



**UNIVERSIDADE FEDERAL DA BAHIA  
INSTITUTO DE FÍSICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA**

**Flávio de Jesus Costa**

**DETERMINAÇÃO DE RIQUEZA E DETECCÃO  
DE SUBESTRUTURAS EM AGLOMERADOS  
DE GALÁXIAS A PARTIR DE DADOS  
FOTOMÉTRICOS**

Salvador-BA

2018

**FLÁVIO DE JESUS COSTA**

**DETERMINAÇÃO DE RIQUEZA E DETECCÃO DE  
SUBESTRUTURAS EM AGLOMERADOS DE GALÁXIAS A  
PARTIR DE DADOS FOTOMÉTRICOS**

Tese apresentada ao Programa de Pesquisa e Pós-graduação em Física, Instituto de Física, Universidade Federal da Bahia, como requisito parcial para obtenção do grau de Doutor em Física.

Orientador: Prof. Dr. Cássio Bruno Magalhães Pigozzo

Coorientador: Prof. Dr. André Luís Batista Ribeiro

SALVADOR-BA  
2018

**FLÁVIO DE JESUS COSTA**

**DETERMINAÇÃO DE RIQUEZA E DETECCÃO DE  
SUBESTRUTURAS EM AGLOMERADOS DE  
GALÁXIAS A PARTIR DE DADOS FOTOMÉTRICOS**

Salvador-BA, 13/04/2018

Comissão Examinadora

---

**Prof. Dr. Cássio Bruno Magalhães Pigozzo**  
UFBA

---

**Prof. Dr. Humberto de Almeida Borges**  
UFBA

---

**Prof. Dra. Denise Nunes Viola**  
UFBA

---

**Prof. Dr. Reinaldo de Carvalho**  
INPE

---

**Prof. Dr. Roberto dos Santos Menezes**  
IFBA

*Dedico este trabalho as pessoas mais especiais da minha vida, à minha família.*

## AGRADECIMENTOS

Agradeço a Deus por mais uma etapa vencida! À minha família, em especial, aos meus pais Valnice e Everaldo, à minha vó Alice e ao meu irmão Fábio Costa por todo o amor e apoio incondicional que me dedicam. Não poderia esquecer (JAMAIS!) de meu tio Nego (em memória) pelos conselhos, pela alegria e, principalmente, por você ter participado de todas as etapas de minha vida. O título que recebo é teu! Como você faz falta! À minha namorada Rafaela Vieira pelo tempo roubado para elaboração deste trabalho. Aos professores IF-UFBA pelos bons conselhos. Aos colegas que tive o prazer de conviver, obrigado pelo incentivo. Em especial, ao meu orientador Cássio Pigozzo e ao meu coorientador André Ribeiro pelos belíssimos momentos intelectuais. Descobri que sou um João Ninguém perto de vocês! Agradeço aos colegas do IFBA, onde trabalhei, que souberam compreender minhas ausências e angústias no decorrer do doutorado, em especial, Eliseu Miranda, Fábio Amaral “Garotinho”, Jorge Dantas e minha coordenadora imediata Marlécia, vocês foram essenciais para que eu conseguisse terminar este ciclo, agradeço de coração. Não poderia esquecer da minha família soteropolitana: Vó Teté, Tia Vera, Tia Virgínia, Fernanda, Ana Paula, Bruno, Débora, Jéssica, Tia Loura (*my aunt*), agradeço por me adotarem como um verdadeiro filho/irmão e, por me suportarem durante minha estada nesta linda capital. Gente, o que dizer da primeira Travessa Lagoa da Vovó? Sobretudo, de suas inconstâncias! É pagode no fundo, é reggae no lado, é Olodum na frente, é Cláudia gritando e crente chamando pelo “sangue do cordeiro”. Enfim, uma superposição de ritmos, crenças, culturas e gostos. Por fim, agradeço aos professores do Laboratório de Astrofísica Teórico e Observacional-LATO por me receberem sempre de braços abertos. Agradeço também aos colegas mestres mais “frescos” da UESC (futuros doutores), Alisson e Allan, pelas tardes descontraídas, pelas dicas, pelas “pimentinhas”, pelas constantes ajudas. Vocês brocam! Pensou que iria esquecer de você, Ualace Lima? Você é o cara! Agradeço demais. Não há palavras para te agradecer, amigo! Presente em todos os momentos! Sou muito grato a você. “Enfim (...)” como diria o amigo Celso (professor do IFBA-Eunapolis) rs ...é isso! Por incrível que pareça, de alguma forma, tudo isso me fez crescer! De coração, meus sinceros agradecimentos!

*“If I have seen further than others,  
it is by standing upon the shoulders of giants.”*  
— SIR ISAAC NEWTON

## RESUMO

Aglomerados de galáxias são uma das maiores estruturas virializadas do universo. O conhecimento acerca de sua evolução, distribuição e de suas propriedades físicas é um tópico central em astrofísica e cosmologia. Para que um catálogo seja útil nessas áreas ele deve fornecer alguma estimativa de massa. Há uma grande discussão na literatura de como se obter a massa dessas estruturas, já que esta não é um observável direto. Neste sentido, faz-se uso das *proxies* de massa, isto é, “substitutos” ou “indicadores” de massa, sendo a riqueza uma possível candidata. A riqueza é uma medida de quantas galáxias se encontram dentro de uma dada abertura em torno do centro de um aglomerado. No presente trabalho, introduzimos uma metodologia para estimar um raio fotométrico, chamado  $R_{dens}$ , que esteja associado a uma escala física do sistema, o  $R_{200}$ . Dentro da abertura definida por  $R_{dens}$  é que se efetua a contagem de galáxias, obtendo sua riqueza,  $N_{dens}$ . A metodologia envolve o uso de um catálogo MOCK constituído por 947 aglomerados para duas finalidades: calibrar um perfil de densidade superficial de galáxias que é utilizado para se obter  $R_{dens}$ , como também os indicadores fotométricos do estado dinâmico dos aglomerados. Analisamos uma amostra composta por 850 aglomerados do SDSS-III fazendo uso das calibrações obtidas com o catálogo MOCK. Nosso estudo revela que é possível determinar  $R_{dens}$  e  $N_{dens}$  de forma robusta, conseguindo relações lineares entre  $N_{dens}$  e  $N_{200}$  com coeficientes de correlação  $> 0,9$ , para diversas amostras de galáxias retiradas dos dados. Além disso, obtivemos uma forte relação entre um indicador fotométrico, a magnitude absoluta da galáxia mais luminosa ( $M_1$ ), e um indicador baseado na forma da distribuição de velocidades das galáxias nos aglomerados, dado pela distância de Hellinger (HD). Mostramos ainda que  $M_1$  se correlaciona com o *gap* de magnitudes  $\Delta M_{14}$ , assim como o número de picos de densidade no campo de galáxias, suavizado por um *kernel* adaptativo. Estabelecemos então um critério, baseado nestes três indicadores fotométricos, para inferir o estado dinâmico de um aglomerado. Aplicado à nossa amostra, esta metodologia permitiu estimar que 22% dos sistemas analisados provavelmente são estruturas relaxadas.

**Palavras-Chave:** Aglomerados de galáxias, raio e riqueza de densidade, subestruturas.

# ABSTRACT

Clusters of galaxies are one of the largest virialized structures in the universe. The knowledge about its evolution, distribution and its physical properties is a central topic in astrophysics and cosmology. For a catalog to be useful in these areas it should provide some mass estimation. There is a great discussion in the literature on how to obtain the mass of these structures, since this is not a direct observable. In this sense, mass proxies are used, ie "substitutes" or "indicators" of mass, with richness being a potential candidate. Richness is a measure of how many galaxies are located within a given aperture around the center of a cluster. In the present work, we introduce a methodology to estimate a photometric radius, called  $R_{dens}$ , that is associated with a physical scale of the system, the  $R_{200}$ . Inside the opening defined by  $R_{dens}$  is that counting of galaxies, obtaining its richness,  $N_{dens}$ . The methodology involves the use of a MOCK catalog consisting of 947 clusters for two purposes. Initially, to calibrate a surface density profile of galaxies that is used to obtain  $R_{dens}$ . Then to calibrate photometric indicators of the dynamic state of the clusters. We analyzed a sample composed of 850 clusters of the SDSS-III using the calibrations obtained with the MOCK catalog. Our study reveals that it is possible to determine  $R_{dens}$  and  $N_{dens}$  in a robust way, obtaining linear relations between  $N_{dens}$  and  $N_{200}$  with correlation coefficients  $> 0.9$ , for several samples of galaxies taken from the data. In addition, we obtained a strong relation between a photometric indicator, the absolute magnitude of the galaxy most luminous ( $M_1$ ), and an indicator based on the shape of the velocity distribution of the galaxies in the clusters, given by Hellinger's distance (HD). We also show that  $M_1$  correlates with the magnitude gap  $\Delta M_{14}$ , as well as the number of peak densities in the field of galaxies, smoothed by an adaptive kernel. We then established a criterion, based on these three photometric indicators, to infer the dynamic state of a cluster. Applied to our sample, this methodology allowed us to estimate that 22% of systems are likely to be relaxed structures.

**Keywords:** Clusters of galaxies, radius and density richness, substructures.



## LISTA DE FIGURAS

Figura 3.1 – À esquerda: o telescópio principal do SDSS com 2,5 m de diâmetro. À direita: o telescópio fotométrico de 0,5 m essencial para calibrar o brilho de um objeto conforme medido com o telescópio principal (GUNN et al., 2006).	15
Figura 4.1 – À esquerda: distribuição de pontos aleatórios, no meio: agrupamento de pontos, à direita: pontos regularmente distribuídos. . . . .	20
Figura 4.2 – Estimador de intensidade de distribuição de pontos. . . . .	22
Figura 4.3 – Exemplo ilustrativo de distintos padrões espaciais. Caso A: agregação; Caso B: regular e Caso C: aleatoriedade espacial. . . . .	26
Figura 4.4 – Definições dos padrões espaciais através das contagens dos <i>quadrats</i> : agrupamento espacial (à esquerda), distribuição regular (meio) e aleatoriedade espacial (à direita). . . . .	27
Figura 4.5 – A técnica determina um campo mal centrado, sobre a área de domínio, o ponto vermelho, e o centro do <i>quadrat</i> central, o ponto azul. . . . .	29
Figura 4.6 – Distâncias das galáxias em relação ao pico de densidade do campo. . . . .	29
Figura 4.7 – Correspondência entre os raio $R_{200}$ e sua respectiva densidade superficial, $\Sigma_{200}$ .	30
Figura 4.8 – Obtenção do raio de densidade a partir do perfil de densidade superficial do aglomerado. . . . .	30
Figura 4.9 – Padrões sistemáticos e os respectivos valores da razão $I_{CE}$ . . . . .	34
Figura 4.10–Comportamento da função densidade de probabilidade em distintos valores do parâmetro de concentração $\kappa$ . . . . .	36
Figura 4.11–Em (a): seleção da região com aglomeramentos espaciais; (b) detecção, por meio de um <i>kernel</i> , das duas maiores densidades do campo; (c) somente os agrupamentos 01 e 02 tiveram densidade maior que limiar de densidade $k$ . . . . .	37
Figura 4.12–Kernel estimator atuando sobre os dados, cuja largura da banda pode ser definida pelo usuário como fixo ou adaptado. . . . .	38
Figura 5.1 – À esquerda: Aglomerado com dois grupos de galáxias. À direita: O método das parcelas identifica que o aglomerado é mal centrado. Veja também que aglomerados com essas características podem apresentar problemas de borda (regiões em amarelo). . . . .	40
Figura 5.2 – Efeito de borda evidenciado pelas galáxias além do limite do <i>frame</i> . A figura mostra um aglomerado “descentralizado” (pico de densidade fora do <i>quadrat</i> central) e, portanto, excluído pelo método. O ponto, em vermelho, representa o pico de densidade do aglomerado. . . . .	40
Figura 5.3 – Exemplo de exclusão do aglomerado 869 devido o seu centro não pertecer ao <i>quadrat</i> central. . . . .	41
Figura 5.4 – Campo centrado no <i>quadrat</i> central. . . . .	41

Figura 5.5 – Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após o método das parcelas. Nenhuma seleção adicional das galáxias é imposta. Pontos cheios são os sistemas utilizados após a remoção de <i>outliers</i> , enquanto os pontos vazios são os <i>outliers</i> . As linhas tracejadas indicam a região dentro de 95% de confiança em torno do ajuste. . . . .	43
Figura 5.6 – Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após método das parcelas com corte em magnitude absoluta $-20, 5$ . . . . .	44
Figura 5.7 – Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após corte em cor $u - r > 2, 22$ . . . . .	46
Figura 5.8 – Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após corte em $Mr \leq -18, 4$ e em cor $u - r > 2, 22$ . . . . .	47
Figura 5.9 – Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após corte em $Mr \leq -20, 5$ e em cor $u - r > 2, 22$ . . . . .	48
Figura 5.10–Correlação entre os raios $R_{dens}$ e $R_{200}$ com BCGs a 50 kpc dos picos de densidade. . . . .	50
Figura 5.11–Relação entre os raios com BCGs inseridas nos raios $R_{200}$ . . . . .	51
Figura 5.12–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ com BCGs a 50 kpc dos picos de densidade e corte em magnitude absoluta em $Mr \leq -18, 4$ . . . . .	52
Figura 5.13–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ com BCGs inseridas nos raios $R_{200}$ e corte em magnitude absoluta $-18, 4$ . . . . .	53
Figura 5.14–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ com BCGs a 50 kpc dos picos de densidade e corte em magnitude absoluta $-20, 5$ . . . . .	55
Figura 5.15–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ com BCGs inseridas nos raios $R_{200}$ e corte em magnitude absoluta $Mr \leq -20, 5$ . . . . .	56
Figura 5.16–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ com BCGs a 50 kpc dos picos de densidades e corte em $u - r > 2, 22$ . . . . .	58
Figura 5.17–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ com BCGs inseridas nos raios $R_{200}$ e corte em $u - r > 2, 22$ . . . . .	59
Figura 5.18–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após corte em $Mr \leq -18, 4$ ; cor $u - r > 2, 22$ e $d_{BCG} \leq 50$ kpc dos centros. . . . .	60
Figura 5.19–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após corte em $Mr \leq -18, 4$ ; cor $u - r > 2, 22$ e BCGs inseridas nos raios $R_{200}$ . . . . .	62
Figura 5.20–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após corte em $Mr \leq -20, 5$ ; cor $u - r > 2, 22$ e $d_{BCG} \leq 50$ kpc. . . . .	63
Figura 5.21–Correlação entre os raios $R_{dens}$ e os raios $R_{200}$ após cortes em $Mr \leq -20, 5$ ; cor $u - r > 2, 22$ e BCGs inseridas nos raios $R_{200}$ . . . . .	64
Figura 5.22–Correlações entre as contagens $N_{dens}$ e $N_{200}$ . . . . .	66
Figura 5.23–Correlações entre $N_{dens}$ e $M_{200}$ . . . . .	67
Figura 5.24–Histogramas de $R_{dens}$ e $N_{200}$ para a amostra tirada do catálogo de Wen e Han (2013). . . . .	69

Figura 5.25–Aplicação do programa <i>pdfCluster</i> ao aglomerado 39 da amostra. À esquerda: a distribuição de galáxias em coordenadas equatoriais. À direita: a identificação do pico de densidade do sistema indicado por um “X” e a determinação do raio de densidade representado pelo círculo verde. . . . .	70
Figura 5.26–Matriz de correlação entre os indicadores de subestruturas. . . . .	73
Figura 5.27–Relação entre $M_1$ e HD. G – gaussianos (vermelho), NG – não-gaussianos (azul) e IND (verde). As regressões para sistemas G e NG são mostradas em amarelo. . . . .	74
Figura 5.28–Relação entre $M_1$ e HD. G – gaussianos (vermelho), NG – não-gaussianos (azul) e IND (verde). As regressões para sistemas G e NG são mostradas em amarelo. . . . .	76
Figura 5.29–Histogramas das distribuições de $M_1$ e $\Delta M_{14}$ para as amostras MOCK e observada. . . . .	77
Figura 5.30–Campo de densidade do aglomerado 39 obtido por kernel adaptativo. Centro e picos de densidade estão indicados pela letra “X”. Um círculo contendo o raio de densidade é superposto à imagem. . . . .	78
Figura 5.31–Campo de densidade do aglomerado 53 obtido por kernel adaptativo. Centro e picos de densidade estão indicados pela letra X. Um círculo contendo o raio de densidade é superposto à imagem. . . . .	79
Figura 6.1 – A conexão entre a massa e riqueza, calculada com duas aberturas de raios: $R_{500}$ (à esquerda) e $R_{200}$ (à direita). Nos painéis inferiores os resíduos $\log(M_{obs}/M_{fit})$ são mostrados. $M_{obs}$ é o valor observado da massa, enquanto $M_{fit}$ é o resultado de regressão linear. . . . .	81
Figura 6.2 – Relação de escala entre a riqueza e a massa extraída do <i>PanStarrs 1+Euclid-like survey</i> . A linha sólida marca a linha de regressão ajustada aos dados observados. A região sombreada marca o 68% de confiabilidade para a regressão. A linha pontilhada vermelha indica a relação de entrada dos dados. . . . .	82
Figura 6.3 – Diagrama mostrando a sequência de passos para a determinação da riqueza de aglomerados. . . . .	84
Figura 6.4 – Diagrama mostrando a sequência de passos para inferir o estado dinâmico dos aglomerados. . . . .	85
Figura A.1 – Representação do catálogo de galáxia simulado. Cada caixa tem 137 Mpc ao longo do cubo. O observador localiza-se no ponto preto, no meio de um dos lados do quadrado do superbox (DUARTE; MAMON, 2014). . . . .	97

Figura A.2–Duas partículas são considerados amigos se o distância entre elas é inferior a uma distância limiar $\varepsilon$ : A e B, B e C, são amigos, mas A e C não são. Amigos de amigos é definida entre dois pontos: se eles são amigos ou se eles estão contidos numa área de interesse (por exemplo, A e C é um amigo de amigo via B). Todos os pontos negros estão em um <i>cluster</i> , enquanto os pontos brancos caracterizam outro <i>cluster</i> . . . . .	98
Figura A.3–Evolução da abundância de aglomerados com desvio para o vermelho. As previsões são exibidas como curvas contínuas como função de $\Omega_m(\sigma_8)$ (BAH-CALL; FAN, 1998). . . . .	99

## LISTA DE TABELAS

Tabela 4.1 – Em uma distribuição aleatória, o valor da razão assume o valor 1. Zero para uma condição de máxima agregação. Já para uma condição de máximo espaçamento (ponto equidistante de outros pontos), $I_{CE}$ assume o valor de 2, 1491. . . . .	34
Tabela 5.1 – Estatística do método das parcelas sobre os aglomerados analisados na amostra completa. . . . .	44
Tabela 5.2 – Estatística do método das parcelas sobre os aglomerados analisados com o corte em magnitude em $-20, 5$ . . . . .	45
Tabela 5.3 – Estatística do método das parcelas sobre os aglomerados analisados com corte em cor $u - r > 2, 22$ . . . . .	46
Tabela 5.4 – Estatística do método das parcelas sobre os aglomerados analisados com corte em $M_r \leq -18, 4$ e em cor $u - r > 2, 22$ . . . . .	47
Tabela 5.5 – Estatística do método das parcelas sobre os aglomerados analisados com corte em $M_r \leq -20, 5$ e em cor $u - r > 2, 22$ . . . . .	49
Tabela 5.6 – Estatística do método <b>BCGs</b> sobre os aglomerados analisados contendo BCGs a 50 kpc dos picos de densidade.. . . .	50
Tabela 5.7 – Estatística do método <b>BCGs</b> sobre os aglomerados analisados om BCGs dentro de $R_{200}$ . . . . .	51
Tabela 5.8 – Estatística do método <b>BCGs</b> sobre os 946 aglomerados analisados. . . . .	53
Tabela 5.9 – Estatística do método <b>BCGs</b> sobre os aglomerados analisados. . . . .	54
Tabela 5.10–Estatística do teste das <b>BCGs</b> sobre os 554 aglomerados analisados. . . . .	55
Tabela 5.11–Estatística do método <b>BCGs</b> sobre os aglomerados analisados com BCGs inseridas nos raios $R_{200}$ e corte em magnitude absoluta $M_r \leq -20.5$ . . . . .	57
Tabela 5.12–Estatística do método <b>BCG</b> sobre os aglomerados analisados com BCGs a 50 kpc dos picos de densidade e corte em cor $u - r > 2, 22$ . . . . .	58
Tabela 5.13–Estatística de método <b>BCG</b> sobre os aglomerados analisados com BCGs nos raios $R_{200}$ e corte em cor $u - r > 2.22$ . . . . .	60
Tabela 5.14–Estatística do método das parcelas sobre os 806 aglomerados analisados com corte em $M_r \leq -18, 4$ ; cor $u - r > 2, 22$ e $d_{BCG} \leq 50$ kpc. . . . .	61
Tabela 5.15–Estatística do método das parcelas sobre os 806 aglomerados analisados com corte em $M_r \leq -18, 4$ ; cor $u - r > 2, 22$ e BCGs inseridas nos raios $R_{200}$ . . . . .	62
Tabela 5.16–Estatística do método das parcelas sobre os 132 aglomerados analisados com corte em $M_r \leq -20, 5$ ; cor $u - r > 2, 22$ e $d_{BCG} \leq 50$ kpc. . . . .	64
Tabela 5.17–Estatística do método das parcelas sobre os aglomerados analisados com corte em $M_r \leq -20, 5$ ; cor $u - r > 2, 22$ e BCGs inseridas nos raios $R_{200}$ . . . . .	65
Tabela 5.18–Propriedades a partir do indicador HD . . . . .	71

Tabela 5.19–Propriedades a partir do indicador AD . . . . .	71
Tabela 5.20–Propriedades a partir do indicador DS . . . . .	72
Tabela 5.21–Propriedades a partir do indicador $M_1$ . . . . .	75

# SUMÁRIO

<b>Lista de figuras</b> . . . . .	<b>ix</b>
<b>Lista de tabelas</b> . . . . .	<b>xiii</b>
<b>1 – Introdução</b> . . . . .	<b>1</b>
<b>2 – Aglomerados de galáxias e suas propriedades físicas</b> . . . . .	<b>4</b>
2.1 Aglomerados de galáxias e suas propriedades físicas . . . . .	4
2.2 Perfis de Densidade . . . . .	6
2.2.1 Perfil de King . . . . .	7
2.2.2 Perfil de Navarro-Frenk-White (NFW) . . . . .	7
2.3 Determinação da massa dos aglomerados . . . . .	8
2.4 Relação massa-riqueza . . . . .	9
2.5 Cinemática em aglomerados . . . . .	11
2.6 Estrutura dos aglomerados . . . . .	12
<b>3 – Dados e Objetivos</b> . . . . .	<b>14</b>
3.1 Dados . . . . .	14
3.1.1 Catálogo Simulado . . . . .	14
3.1.2 Catálogo Observado . . . . .	14
3.1.3 <i>Sloan Digital Sky Survey</i> (SDSS) . . . . .	15
3.2 Objetivos . . . . .	16
<b>4 – Metodologia</b> . . . . .	<b>17</b>
4.1 Descrição geral . . . . .	17
4.2 <b>O ambiente R</b> . . . . .	19
4.2.1 <b>Spatstat</b> . . . . .	19
4.3 Processo Espacial . . . . .	20
4.3.1 Estimador de Intensidade . . . . .	21
4.3.2 Processo de Poisson . . . . .	22
4.3.3 Processo de Poisson não-homogêneo . . . . .	24
4.4 Método <i>Quadrats</i> (ou Método das parcelas) . . . . .	24
4.4.1 Método das parcelas em Aglomerados de Galáxias . . . . .	28
4.5 Método do Raio de densidade ( $R_{dens}$ ) . . . . .	29
4.6 Análise de subestruturas: métodos espectroscópicos . . . . .	31
4.6.1 Teste de Anderson-Darling (AD) . . . . .	31
4.6.2 A distância de Hellinger (HD) . . . . .	32

4.6.3	O teste de Dressler-Schectman (DS)	32
4.7	Análise de subestruturas: métodos fotométricos	33
4.7.1	Posição da BCG e <i>gap</i> de magnitude	33
4.7.2	Índice de Clark e Evans	34
4.7.3	Distribuição de Von Mises	35
4.7.4	Programa <i>pdfCluster</i>	36
<b>5</b>	<b>– Resultados das Análises</b>	<b>39</b>
5.1	Método das parcelas e o catálogo simulado	39
5.1.1	<i>Quadrats</i> e Efeito de borda	39
5.1.2	Raios de densidade ( $R_{dens}$ ): Apresentação geral e aplicação nos catálogos simulados	42
5.2	Critérios de seleção de galáxias aplicados ao catálogo simulado	42
5.2.1	Conclusão sobre os testes realizados	65
5.3	Estimando a riqueza	66
5.3.1	Aplicação da metodologia sobre a amostra observada	68
5.4	Identificação de subestruturas nos catálogos simulado e real	70
5.4.1	Diagnósticos de HD, AD e DS sobre o catálogo simulado	70
5.4.2	Calibração do método fotométrico	73
5.4.3	Aplicação do método $M_1 - \Delta_{M14} - N_{picos}$ aos dados observados	76
5.4.3.1	Aglomerado 39	77
5.4.3.2	Aglomerado 53	79
<b>6</b>	<b>– Discussão Final</b>	<b>80</b>
6.1	Contexto da tese	80
6.2	Síntese	83
6.2.1	Estimando riqueza	83
6.2.2	Estimando o estado dinâmico	83
6.3	Conclusões e perspectivas	85
<b>Referências</b>		<b>87</b>
<b>Apêndices</b>		<b>95</b>
<b>APÊNDICE A – Apêndice</b>		<b>96</b>
A.1	Extração de catálogos de grupos de galáxias	96
A.1.1	<b>Friends-of-Friends (FoF)</b>	97
A.1.2	Evolução na Densidade Numérica de Aglomerados	98
A.1.3	Função de massa	99



A.1.4	Função de massa e Aglomerados de galáxias . . . . .	103
-------	---	-----

# 1 INTRODUÇÃO

Aglomerados de Galáxias (AGs) são as maiores estruturas gravitacionalmente ligadas do universo observável. São, portanto, excelentes traçadores da distribuição de matéria em grande escala. Eles também são essenciais no estudo de formação e evolução de galáxias, assim como podem ser utilizados para a determinação de parâmetros cosmológicos fundamentais como a constante de Hubble  $H_0$ , a densidade de matéria  $\Omega_m$ , a normalização da flutuação de densidade  $\sigma_8$  e o parâmetro da equação de estado da energia escura  $\omega$ . Essas quantidades podem ser medidas a partir de alguns métodos: razão massa-luminosidade, fração de bárions em aglomerados de galáxias, estudo de processos de emissão (raios-X) e absorção (efeito Sunyaev-Zel'Dovich) no gás intra-aglomerado, entre outros (White et al. (1993b) e Evrard (1997)). White et al. (1993a) e Liddle et al. (1996) preveem, por exemplo, uma relação degenerada entre  $\Omega_m$  e  $\sigma_8$  (a determinação únivoca de um parâmetro depende de uma determinação precisa do outro). Outro aspecto importante relacionado a aglomerados é o uso da sua função de massa para discriminar modelos cosmológicos (RAHMAN; SHANDARIN, 2001). Tal função estima, para um dado volume e num certo intervalo de *redshift*, a quantidade de objetos num certo intervalo de massa e tem importância fundamental para a astrofísica e para cosmologia em virtude de sua dependência com ingredientes da teoria de formação de estruturas (espectro de potência, fator de crescimento e o contraste de densidade crítico). Isto faz com que a contagem de aglomerados, ou seja, a determinação de sua abundância, seja um dos temas mais estudados atualmente.

