hao L., Bode P., Annis J., Gunn J. E., Schneider D. P., 2003, , [599](#), [814](#)
- Barrow J. D., Bhavsar S. P., Sonoda D. H., 1985, , [216](#), [17](#)
- Bartelmann M., 1996, , [313](#), [697](#)
- Beuret M., Billot N., Cambrésy L., Eden D. J., Elia D., Molinari S., Pezzuto S., Schisano E., 2017, , [597](#), [A114](#)
- Borvka O., 1926
- Campana R., Massaro E., Gasparrini D., Cutini S., Tramacere A., 2008, , [383](#), [1166](#)
- Cartwright A., Whitworth A. P., 2009, , [392](#), [341](#)
- Collins C. A., et al., 2000, , [319](#), [939](#)
- Cora S. A., Hough T., Vega-Martínez C. A., Orsi Á. A., 2019, , [483](#), [1686](#)
- Croft R. A. C., Dalton G. B., Efstathiou G., Sutherland W. J., Maddox S. J., 1997, , [291](#), [305](#)
- Davis M., Peebles P. J. E., 1983, , [267](#), [465](#)
- Estrada J., Sefusatti E., Frieman J. A., 2009, , [692](#), [265](#)
- Fabian A. C., 1994, , [32](#), [277](#)
- Goodstein J., 1977, [Historia Mathematica - HIST MATH](#), 4
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, , [622](#), [759](#)
- Graham R., Hell P., 1985, [Annals of the History of Computing](#), 7, 43
- Hansen S. M., Sheldon E. S., Wechsler R. H., Koester B. P., 2009, , [699](#), [1333](#)
- Hao J., et al., 2009, , [702](#), [745](#)
- Hao J., et al., 2010, , [191](#), [254](#)
- Hogg D. W., Cohen J. G., Blandford R., 2000, , [545](#), [32](#)
- Hong T., Han J., Wen Z., Sun L., Zhan H., 2012, [The Astrophysical Journal](#), 749
- Johnston D. E., et al., 2007, arXiv e-prints, p. [arXiv:0709.1159](#)

- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, [Monthly Notices of the Royal Astronomical Society](#), 457, 4340
- Knebe A., et al., 2018, , [474](#), [5206](#)
- Koester B. P., et al., 2007a, , [660](#), [221](#)
- Koester B. P., et al., 2007b, , [660](#), [239](#)
- Kruskal J. B., 1956, Proceedings of the American Mathematical society, 7, 48
- Krzewina L. G., Saslaw W. C., 1996, , [278](#), [869](#)
- Lee S., Park C., 2002, [Journal of Korean Astronomical Society](#), 35, 111
- Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Harnden F. R. J., Primini F. A., Payne H. E., eds, Astronomical Society of the Pacific Conference Series Vol. 238, Astronomical Data Analysis Software and Systems X. p. 269 ([arXiv:astro-ph/0101420](#))
- Maddox S. J., Sutherland W. J., Efstathiou G., Loveday J., 1990, , [243](#), [692](#)
- Merritt D., 1985, , [289](#), [18](#)
- Mo H., Van den Bosch F., White S., 2010, Galaxy formation and evolution. Cambridge University Press
- Naidoo K., 2019, [The Journal of Open Source Software](#), 4, 1721
- Naidoo K., Whiteway L., Massara E., Gualdi D., Lahav O., Viel M., Gil-Marín H., Font-Ribera A., 2020, , [491](#), [1709](#)
- Naidoo K., Massara E., Lahav O., 2021, arXiv e-prints, p. [arXiv:2111.12088](#)
- Navarro J. F., Frenk C. S., White S. D. M., 1997, , [490](#), [493](#)
- Nelson D., et al., 2015, [Astronomy and Computing](#), 13, 12–37
- Ostriker J. P., Tremaine S. D., 1975, , [202](#), [L113](#)
- Pereyra L. A., Sgró M. A., Merchán M. E., Stasyszyn F. A., Paz D. J., 2020, , [499](#), [4876](#)
- Planck Collaboration et al., 2020, , [641](#), [A6](#)
- Prim R. C., 1957, [Bell System Technical Journal](#), 36, 1389
- Riebe K., et al., 2013, [Astronomische Nachrichten](#), 334, 691
- Ruiz A. N., Paz D. J., Lares M., Luparello H. E., Ceccarelli L., Lambas D. G., 2015, , [448](#), [1471](#)
- Schechter P., 1976, , [203](#), [297](#)
- Shectman S. A., Landy S. D., Oemler A., Tucker D. L., Lin H., Kirshner R. P., Schechter P. L., 1996, , [470](#), [172](#)
- Silk J., 1976, , [208](#), [646](#)
- Suzuki N., et al., 2012, , [746](#), [85](#)
- Szapudi I., Szalay A. S., 1998, , [494](#), [L41](#)
- Wen Z. L., Han J. L., Liu F. S., 2009, , [183](#), [197](#)

York D. G., et al., 2000, , [120](#), [1579](#)

da Costa L. N., et al., 1998, , [116](#), [1](#)

Los miembros del tribunal certifican que las correcciones fueron realizadas y el presente trabajo esta en condiciones de ser presentado.