Entre as várias questões relativas aos AGs, encontra-se uma realmente básica: como medir a massa de aglomerados de galáxias? Dos métodos mais utilizados, encontram-se estudos baseados em: lentes gravitacionais, raios-X, estimativa de massa dinâmica e *mass proxies*, sendo o último baseado em relações de escala existentes entre observáveis diretos (a riqueza) e a massa do aglomerado, (ver Rozo et al. (2011) e Andreon e Bergé (2012)). Um dos problemas iniciais neste estudo é o fato de que a massa de aglomerados não configura um observável e, portanto, uma alternativa é a busca de outras quantidades que melhor traduzem a estimativa da função de massa. Assim, observáveis que possuem uma forte correlação com a massa (e cujas dispersões e degenerescências com parâmetros cosmológicos sejam razoavelmente compreendidas) são altamente desejáveis por permitirem maior precisão ao mensurar a função de massa e devem ter maior eficiência quando usar o produto disso para estimar parâmetros, por consistência (ROZO et al., 2009c).

Teoricamente, espera-se que a abundância de halos maciços seja exponencialmente sensível à amplitude de flutuações de matéria (PRESS; SCHECHTER, 1974). Esta base teórica fora prevista e confirmada por simulações numéricas, após uma cuidadosa calibração de abundância de halos, em função da massa, para várias cosmologias (ver, por exemplo, Evrard et al. (2002) e Warren et al. (2006)). No entanto, a relação de indicadores observáveis de massa com a massa

real dos halos de matéria introduz incertezas sistemáticas nas análises, sendo normalmente a principal fonte de erro nesses estudos (por exemplo, Henry et al. (2009)).

No presente trabalho, pretendemos estudar dois pré-requisitos importantes para o estudo de AGs no que se refere à obtenção de suas massas. O primeiro deles se refere à determinação da riqueza de aglomerados, um estimador que, em geral, é dado pela contagem de galáxias numa certa região central do sistema. Se conhecemos a riqueza de um conjunto de aglomerados, podemos ajustar uma relação massa-riqueza que, se bem calibrada para uma amostra onde conhecemos as massas, pode ser usada como um indicador de massa para sistemas onde apenas a riqueza pode ser obtida (vide (ROZO et al., 2009a) para uma ampla discussão do assunto). Entretanto, a obtenção de riqueza e sua relação com a massa enfrenta uma série de dificuldades. Uma delas relaciona-se com a crescente aquisição de dados fotométricos para os quais a determinação do raio de contagem, assim como dos membros físicos (gravitacionalmente ligados) dos aglomerados é altamente incerta (vide (ANDREON; HURN, 2010) e (BAXTER et al., 2016)). Isto estabelece um problema na determinação da riqueza que pode dificultar ou mesmo inviabilizar a calibração da relação massa-riqueza. Outro problema está relacionado aos possíveis vieses no cálculo da massa de aglomerados quando esses sistemas não estão em equilíbrio (e portanto não obedecem ao teorema do virial) e também não seguem distribuições regulares em seus espaços de fase projetados (vide (RIBEIRO et al., 2010) e (RIBEIRO et al., 2011)). Sistemas com distribuição de velocidades não-gaussiana e/ou com a presença de subestruturas podem afetar o uso de diversos estimadores de massa e aumentar o espalhamento da relação massa-riqueza ((LOPES et al., 2009); (ANDREON; HURN, 2010); (RIBEIRO et al., 2013) e (CARVALHO et al., 2017)).

Esses dois problemas são investigados nesta tese, fazendo-se uso de um catálogo simulado (*Mock*) e também um conjunto de dados reais, observados. Nossos objetivos são dois: (i) encontrar um raio de abertura fotométrico que tenha sentido físico e, dentro dele, realizar a contagem de galáxias que serão usadas na determinação da riqueza do sistema; e (ii) encontrar a melhor combinação de indicadores fotométricos de subestruturas em aglomerados a partir de indicadores espectroscópicos. Na primeira etapa do trabalho introduziremos alguns métodos que determinam os raios de contagem e as riquezas estatísticas: o método *quadrats* (ou método das parcelas), e a obtenção de um raio característico dos sistemas a partir das densidades superficiais dos aglomerados. Após essa etapa, utilizaremos o programa *pdfCluster* (e alguns outros indicadores auxiliares) a fim de identificar possíveis subestruturas nas regiões internas aos raios de densidade. Esse indicador será comparado com alguns métodos tradicionais empregados para determinar se um aglomerado está ou não em equilíbrio: teste de Dressler-Schectman (DRESSLER; SHECTMAN, 1988); o teste de Anderson-Darling ((HOU et al., 2009), ((HOU et al., 2009) e (HOU et al., 2012)) e a distância de Hellinger ((RIBEIRO et al., 2013) e (CARVALHO et al., 2017)).

A tese tem a seguinte organização: no Capítulo 2 faremos uma breve revisão sobre

---

Aglomerados de galáxias; no Capítulo 3 apresentaremos os dados utilizados, bem como os objetivos específicos deste trabalho; no Capítulo 4 descreveremos a metodologia estatística; no Capítulo 5 apresentaremos os resultados da metodologia estatística quando aplicada sobre os catálogos simulado e real; no Capítulo 6 apresentaremos as conclusões e considerações finais, bem como, as perspectivas deste trabalho.

## 2 AGLOMERADOS DE GALÁXIAS E SUAS PROPRIEDADES FÍSICAS

Nesse capítulo apresentaremos como os aglomerados de galáxias se distribuem no universo e quais as suas principais características físicas.

### 2.1 Aglomerados de galáxias e suas propriedades físicas

Um dos grandes objetivos em astrofísica e em cosmologia é verificar como se comporta o universo em larga escala e como as galáxias estão distribuídas. O estudo sobre a distribuição de galáxias revela que esses objetos, em sua maioria, estão reunidos entre si, interagindo gravitacionalmente formando **Aglomerados de Galáxias**. Portanto, estes sistemas são excelentes traçadores da estrutura em grande escala (tipicamente da ordem de bilhões de anos-luz). Em um universo dominado, recentemente, por uma constante cosmológica e matéria escura fria (no contexto do modelo *Lambda Cold Dark Matter*- $\Lambda$ CDM<sup>1</sup>), halos de matéria escura se formam através de instabilidade gravitacional a partir de flutuações quânticas primordiais no campo de densidade de matéria. A amplitude dessas flutuações aumenta conforme as mesmas deixam de expandir com o fluxo de Hubble, colapsando e virializando, formando estruturas densas e relaxadas<sup>2</sup>. Essas estruturas formarão estruturas maiores, através de fusões, até originarem aglomerados de galáxias e ainda aglomerados de aglomerados de galáxias (os superaglomerados)<sup>3</sup> em um estágio final.

Quanto às suas propriedades, os aglomerados podem ser classificados segundo as seguintes características:

- **Riqueza:** número de membros num determinado raio;
- **Forma:** regular (esférica, achatada) ou irregular;
- **Conteúdo de galáxias:** rico em espirais, pobre em espirais, rico em elípticas.

---

<sup>1</sup>Representa o modelo de concordância da teoria do “Big Bang” que explica as observações cósmicas realizadas sobretudo a partir da Radiação Cósmica de Fundo em Microondas (RCFM).

<sup>2</sup>Consiste no estado dinâmico onde se estabelece o equilíbrio hidrostático do sistema. Processos de relaxamento naturalmente levam à formação de núcleos físicos (King (1972); Lynden-Bell (1967) e Peebles (1993)). Para Colless e Dunn (1995) aglomerados relaxados devem ter uma distribuição gaussiana ao longo do *redshift* para galáxias membros.

<sup>3</sup>O astrônomo francês Gérard de Vaucouleurs em 1953 demonstrou, pela primeira vez, que aglomerados de galáxias podem se agrupar e formar estruturas ainda maiores denominadas **superaglomerados de galáxias**. A existência dessas estruturas revela que a distribuição de galáxias no universo não é uniforme (DICKINSON, 1998). No entanto, a formação e definição dos superaglomerados são, ainda, motivos de debate na literatura, ver por exemplo Hamilton (1998).

Quanto à sua riqueza, isto é, o número de galáxias inseridas em algum raio dos AGs, podem ser classificados como:

- **Aglomerados ricos:** quando há várias centenas ou até milhares de galáxias.
- **Aglomerados pobres:** quando há entre algumas dezenas até algumas centenas de galáxias.

No caso de haver mais que três e até cerca de uma dezena de galáxias, estas associações são chamadas de **grupos de galáxias**, cuja massa virial é tipicamente de  $10^{12}$  a  $10^{13} M_{\odot}$ , com raio próximo a 1 Mpc e dispersão de velocidades, em média, de  $300 \text{ km s}^{-1}$  na linha de visada.

O estudo sobre a distribuição de galáxias em aglomerados é um tema bastante investigado, sobretudo à dificuldade de se estabelecer uma clara fronteira geométrica para essas estruturas. Neste sentido, seu tamanho não é uma propriedade definida de maneira unívoca, havendo inúmeras técnicas e definições quando se diz respeito as suas dimensões espaciais. Abaixo, apresentamos algumas definições para raios de aglomerados:

- **Raio Gravitacional ( $R_G$ ):** é o raio no qual a energia cinética das galáxias aproximadamente se iguala a energia gravitacional.

$$R_G \equiv \frac{2GM}{3\sigma_v^2} \equiv 1.5 \left( \frac{M}{5 \times 10^{14} M_{\odot}} \right) \times \left( \frac{\sigma_v}{10^3 \text{ km/s}} \right)^{-2} h^{-1} \text{ Mpc}, \quad (2.1)$$

onde  $M$  é a massa do aglomerado e  $\sigma_v$  é a sua dispersão de velocidade, observada na direção radial.

- **Raio Central ou do caroço ( $R_c$ ):** é o raio no qual a densidade de galáxias é igual à metade da densidade central ( $\rho(R_c) = \rho_0/2$ ). Um estudo realizado por Bahcall (1975), determinando a distribuição projetada e o centro de 15 aglomerados regulares ricos, revelou que a estimativa deste raio fica em torno de

$$R_c = (0,25 \pm 0,04) h^{-1} (\text{Mpc}), \quad (2.2)$$

onde  $h$  é uma normalização da constante de Hubble, denominada de constante de Hubble reduzida<sup>4</sup>.

- **Raio Harmônico ( $R_H$ ):** conhecido também como raio do virial projetado, é o raio que escreve a separação média projetada das galáxias (GIRARDI et al., 1994)

$$R_H = \frac{N(N-1)/2}{\sum_i \sum_{j>i} R_{ij}^{-1}}, \quad (2.3)$$

onde  $N$  é o número de galáxias membros e  $R_{ij}$  é a distância projetada entre os membros  $i$  e  $j$ .

<sup>4</sup>veja sua definição no apêndice, seção A.1.9, especificadamente equação A.45.

- **Raio 200** ( $R_{200}$ ): é uma outra classe de estimativas de raio que baseia-se em sobredensidades de modelos esféricos. Nesses modelos, os aglomerados deverão ser virializados em regiões onde a média da densidade de massa seja superior à densidade crítica do universo por um fator  $\Delta \sim 200$ . O raio em que esta densidade é atingida,  $R_{\Delta=200}$ , é usado como raio característico do aglomerado. Como exemplo, segue a definição dada por Carlberg et al. (1997)

$$R_{200} = \frac{\sqrt{3}\sigma_v}{10H(z)}, \quad (2.4)$$

onde  $\sigma_v$  é a dispersão de velocidade e  $H(z)$  é a constante de Hubble no *redshift*  $z$ .

- **Raio do Halo** ( $R_h$ ): Conhecida como o raio de *cutoff* (ou raio de truncamento) de uma esfera isotérmica. Na prática é determinado via uma função ajustada ao perfil de densidade. Em geral, sua aproximação:  $R_h \cong R_c$

Os AGs podem conter desde algumas dezenas de galáxias até centenas de galáxias num raio típico de  $1.5h^{-1}$  Mpc, com raio central  $R_c \sim (0.1 - 0.25)h^{-1}$  Mpc. Nos aglomerados, as galáxias possuem movimentos peculiares aleatórios da ordem de  $\sim 750\text{km/s}$  dentro de  $1.5h^{-1}$  Mpc, isto é, a dispersão de velocidade ao longo da linha de visada. Assim, a massa dinâmica obtida a partir do teorema do virial considerando uma esfera de raio  $1.5h^{-1}$  Mpc, tem a ordem de  $(0.1 - 2) \times 10^{15}h^{-1}M_\odot$ . Contudo, a luminosidade na banda B é tipicamente  $(0.6 - 6) \times 10^{12}h^{-1}L_\odot$ . Portanto, a razão massa-luminosidade típica de aglomerados no universo local é

$$(M/L_B)_{AGs} = (0.16 - 0.33) \times 10^3 h \frac{M_\odot}{L_\odot} \quad (2.5)$$

$$(M/L_B)_{AGs} \sim 300h \frac{M_\odot}{L_\odot} \quad (2.6)$$

## 2.2 Perfis de Densidade

Os perfis de densidade produzem informações a respeito da distribuição radial de galáxias nos aglomerados. Tais informações podem ser relacionadas às teorias de formação e evolução de aglomerados. Em geral, o perfil é descrito numa função com 3 parâmetros:

$$\rho(r) = \rho_0 f_1(r, R_c, R_h) \quad (2.7)$$

$$\sigma(r) = \sigma_0 f_2(r, R_c, R_h), \quad (2.8)$$

onde  $\rho(r)$  e  $\sigma(r)$  são, respectivamente, os perfis espaciais e projetados (de número, massa ou brilho);  $f_1$  e  $f_2$  são as funções que melhor se ajustam aos perfis.

Quanto aos parâmetros ajustáveis:

- densidade central  $\rho_0$  e  $\sigma_0$ ;
- uma escala central de comprimento  $R_c$ ;
- uma medida para os limites do aglomerado  $R_h$  (raio do halo).

Distintas formas funcionais foram propostas para descrever os perfis. Abaixo, descreveremos os perfis de **King** e de **Navarro-Frenk-White (NFW)**.

### 2.2.1 Perfil de King

King (1966) propôs o perfil que recebe seu nome. Tal modelo basea-se na distribuição de luminosidade do aglomerado projetada no plano do céu. É um dos perfis mais usados na literatura. Sua origem se deu em estudos na distribuição do brilho em aglomerados globulares como também em galáxias elípticas. Esse perfil funciona muito bem quando aplicado ao centro dos aglomerados de galáxias. Seu perfil tem a forma

$$\rho = \rho_0(1 + (r/R_c)^2)^{-\frac{3}{2}}, \quad (2.9)$$

onde  $R_c$  é o raio central.

### 2.2.2 Perfil de Navarro-Frenk-White (NFW)

Navarro et al. (1997) propuseram um perfil que recebe seus nomes a fim de ajustar os dados de simulações de  $N$  corpos seguindo o modelo *Cold Dark Matter* (CDM)<sup>5</sup>. O perfil é dado por:

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}, \quad (2.10)$$

onde  $\rho_s$  e  $r_s$  são parâmetros de escala. Esse perfil representa muitas vezes o melhor ajuste do modelo aos dados observados de aglomerados de galáxias, como também a não inclusão dos efeitos de larga escala. O perfil de NFW tem uma expressão analítica para o perfil de densidade superficial  $\Sigma(R)$  dada por Bartelmann (1996)

$$\Sigma(R) = 2\rho_s r_s F(R/r_s), \quad (2.11)$$

onde

$$F(X) = \begin{cases} \frac{1}{X^2} \left( 1 - \frac{2}{\sqrt{1-X^2}} \operatorname{arctanh} \sqrt{\frac{1-X}{1+X}} \right), & (X < 1) \\ \frac{1}{3}, & (X = 1) \\ \frac{1}{X^2} \left( 1 - \frac{2}{\sqrt{X^2-1}} \operatorname{arctan} \sqrt{\frac{X-1}{X+1}} \right), & (X > 1). \end{cases} \quad (2.12)$$

<sup>5</sup>Descrevemos alguns modelos no Apêndice. Veja as seções A.1.10 e A.1.11



O perfil NFW tem uma divergência, não física, na origem, variando com  $r^{-1}$  nas regiões do interior e com  $r^{-3}$  nas exteriores, o que implica uma outra propriedade irrealista de uma massa total infinita. Uma maneira de contornar esta divergência é truncar o perfil, em um raio máximo, por exemplo, o raio do virial. Uma generalização comum do perfil NFW é obtida por um ajuste da inclinação como um parâmetro livre  $\alpha$  (com  $\alpha = +1$ ) por Jing e Suto (2000)

$$\rho(r) = \frac{\rho_s}{(r/r_s)^\alpha (1 + r/r_s)^{3-\alpha}}. \quad (2.13)$$

## 2.3 Determinação da massa dos aglomerados

Para a cosmologia o parâmetro mais fundamental dos AGs é sua massa. São três os principais componentes de massa de AGs: as galáxias propriamente ditas, que juntas formam menos de 5% da massa total do aglomerado, o meio intra-aglomerado (MIA), composto basicamente pelo gás difuso, colapsado e com alta temperatura (ao redor de 15% da massa total), e matéria escura, que ocupa cerca de 80% da matéria existente no aglomerado e é, conseqüentemente, a principal responsável pelo confinamento gravitacional de todos os componentes. O MIA que ocupa o espaço entre as galáxias é quente ( $\sim 10^7$  K), difuso e altamente ionizado. Este gás emite fortemente em raio-X por efeitos de radiação Bremsstrahlung<sup>6</sup>. Através da análise espectral de raios-X, pode-se determinar a temperatura e a metalicidade do aglomerado. Como o tempo de virialização do aglomerado é significativamente menor que a idade do universo, tradicionalmente se assumia, para cálculo da massa dos aglomerados, que eles já teriam tido tempo de atingir o equilíbrio hidrostático. Com essa suposição, as variações radiais da temperatura e densidade do gás intraglomerado nos fornece o perfil radial de massa diretamente. Por outro lado, há a hipótese de que o equilíbrio hidrostático só é válido se os termos relacionados a variações espacial e temporal têm a velocidade do gás nula. Nesse caso, o único componente de pressão exercida é o térmico. Sabe-se hoje, através de observações do meio intergalático em AGs, que os aglomerados de galáxias são dinamicamente ativos e frequentemente saem do equilíbrio e da situação de simetria esférica que os atuais modelos de massa assumem. A própria formação dos aglomerados induz o afastamento da situação de equilíbrio. O modelo de formação de estruturas do universo é o chamado bottom-up. Nele, os AGs são formados gravitacionalmente pela coalescência de sistemas menores. Esse processo leva à criação de subestruturas em temperatura, densidade e velocidade do gás ((EVARD et al., 1995); (KATZ; WHITE, 1993); (NOVIKOV et al., 1999); (ZAROUBI et al., 1998) e (RICKER; SARAZIN, 2001)), que podem ser observados em raios-X (e.g. (DUPKE; BREGMAN, 2006) e (MARKEVITCH et al., 2004)). Logo, a hipótese de equilíbrio hidrostático não é completamente válida. É fundamental saber até que ponto pode-se aplicá-la para determinação da massa dos aglomerados. De maneira semelhante, a determinação

<sup>6</sup>O termo Bremsstrahlung tem origem alemã e significa “radiação de freiamento”, emitida por cargas em movimento sujeito à ação de um campo elétrico. A emissão dessa radiação, num contexto astrofísico, está ligado a colisões coulombianas de elétrons em plasmas quentes. Durante as colisões com os íons no plasma, os elétrons emitem radiação.

da massa de aglomerados via lentes gravitacionais também tem incertezas sistemáticas relacionadas às assimetrias da distribuição da massa. Por exemplo, distribuições triaxiais com o eixo maior na linha de visada aumentaria artificialmente a massa projetada, já que a massa obtida por lentes gravitacionais é a massa projetada.

A distribuição de galáxias e como esta distribuição evolui com o tempo são testes importantes de modelos cosmológicos: por exemplo, diferentes modelos predizem números diferentes de aglomerados de galáxias para diferentes *redshifts*. Daí, o interesse da cosmologia no estudo de aglomerados de galáxias, a fim de desvendar enigmas que ainda prevalecem neste “laboratório” de processos físicos. Isto é, podemos utilizar aglomerados para determinação de parâmetros cosmológicos, tais como,  $\Omega_m$  (baseado na razão massa-luminosidade, e na fração de bárions) e  $\sigma_8$  (flutuação de massa medida em esferas de raio  $8h^{-1}$  Mpc). Esse estudo influi diretamente nas distribuições e evolução desses sistemas, as quais constituem o objetivo fundamental da cosmologia moderna. Existem alguns métodos independentes para estimar a massa dos aglomerados de galáxias:

- Estimativa de massa dinâmica;
- Raio-X;
- Método da cáustica;
- Lentes gravitacionais;
- Substitutos de massa (*Mass proxies*).

O presente trabalho se dedicará a problemas associados ao último método.

## 2.4 Relação massa-riqueza

A propriedade dos aglomerados de galáxias mais fácil de se observar pelas *surveys* óticas é a riqueza, isto é, um número que caracteriza a quantidade de galáxias membros. Contudo, existem diversas maneiras de se definir a riqueza de um aglomerado seguindo distintos critérios. Por exemplo, Yang et al. (2008) calculam a densidade do espaço de fase para cada aglomerado e atribuem galáxias acima de um valor crítico (*threshold*) a cada aglomerado-candidato. A riqueza, neste caso, é o número de galáxias contabilizado, respeitando o *threshold*. Uma outra maneira de se definir riqueza estabelece a contagem de galáxias dentro da sequência vermelha, para uma certa abertura, subtraindo alguma estimativa de contaminação pelo campo. Neste método há uma dependência de informação espectroscópica (GLADDERS; YEE, 2005). Outro método possível leva em conta apenas a contagem de galáxias dentro de um raio estatístico definido a partir de contraste entre o aglomerados e uma distribuição de Poisson não-homogênea (COSTA, 2011).

Um problema relacionado com medidas de riqueza deve-se ao fato de que as galáxias são contadas dentro de um raio que precisa ser previamente definido, algumas vezes chamado de raio de abertura (LOPES et al., 2006). O ideal seria o raio estar de alguma forma relacionado ao estado dinâmico do sistema, mas isto exigiria dados adicionais sobre o aglomerado, o que tornaria menos vantajoso o uso de um *proxy* de massa. Para evitar este problema, pode-se definir um raio de abertura fixo, independente da massa. A otimização da escolha deste raio representa também uma dificuldade, pois tem dependência com o centróide do aglomerado e com o *redshift* (LOPES et al., 2006). Por esta razão, a riqueza de aglomerados algumas vezes é comparada com simulações em que a massa dos aglomerados é medida dentro um raio comóvel<sup>7</sup> (BODE et al., 2001). Contudo, a calibração da relação massa-riqueza depende neste caso de uma forma paramétrica, análoga à função de Jenkins (JENKINS et al., 2001), dando a função de massa definida para um raio fixo. Uma vez que não existe uma forma paramétrica simples para essa função, não é claro como fazer uma boa auto-calibração da relação massa-riqueza, mesmo dispondo de grandes quantidades de dados.

Obter a riqueza como *mass proxy* exige um conhecimento adicional: informações sobre  $R_{200}$  ou algum outro raio de referência, utilizado para fixar uma escala de contagem de galáxias (galáxias membros). Há autores que discordam de tal raciocínio (ver (LOPES et al., 2009)). Estimar a riqueza nessa escala de referência requer saber como essas escalas são medidas. Há alguns meios para tal

- **Dispersão da massa**

Sabermos  $\sigma_m$ , ou seja, a dispersão intrínseca da massa em dados de riqueza. Esse pode ser de interesse para quem quer conhecer a parte da dispersão observada.

- **Incerteza da massa**

A incerteza da massa estimada a partir da riqueza do aglomerado. Esta é, por exemplo, o caso quando se tem um ou alguns aglomerados com uma medida de riqueza, a fim de estimar sua(s) massa(s).

---

<sup>7</sup>Na cosmologia padrão, considerando o espaço em expansão, a distância é uma grandeza dinâmica que altera-se no tempo. A distância comóvel é definida numa curva de conexão entre dois pontos (métrica) como uma curva de tempo cosmológico. As distâncias comóveis não podem ser medidas por um observador com as limitações da Terra. Essas distâncias (objetos distantes) são determinadas por meio da luminosidade de velas padrão (objetos cuja luminosidade intrínseca é conhecida) ou pelo *redshift* de galáxias distantes convertendo-se em medidas baseadas em algum modelo particular de espaço-tempo.

## 2.5 Cinemática em aglomerados

Considerando um aglomerado com  $N$  galáxias de massa  $m_i$  e posição  $x_i$ . Defini-se a energia cinética total do sistema como

$$K = \frac{1}{2} \sum_i^N m_i |\dot{x}_i|^2 = \frac{1}{2} M \langle v^2 \rangle, \quad (2.14)$$

sendo a massa total  $M = \sum_i^N m_i$  e a velocidade quadrática média das galáxias em um aglomerado  $\langle v^2 \rangle \equiv \frac{1}{M} \sum_i^N m_i |\dot{x}_i|^2$ .

Já a energia potencial do sistema

$$U = -\frac{G}{2} \sum_{i,j} \frac{m_i m_j}{|\vec{x}_j - \vec{x}_i|}, \quad (2.15)$$

também definimos o momento de inércia

$$I \equiv \sum_i^N m_i |\vec{x}_i|^2, \quad (2.16)$$

obtendo a derivada de segunda ordem em relação ao tempo

$$\ddot{I} = 2 \sum_i^N m_i (x_i \ddot{x}_i + \dot{x}_i \dot{x}_i), \quad (2.17)$$

substituindo as equações das energias cinética e potencial em 2.17, obtemos o **Teorema do Virial**

$$\ddot{I} = 2U + 4K. \quad (2.18)$$

Para um sistema estacionário ( $I = cte$ ), tem

$$K = -\frac{U}{2}. \quad (2.19)$$

Ainda assim, a energia potencial de um aglomerado pode ser escrita como

$$U = -\alpha \frac{GM^2}{R_H}, \quad (2.20)$$

onde  $\alpha$  é um fator que depende do tipo de perfil de densidade do aglomerado e  $R_h$  é o raio harmônico. Agora, combinando a energia potencial com o Teorema do Virial, obtemos:

$$\frac{1}{2} M \langle v^2 \rangle = \frac{\alpha}{2} \frac{GM^2}{R_H}, \quad (2.21)$$

onde a massa do virial ( $M_V$ )

$$M_V = \frac{\langle v^2 \rangle R_H}{\alpha G}. \quad (2.22)$$

## 2.6 Estrutura dos aglomerados

Em um cenário hierárquico, as maiores estruturas do universo se formam por meio de fusões de estruturas menores. Nesse contexto, os aglomerados de galáxias acabam por apresentar regiões com propriedades dinâmicas distintas quando comparadas ao aglomerado como um todo, sobretudo, nas zonas mais periféricas. Estas regiões são chamadas de **subestruturas**. A existência de subestruturas em aglomerados sugere que estas estruturas são objetos ainda jovens, cujo cenário de formação aponta para grandes fusões recentes.

Nessa perspectiva, nos aglomerados ricos, com picos de densidade central, o percentual de subestruturas, em regiões internas, é cada vez menor. Este fato é confirmado no trabalho de Solanes et al. (1998), em que apenas 10% de uma amostra de 67 aglomerados, confirma-se a presença de subestruturas nas regiões internas (em  $R_{200}$ ) aumentando sua frequência em regiões mais periféricas. Da mesma forma, (AGUERRI; SANCHEZ-JANSSEN, 2010) também encontram subestruturas em regiões inferiores a  $R_{200}$ , com percentual de 11% para galáxias mais brilhantes considerando a magnitude absoluta  $M_r < -20$  e aumentando para 33% em magnitude absoluta  $M_r < -19$ .

A detecção de subestruturas em aglomerados pode ser realizado mediante o conhecimento de algumas propriedades físicas. Dentre eles: a distribuição de velocidades (teste 1D); as posições das galáxias no aglomerado (teste 2D), embora podendo sofrer efeitos de projeção. O teste 3D é a combinação das anteriores.

Os testes 1D pressupõem que as velocidades das galáxias num aglomerado relaxado devam ter uma distribuição normal. Estudos recentes indicam o poder estatístico de testes como o de Anderson-Darling (HOU et al., 2009) e a distância de Hellinger ((RIBEIRO et al., 2013) e (CARVALHO et al., 2017)). Estes testes indicam que 20 – 40% dos AGs possuem distribuições de velocidades com desvios significativos da normalidade.

Um exemplo para testes 2D é descrito por (WEST et al., 1988), conhecido como teste  $\beta$  onde mede a assimetria das distribuição das galáxias no aglomerado em relação ao seu centro, sendo uma subestrutura vista como uma assimetria localizada. O teste  $\beta$  estima a distância média entre uma galáxia  $i$  e 5 galáxias vizinhas, como também calcula a densidade local em relação a um ponto  $O$  diametralmente oposto à galáxia  $i$  e, por fim, calcula o parâmetro  $\beta$

$$\beta = \log_{10} \frac{d_i}{d_O}, \quad (2.23)$$

sendo  $d_i$  é a distância média para galáxia em relação as 5 galáxias vizinhas e  $d_O$  é a distância média entre as 5 galáxias vizinhas no ponto  $O$ . Caso o valor médio de  $\beta$ , para todas as galáxias tender a 0, a distribuição de galáxias é simétrica.

Uma outra técnica utilizada por (DRESSLER; SHECTMAN, 1988) para identificar subestruturas é o teste- $\Delta$  ou teste Delta de DS. O teste procura regiões que tenham maiores valores de velocidade (local),  $v_l$ , e dispersão de velocidades,  $\sigma_l$ , que os valores medidos no

aglomerado como todo. O estudo Pinkney et al. (1996), comparando diversos métodos para a determinação de subestruturas, sugere que o teste DS é aquele com maior grau de confiabilidade. Uma descrição mais detalhada deste método é apresentada no Capítulo 4.

## 3 DADOS E OBJETIVOS

### 3.1 Dados

A seguir apresentaremos os dois conjuntos de dados utilizados essencialmente para o teste da metodologia estatísticas (*Quadrats* e *PdfCluster*), bem como para os indicadores espectrocópicos<sup>1</sup> e fotométricos<sup>2</sup>: Catálogo simulado e catálogo real.

#### 3.1.1 Catálogo Simulado

A amostra é constituída pelos 947 aglomerados de galáxias dos catálogos simulados utilizados no trabalho de Duarte e Mamon (2014). Estes catálogos foram construídos a partir de saídas de simulações de galáxias reais do *Millennium-II* (BOYLAN-KOLCHIN et al., 2009) e a seleção de grupos de galáxias com base no *algoritmo Friends-of-Friends FoF* (ver (GARDNER et al., 2007a) e (GARDNER et al., 2007b)). Os dados são compostos de posições espaciais reais, velocidades comóveis, massas e luminosidades na banda  $r$ , com perfis de densidade reais e, principalmente, obedecendo às relações de escala observadas.

#### 3.1.2 Catálogo Observado

Utilizamos também os 132.684 aglomerados de galáxias identificadas por (WEN et al., 2012) utilizando *redshift* fotométrico do *Sloan Digital Sky Survey III* (SDSS-III). A identificação dos aglomerados se deu numa faixa de *redshift*  $0.05 \leq z < 0.8$ , com incerteza de  $\sim 0,025 - 0,030$ , para  $z < 0,45$ , tornando-se maior em *redshifts* maiores. A seleção de galáxias membro desse catálogo, se deu com relação à incerteza do *redshifts* fotométricos, na forma  $\sigma_z = \sigma_0(1+z)$  para todas as galáxias. Os objetos com grandes erros fotométricos a partir de  $0,08(1+z)$ , isto é, cerca de  $3\sigma_z$ , foram removidos, justificados pela má fotometria ou contaminação de estrelas. Este procedimento removeu 20% de objetos, a maioria objetos menos luminosos (magnitude absoluta de  $r > 21$ ). A seleção das *Bright Central Galaxies* (BCGs<sup>3</sup>) foi feita para erros nos *redshifts* fotométricos ( $z_{phot}$ ) tais que  $z_{err} \leq 0.055(1+z)$ , considerando objetos com elptiçidades maiores que 0,7 na banda  $r$  (WEN et al., 2012). As elpsidades foram obtidas através de um *cross-match* com o banco de dados do KIAS (Korea Institute for Advanced Study Value-Added Galaxy Catalog, baseado no SDSS-DR7). Os dados de  $z_{phot}$  e magnitudes absolutas  $M_r$ , correções  $K$  são obtidas da Tabela *Photoz*, sendo que os  $z_{phot}$  são estimados pelo método de Csabai et al. (2007). As magnitudes absolutas obtidas com  $z_{phot}$  foram corrigidas para evolução

---

<sup>1</sup>Descritos na seção 4.6

<sup>2</sup>Descritos na seção 4.7

<sup>3</sup>Descrita na seção 4.7.1

e obtidas como:  $M_r^{corr} = M_r(z) + Qz$ , onde  $M_r$  é magnitude aparente na banda  $r$  de Petrosyan, e  $Q$  é o fator de evolução, definido como  $Q = 1, 62$ , seguindo Blanton et al. (2003).

### 3.1.3 Sloan Digital Sky Survey (SDSS)

É um dos levantamentos mais ambiciosos da história astronômica. Este projeto está mapeando, com detalhes, um quarto de todo o céu, determinando as posições e brilhos absolutos de centenas de milhares de objetos celestes, além de distâncias de mais de um milhão de galáxias e quasares. O projeto conta com três operações, sendo que duas já concluídas (SDSS-I, 2000-2005; SDSS-II, 2005-2008) e a terceira (SDSS- III, 2008-2014). Os resultados do SDSS estão eletronicamente disponíveis à comunidade científica e ao público em geral, tanto na forma de imagens quanto em catálogos precisos de todos os objetos descobertos. Seus telescópios estão localizados no *Apache Point Observatory* (APO) em Sunspot, Novo México, Estados Unidos. São eles: um telescópio principal com lentes de 2,5 m de diâmetro com a capacidade de produzir imagens com foco sobre um grande campo de visão (feita, especialmente, para mapear objetos mais distantes e de menor brilho) e um telescópio fotométrico de 0,5 m com a função de monitorar pequenas mudanças na temperatura e na pressão da atmosfera ao longo das observações (monitor de *seeing*) (ver figura 3.1).

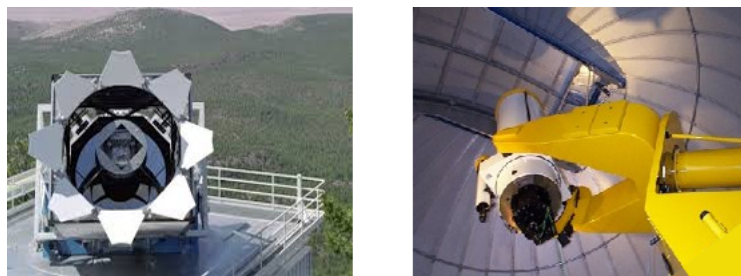


Figura 3.1 – À esquerda: o telescópio principal do SDSS com 2,5 m de diâmetro. À direita: o telescópio fotométrico de 0,5 m essencial para calibrar o brilho de um objeto conforme medido com o telescópio principal (GUNN et al., 2006).

O *Sloan* utiliza modernos detectores digitais para sua coleta de dados. A câmara de imageamento utilizada é composta por dois mosaicos de CCDs, sendo um de imageamento e outro destinado para a astrometria. O mosaico de imageamento consiste em 30 CCDs, cada um com resolução de 2048 por 2048 pixels (elementos de imagem), formando um arranjo de seis colunas e cinco linhas. Assim, cada linha do mosaico detecta um intervalo de comprimento de onda diferente, numa das cinco bandas ópticas,  $u, g, r, i, z$ . Além disso, o SDSS é dotado de um par de espectrógrafos alimentados por fibras ópticas a fim de coletar espectros. O telescópio é capaz de gravar 640 espectros ao mesmo tempo. Utilizando placas de alumínio no plano focal com orifícios nas posições dos objetos. Cada espectro é medido de 3800 (azul) até 9200 (infravermelho). Por noite cerca de seis a nove placas são usadas para gravar espectros. Toda noite



o telescópio produz cerca de 200 GB de dados. Além disso, o projeto também conta com uma estrutura totalmente automatizada para as operações de levantamento e processamento de dados. As imagens, espectros e os catálogos são fornecidos pelo arquivo de dados DAS (*Data Archive Server*), cujo objetivo é dispor os arquivos produzidos pelo processamento de pipelines de dados do SDSS. Acessando o banco de dados de imagens disponível no projeto, o SDSS apresentou imageamento de área total da DAS correspondendo a 45 mil graus quadrados (1,3 milhões de frames/filtro), um volume de dados de imagens de 15,7 TB e razoáveis limites de magnitudes em cinco filtros com precisão de calibração em torno de 2–3%. Já espectroscopicamente, apresentou uma área total de 9380 graus quadrados, um volume de dados (com calibração de espectros) de 3,3 TB e com limites de magnitude petrosiana para amostras de galáxias. Com base nestes dados, o SDSS parece ser mais completo fotometricamente do que espectroscopicamente. Atualmente o projeto encontra-se em sua décima terceira liberação de dados DR13 (Thirteenth Data Release) que inclui cinco tipos de dados: imagens, espectros ópticos, espectros de infravermelhos, espectros IFU e dados de catálogos. Esta versão de dados do SDSS apresenta melhorias nas calibrações fotométricas e no método de redução espectral oferecendo maiores extensões em dados fotométricos (vide ALBARETI et al. (2017)).

## 3.2 Objetivos

A seguir, apresentamos os dois objetivos fundamentais que conduzirão nossos estudos. A partir de dados fotométricos, esperamos:

- Determinar raios e riquezas representativos dos aglomerados.
- Identificar subestruturas dentro dos raios representativos, e classificar os sistemas como regulares (relaxados) ou irregulares (não-relaxados).

Ao final do trabalho, discutiremos os principais resultados, o potencial dos métodos empregados, assim como possíveis melhorias e desdobramentos futuros.

## 4 METODOLOGIA

### 4.1 Descrição geral

Neste trabalho, temos dois objetivos centrais. O primeiro deles refere-se à determinação de uma região, em torno do centróide de um aglomerado, que seja representativa de uma escala física do sistema, por exemplo, um múltiplo do seu raio virial. Essa tarefa não é simples se os dados disponíveis são apenas fotométricos, o que implica em não sabermos ao certo que galáxias são membros de um dado aglomerado. Usualmente, diferentes seleções de objetos são utilizadas quando não temos uma correta identificação dos membros. Objetos mais vermelhos e mais luminosos em geral são os escolhidos para tentar separar o aglomerado do seu *background*

<sup>1</sup> Cada uma dessas seleções de galáxias, de um lado melhoram a pureza do campo, por outro, podem introduzir outros vieses tanto astrofísicos como estatísticos, ambos ligados a variabilidade dos dados. A alternativa de tentar levar em conta todos os objetos até o limite de detecção fotométrica é menos sensível à variabilidade dos dados, mas precisa lidar com maior impureza nos dados.

Independentemente de como se escolha os objetos que serão levados em conta numa análise fotométrica, seria desejável estabelecer um método objetivo para estimar a riqueza dos aglomerados. Se determinarmos uma região que se correlacione com a região virializada em tamanho (ou raio) podemos efetuar uma contagem de galáxias em seu interior, que pode ser utilizada como uma estimativa da riqueza do sistema.

O segundo objetivo consiste em, dada a região central determinada através de um raio, definimos um indicador das possíveis subestruturas na região. Ou seja, nossos propósitos são: (i) determinar riqueza; e (ii) investigar subestruturas em dados fotométricos de aglomerados de galáxias.

Os métodos empregados neste trabalho se dividem da seguinte maneira. Na primeira fase, realizamos os seguintes passos.

1. Inicialmente, aplicamos o método de parcelas, cujo objetivo é eliminar campos que não se distinguem de um processo espacial aleatório e/ou encontram-se mal centrados. Ao final do processo, permanecem na amostra os campos que passam à etapa seguinte.
2. Método de determinação do raio via densidade superficial. Nesta etapa usamos a informação de  $R_{200}$  vinda do catálogo com espectroscopia e membros determinados pelo FoF. Para cada aglomerado definimos sua densidade superficial fotométrica equivalente a  $R_{200}$ , que chamamos de  $\Sigma_{200}$ . Isto é feito para toda a amostra, estabelecendo uma relação entre

---

<sup>1</sup>vide Popesso et al. (2004) para uma discussão sobre o problema.

as duas variáveis. A relação é usada em seguida para, dada a  $\Sigma_{200}$ , obtemos um raio fotométrico, que chamamos de  $R_{dens}$  ou raio de densidade <sup>2</sup>. O programa que realiza este procedimento foi desenvolvido neste trabalho em linguagem *R*. São testadas várias seleções de galáxias dos campos, em termos de suas magnitudes absolutas, cores, presença ou não de BCGs próximas ao centróide de cada campo.

3. Para cada seleção de galáxias, fazemos uma validação comparando os raios de densidade obtidos com os  $R_{200}$  do catálogo simulado (MOCK). Note-se que, inicialmente, o objetivo é calibrar o método, para posteriormente aplicá-lo sobre a amostra observada. A validação é feita através de regressões lineares e coeficientes de correlação entre os raios estimado e esperado. A melhor regressão indica qual a seleção de galáxias produz melhores resultados para o método.
4. Em seguida, usamos o método calibrado para determinar os raios de densidade para o catálogo de aglomerados com a disponibilidade apenas de dados fotométricos.

Na segunda fase do trabalho, desejamos estudar a estrutura dos aglomerados cujos raios foram determinados. Seguimos então os seguintes passos:

1. Mais uma vez utilizamos o catálogo MOCK, inicialmente utilizando a informação dos membros via FoF. Selecionamos membros dentro de  $R_{200}$  e realizamos um conjunto de diagnósticos sobre a regularidade e possível condição de equilíbrio do sistema. São aplicados sobre as amostras o teste de Anderson-Darling (AD) sobre a distribuição de velocidades, o cálculo da distância de Hellinger (HD), que também mede desvios da normalidade, e o teste de Dressler-Schectman para a presença de subestruturas (DS). Esses diagnósticos são chamados aqui de espectroscópicos, uma vez que requerem o conhecimento das velocidades radiais das galáxias. Adicionamos a essa lista um conjunto de diagnósticos fotométricos: a magnitude absoluta da BCG ( $M_1$ ), ou seja, as galáxias mais brilhantes do aglomerado; os *gaps* de magnitude  $\Delta M_{12} = M_2 - M_1$  e  $\Delta M_{14} = M_4 - M_1$ , respectivamente, a diferença de magnitude entre a segunda e primeira (assim como a quarta e a primeira) galáxia mais brilhante do aglomerado; a distância e a velocidade peculiar da BCG ao centro do aglomerado, respectivamente,  $D_{BCG}$  e  $V_{BCG}$ , o índice de agregação de Clark-Evans ( $I_{CE}$ ), o coeficiente de concentração da distribuição de von Mises (ou índice de agregação circular -  $I_{AC}$ ), e o número de picos de densidade da distribuição suavizada via kernel adaptativo.
2. Correlacionamos então os testes para ver que combinações revelam melhor concordância entre diagnósticos espectroscópicos e fotométricos. Isto permite uma escolha de quais ferramentas serão usadas nos dados observados para determinar a presença de subestruturas em aglomerados quando dispomos apenas de dados fotométricos.

---

<sup>2</sup>O raio é melhor descrito na subseção 5.1.2

3. Finalmente, para a escolha efetuada, fazemos a análise da amostra observada.

## 4.2 O ambiente R

A metodologia descrita acima foi aplicada fazendo-se uso do ambiente  $R^3$ . O R é um conjunto integrado de recursos de *software* para manipulação de dados, cálculo, estatística e exibição gráfica. Ele inclui:

- tratamento eficaz de dados e instalação de armazenamento;
- um conjunto de operadores para cálculos em matrizes;
- uma grande coleção coerente e integrada de ferramentas para análise dos dados;
- facilidades gráficas para análise de dados e exibir em tela ou em cópia impressa;
- uma linguagem de programação bem desenvolvida, simples e eficaz que inclui condicionais, *loops*, funções definidas pelo usuário e facilidades de entrada e saída.

O ambiente é projetado em torno de uma linguagem de computação que permite aos usuários adicionarem funcionalidades definindo novas funções. A linguagem R é semelhante a outras linguagens de programação, como C++ e Java, o que ajuda ao usuário executar uma ampla variedade de tarefas de computação ao lhes fornecer acesso a diversos comandos inclusive orientação a objetos. Além disso, sua distribuição é gratuita e vem na forma de programa de fonte aberta.

Neste trabalho usamos diversas funções do pacote *Spatstat* que contém ferramentas desenvolvidas para estatística espacial.

### 4.2.1 Spatstat

*Spatstat* (BADDELEY A., 2008), é um dos vários pacotes da linguagem R que serve para analisar padrões de pontos em duas dimensões. O pacote *Spatstat* é de autoria de Adrian Baddeley e Rolf Turner e possui distribuição como fonte aberta nas condições da licença de grande público (*General Public License-GNU*). O pacote suporta

- criação e manipulação de padrão de pontos;
- análise exploratória de dados;
- simulação de modelos de processo pontual;

---

<sup>3</sup>R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

- modelo de ajuste paramétrico;
- testes de hipóteses e diagnósticos.

O *Spatstat* usa os recursos orientados a objeto do *R* (classes e métodos) para facilitar, manipular, analisar os conjuntos de dados, e gerar gráficos de qualidade. As funções disponíveis no *Spatstat*, para tratar padrões de pontos, são ideais para a implementação do método das parcelas (vide Seção 4.6), bem como para a preparação de dados que serão posteriormente analisados em outro pacote, o *pdfCluster* também orientado ao estudo de distribuições espaciais. Além destes, utilizaremos ainda o pacote *circular*, para efetuar uma análise da simetria dos sistemas estudados, e os pacotes *distrEx*, que efetua o cálculo da distância de Hellinger, e *nortest*, que realiza o teste de Anderson-Darling. As regressões lineares são realizadas fazendo-se uso dos pacotes *MASS*, *car* e *robustbase*.

### 4.3 Processo Espacial

Neste trabalho, usamos algumas técnicas da estatística espacial, que considera uma distribuição de pontos em um plano como um processo espacial. Portanto, cada campo centrado em um aglomerado de galáxias será tratado como tal processo. Antes de apresentar os métodos específicos que serão usados, fazemos uma breve revisão de conceitos gerais a respeito de processos espaciais, que são usados explícita ou implicitamente em nossas ferramentas de análise.

Processos espaciais são fenômenos expressos através de ocorrências identificadas como pontos localizados no espaço. Uma das características mais importantes de um padrão espacial é a localização dos pontos e a relação entre eles. Um dos objetivos mais comuns da análise de um padrão espacial é verificar se os eventos observados apresentam algum tipo de comportamento sistemático, ao invés de estarem distribuídos aleatoriamente, isto é, verificar para um dado padrão observado se é aleatório ou se apresenta em aglomerados, ou ainda, se os pontos estão regularmente distribuídos.

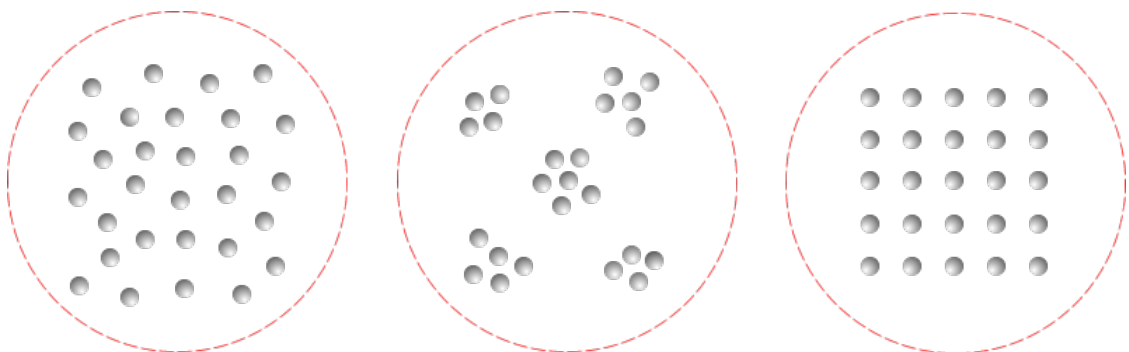


Figura 4.1 – À esquerda: distribuição de pontos aleatórios, no meio: agrupamento de pontos, à direita: pontos regularmente distribuídos.

Um processo espacial consiste em uma série de coordenadas de pontos, como exemplo,  $(P_1, P_2, \dots)$  dos eventos de interesse dentro da área de estudo. Neste contexto, o termo **evento** refere-se a qualquer tipo de fenômeno localizável no espaço que, dentro de uma referida escala de investigação, possa estar associado a uma representação pontual. Assim, dentro de uma distribuição de pontos busca-se detectar a existência de algum padrão espacial, por exemplo, aglomerações em relação a um processo de Poisson. Sob a visão estatística, os processos espaciais são gerados por um mecanismo estocástico, isto é, um processo cujo comportamento é não-determinístico, no sentido em que cada estado desse processo não determina completamente qual será o seu estado seguinte. Tal mecanismo pode ser descrito em termos de efeito de primeira e segunda ordem. Os efeitos chamados de primeira ordem são aqueles considerados globais ou de larga escala. Eles descrevem a variação do valor médio do processo. Neste caso, o foco de maior interesse está no número de eventos por área. Já os efeitos de segunda ordem são aqueles considerados locais ou de pequena escala. Estes representam a dependência espacial no processo associada ao relacionamento entre pares de eventos no espaço, por unidade de área. Consideremos um conjunto de pontos  $(P_1, P_2, \dots)$  em certa região  $A$  onde ocorrem os eventos. A modelagem do processo espacial é considerada quando se obtêm subregiões  $S$  em  $A$  através de sua expectativa ou valor esperado,  $E[N(S)]$  e a covariância  $C[N(S_i)N(S_j)]$  onde  $N(S)$  significa o número de eventos em  $S$ . O objetivo da estatística espacial é estimar as prováveis localizações de ocorrência destes eventos, levando-se em conta o valor limite da quantidade de eventos por área. Este valor corresponde à esperança de  $N(S)$  para uma região infinitesimal  $dS$ , quando essa tende a zero. Cada ponto é definido nas coordenadas  $(x, y)$ . Essa média é denominada *intensidade* (propriedade de primeira ordem), sendo definida como:

$$\lambda(x, y) = \lim_{dS \rightarrow 0} \left[ \frac{E(N(dS))}{|dS|} \right]. \quad (4.1)$$

As propriedades de segunda ordem são definidas da mesma forma. Considera-se a intensidade conjunta  $\lambda((x_1, y_1)(x_2, y_2))$  entre duas regiões infinitesimais  $|dS_1|$  e  $|dS_2|$ .

$$\lambda((x_1, y_1)(x_2, y_2)) = \lim_{dS_1, dS_2 \rightarrow 0} \left[ \frac{E(N(dS_1), N(dS_2))}{dS_1, dS_2} \right]. \quad (4.2)$$

Para um processo, a intensidade  $\lambda(x, y)$  passa a ser uma constante, isto é,  $\lambda(x, y) = \lambda$ , além disso, se for também isotrópico,  $\lambda((x_1, y_1)(x_2, y_2))$  se reduz a  $\lambda(|h|)$ , sendo  $|h|$  a distância entre dois pontos arbitrários. Se o processo é não estacionário, a intensidade média do processo varia na região delimitada  $A$ . De tal modo, as técnicas de análise de distribuição de pontos, em sua maior parte, supõem, explícita ou implicitamente, o comportamento de estacionaridade e isotropia de processo aleatório dos eventos analisados.

#### 4.3.1 Estimador de Intensidade

A análise do comportamento de padrões de pontos é estimado pela intensidade pontual do processo numa área de influência. Para tal, pode-se ajustar uma função bidimensional sobre

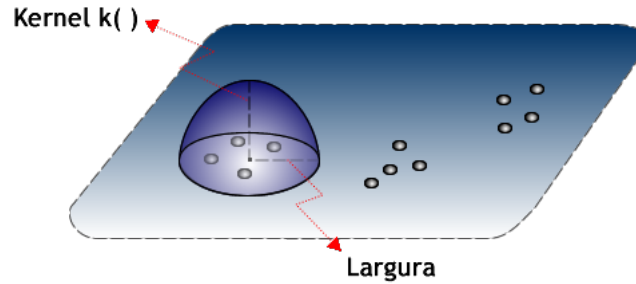


Figura 4.2 – Estimador de intensidade de distribuição de pontos.

os eventos a serem considerados, compondo uma superfície cujo valor será proporcional à intensidade de amostras por unidade de área. Essa função realiza a contagem de todos os pontos numa determinada área, podendo-os pela distância de cada um a uma localização de interesse. Ao supor  $s_1, \dots, s_n$  localizações de  $n$  eventos observados em uma região  $A$  e que  $s$  represente uma localização genérica a ser estimada, o estimador de intensidade é computado a partir dos  $j$  eventos  $s_i, \dots, s_{i+j-1}$  contidos num raio de tamanho  $\tau$  em torno de  $s$  e da distância  $d$  entre a posição de  $s$  e a  $i$ -ésima amostra, a partir de funções cuja forma geral é:

$$\hat{\lambda}_\tau(s) = \frac{1}{\tau^2} \sum_{i=1}^n k\left(\frac{d(s_i, s)}{\tau}\right), d(s_i, s) \leq \tau. \quad (4.3)$$

Este estimador é conhecido *kernel estimator* (DUONG, 2004) e os parâmetros básicos são:

- Um *raio de influência*<sup>4</sup> ( $\tau \geq 0$ ) que define a vizinhança do ponto a ser interpolado e controla a “suavização” da superfície gerada;
- Uma função de estimação com propriedades de suavização do fenômeno.

Já para a função de interpolação  $k()$ , é comum usar funções de terceira ou quarta ordem, como:

$$k(r) = \frac{3}{\pi}(1 - r^2), \quad (4.4)$$

ou um *kernel gaussiano*,

$$k(r) = \frac{1}{2\pi\tau} \exp\left(-\frac{r^2}{2\tau^2}\right). \quad (4.5)$$

Nestes estimadores,  $r$  representa a distância entre a localização em que desejamos calcular a função e o evento observado.

### 4.3.2 Processo de Poisson

Na natureza existem dois tipos de fenômenos: aqueles que são realizados sob as mesmas condições apresentando os mesmos resultados e aqueles, que apesar de serem realizados nas

<sup>4</sup>O raio de influência define a área centrada no ponto de estimação  $s$  que indica quantos eventos  $s_i$  contribuem para a estimativa da função intensidade  $\lambda$ .

mesmas condições, não apresentam necessariamente os mesmos resultados. O primeiro tipo de fenômeno é denominado determinístico. O segundo, conhecido como fenômeno estocástico ou aleatório, será objeto de estudo. Assim, na vida cotidiana, em algumas circunstâncias utilizam-se fatores probabilísticos acerca de alguns fenômenos, como por exemplo: o número de veículo em postos de pedágio, o número de acidentes em rodovias, a chegada de passageiros ao *check-in* de um aeroporto e etc. O melhor processo probabilístico a ser aplicado para estimar o número de pontos em uma área onde ocorrem os eventos é conhecido como processo de Poisson, também chamado como lei dos acontecimentos raros (NOVAES, 1989).

Ele expressa, por exemplo, a probabilidade de um certo número de eventos,  $k$ , ocorrerem numa dada localidade ou região delimitada  $A$ . Isto é

$$P(N(A) = k) = \frac{e^{-\lambda|A|}(\lambda|A|)^k}{k!}, \quad (4.6)$$

onde  $A$  é uma região delimitada.  $N(A)$  é o número de eventos contidos em  $A$ .  $k$  é o número de eventos e  $\lambda$  é a intensidade do processo. Considerando uma configuração espacial de pontos no plano, e que esses pontos pertencem a um mesmo espaço  $A$ , subconjunto de  $R^d$  ( $d$  representa a dimensão), constitui o processo de Poisson se as seguintes proposições existirem:

- $N(A) \sim Poisson(\lambda(A))$  – segue um processo de Poisson;
- $N(A)$  é independente de  $N(B)$  para  $A$  e  $B$  disjuntos.

E satisfizerem os seguintes axiomas:

- Se  $A_1, A_2, \dots$ , são regiões disjuntas, então  $N(A_1), N(A_2), \dots, N(A_n)$  são independentes e  $N(A_1 \cup A_2 \cup \dots \cup A_n) = N(A_1) + N(A_2) + \dots + N(A_n)$ ;
- A distribuição de probabilidade de  $N(A)$  depende do conjunto  $A$ , só através do seu tamanho  $|A|$ ;
- Existe  $\lambda > 0$  tal que  $P(N(A)) \geq \lambda|A| + ordem(|A|)$
- Probabilidade zero de pontos de sobreposição:

$$\lim_{|A| \rightarrow 0} \frac{P(N(A) \geq 1)}{P(N(A) = 1)} = 1. \quad (4.7)$$

caso essa condição seja satisfeita, tem

$$P(N(A) = k) = \frac{e^{-\lambda|A|}(\lambda|A|)^k}{k!} \quad (4.8)$$

para  $k = 0, 1, 2, \dots$



### 4.3.3 Processo de Poisson não-homogêneo

É o processo que ocorre de maneira mais realista, isto é, um processo que leva em conta especificidades locais do processo espacial. Diferentemente do que ocorre no processo de Poisson homogêneo, no não-homogêneo a intensidade do processo *varia* no espaço  $\lambda(x, y)$ .

Sendo  $\lambda(x, y) > 0$  qualquer região plana infinita  $A$ ,  $N(A)$  (o número de eventos com vetores correspondentes em  $A$ ) segue a **distribuição de Poisson** com média:

$$\int_A \lambda(x, y) dx dy. \quad (4.9)$$

A distribuição espacial do número de eventos,  $N(A)$ , segue uma função de densidade:

$$f(x, y) = \lambda(x, y) / \lambda. \quad (4.10)$$

O processo de Poisson não-homogêneo, através da média esperada, torna-se relevante na determinação de algum tipo de padrão sistemático no processo espacial. Se um padrão de eventos pontuais apresentarem desvios significativos do comportamento esperado para uma distribuição de Poisson, isto indica a existência de uma distribuição espacial que merece ser objeto de maior análise. A identificação dessas regiões de transição de um processo de Poisson para outro tipo de processo, no caso de um campo de galáxias, pode significar a identificação de “raios estatísticos”, que estariam possivelmente correlacionados com alguma escala física dos aglomerados. Este tipo de análise não foi feita no presente estudo, mas fica como uma possível ferramenta a ser utilizada em trabalhos futuros.

A seguir, apresentaremos a metodologia da contagem das parcelas, essencial para determinação dos raios de densidade.

## 4.4 Método *Quadrats* (ou Método das parcelas)

Neste método a verificação da hipótese nula, isto é, da aleatoriedade espacial, é baseada na divisão da área em estudo em parcelas (ou *quadrats*), não necessariamente de mesmo tamanho. O objetivo do método é realizar contagens dos objetos em partes, e assim verificar a hipótese de as contagens seguirem uma distribuição de Poisson (aleatória). Após a realização das contagens, pode-se fazer uso de diversos índices a fim de testar a aleatoriedade espacial, tais como o índice de Fisher, o índice de Morisita, o índice de David & Mores, a lei da potência de Taylor e os índices de agregação de Lloyd e Iwao (RUIZ et al., 2003). No entanto, alguns desses índices, além de ignorarem a localização espacial das amostras, podem depender fortemente do tamanho

das unidades amostrais (UPTON; FINGLETON, 1985). O meio termo para o problema seria utilizar diferentes índices de dispersão antes de se emitir uma conclusão a respeito da disposição espacial de uma determinada variável (MOLLET et al., 1984). Podemos entender o método como uma contagem do número de pontos em  $n$  *quadrats*, por exemplo,  $(u_1, \dots, u_n)$ , onde a aleatoriedade pode ser testada com base na ideia de que, se essas contagens seguem uma distribuição de Poisson, espera-se alcançar a igualdade de média e variância das contagens. A média e a variância do número de pontos em  $n$  *quadrats*, dada uma distribuição espacial de pontos são, respectivamente

$$\bar{U} = \frac{\sum_{i=1}^n u_i}{n} \quad (4.11)$$

$$S^2 = \frac{\sum_{i=1}^n (u_i - \bar{U})^2}{n - 1}, \quad (4.12)$$

onde

- $u_i$  é o número de pontos no *quadrats*  $i$ ;
- $n$  é o número de *quadrats*;
- $\bar{U}$  é a média das contagens em cada *quadrat*.

Para verificar o padrão de distribuição dos pontos pode-se calcular o índice de dispersão, utilizado neste trabalho

$$I = \frac{S^2}{\bar{U}}, \quad (4.13)$$

em que  $S^2$  é a variância das contagens em cada *quadrat*. A interpretação do índice é:

- $I > 1$  indica presença de agregação;
- $I = 1$  indica que a distribuição espacial é aleatória;
- $I < 1$  indica a existência de regularidade.

Na figura 4.3, apresentamos um exemplo ilustrativo de uma amostra distribuída de acordo com os três padrões espaciais. Neste exemplo é possível imaginar a distribuição de 30 pontos em unidades amostrais escolhidas ao acaso.

No caso A, temos: média = 3, variância = 17,2 e  $I_A = 17,2/3 = 5,7$ . Para o caso B: média = 3; variância = 0,2 e  $I_B = 0,2/3 = 0,1$ . Já para o caso C: média = 3; variância = 2,7 e  $I_C = 2,7/3 = 0,9$ .

Como poderíamos estimar a distribuição espacial dos pontos em cada caso? Caso contássemos novamente as dez unidades amostrais, de cada caso, dificilmente obteríamos os mesmos

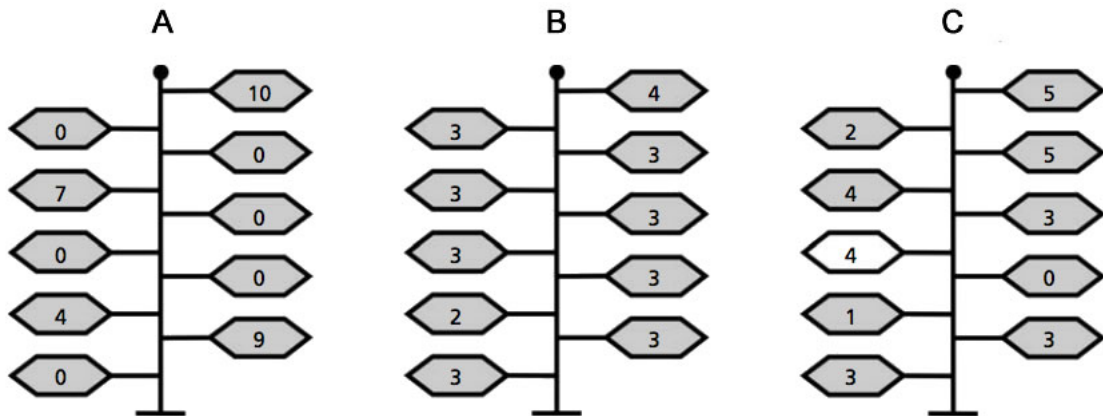


Figura 4.3 – Exemplo ilustrativo de distintos padrões espaciais. Caso A: agregação; Caso B: regular e Caso C: aleatoriedade espacial.

valores da média e variância, mesmo que a distribuição espacial dos casos *A*, *B* e *C* permanecessem imutáveis. Tal fato ocorre devido ao acaso. Além da estimativa de *I*, precisamos de uma estimativa da incerteza associada a este valor que é a probabilidade desse valor ter-se desviado de  $I = 1$ , devido ao acaso envolvido no processo de amostragem. Essa expectativa é a conhecida **hipótese nula**<sup>5</sup> Para testar a significância do valor de *I* observado, pode-se utilizar a distribuição do qui-quadrado ( $\chi^2$ ), que descreve as probabilidades esperadas de valores de *I*. O teste para o índice *I* é um teste de  $\chi^2$  bicaudal (isto é, utiliza os dois lados da distribuição: serão significativos tanto os valores menores como os valores maiores que o esperado), com  $N - 1$  graus de liberdade, da seguinte forma:

$$\chi^2 = I(N - 1). \quad (4.14)$$

Os pequenos valores serão indicativos de uma distribuição regular (*I* menor que 1), por outro os maiores valores significativos terão indicativos de uma distribuição agregada (*I* maior que 1). No exemplo ilustrativo apresentado, teremos:

- Caso *A*:  $\chi^2 = 5,7(10 - 1) = 51,3$ .
- Caso *B*:  $\chi^2 = 0,1(10 - 1) = 0,9$ .
- Caso *C*:  $\chi^2 = 0,9(10 - 1) = 8,1$ .

Ao se fazer uso do método da contagem dos *quadrats* sobre uma certa distribuição de pontos, observa-se:

- uma população com distribuição agregada, alguns *quadrats* conterão muitos pontos (os que caírem próximo ou em cima dos agregados), enquanto outros *quadrats* conterão muito

<sup>5</sup>A hipótese nula é também conhecida como *Completa Aleatoriedade Espacial* (CAE), o processo compara os distintos padrões espaciais em análise rejeitando àqueles que diferem ao da aleatoriedade.

poucos ou nenhum (os pontos que caírem fora). A variação do número de pontos por *quadrats* será grande. Poucos *quadrats* terão um número próximo da média;

- uma população com distribuição aleatória, os *quadrats* conterão um número de pontos variável, alguns com mais pontos que a média, outros com menos. Um número maior de *quadrats* terá valores próximos ao da média;
- uma população com distribuição regular, os *quadrats* conterão um número semelhante de pontos. A variação entre as contagens nos *quadrats* será pequena. Quase todos os *quadrats* terão valores próximos da média.

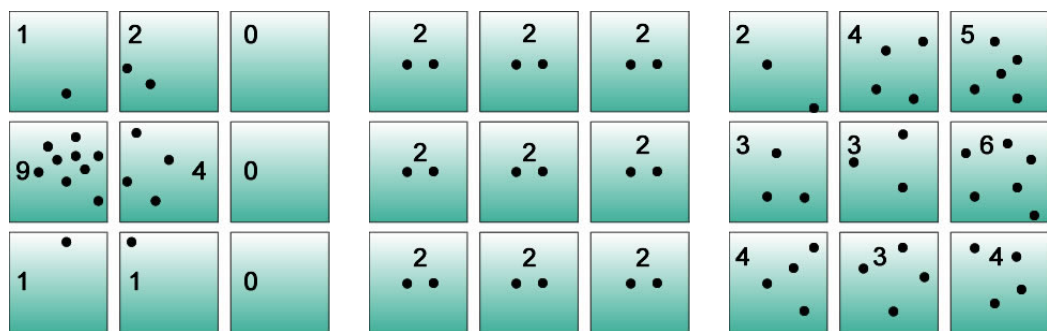


Figura 4.4 – Definições dos padrões espaciais através das contagens dos *quadrats*: agrupamento espacial (*à esquerda*), distribuição regular (*meio*) e aleatoriedade espacial (*à direita*).

O método em questão tornou-se fundamental para os resultados deste trabalho, uma vez que também nos fornece a contabilidade das distâncias entre os objetos em estudo. Em relação aos métodos de distância, a unidade básica de amostragem é um ponto, que por sua vez é escolhido aleatoriamente na região de estudo, e a informação é registrada na forma de distâncias aos eventos vizinhos, como, por exemplo, as distâncias ao primeiro, segundo, terceiro, ..., k-ésimo vizinho mais próximo. Há na literatura vários estimadores  $\lambda$  baseados nos métodos de distância (DIGGLE, 1983). Contudo, em configurações que espera-se agregação espacial, utilizar-se-á o estimador proposto por Byth (1982). Nesse tipo de distribuição o modelo probabilístico envolvido é o de um processo de Poisson homogêneo observado na região de interesse. Com base no tamanho da amostra  $n$ , o estimador proposto por Byth é:

$$\hat{\lambda}_1 = \frac{3n}{\pi \sum_{i=1}^n x_{3i}^2}, \tag{4.15}$$

onde  $x_{3i}^2$  é a distância do i-ésimo ponto a seu terceiro vizinho mais próximo. Ou ainda, ao estimar o total de pontos da região em estudo, temos:

$$\hat{\lambda}_1 = \frac{3n|A|}{\pi \sum_{i=1}^n x_{3i}^2}, \tag{4.16}$$

onde  $|A|$  é a área da região.

#### 4.4.1 Método das parcelas em Aglomerados de Galáxias

Aqui utilizamos os *quadrats* ou parcelas com duas finalidades: (i) testar a aleatoriedade dos dados; e (ii) determinar se os aglomerados estão bem centrados nos campos de galáxias em que estão inseridos. Testar a aleatoriedade significa verificar se, para a seleção de galáxias que tenha sido feita, o aglomerado ainda é identificável estatisticamente. Testar se o sistema se encontra no centro do campo visa a evitar efeitos de borda. Em ambos os casos, se o aglomerado se torna indistinto de um processo de Poisson, ou se o centro se aproxima de uma das bordas do campo, tal sistema é eliminado da amostra. A seguir, descrevemos em linhas gerais como o método funciona.

Inicialmente, executamos a função *quadrat.test* [*spatstat*] para verificar a aleatoriedade espacial. Se o programa rejeita a hipótese nula de o campo ser um processo de Poisson, passamos ao estudo das parcelas. Nossos dados foram contabilizados a fim de quantificar o número de objetos inseridos em cada *quadrat*. Para tal, utilizamos a função *quadratcount* ( $X$ ,  $nx = 5$ ;  $ny = nx$ ) da biblioteca *spatstat*. Os argumentos desta função são, respectivamente:

- $X$ , os dados que serão verificados para identificar distintos padrões espaciais;
- $nx$  e  $ny$ , os números de parcelas em direções  $x$  e  $y$ .

Escolhemos, arbitrariamente, a mesma quantidade de parcelas ( $nx = ny = 5$ ), um número ímpar para gerar um *quadrat* central. A escolha, embora arbitrária, mostrou-se bastante estável numa série de testes que realizamos sobre amostras de controle. Uma vez que o comando é executado, a janela é dividida em  $nx \times ny$  grades ou parcelas (*quadrats*). Assim, o número de objetos é contabilizado em cada *quadrat*, o que representa, neste método, a riqueza de cada parcela do campo de galáxias.

Em seguida, a fim de determinarmos problemas de borda, introduzimos um teste que verifica se a contagem da parcela central contém o centro nominal do aglomerado. A figura 4.5 mostra um exemplo ilustrativo da técnica que, a partir de então, será chamada de **método das parcelas**.

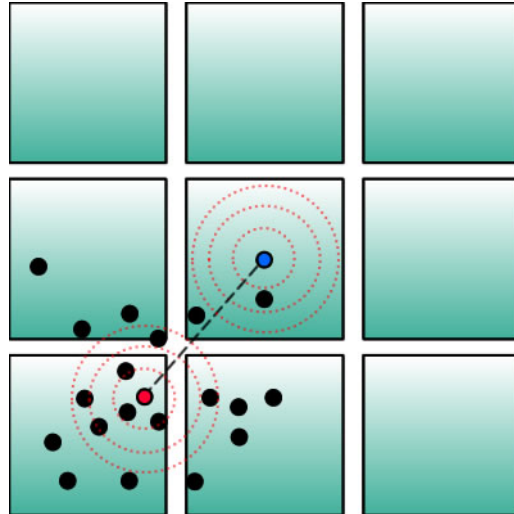


Figura 4.5 – A técnica determina um campo mal centrado, sobre a área de domínio, o ponto vermelho, e o centro do *quadrat* central, o ponto azul.

#### 4.5 Método do Raio de densidade ( $R_{dens}$ )

Após a seleção dos aglomerados pelo método das parcelas, implementamos um *script*, em linguagem R, estimando as distâncias de todas as galáxias (e BCGs) com relação ao pico de densidade do campo em análise gerando, por conseguinte, o perfil de densidade superficial do aglomerado  $\Sigma$ . Esse passo foi realizado para cada aglomerado (ver figura 4.6).

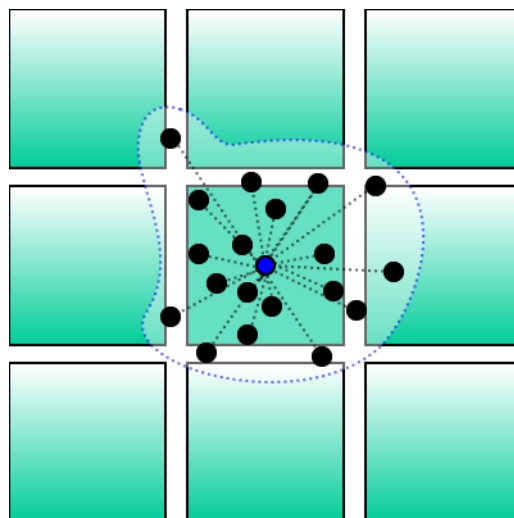


Figura 4.6 – Distâncias das galáxias em relação ao pico de densidade do campo.

A próxima etapa destina-se à utilização das distâncias projetadas de cada galáxia (anteriormente calculadas) normalizadas pelo raio  $R_{200}$  para cada aglomerado. Os valores de  $R_{200}$  foram obtidos do próprio catálogo simulado com suas propriedades físicas tabeladas (coordenadas

equatoriais, massa e raios viriais e número de membros). A obtenção das densidades superficiais normalizadas se deu pelo valor correspondente de cada distância projetada (normalizada).

Para melhor exemplificar esta etapa mostraremos a figura 4.7, uma simulação do raio  $R_{200}$  cruzando com a densidade superficial fotométrica correspondente, a qual denominamos  $\Sigma_{200}$ .

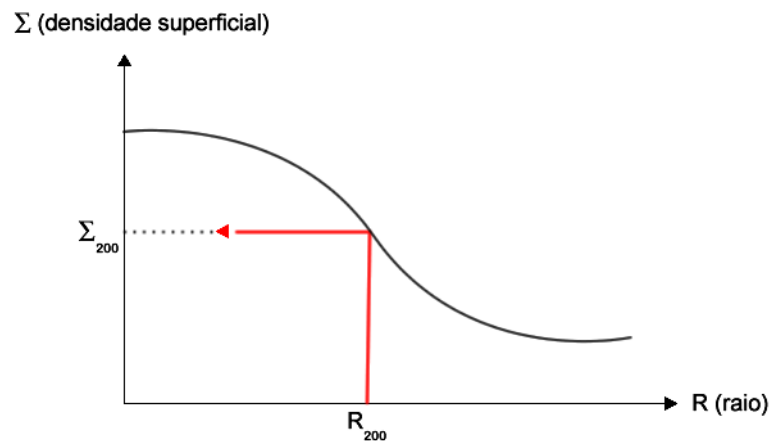


Figura 4.7 – Correspondência entre os raio  $R_{200}$  e sua respectiva densidade superficial,  $\Sigma_{200}$ .

A relação entre  $R_{200}$  e  $\Sigma_{200}$  é então estabelecida. Por fim, o programa executa o caminho inverso do passo anterior. Essencialmente, o programa parte do cálculo da densidade do perfil calibrado a fim de recuperar (ou estimar) os raios de cada campo. Estes raios, de agora em diante, serão chamados de raios de densidade,  $R_{dens}$  (ver figura 4.8).

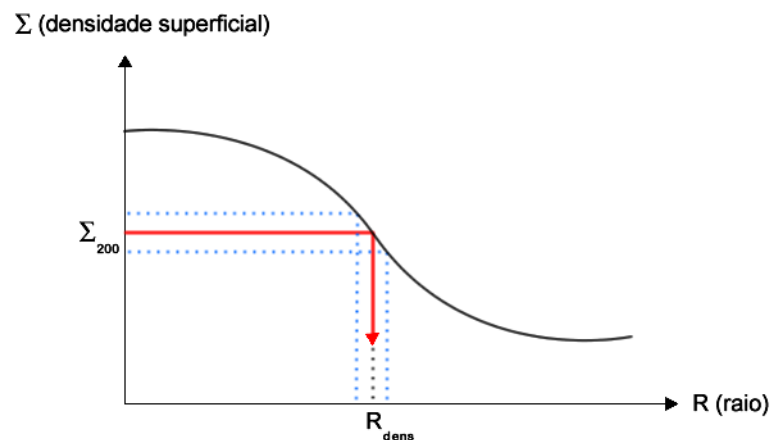


Figura 4.8 – Obtenção do raio de densidade a partir do perfil de densidade superficial do aglomerado.

As aplicações dos métodos das parcelas e do raio de densidade estão relacionados à primeira etapa de nosso trabalho, que tem como objetivo efetuar uma contagem de galáxias dentro de uma escala típica do sistema. As demais metodologias empregadas no trabalho referem-se à etapa seguinte: inferir a presença de subestruturas na região central dos aglomerados. Os métodos foram agrupados de acordo com a informação necessária para seu uso. Na próxima seção, apresentamos os métodos espectroscópicos.

## 4.6 Análise de subestruturas: métodos espectroscópicos

### 4.6.1 Teste de Anderson-Darling (AD)

O teste AD vem sendo usado sistematicamente em estudos de aglomerados de galáxias (vide (HOU et al., 2009); (HOU et al., 2012); (HOU et al., 2013) e (ROBERTS; PARKER, 2017)) com resultados satisfatórios. O teste é utilizado para verificar se um dado conjunto de dados  $x_1, x_2, \dots, x_n$  é proveniente de uma determinada distribuição de probabilidade  $F(x)$ . As duas hipóteses para o teste de AD para uma distribuição normal são dadas por

- $H_0$  : os dados seguem uma distribuição de probabilidade normal  $F(x)$ .
- $H_1$  : os dados não seguem uma distribuição de probabilidade normal  $F(x)$ .

A hipótese nula é que os dados são normalmente distribuídos e a hipótese alternativa é que os dados não seguem uma distribuição normal. Para um nível de significância (*p-value*) menor que  $p \leq 0,05$ , rejeita-se a hipótese nula e conclui que os dados analisados não seguem uma distribuição normal.

Para o teste, Anderson e Darling (1954) propuseram a seguinte estatística:

$$A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]}{F(x)(1 - F(x))} dF(x), \quad (4.17)$$

onde  $F_n(x)$  é a função de distribuição acumulada empírica, definida como

$$F_n(x) = \begin{cases} 0, & \text{se } x < x_{(1)} \\ \frac{k}{n}, & \text{se } x_{(k)} \leq x < x_{(k+1)} \\ 1, & \text{se } x \geq x_{(n)} \end{cases} \quad (4.18)$$

também,  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ , são as estatísticas de ordem da amostra aleatória. Por outro lado a estatística  $A^2$  pode ser colocada numa forma equivalente:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1) \ln(F(x_{(i)})) + (2(n - i) + 1) \ln(1 - F(x_{(i)}))]. \quad (4.19)$$

A transformação  $F(x_{(i)})$  leva  $x_{(i)}$  em  $U_{(i)}$  de uma amostra de tamanho  $n$  com distribuição uniforme. Logo,

$$A^2 = -n - \frac{1}{n} D, \quad (4.20)$$



onde  $D$  é dado por

$$D = \sum_{i=1}^n [(2i - 1)(\ln(U_{(i)}) + (2(n - i) + 1) \ln(1 - U_{(i)}))]. \quad (4.21)$$

Para calcular o valor da estatística  $A^2$  procede-se:

- a ordenação dos valores da amostra:  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ ;
- a estimação dos parâmetros da distribuição de interesse;
- o cálculo de  $U_i = F(x_{(i)})$ , como também o valor da estatística de Anderson-Darling

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i - 1)(\ln(U_{(i)}) + (2(n - i) + 1) \ln(1 - U_{(i)}))]. \quad (4.22)$$

#### 4.6.2 A distância de Hellinger (HD)

Trata-se de um método que mede desvios da normalidade de uma distribuição de velocidades em grupos e aglomerados de galáxias. Foi introduzido pela primeira vez em astronomia por (RIBEIRO et al., 2013). A distância de Hellinger (HD) define um sistema galáctico relaxado como aquele com distribuição de velocidade unimodal com desvio de normalidade abaixo de um valor crítico  $HD < v_c$ , que, em geral, depende do número de objetos analisados. CARVALHO et al. (2017) calibram uma relação entre HD e N, corrigindo esse viés e permitindo uma classificação de sistemas galácticos em função dos valores de HD e de uma curva que indica quando os desvios de uma normal são significativos.

O valor HD quantifica os desvios da normalidade, uma aproximação estável da métrica de informações de Fisher (S., 1985):

$$HD^2(p, q) = 2 \sum_{x \in X} \sqrt{[p(x) - \sqrt{q(x)}]^2}, \quad (4.23)$$

sendo  $p$  e  $q$  distribuições de probabilidade e  $x$  uma variável aleatória. Os possíveis valores de HD estão na faixa de  $[0, \sqrt{2}]$ , com valores possíveis de  $p$  e  $q$  para  $[0, 1]$ .

O indicador HD utiliza um estimador da distância entre a distribuição de velocidades empíricas das galáxias num grupo e a função de distribuição gaussiana esperada teoricamente. Como resultado o teste classifica as distribuições em análise com as seguintes saídas: (0), (1) e (2), respectivamente, Gaussiana (G), Não Gaussiana (NG) ou Indeterminado (Ind). A última representa os diagnósticos com baixa confiabilidade do método (vide (CARVALHO et al., 2017)).

#### 4.6.3 O teste de Dressler-Schechtman (DS)

O teste procura regiões que tenham maiores valores de velocidade (local),  $v_l$ , e dispersão de velocidades,  $\sigma_{v_l}$ , que os valores medidos no aglomerado como todo. Essas quantidades são

calculadas, no espaço de posições, para cada galáxia  $i$  e suas vizinhas mais próximas  $N_v$ . Para tal, Dressler e Shectman utilizam 10 galáxias vizinhas. Para cada galáxia, os valores de  $\delta_i$  são definidos como

$$\delta_i^2 = \frac{(N_v)}{\sigma_g^2} [(v_l - v_g)^2 + (\sigma_{v_l} - \sigma_{v_g})^2], \quad (4.24)$$

onde  $\sigma_{v_g}$  e  $v_g$  são respectivamente os valores da dispersão de velocidades e o valor da velocidade global do aglomerado. Com base no cálculo sobre a soma de todos os valores dos desvios cinemáticos  $\delta_i$  estima-se um valor final para  $\Delta$ , como:

$$\Delta_{DS} = \sum \delta_i. \quad (4.25)$$

Quanto maior o valor de  $\Delta_{DS}$ , maior será o desvio dos parâmetros de uma certa região quando comparado aos parâmetros globais. Os valores de  $\Delta_{DS}$  são calibrados por meio de simulações de Monte Carlo. O índice de subestruturas, dado por um valor probabilístico  $p$ , é quantificado pela razão

$$\frac{N(\Delta_{sim} > \Delta_{obs})}{N_{sim}}, \quad (4.26)$$

onde  $N$  é o número de galáxias do aglomerado e  $N_{sim}$  é o número total de simulações. O menor valor de  $p$ , representa maior probabilidade de se encontrar subestruturas baseada nos valores de  $\Delta$  simulado e observado (respectivamente,  $\Delta_{sim}$  e  $\Delta_{obs}$ ). Outros estudos, por exemplo (KNEBE; MÜLLER, 1999), propõe que  $\Delta_{DS}/N_{gal} \geq 1.4$  é um critério de detecção de subestruturas.

Em seguida, apresentamos os indicadores fotométricos de subestruturas usados neste trabalho.

## 4.7 Análise de subestruturas: métodos fotométricos

### 4.7.1 Posição da BCG e *gap* de magnitude

Alguns indicadores de subestruturas, ou de um grau maior ou menor de evolução de aglomerados de galáxias, estão ligados a sua galáxia central, a BCG (*Brightest Central Galaxy*). Geralmente, assume-se que a galáxia central é a mais luminosa e mais massiva do sistema, e que reside em repouso no centro do halo de matéria escura. Isto, em linhas gerais, define o que se chama de Paradigma da Galáxia Central (vide Bosch et al. (2005) e Skibba et al. (2010)). Se este paradigma for correto, sistemas mais evoluídos devem possuir suas BCGs com pequenas distâncias ao centro do aglomerado, baixas velocidades peculiares, e consideráveis *gaps* de magnitude,  $\Delta M_{12}$  ou  $\Delta M_{14}$ , entre a primeira e a segunda galáxias mais luminosas. Se isto não acontece, podemos interpretar o sistema como menos evoluído, ou perturbado, ou possuindo subestruturas. Estas quantidades serão usadas neste trabalho como indicadores do estado dinâmico do sistema.

### 4.7.2 Índice de Clark e Evans

Um outro método empregado ao processo espacial definido pelo campo de galáxias que contém o aglomerado é o índice de agregação de Clark e Evans (CLARK; EVANS, 1954) (aqui denominado  $I_{CE}$ ). O índice quantifica o grau de agregação dos pontos em um processo espacial. Nossa expectativa para um sistema mais evoluído é a de um maior grau de concentração, que será medido pelo índice.

Clark e Evans (1954) propuseram uma medida estatística a qual uma distribuição espacial numa determinada área bidimensional pode se desviar de uma distribuição aleatória. Neste teste, a estimativa de espaçamento entre um objeto e um vizinho é realizada seguindo o método do vizinho mais próximo. As distâncias, entre todos os pontos da distribuição ou de uma amostra de pontos selecionada aleatoriamente (com relação a seus vizinhos mais próximos), é mensurada para o conjunto de observações. Por outro lado, a distância média ao vizinho mais próximo que deverá ser esperada se os objetos forem aleatoriamente distribuídos também é calculada. Portanto, a razão entre a distância média observada e a distância média esperada serve como uma medida do desvio da aleatoriedade. Numa distribuição de  $N$  pontos com uma densidade especificada  $\rho$ , a distância  $r$  de cada ponto a seu vizinho mais próximo é medida. A média dessas distâncias pode ser representada por:  $\bar{r}_A = \sum_{i=1}^N r_i/N$ . A distância média esperada, se esta população for distribuída ao acaso, terá um valor  $\bar{r}_E = 1/2\rho^{1/2}$ . Assim, a razão  $I_{CE} = \bar{r}_A/\bar{r}_E$  é utilizada como sendo a medida do grau a qual uma distribuição observada aproxima-se ou afasta-se da aleatoriedade. Essa quantidade é chamada de **Índice de Clark & Evans**,  $I_{CE}$ . A tabela 4.1 e a figura 4.9 revelam o comportamento dos padrões sistemáticos a partir dos valores da razão  $I_{CE}$ .

Valores de $I_{CE}$	Padrão da distribuição
1	Aleatório
0	Agrupamento
2,1491	Regular

Tabela 4.1 – Em uma distribuição aleatória, o valor da razão assume o valor 1. Zero para uma condição de máxima agregação. Já para uma condição de máximo espaçamento (ponto equidistante de outros pontos),  $I_{CE}$  assume o valor de 2,1491.

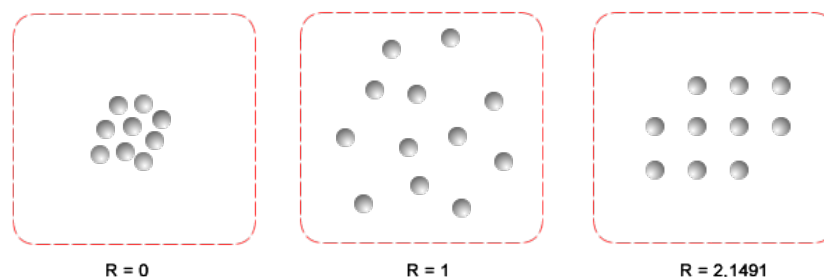


Figura 4.9 – Padrões sistemáticos e os respectivos valores da razão  $I_{CE}$ .

Clark e Evans (1954) ainda apresentam uma variável padrão da curva normal  $c = (\bar{r}_A - \bar{r}_E / \sigma_{\bar{r}_E})$ , onde  $\sigma_{\bar{r}_E}$  é o desvio padrão da distância média ao vizinho mais próximo em uma distribuição aleatória de mesma densidade da população observada. Esta variável é recomendada para testar a significância do desvio da distância observada em relação à distância esperada.

### 4.7.3 Distribuição de Von Mises

Um outro método utilizado neste trabalho consiste em analisar a distribuição de galáxias do ponto de vista da estatística circular. A distribuição é interpretada como um conjunto de dados direcionais que se projetam sobre uma circunferência de raio unitário. O objetivo é verificar se as direções se aglomeram ou não. Se elas se aglomeram, o sistema é assimétrico. Caso elas se espalhem uniformemente no círculo, o sistema é simétrico. Aglomerados com assimetria espacial são interpretados como sendo menos evoluídos que aqueles mais simétricos no espaço (vide West et al. (1988)).

A distribuição de Von Mises, também conhecida como distribuição normal circular, foi proposta em 1918 pelo físico alemão R. von Mises, ao investigar desvios dos pesos atômicos de valores inteiros. Essa distribuição pode ser aplicada a dados que podem ser convertidos em dados circulares.

Uma variável aleatória circular  $x$  tem distribuição von Mises com parâmetros  $\mu$  e  $\kappa$  se sua função densidade de probabilidade for dado por Mardia e Jupp (2009)

$$f(x|\mu\kappa) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)}, \quad (4.27)$$

com  $0 \leq x < 2\pi$ ,  $0 \leq \mu < 2\pi$  e  $\kappa > 0$ . Os parâmetros  $\mu$  e  $\kappa$  são, respectivamente, a média circular de  $x$  e uma medida de concentração.  $I_0(\kappa)$  representa a função de Bessel modificada de ordem zero, dada pela forma

$$I_0(\kappa) = \sum_{r=0}^{\infty} \frac{1}{r!^2} \left(\frac{1}{2}\kappa\right)^{2r}. \quad (4.28)$$

Os parâmetros  $\mu$  e  $1/\kappa$  são análogos a  $\mu$  e  $\sigma^2$  (a média e variância) de uma distribuição normal:

- $\mu$  é uma medida de localização (a distribuição é agrupada em torno de  $\mu$ );
- $\kappa$  é uma medida de concentração (uma medida recíproca da dispersão, então  $1/\kappa$  é análogo a  $\sigma^2$ ).

A seguir descrevemos o comportamento da função densidade de probabilidade em termos dos parâmetros  $\mu$  e  $\kappa$  (ver a figura 4.10):

- para  $\kappa = 0$ , a distribuição é uniforme (e para  $\kappa$  suficientemente pequeno a distribuição é aproximadamente uniforme).
- para  $\kappa$  grande, a distribuição é muito concentrada sobre o ângulo  $\mu$  com  $\kappa$  sendo uma medida de concentração. Assim, a distribuição se aproxima de uma distribuição normal em  $x$  com média  $\mu$  e variância  $1/\kappa$ .

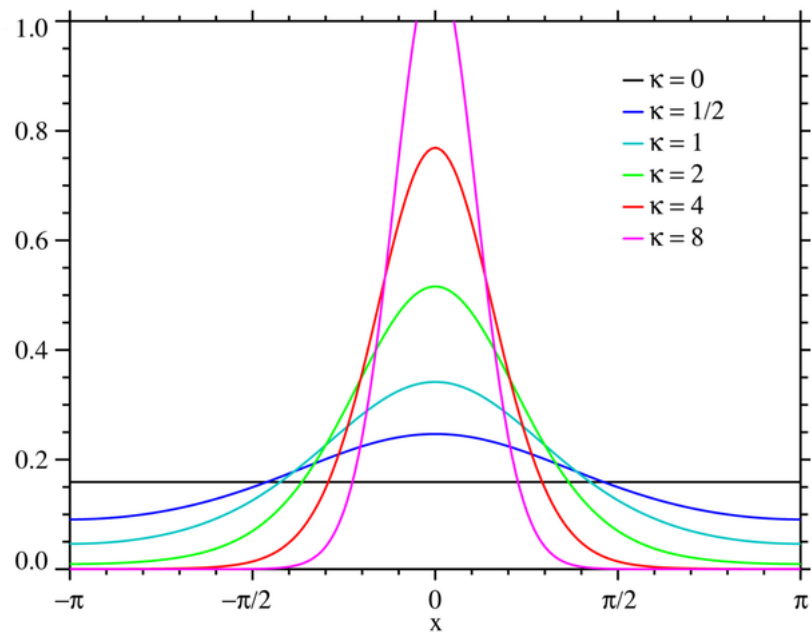


Figura 4.10 – Comportamento da função densidade de probabilidade em distintos valores do parâmetro de concentração  $\kappa$ .

#### 4.7.4 Programa *pdfCluster*

Finalmente, introduzimos o método que identifica subestruturas mais diretamente, através da localização e número de picos de densidade na distribuição projetada de galáxias. O pacote é dotado apenas por uma função que recebe seu próprio nome: o *pdfCluster()*. O programa analisa aglomeramentos espaciais por meio da estimativa de densidade feita por distintos métodos de um *kernel* (AZZALINI; TORELLI, 2007), ver a ilustração da figura 4.11(a). Estas estruturas são classificadas em dois diferentes conjuntos seguindo uma partição do espaço amostral: o primeiro, com uma densidade limite  $k$ , e o segundo, com uma densidade superior a este limiar. Os aglomeramentos espaciais correspondem às máximas regiões de densidades associadas ao último conjunto (figuras 4.11(b) e 4.11(c)).

A estimativa da densidade é realizada por meio de um *kernel* estimador usando a função

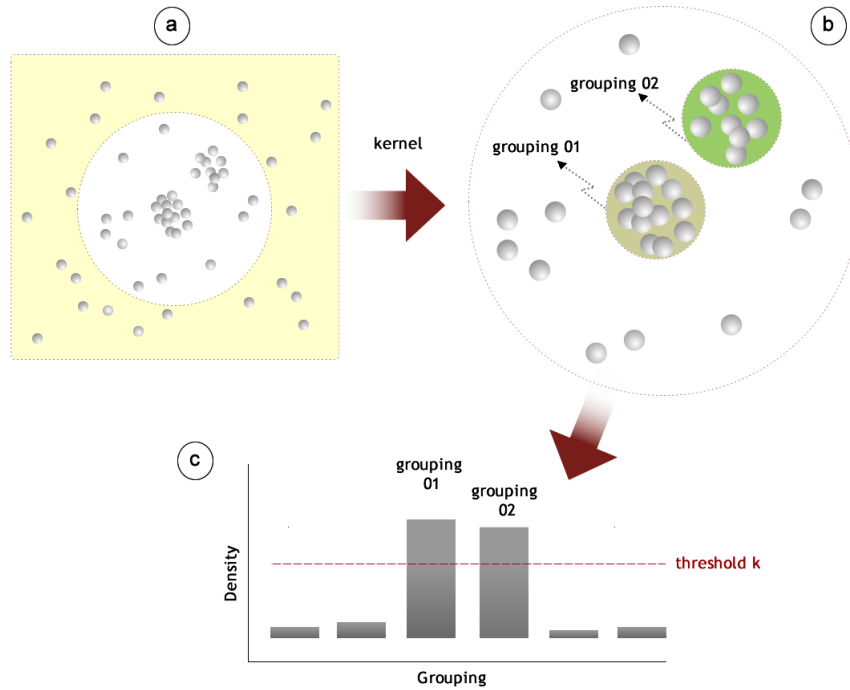


Figura 4.11 – Em (a): seleção da região com aglomeramentos espaciais; (b) detecção, por meio de um *kernel*, das duas maiores densidades do campo; (c) somente os agrupamentos 01 e 02 tiveram densidade maior que limiar de densidade  $k$ .

$kepdf()$ <sup>6</sup>. O estimador utilizado tem a forma

$$\hat{f}(y) = \sum_{i=1}^n \frac{1}{nh_{i,1} \cdots h_{i,d}} \prod_{j=1}^d K\left(\frac{y_j - x_{i,j}}{h_{i,j}}\right), \quad (4.29)$$

onde  $K$  é a função de interpolação responsável pelo tipo do *kernel* utilizado, como por exemplo, pode optar-se pelo *kernel* gaussiano. O parâmetro  $y_i$  são as localizações de  $n$  eventos esperados,  $x_i$  é a localização dos eventos observados na região em estudo e  $h_i$  é a largura da banda analisada pelo *kernel*. O usuário, a fim de estimar a densidade do campo, pode optar por escolher a forma da largura da banda: **fixa** ou **adaptada**, definida, respectivamente, como  $h_i = h$  (este caso, gera um vetor de suavização constante para as observações  $x_i$ ) ou  $h_i = (h_{i,1}, h_{i,d})^\top$  (por sua vez, especifica um vetor de suavização de largura de acordo com os eventos observados  $x_i$ ). Neste trabalho utilizamos o argumento *bwtype*="adaptive". A figura 4.12 ilustra a ação do *kernel* sobre os dados.

<sup>6</sup>Esta função permite o usuário escolher distintos métodos do *kernel*.

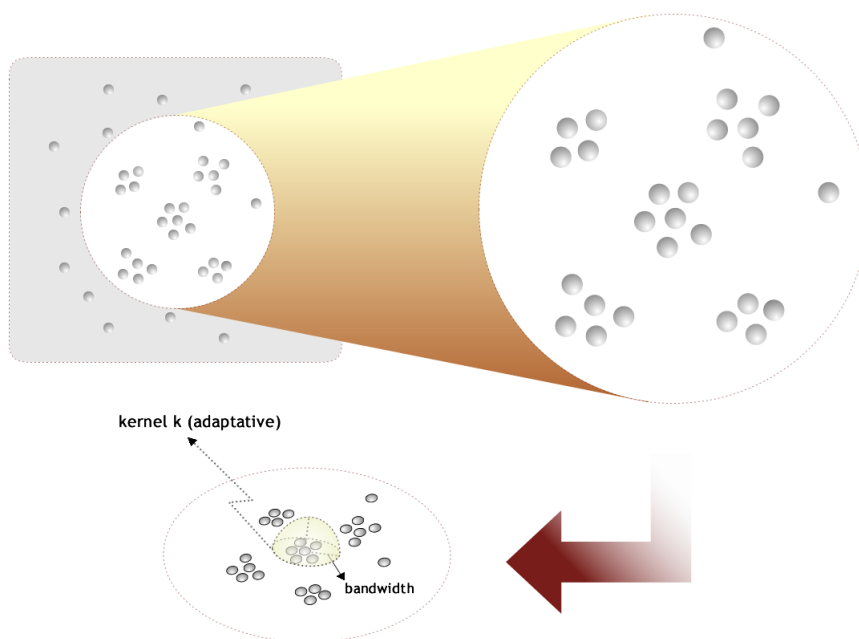


Figura 4.12 – Kernel estimator atuando sobre os dados, cuja largura da banda pode ser definida pelo usuário como fixo ou adaptado.

## 5 RESULTADOS DAS ANÁLISES

Neste capítulo, apresentamos os resultados referentes à aplicação das técnicas descritas no Capítulo 4. Inicialmente, apresentamos os resultados decorrentes dos métodos das parcelas e raio de densidade. O objetivo aqui, além da aplicação do método em si, consiste em determinar qual a seleção de galáxias torna o uso do método mais efetivo.

### 5.1 Método das parcelas e o catálogo simulado

Nossa amostra, nesta etapa, é constituída por 947 aglomerados de galáxias dos catálogos simulados utilizados no trabalho de Duarte e Mamon (2014). Estes catálogos foram construídos a partir de saídas de simulações de galáxias reais do *Millennium-II* (BOYLAN-KOLCHIN et al., 2009) e a seleção de grupos de galáxias com base no algoritmo *Friends-of-Friends-FoF* (vide subsecção 3.1.2).

Os primeiros resultados deste trabalho vêm da aplicação do método das parcelas, em particular, da análise do centro nominal dos 947 aglomerados em relação ao *quadrat* central de cada campo, verificando ainda se a coordenada do pico de densidade de cada aglomerado (obtido pelo *pdfCluster* dentro da parcela de maior contagem) corresponde ao centro que consta no catálogo. Problemas associados com a posição do centro do aglomerado levam à sua eliminação da amostra.

#### 5.1.1 *Quadrats* e Efeito de borda

Um dos pontos centrais da utilização do método das parcelas é o posicionamento do centro de cada campo em relação ao *quadrat* central. Espera-se que essa parcela contenha o centro do campo em análise. Na figura 5.1 mostramos um diagrama ilustrando o método. O campo nesta figura apresenta dois picos de densidade (1 e 2), e nenhum coincide, espacialmente, com o *quadrat* central.

Aglomerados com essas características são descartados pela técnica, para evitar os efeitos de borda. Na figura 5.2 vemos a ilustração de um grupo de galáxias, com alguns membros fora dos limites do *frame*, não sendo contabilizados (erroneamente) na análise. Estudos de catálogos simulados, como por exemplo Berlind et al. (2006) revelam o quanto o efeito de borda pode afetar a riqueza estatística, sendo mais acentuado em grupos ricos de galáxias.



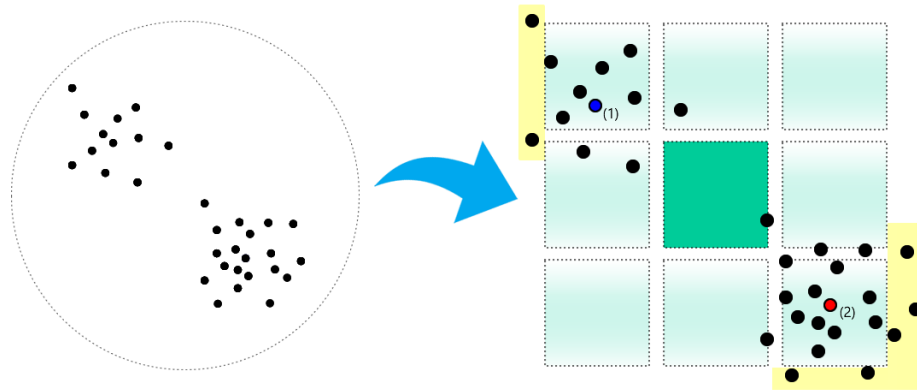


Figura 5.1 – À esquerda: Aglomerado com dois grupos de galáxias. À direita: O método das parcelas identifica que o aglomerado é mal centrado. Veja também que aglomerados com essas características podem apresentar problemas de borda (regiões em amarelo).

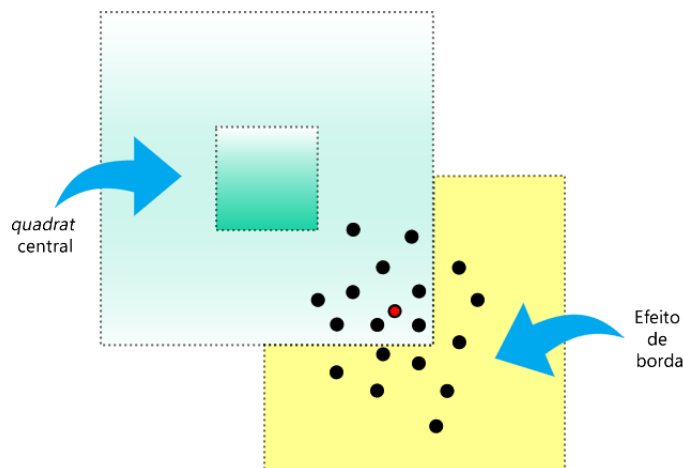


Figura 5.2 – Efeito de borda evidenciado pelas galáxias além do limite do *frame*. A figura mostra um aglomerado “descentralizado” (pico de densidade fora do *quadrat* central) e, portanto, excluído pelo método. O ponto, em vermelho, representa o pico de densidade do aglomerado.

Agora mostraremos a aplicação da técnica sobre dois casos de nossa amostra, a começar pelo aglomerado 869. Na figura 5.3, vemos que o método contabiliza o número de galáxias por *quadrat*. A riqueza é representada pelos números, em vermelho, em cada *quadrat*. As galáxias são representadas pelos pontos cheios. Da figura, vemos com clareza que o aglomerado 869 é um campo mal centrado. Seu pico de densidade contém 276 galáxias, enquanto seu *quadrat* central possui apenas 92 galáxias. Já na figura 5.4 vemos o exemplo do aglomerado 03. Observe que neste campo o pico de densidade coincide com a área de domínio do *quadrat* central. O campo 03, portanto, é mantido na amostra. A probabilidade de um efeito de borda em torno do aglomerado 869 faz com que ele seja excluído da amostra. A outra possível causa da exclusão

de campos de nossa amostra seria a indistinguibilidade entre os aglomerados e um processo de Poisson. Esta causa, no entanto, não se verificou em nenhum campo estudado neste trabalho.

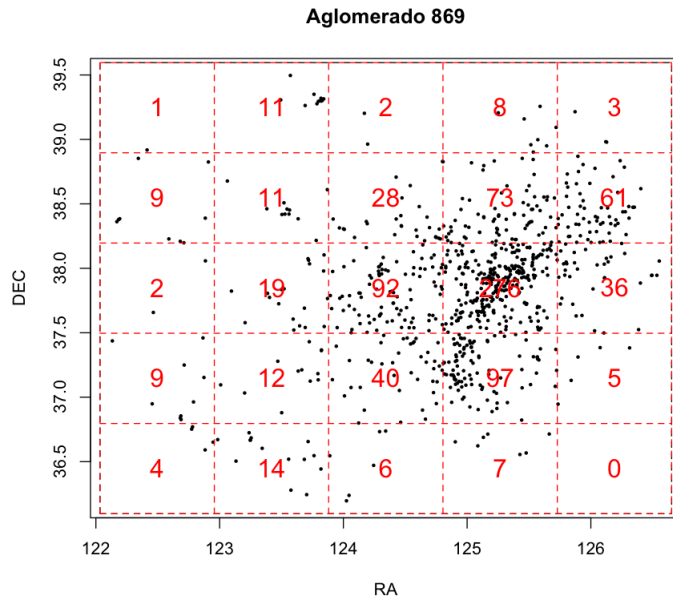


Figura 5.3 – Exemplo de exclusão do aglomerado 869 devido o seu centro não pertecer ao *quadrat* central.

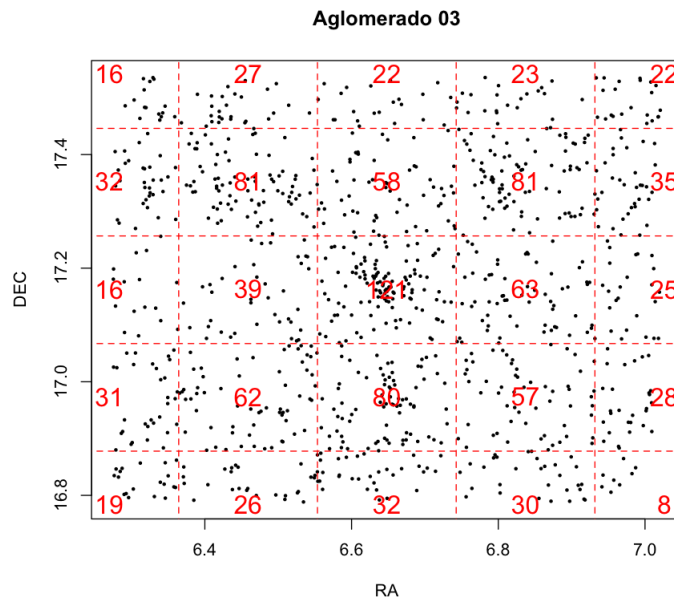


Figura 5.4 – Campo centrado no *quadrat* central.

### 5.1.2 Raios de densidade ( $R_{dens}$ ): Apresentação geral e aplicação nos catálogos simulados

Após a seleção dos aglomerados pelo método das parcelas, implementamos um *script*, em linguagem R, para obter os raios de densidade, conforme a descrição feita na Seção 4.5. O método obtém uma relação entre  $R_{200}$  e  $\Sigma_{200}$  e, em seguida, calibra esta relação, que pode ser utilizada para converter as densidades superficiais fotométricas, dadas por  $\Sigma_{200}$ , em um raio, que denominamos de “raio de densidade”.

Tanto o método de parcelas quanto o de raios de densidade dependem da seleção de galáxias usada como entrada nos respectivos programas. Na próxima seção, essa dependência é avaliada para diferentes critérios de seleção aplicados sobre os dados.

## 5.2 Critérios de seleção de galáxias aplicados ao catálogo simulado

Apresentaremos, a seguir, a relação entre os raios de densidade com os raios  $R_{200}$  apenas para os aglomerados previamente selecionados pelo método das parcelas. Para verificar o quão dependente é o método da seleção de galáxias efetuada, assim como possíveis vieses com relação à posição da BCG, executamos o programa para amostras selecionadas de acordo com as definições:

- Todos os aglomerados não eliminados pelo método de parcelas;
- Apenas aglomerados cujas BCGs distam  $d_{BCG} \leq 50$  kpc ou  $d_{BCG} \leq R_{200}$ ;
- Seleção de galáxias membros em magnitude absoluta na banda  $r$  em  $M_r \leq -18.4$  ou  $M_r \leq -20, 5$ ;
- Seleção de galáxias vermelhas em  $u - r > 2, 22$ . Este separador de cor óptico foi definido no trabalho de Strateva et al. (2001), apenas para obter galáxias mais vermelhas de cada aglomerado;
- Algumas combinações das seleções anteriores.

A eficiência do método é inferida através de regressões lineares entre  $R_{200}$  e  $R_{dens}$ . O coeficiente de correlação de Spearman, com e sem correção residual, é utilizado para comparar as regressões. A correção residual é feita para levar em consideração os *outliers* que podem afetar o ajuste linear. O procedimento envolve ajustes sucessivos, sendo que, em cada um deles, o valor-p de Bonferroni <sup>1</sup> (ver (BONFERRONI, 1936)) para a observação mais extrema é calculado. O

<sup>1</sup>A correção de Bonferroni testa cada hipótese individual em um nível de significância de  $\alpha/m$ . Onde  $\alpha$  é o nível global desejado e  $m$  é o número de hipóteses.

procedimento é repetido até que nenhum *outlier* seja detectado a 95% de confiança (vide Cook e Weisberg (1982)). A regressão obtida após a eliminação dos *outliers* tem novamente o coeficiente de correlação calculado. O coeficiente com correção residual é um indicador “robusto” da relação entre  $R_{200}$  e  $R_{dens}$ .

- **Amostra completa** (eliminando apenas os campos mal centrados):

Na figura 5.5, vemos a correlação entre os nossos raios estatísticos  $R_{dens}$  e os raios  $R_{200}$ , para os 806 (dos 947) aglomerados selecionados através do método das parcelas.

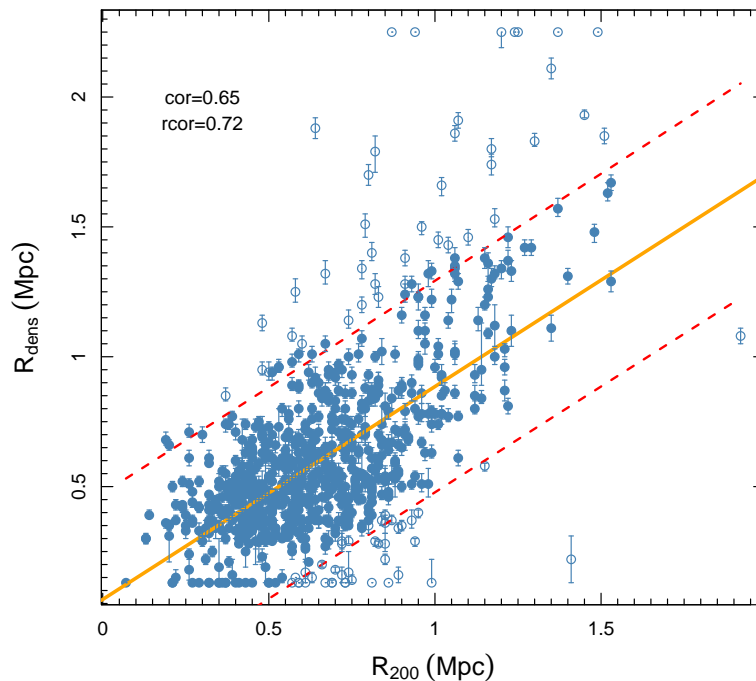


Figura 5.5 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após o método das parcelas. Nenhuma seleção adicional das galáxias é imposta. Pontos cheios são os sistemas utilizados após a remoção de *outliers*, enquanto os pontos vazios são os *outliers*. As linhas tracejadas indicam a região dentro de 95% de confiança em torno do ajuste.

O resultado revela uma correlação razoável entre as duas quantidades, com coeficiente de correlação igual a 0,65 e, após a remoção dos *outliers*, o coeficiente robusto chegando a 0,72. A tabela 5.2 mostra a funcionalidade da técnica, como também a estatística gerada após sua aplicação.

Nome do teste:	Método das parcelas (amostra completa)
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	947 aglomerados
Selecionados:	806 aglomerados
Excluídos:	141 aglomerados
Outliers:	82 aglomerados
Amostra:	724 aglomerados
Resultado:	$rcor=0,72$

Tabela 5.1 – Estatística do método das parcelas sobre os aglomerados analisados na amostra completa.

- **Seleção de membros em  $Mr \leq -20,5$ :**

A figura 5.6 mostra a correlação entre  $R_{dens}$  e  $R_{200}$ , com os aglomerados selecionados através do método das parcelas, agora com corte em  $Mr \leq -20,5$ .

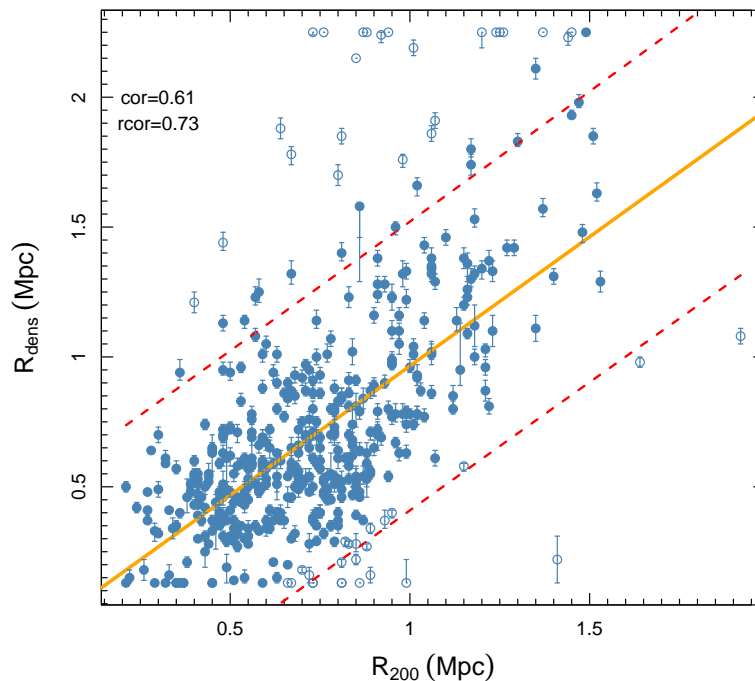


Figura 5.6 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após método das parcelas com corte em magnitude absoluta  $-20,5$ .

Este resultado é aproximadamente o mesmo que o obtido sem a aplicação de nenhum critério de seleção de galáxias. Devemos notar, porém, que o número de aglomerados é bem menor. Isto não acontece em virtude do método de parcelas, que exclui apenas 82 aglomerados, ou via correção residual, que elimina 48 aglomerados. A redução mais drástica está no total de dados inicial, que requer em cada campo um mínimo de 20 galáxias para efetuar a análise. Ao selecionar objetos mais luminosos, o número total de objetos por campo diminui, e a amostra fica menor antes mesmo da aplicação do método.<sup>2</sup>

Nome do teste:	método das parcelas (corte em magnitude absoluta $-20,5$ )
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	554 aglomerados
Selecionados:	472 aglomerados
Excluídos:	82 aglomerados
Outliers:	48 aglomerados
Amostra:	424 aglomerados
Resultado:	rcor=0,73

Tabela 5.2 – Estatística do método das parcelas sobre os aglomerados analisados com o corte em magnitude em  $-20,5$ .

- **Seleção de membros em  $u - r > 2,22$ :**

A figura 5.7 mostra a relação entre os raios  $R_{dens}$  e  $R_{200}$  para o critérios de seleção de galáxias vermelhas.

<sup>2</sup>A seleção de galáxias em  $M_r \leq -18.4$  não altera o tamanho da amostra e reproduz o resultado da amostra completa. Por esta razão não foi apresentado.

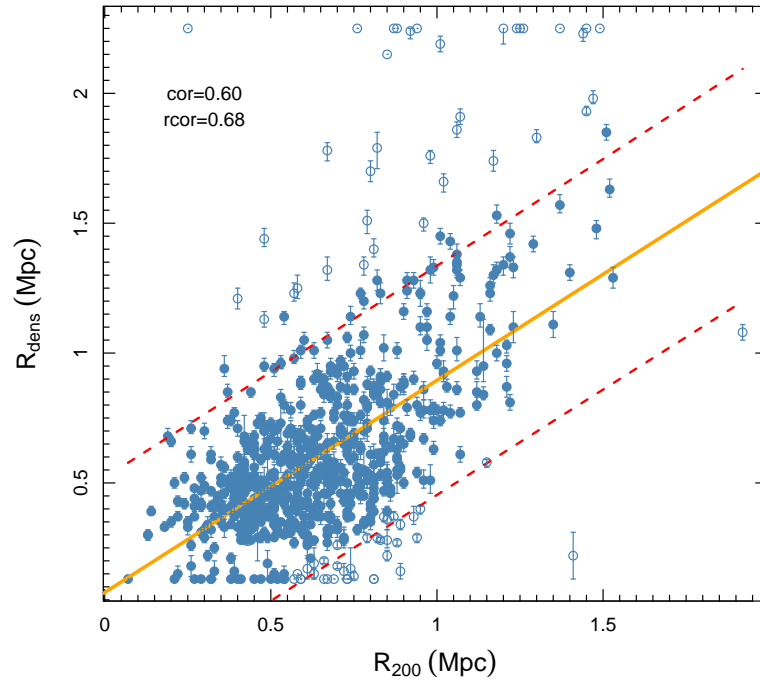


Figura 5.7 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após corte em  $cor\ u - r > 2,22$ .

O resultado é um pouco pior que os outros critérios apresentados anteriormente. Notemos que o número de aglomerados na amostra final é um pouco maior, indicando que a seleção de galáxias vermelhas leva a menos exclusões pelo método de parcelas.

Nome do teste:	método das parcelas (corte em $cor\ u - r > 2,22$ )
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	947 aglomerados
Selecionados:	865 aglomerados
Excluídos:	128 aglomerados
Outliers:	74 aglomerados
Amostra:	791 aglomerados
Resultado:	$rcor=0,68$

Tabela 5.3 – Estatística do método das parcelas sobre os aglomerados analisados com corte em  $cor\ u - r > 2,22$ .

- **Seleção de membros em  $Mr \leq -18,4$  e  $u - r > 2,22$ :**

A figura 5.8 mostra a relação entre os raios  $R_{dens}$  e  $R_{200}$  para o critérios de seleção em magnitude e de galáxias vermelhas.

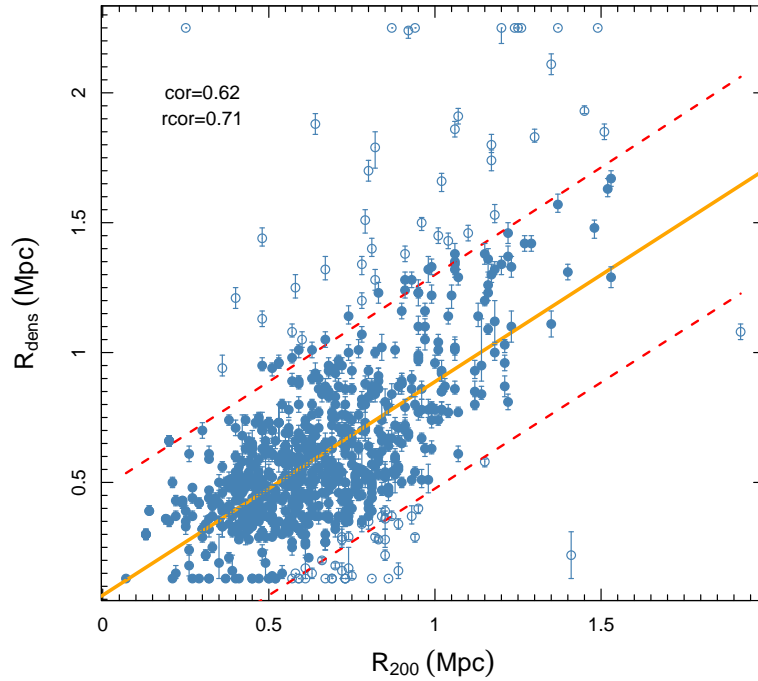


Figura 5.8 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após corte em  $Mr \leq -18,4$  e em  $cor\ u - r > 2,22$ .

A tabela 5.4 mostra a estatística para esta seleção, mais uma vez com resultados semelhantes aos obtidos com a amostra completa.

Nome do teste:	método das parcelas (corte em $Mr \leq -18,4$ e em $cor\ u - r > 2,22$ )
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	947 aglomerados
Selecionados:	806 aglomerados
Excluídos:	141 aglomerados
Outliers:	82 aglomerados
Amostra:	724 aglomerados
Resultado:	$rcor=0,71$

Tabela 5.4 – Estatística do método das parcelas sobre os aglomerados analisados com corte em  $Mr \leq -18,4$  e em  $cor\ u - r > 2,22$ .



- Seleção de membros em  $Mr \leq -20,5$  e  $u - r > 2,22$ :

A figura 5.9 mostra a relação entre os raios  $R_{dens}$  e  $R_{200}$  para o critérios de seleção em magnitude absoluta  $Mr \leq -20,5$  e de galáxias vermelhas.

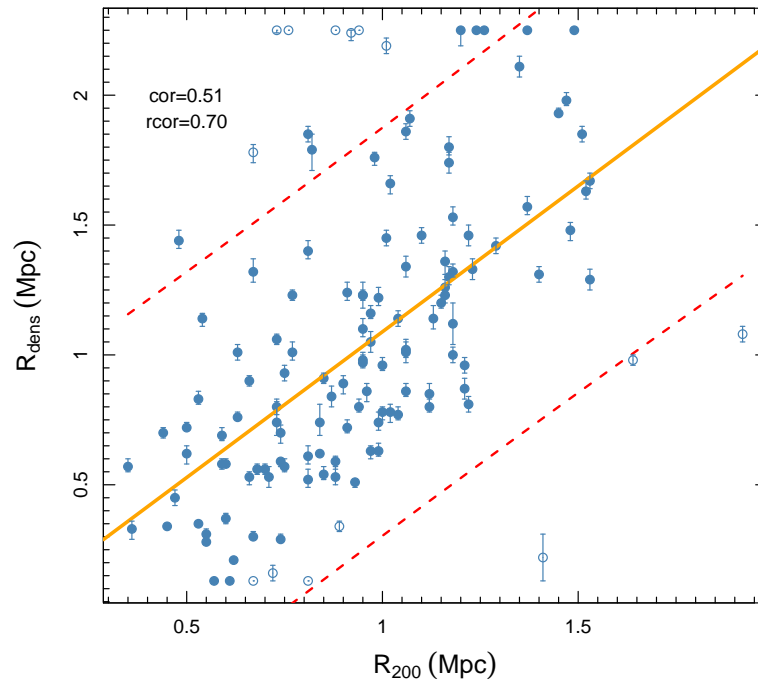


Figura 5.9 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após corte em  $Mr \leq -20,5$  e em cor  $u - r > 2,22$ .

A tabela 5.5 mostra a estatística dos critérios de corte. Notemos a drástica redução de sistemas na amostra, sem haver melhora no resultado da correlação.

Nome do teste:	método das parcelas (corte em $Mr \leq -20,5$ e em cor $u - r > 2,22$ )
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	158 aglomerados
Selecionados:	132 aglomerados
Excluídos:	26 aglomerados
Outliers:	14 aglomerados
Amostra:	118 aglomerados
Resultado:	rcor=0,70

Tabela 5.5 – Estatística do método das parcelas sobre os aglomerados analisados com corte em  $Mr \leq -20,5$  e em cor  $u - r > 2,22$ .

A partir de agora, além de aplicarmos os critérios de seleção ( $Mr \leq -18,2$ ,  $Mr \leq -20,5$  e  $u - r > 2,22$ ), investigaremos como os raios de densidade  $R_{dens}$  se comportam numa seleção de aglomerados cujas BCGs se encontrem a uma distância (i)  $d_{BCG} \leq 50$  kpc do pico de densidade do campo; e (ii) no interior de  $R_{200}$ .

- **Seleção de aglomerados com  $d_{BCG} \leq 50$  kpc ao centro:**

A figura 5.10 mostra a correlação entre os raios de densidade  $R_{dens}$  e os raios  $R_{200}$  para os 162 aglomerados selecionados através desta primeira seleção baseada nas **BCGs**.

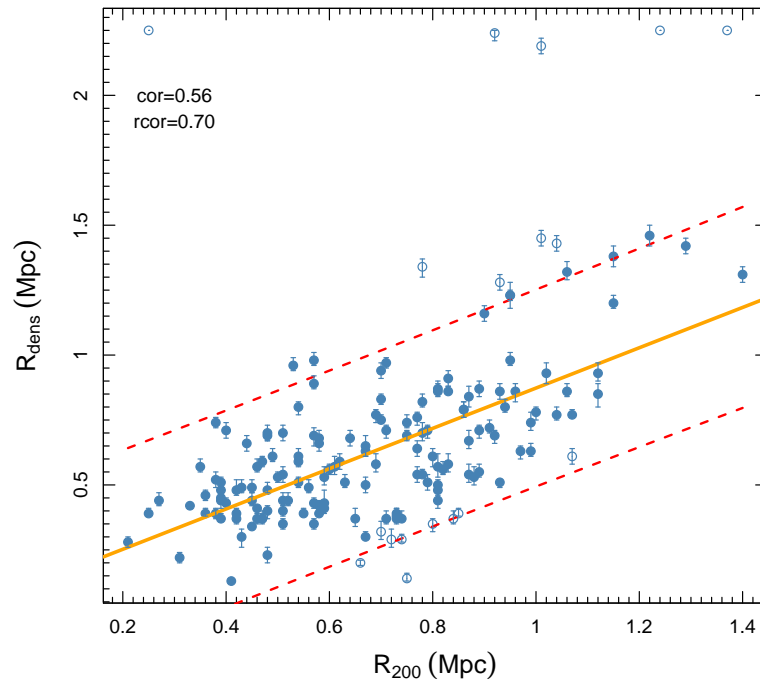


Figura 5.10 – Correlação entre os raios  $R_{dens}$  e  $R_{200}$  com BCGs a 50 kpc dos picos de densidade.

A tabela 5.6 apresenta a estatística do método das BCGs após a sua aplicação. Notemos o pequeno número de sistemas que respeitam este critério. Notemos ainda que o resultado da correlação não melhora em relação à análise da amostra completa.

Nome do teste:	Método <i>BCGs</i> aos picos de densidade
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	947 aglomerados
Selecionados:	162 aglomerados
Excluídos:	785 aglomerados
Outliers:	18 aglomerados
Amostra:	144 aglomerados
Resultado:	$rcor=0,70$

Tabela 5.6 – Estatística do método **BCGs** sobre os aglomerados analisados contendo BCGs a 50 kpc dos picos de densidade..

- Seleção de aglomerados com BCGs dentro de  $R_{200}$ :

Já a figura 5.2 mostra a correlação entre os raios de densidade  $R_{dens}$  e os raios  $R_{200}$  para os 576 aglomerados selecionados pelo método BCG inseridas em  $R_{200}$ . Observa-se que o índice de correlação de *Pearson* robusto aumenta um pouco em relação aos resultados anteriores.

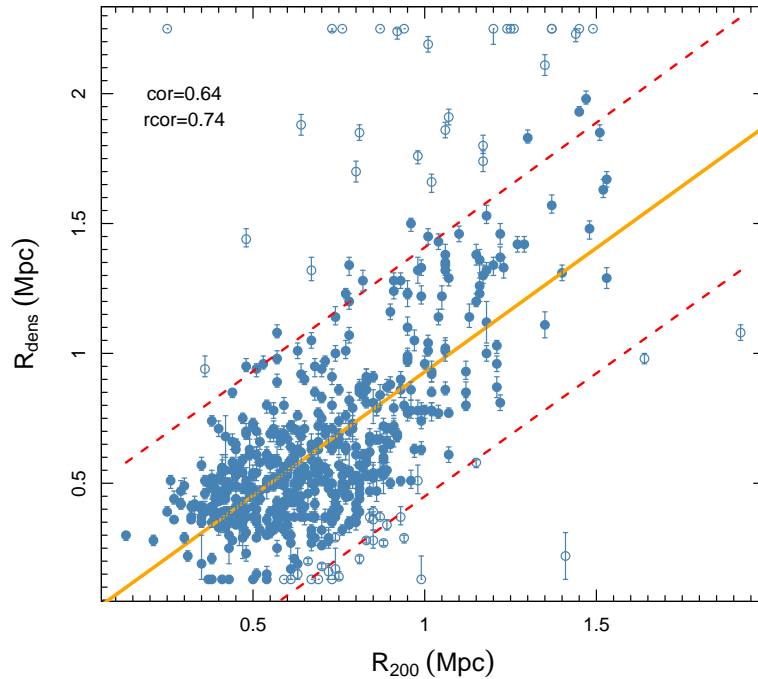


Figura 5.11 – Relação entre os raios com BCGs inseridas nos raios  $R_{200}$ .

Abaixo, mostramos a tabela 5.7 descrevendo a estatística do método com este critério de seleção.

Nome do teste:	Método <i>BCGs</i> nos raios $R_{200}$
O que o teste faz:	Seleciona campos cujas distâncias entre as <b>BCGs</b> e os picos de densidade, de cada aglomerado, estejam inseridas no $R_{200}$
Total de dados:	947 aglomerados
Selecionados:	576 aglomerados
Excluídos:	371 aglomerados
Outliers:	57 aglomerados
Amostra:	519 aglomerados
Resultado:	rcor=0,74

Tabela 5.7 – Estatística do método **BCGs** sobre os aglomerados analisados om **BCGs** dentro de  $R_{200}$ .

- Seleção de aglomerados com  $d_{BCG} \leq 50$  kpc e galáxias com  $M_r \leq -18.4$ :

A figura 5.12 mostra o comportamento entre os raios de densidade  $R_{dens}$  e os raios  $R_{200}$  com seleção de galáxias membros com  $M_r \leq -18,4$  e sistemas cujas BCGs encontram-se a distâncias  $\leq 50$  kpc. O teste selecionou 156 aglomerados.

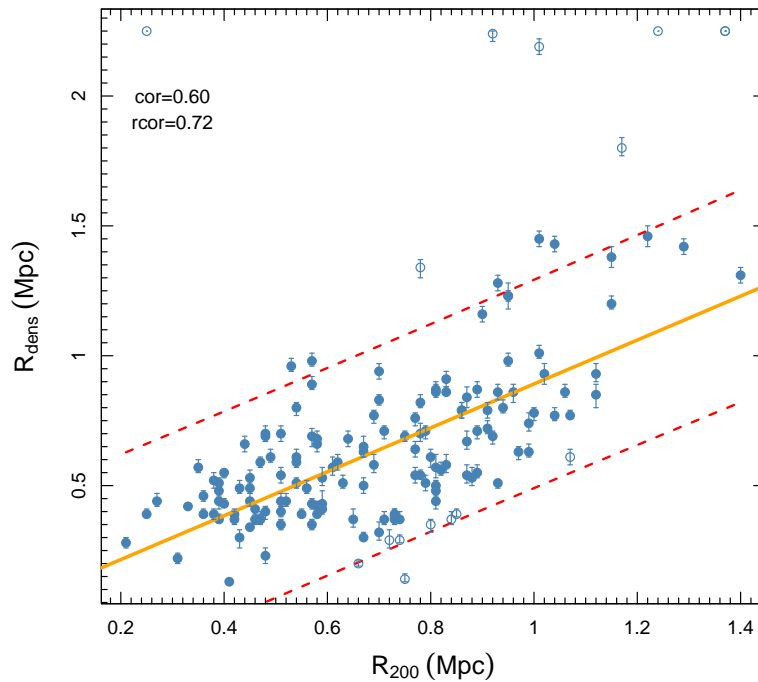


Figura 5.12 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com BCGs a 50 kpc dos picos de densidade e corte em magnitude absoluta em  $M_r \leq -18,4$ .

A tabela 5.8 mostra a estatística do método, que seleciona um número pequeno de aglomerados, sem melhoria na correlação robusta.

Nome do teste:	Método <i>BCGs</i> nos raios $R_{200}$
O que o teste faz:	Seleciona campos cujas distâncias entre as <b>BCGs</b> e os picos de densidade, de cada aglomerado, estejam inseridas no $R_{200}$
Total de dados:	946 aglomerados
Selecionados:	156 aglomerados
Excluídos:	790 aglomerados
Outliers:	16 aglomerados
Amostra:	140 aglomerados
Resultado:	$rcor=0,72$

Tabela 5.8 – Estatística do método **BCGs** sobre os 946 aglomerados analisados.

- Seleção de aglomerados com **BCGs** dentro de  $R_{200}$  e galáxias com  $Mr \leq -18.4$ :

A figura 5.13 mostra a correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com seleção de galáxias membros em  $Mr \leq -18,4$  e sistemas com **BCGs** inseridas em  $R_{200}$ . O teste selecionou 577 aglomerados.

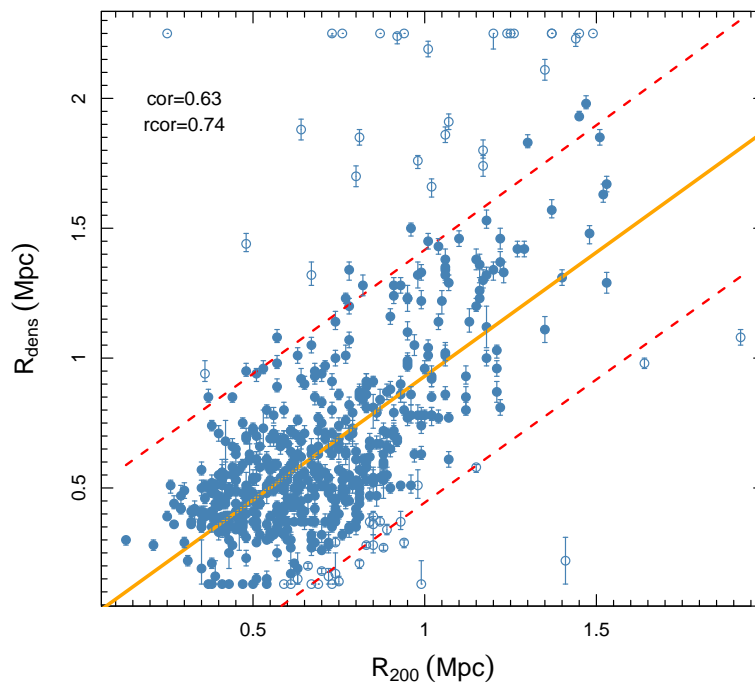


Figura 5.13 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com **BCGs** inseridas nos raios  $R_{200}$  e corte em magnitude absoluta  $-18,4$ .

A tabela 5.9 revela a estatística do método aplicado, que selecionou um número intermediário de aglomerados, 577, obtendo uma correlação robusta um pouco melhor do que aquela obtida para a amostra completa.

Nome do teste:	Método <i>BCGs</i> nos raios $R_{200}$ e corte de magnitude $-18.4$
O que o teste faz:	Seleciona campos cujas distâncias entre as <b>BCGs</b> e os picos de densidade, de cada aglomerado, estejam inseridas no $R_{200}$
Total de dados:	947 aglomerados
Selecionados:	577 aglomerados
Excluídos:	369 aglomerados
Outliers:	58 aglomerados
Amostra:	519 aglomerados
Resultado:	$rcor=0,74$

Tabela 5.9 – Estatística do método **BCGs** sobre os aglomerados analisados.

- **Seleção de aglomerados com  $d_{BCG} \leq 50$  kpc dos centros e galáxias em  $M_r \leq -20.5$ :**

A figura 5.14 mostra a correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$ , agora com seleção de galáxias com  $M_r \leq -20,5$  e distância da BCG ao centro  $\leq 50$  kpc. O teste selecionou apenas 104 aglomerados, e houve uma piora da correlação robusta.

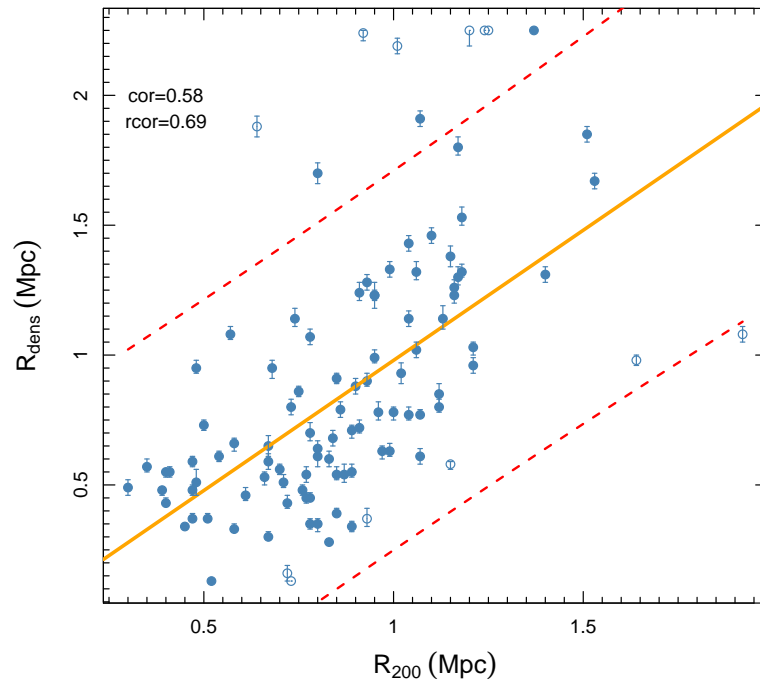


Figura 5.14 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com BCGs a 50 kpc dos picos de densidade e corte em magnitude absoluta  $-20,5$ .

A tabela 5.10 mostra a estatística do método.

Nome do teste:	Teste <i>BCGs</i> a 50 kpc dos picos de densidade e corte em magnitude absoluta $-20,5$
O que o teste faz:	Seleciona campos cujas distâncias entre as <b>BCGs</b> e os picos de densidade, de cada aglomerado, não ultrapassem 0,050 Mpc
Total de dados:	554 aglomerados
Selecionados:	104 aglomerados
Excluídos:	450 aglomerados
Outliers:	12 aglomerados
Amostra:	92 aglomerados
Resultado:	$rcor=0,69$

Tabela 5.10 – Estatística do teste das **BCGs** sobre os 554 aglomerados analisados.

- Seleção de aglomerados com BCGs dentro de  $R_{200}$  e galáxias com  $M_r \leq -20,5$ :



A figura 5.15 apresenta a correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com seleção de galáxias em  $M_r \leq -20,5$  e sistemas com BCGs inseridas em  $R_{200}$ . O teste selecionou 320 aglomerados.

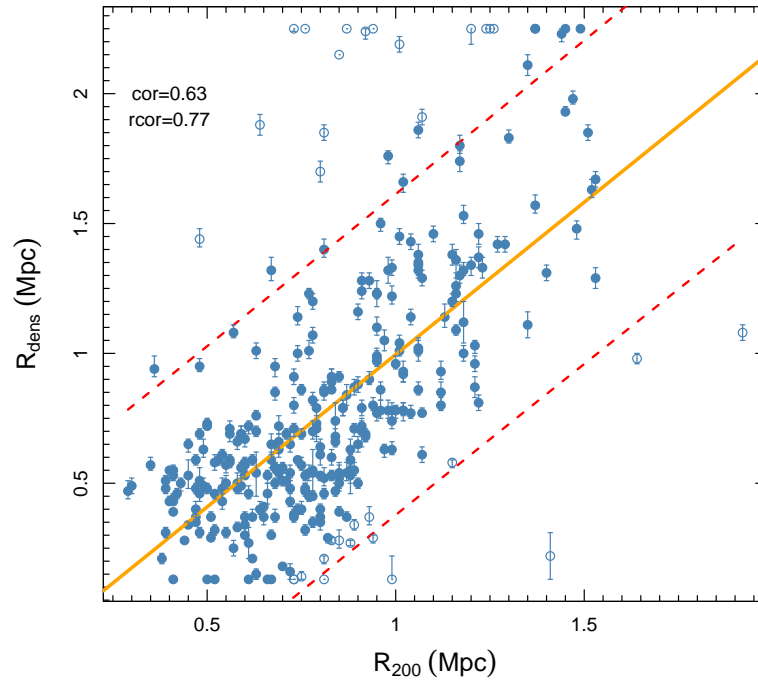


Figura 5.15 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com BCGs inseridas nos raios  $R_{200}$  e corte em magnitude absoluta  $M_r \leq -20.5$ .

A tabela 5.11 mostra a estatística do método. Notemos a significativa melhora na correlação robusta.

Nome do teste:	Método <i>BCGs</i> nos raios $R_{200}$ e corte de magnitude $-20.5$
O que o teste faz:	Seleciona campos cujas distâncias entre as <b>BCGs</b> e os picos de densidade, de cada aglomerado, estejam inseridas no $R_{200}$
Total de dados:	554 aglomerados
Selecionados:	320 aglomerados
Excluídos:	234 aglomerados
Outliers:	32 aglomerados
Amostra:	288 aglomerados
Resultado:	$rcor=0,77$

Tabela 5.11 – Estatística do método **BCGs** sobre os aglomerados analisados com **BCGs** inseridas nos raios  $R_{200}$  e corte em magnitude absoluta  $M_r \leq -20.5$ .

- **Seleção de aglomerados com  $d_{BCG} \leq 50$  kpc e galáxias em  $u - r > 2,22$ :**

A figura 5.16 mostra a correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$ , agora com seleção de galáxias em apresenta a correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com seleção de galáxias em  $u - r > 2,22$  e distância da BCG ao centro  $\leq 50$  kpc. O teste selecionou 251 aglomerados.

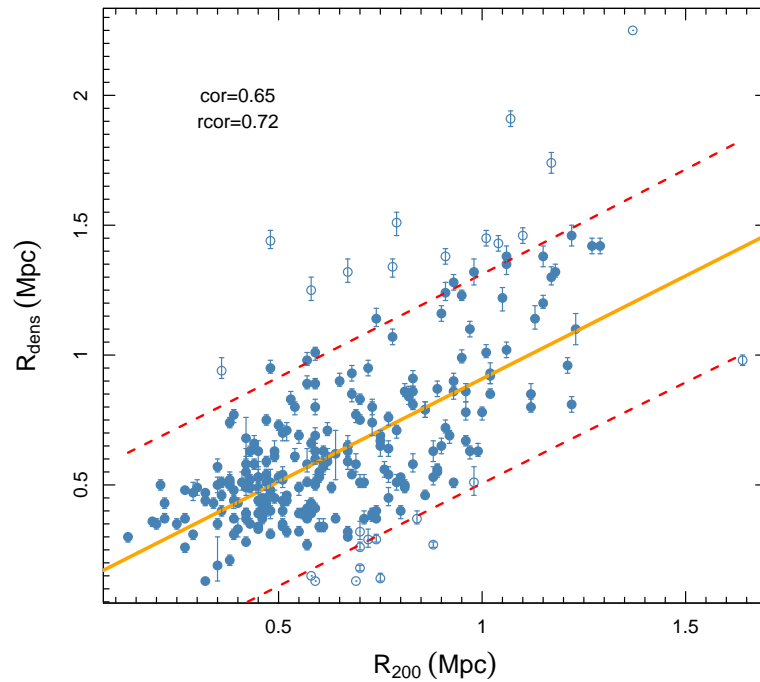


Figura 5.16 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com BCGs a 50 kpc dos picos de densidades e corte em  $u - r > 2,22$ .

A tabela 5.12 mostra a estatística do método. Esta seleção não indica uma melhora em relação à análise feita com amostra completa.

Nome do teste:	Método BCGs a 50 kpc dos picos de densidade e corte em cor $u - r > 2,22$
O que o teste faz:	Seleciona campos cujas distâncias entre as <b>BCGs</b> e os picos de densidade, de cada aglomerado, não ultrapassem 0,050 Mpc
Total de dados:	865 aglomerados
Selecionados:	251 aglomerados
Excluídos:	614 aglomerados
Outliers:	26 aglomerados
Amostra:	225 aglomerados
Resultado:	$rcor=0,72$

Tabela 5.12 – Estatística do método **BCG** sobre os aglomerados analisados com BCGs a 50 kpc dos picos de densidade e corte em cor  $u - r > 2,22$ .

- Seleção de aglomerados com BCGs dentro de  $R_{200}$  e galáxias com  $u - r > 2,22$ :

A figura 5.17 revela a correlação entre os nossos raios  $R_{dens}$  e os raios  $R_{200}$  com seleção de galáxias com  $u - r > 2,22$  e BCGs inseridas em  $R_{200}$ . O teste selecionou 663 aglomerados.

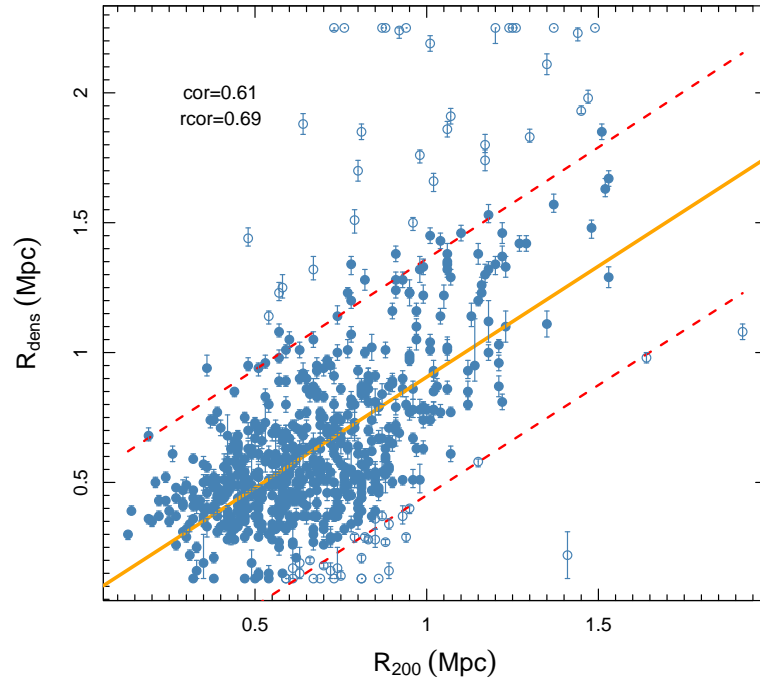


Figura 5.17 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  com BCGs inseridas nos raios  $R_{200}$  e corte em  $u - r > 2.22$ .

A tabela 5.13 mostra a estatística do método. Mais uma vez, a seleção não produziu melhora na correlação.

Nome do teste:	Método BCGs nos raios $R_{200}$ e corte em $cor\ u - r > 2.22$
O que o teste faz:	Seleciona campos cujas distâncias entre as <b>BCGs</b> e os picos de densidade, de cada aglomerado, estejam inseridas no $R_{200}$
Total de dados:	865 aglomerados
Selecionados:	663 aglomerados
Excluídos:	202 aglomerados
Outliers:	68 aglomerados
Amostra:	595 aglomerados
Resultado:	$rcor=0,69$

Tabela 5.13 – Estatística de método **BCG** sobre os aglomerados analisados com BCGs nos raios  $R_{200}$  e corte em  $cor\ u - r > 2.22$

- **Seleção de galáxias em  $Mr \leq -18,4$ ;  $u - r > 2,22$  e sistemas com  $d_{BCG} \leq 50$  kpc:**

A figura 5.18 mostra a relação entre os raios  $R_{dens}$  e  $R_{200}$  para o critérios de seleção de galáxias em  $Mr \leq -18,4$ ;  $u - r > 2,22$  e sistemas com  $d_{BCG} \leq 50$  kpc dos centros.

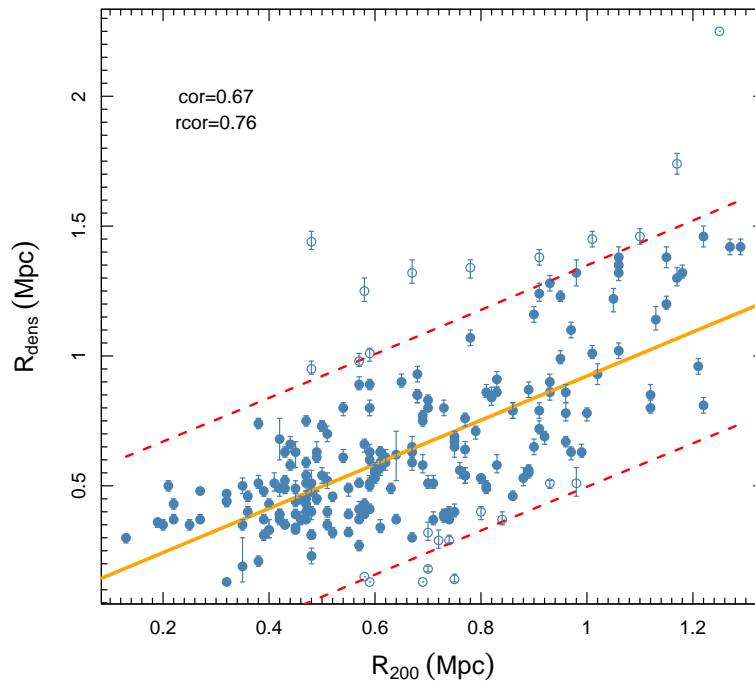


Figura 5.18 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após corte em  $Mr \leq -18,4$ ;  $cor\ u - r > 2,22$  e  $d_{BCG} \leq 50$  kpc dos centros.

A tabela 5.14 mostra a estatística para esta seleção. Notemos uma significativa melhora na correlação robusta. Sendo, até o momento, o segundo melhor resultado obtido, apesar do pequeno número de sistemas analisados ao final.

Nome do teste:	Seleção em $Mr \leq -18,4$ ; cor $u - r > 2,22$ e $d_{BCG} \leq 50$ kpc)
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	806 aglomerados
Selecionados:	223 aglomerados
Excluídos:	583 aglomerados
Outliers:	24 aglomerados
Amostra:	199 aglomerados
Resultado:	rcor=0,76

Tabela 5.14 – Estatística do método das parcelas sobre os 806 aglomerados analisados com corte em  $Mr \leq -18,4$ ; cor  $u - r > 2,22$  e  $d_{BCG} \leq 50$  kpc.

- **Seleção de galáxias em  $Mr \leq -18,4$ ;  $u - r > 2,22$  e sistemas com BCGs inseridas em  $R_{200}$ :**

A figura 5.19 mostra a relação entre os raios  $R_{dens}$  e  $R_{200}$  para o critérios de seleção de galáxias em  $Mr \leq -18,4$ ;  $u - r > 2,22$  e sistemas com BCGs inseridas em  $R_{200}$ .

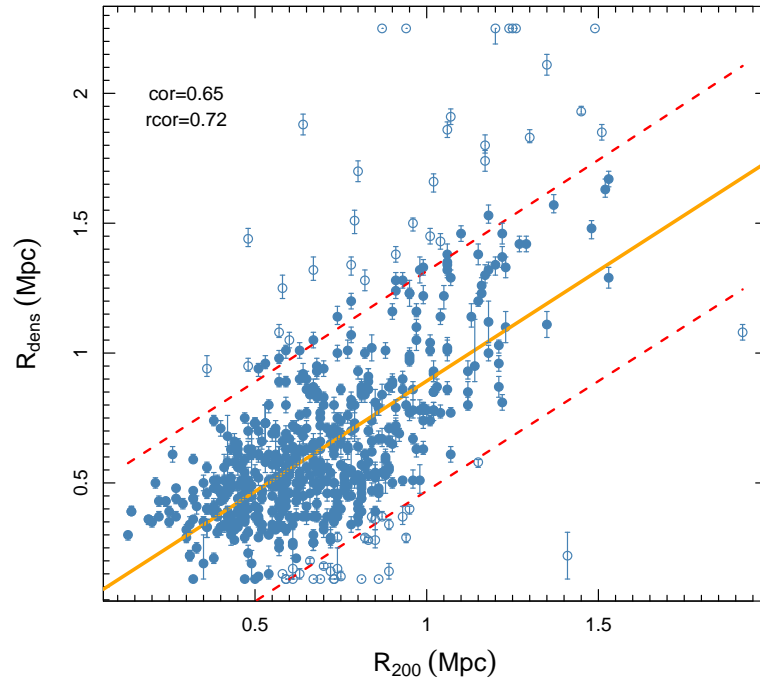


Figura 5.19 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após corte em  $Mr \leq -18,4$ ;  $cor$   $u - r > 2,22$  e BCGs inseridas nos raios  $R_{200}$ .

A tabela 5.15 mostra a estatística dos critérios de corte. A amostra final é maior que no caso anterior, contudo a correlação não apresenta melhora em relação à análise feita com a amostra completa.

Nome do teste:	método das parcelas (corte em $Mr \leq -18,4$ ; $cor$ $u - r > 2,22$ e $d_{BCG} \leq 50$ kpc das BCGs)
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	806 aglomerados
Selecionados:	623 aglomerados
Excluídos:	183 aglomerados
Outliers:	64 aglomerados
Amostra:	559 aglomerados
Resultado:	$rcor=0,72$

Tabela 5.15 – Estatística do método das parcelas sobre os 806 aglomerados analisados com corte em  $Mr \leq -18,4$ ;  $cor$   $u - r > 2,22$  e BCGs inseridas nos raios  $R_{200}$ .

- **Seleção de galáxias em  $Mr \leq -20,5$ ;  $u - r > 2,22$  e sistemas com  $d_{BCG} \leq 50$  kpc:**

A figura 5.20 mostra a relação entre os raios  $R_{dens}$  e  $R_{200}$  para o critérios de seleção de galáxias em  $Mr \leq -20,5$ ;  $u - r > 2,22$  e sistemas com  $d_{BCG} \leq 50$  kpc. Neste caso, além de uma redução severa do número de aglomerados, houve uma piora significativa da correlação.

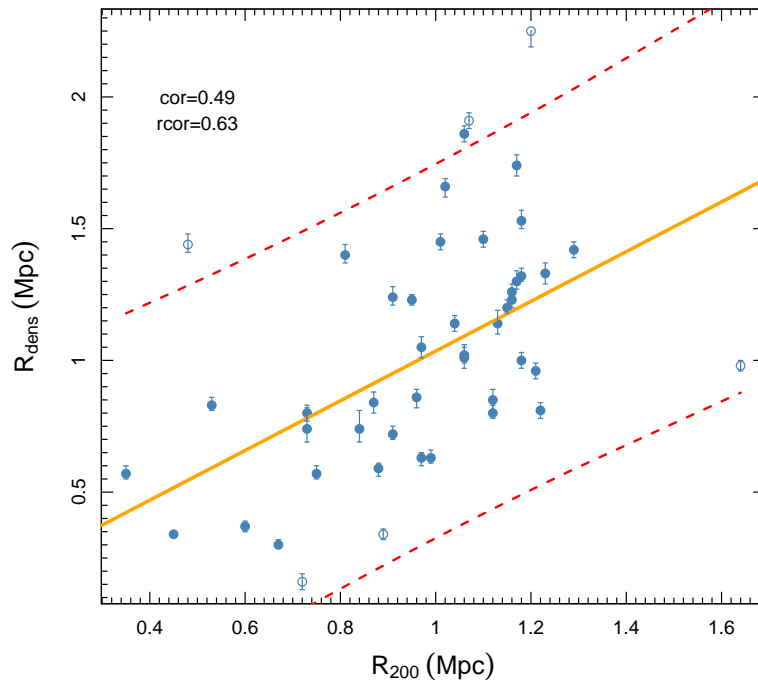


Figura 5.20 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após corte em  $Mr \leq -20,5$ ;  $cor$   $u - r > 2,22$  e  $d_{BCG} \leq 50$  kpc.

A tabela 5.16 mostra a estatística dos critérios de corte.



Nome do teste:	método das parcelas (corte em $Mr \leq -20,5$ ; cor $u - r > 2,22$ e $d_{BCG} \leq 50$ kpc das BCGs)
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	132 aglomerados
Selecionados:	47 aglomerados
Excluídos:	85 aglomerados
Outliers:	6 aglomerados
Amostra:	41 aglomerados
Resultado:	rcor=0,63

Tabela 5.16 – Estatística do método das parcelas sobre os 132 aglomerados analisados com corte em  $Mr \leq -20,5$ ; cor  $u - r > 2,22$  e  $d_{BCG} \leq 50$  kpc.

• **Seleção de membros em  $Mr \leq -20,5$ ;  $u - r > 2,22$  e BCGs inseridas em  $R_{200}$ :**

Por fim, a figura 5.21 mostra a relação entre os raios  $R_{dens}$  e  $R_{200}$  para o critérios de seleção em magnitude  $Mr \leq -20,5$ , galáxias vermelhas e BCGs inseridas nos raios  $R_{200}$ .

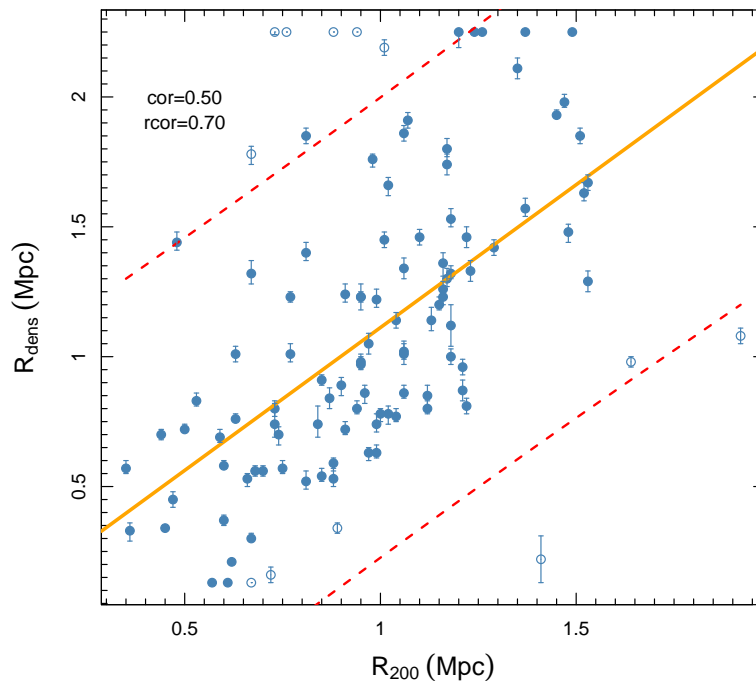


Figura 5.21 – Correlação entre os raios  $R_{dens}$  e os raios  $R_{200}$  após cortes em  $Mr \leq -20,5$ ; cor  $u - r > 2,22$  e BCGs inseridas nos raios  $R_{200}$ .

A tabela 5.17 mostra a estatística dos critérios de corte. Além da redução severa da amostra, a seleção não resultou em melhora da correlação.

Nome do teste:	método das parcelas (corte em $M_r \leq -20,5$ ; cor $u - r > 2,22$ e $d_{BCG} \leq 50$ kpc das BCGs)
O que o teste faz:	Seleciona os campos com picos de densidade no <i>quadrat</i> central
Total de dados:	132 aglomerados
Selecionados:	110 aglomerados
Excluídos:	22 aglomerados
Outliers:	12 aglomerados
Amostra:	98 aglomerados
Resultado:	$rcor=0,70$

Tabela 5.17 – Estatística do método das parcelas sobre os aglomerados analisados com corte em  $M_r \leq -20,5$ ; cor  $u - r > 2,22$  e BCGs inseridas nos raios  $R_{200}$ .

### 5.2.1 Conclusão sobre os testes realizados

A escolha do melhor resultado entre os testes apresentados deve ser um compromisso entre o grau de correlação dos raios  $R_{dens}$  e  $R_{200}$ , assim como do tamanho da amostra resultante após a aplicação dos critérios de seleção. Se fôssemos ranquear os testes com base apenas nos coeficientes de correlação robustos, teríamos os três primeiros:

- 1° Seleção de galáxias com  $M_r \leq -20,5$  e BCGs dentro de  $R_{200}$  ( $rcor = 0,77$ ; 288 aglomerados).
- 2° Seleção de galáxias com  $M_r \leq -18,4$  e BCGs dentro de 50 kpc dos centros ( $rcor = 0,76$ ; 199 aglomerados).
- 2° Seleção de sistemas com BCGs dentro de  $R_{200}$  ( $rcor = 0,74$ ; 519 aglomerados).

Notemos que o primeiro classificado reduz a amostra a  $\sim 30\%$  do tamanho original. O segundo classificado a reduz ainda mais, usando na regressão robusta apenas  $\sim 20\%$  da amostra. O terceiro classificado utiliza  $\sim 55\%$  da amostra original na regressão robusta. E ainda, devemos lembrar que a análise feita sobre a amostra completa, eliminando apenas os campos mal centrados, considera 724 aglomerados,  $\sim 76\%$  da amostra inicial, com um coeficiente  $rcor = 0,72$ . Todos estes resultados apresentaram  $rcor > 0,7$ . Coeficientes de correlação entre 0,7 e 0,9 são usualmente tomados como indicadores de um grau de correlação forte entre as variáveis envolvidas (vide, por exemplo, Weisberg (1987)). Ou seja, utilizar a amostra completa

produz um resultado com mesma interpretação estatística que aqueles apresentados em nosso ranqueamento das melhores seleções. Isto sugere que poderíamos fazer uso da amostra completa na análise subsequente, sem perda significativa de qualidade. Por essa razão, na próxima seção, avaliaremos a riqueza estimada usando, além dos três melhores critérios de seleção, a amostra completa.

### 5.3 Estimando a riqueza

Apresentamos agora as riquezas obtidas pelo nosso método. Inicialmente, avaliamos as regressões entre  $N_{200}$ , informação disponível no catálogo MOCK, e  $N_{dens}$  que obtivemos através da contagem de galáxias dentro de  $R_{dens}$ . A figura 5.22 mostra quatro painéis onde as regressões referentes aos modelos selecionados na seção anterior. Nesta figura vemos que as correlações entre  $N_{200}$  e  $N_{dens}$  são muito fortes, para qualquer caso considerado. Este resultado sugere que o método empregado conduz a um bom *proxy* de  $N_{200}$ . O valor de  $N_{dens}$  pode então ser usado para tentar estabelecer uma relação massa-riqueza.

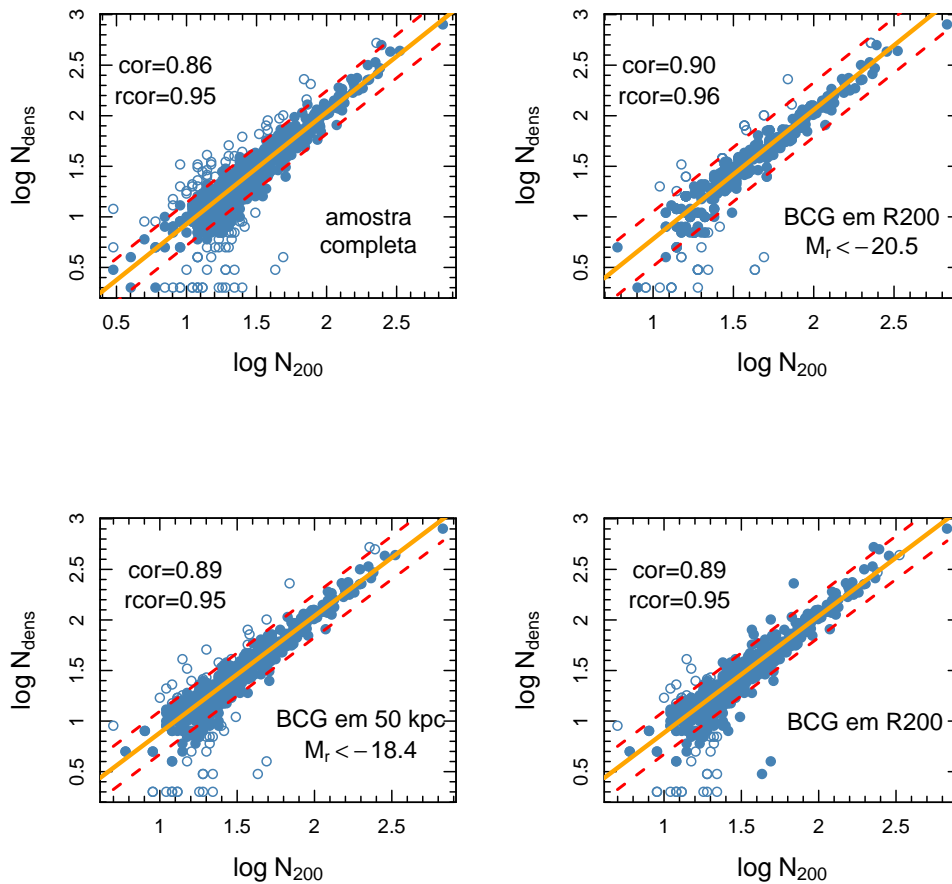


Figura 5.22 – Correlações entre as contagens  $N_{dens}$  e  $N_{200}$ .

Devemos notar, mais uma vez, que o uso da amostra completa, sem qualquer corte, mostra-se tão bom quanto nos casos onde alguma seleção é feita. Isto reforça a ideia de usarmos diretamente a amostra completa, uma vez que esta pressupõe menos informação para a aplicação do método. Basicamente, só precisamos das posições das galáxias.

Podemos ainda verificar as regressões entre  $N_{dens}$  e  $M_{200}$ . Na figura 5.23 vemos os quatro painéis correspondentes. A simples inspeção visual indica que para este par de variáveis as relações possuem maior espalhamento, o que de fato se verifica nos coeficientes de correlação, todos com  $rcor < 0,7$ . Ou seja, um resultado mais modesto que o anterior. Algumas ressalvas devem ser feitas, no entanto. Em primeiro lugar,  $M_{200}$  é determinada a partir de membros selecionados pelo FoF, com significativo grau de contaminação. A análise virial subsequente é afetada pela presença de falsos membros e isto pode gerar um espalhamento em  $M_{200}$ . Em trabalhos futuros, refaremos a identificação dos membros usando métodos específicos de remoção de objetos "intrusos" (por exemplo, o *shifting gapper*), antes de proceder a estimativa da massa, seja por métodos baseados no teorema do virial ou não.

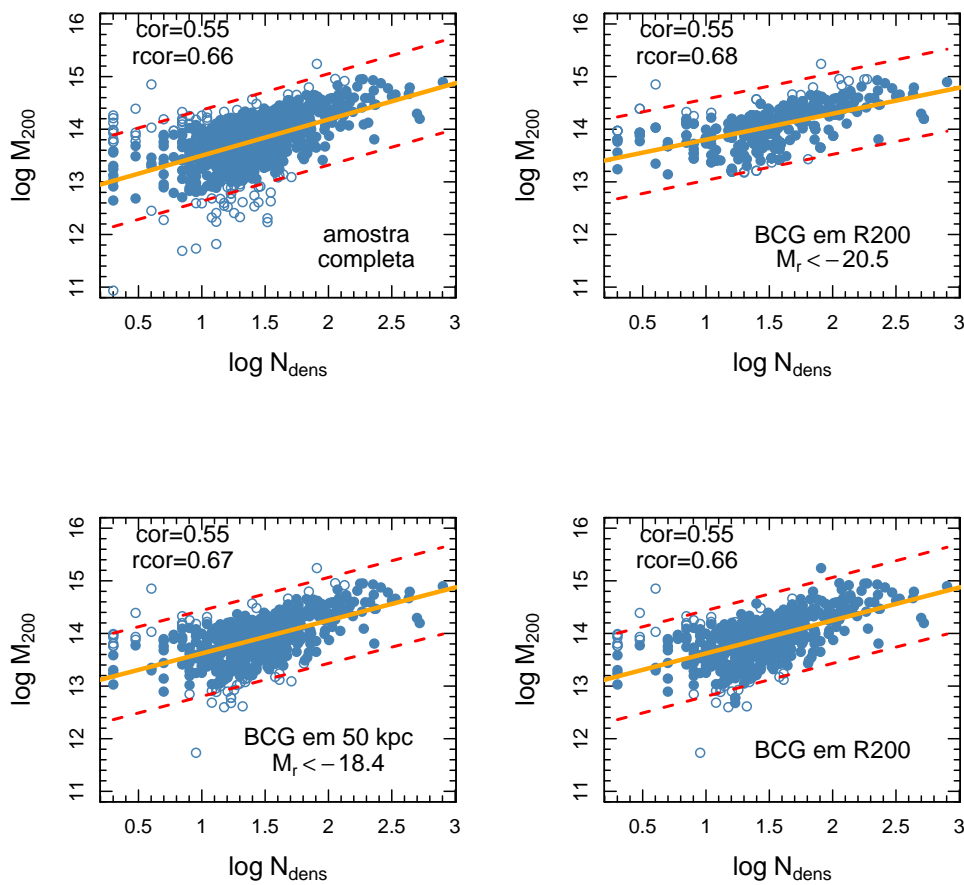


Figura 5.23 – Correlações entre  $N_{dens}$  e  $M_{200}$ .

Resumindo, podemos concluir o seguinte:

- O método das parcelas somado à obtenção de  $R_{dens}$  levam a uma boa predição de  $N_{200}$ , dado por  $N_{dens}$ .
- O uso de uma amostra fotométrica, sem imposição de qualquer tipo de corte, produz resultados tão bons quanto aqueles obtidos para diferentes critérios de seleção.
- O método não é tão eficiente para prever  $M_{200}$ .

### 5.3.1 Aplicação da metodologia sobre a amostra observada

Finalmente, aplicamos o método *quadrats*+ $R_{dens}$  sobre a amostra de Wen e Han (2013). Dos 132.684 aglomerados listados no catálogo original, fizemos uma redução para 819 sistemas a  $z < 0.4$ , incluídos em campos de aproximadamente  $1,5^\circ \times 1,5^\circ$ , contendo ao menos 150 galáxias. Este corte tem como objetivo excluir da análise campos excessivamente pobres, muito próximos do ruído estatístico. O tamanho total da amostra é comparável com o catálogo MOCK.

Após a execução dos *scripts* de análise, sem impor nenhuma seleção particular de galáxias, obtivemos 689 aglomerados para os quais foi possível estimar  $R_{dens}$  e  $N_{dens}$ . Na figura 5.24, vemos os histogramas referentes às distribuições de  $R_{dens}$  e  $N_{dens}$  para a amostra real comparadas com as distribuições de  $R_{200}$  e  $N_{200}$  da amostra simulada. Os raios apresentam valores  $\langle R_{200} \rangle = 0.61 \pm 0.25$  Mpc e  $\langle R_{dens} \rangle = 0.73 \pm 0.36$  Mpc.<sup>3</sup> Ou seja, os raios de densidade são ligeiramente maiores, diferença esta que é confirmada pelo teste de permutação (permTS do R), que acusa um valor-p de 0.0023. Isto provavelmente se deve a objetos com  $R_{dens} > 2.0$  Mpc, que aparecem apenas na amostra observada. Já as contagens apresentam uma característica interessante: para objetos com o logaritmo dos raios menor que 1.5,  $R_{dens} > R_{200}$ , e a partir desse raio, a tendência se inverte, com  $R_{200} > R_{dens}$ . Essas tendências podem significar algum erro sistemático dentro do procedimento. Contudo, fazemos a ressalva de que o catálogo MOCK possui contaminação de membros introduzida pelo FoF. Em estudos futuros, redefinir os membros do aglomerado será importante para a verificação de possíveis erros sistemáticos introduzidos pelo procedimento.

<sup>3</sup>A distância em Mpc foi obtida a partir do *redshift* fotométrico indicado no trabalho de Wen e Han (2013).

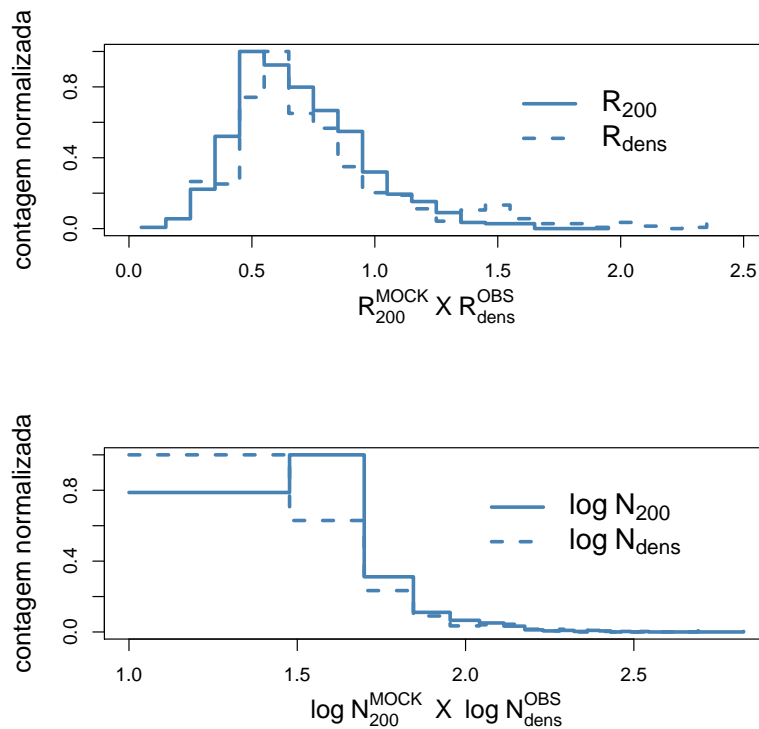


Figura 5.24 – Histogramas de  $R_{\text{dens}}$  e  $N_{200}$  para a amostra tirada do catálogo de Wen e Han (2013).

A seguir apresentamos o exemplo de um campo rico da amostra observada. Na figura 5.25 vemos, no painel da esquerda, a distribuição de galáxias na região do aglomerado. Sobre esta distribuição é que o método *quadrats*+ $R_{\text{dens}}$  é aplicado. Inicialmente, *quadrats* testa a hipótese de aleatoriedade espacial e verifica se o sistema é mal centrado. Uma vez que nenhuma das duas situações é verificada, o campo é mantido na análise e passamos ao ponto seguinte, ilustrado pelo painel da direita, onde vemos a suavização do campo realizada pelo programa *pdfCluster* através de um *kernel* adaptativo. O programa identifica ainda o pico de densidade máxima, que é atribuído ao centro do sistema. Na figura, este ponto é indicado por um "X". Em seguida, distâncias a esse centro são calculadas e o programa traça o perfil de densidade superficial, que é comparado com a calibração feita com o catálogo MOCK, a fim de determinar o raio correspondente a  $\Sigma_{200}$ . A esse valor está associada uma escala de tamanho. No exemplo apresentado,  $R_{\text{dens}} = 0.28$  graus ou, usando a informação do *redshift* fotométrico do aglomerado,  $z = 0.13$ , concluímos que  $R_{\text{dens}} = 1.84$  Mpc. Dentro desse raio encontram-se 353 galáxias, que corresponde a  $N_{\text{dens}}$ .

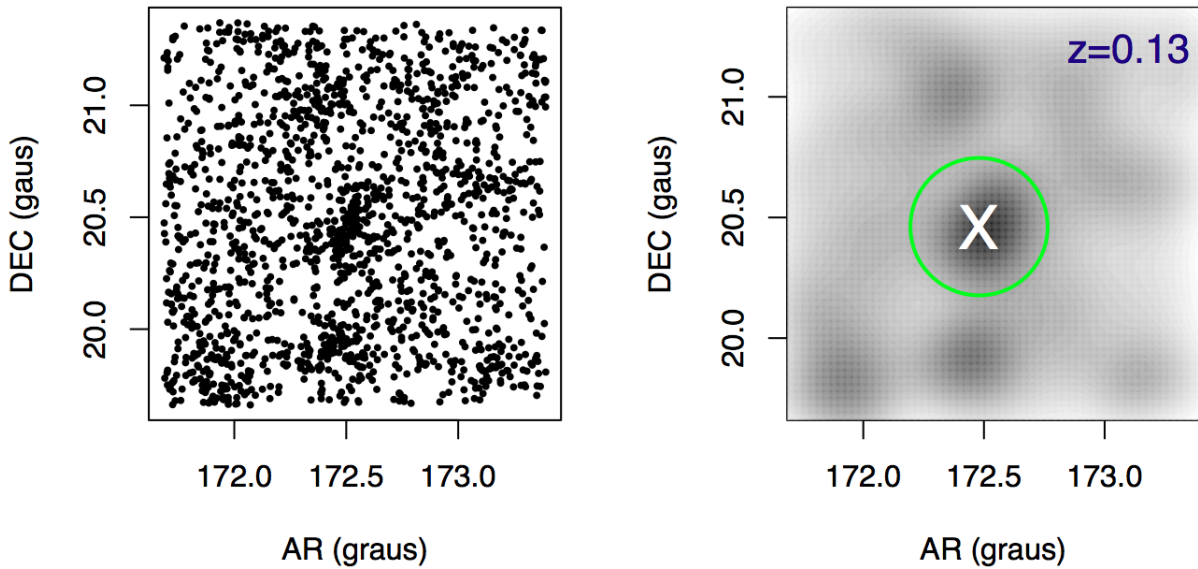


Figura 5.25 – Aplicação do programa *pdfCluster* ao aglomerado 39 da amostra. À esquerda: a distribuição de galáxias em coordenadas equatoriais. À direita: a identificação do pico de densidade do sistema indicado por um “X” e a determinação do raio de densidade representado pelo círculo verde.

O próximo passo do trabalho é tentar verificar se dentro de  $R_{dens}$  existem subestruturas. Esta etapa do trabalho é descrita na próxima seção.

## 5.4 Identificação de subestruturas nos catálogos simulado e real

Apresentamos agora os resultados de três indicadores estatísticos (Distância de Hellinger (HD), teste de Anderson-Darling (AD) e teste de Dressler-Schectman (DS) – descritos na Seção 4.6) sobre os dados do catálogo simulado, a fim de obter os respectivos diagnósticos baseados em espectroscopia. Em seguida, calculamos os indicadores fotométricos  $\Delta M_{12}$ ,  $\Delta M_{14}$ ,  $M_1$ ,  $I_{CE}$ ,  $I_{AC}$ ,  $D_{BCG}$ ,  $V_{BCG}$  e  $N_{picos}$  (descritos nas Seções 4.1 e 4.7). A comparação entre os métodos espectroscópicos e fotométricos é discutida na sequência, onde critérios para analisar dados unicamente fotométricos são definidos. Finalmente, estes critérios são aplicados sobre a amostra de Wen e Han (2013).

### 5.4.1 Diagnósticos de HD, AD e DS sobre o catálogo simulado

Inicialmente, aplicamos os testes estatísticos de HD, AD e DS aos aglomerados selecionados por meio do método das parcelas, ou seja, deixamos na amostra os sistemas bem centrados. Em seguida, aplicamos os testes sobre os membros espectroscópicos confirmados dentro de  $R_{200}$ . Apenas sistemas com 20 galáxias ou mais dentro de  $R_{200}$  foram analisados. Essa imposição

Propriedade	G	NG	valor-p
$\Delta M_{12}$	$0.547 \pm 0.027$	$0.342 \pm 0.037$	0.0003
$\Delta M_{14}$	$0.959 \pm 0.030$	$0.737 \pm 0.047$	0.0006
$M_1$	$-22.316 \pm 0.035$	$-22.067 \pm 0.055$	0.0009
$I_{CE}$	$0.517 \pm 0.060$	$0.486 \pm 0.014$	0.0299
$I_{AC}$	$4.113 \pm 2.270$	$2.387 \pm 0.975$	0.0003
$D_{BCG}$	$0.404 \pm 0.050$	$0.516 \pm 0.095$	0.3069
$V_{BCG}$	$139.101 \pm 10.958$	$171.244 \pm 31.090$	0.2271
$A^2$	$0.262 \pm 0.005$	$0.895 \pm 0.066$	< 0.0001
$\Delta_{DS}$	$1.239 \pm 0.018$	$2.188 \pm 0.073$	< 0.0001

Tabela 5.18 – Propriedades a partir do indicador HD

Propriedade	G	NG	valor-p
$\Delta M_{12}$	$0.496 \pm 0.018$	$0.387 \pm 0.033$	0.0203
$\Delta M_{14}$	$0.908 \pm 0.020$	$0.764 \pm 0.040$	0.0072
$M_1$	$-22.148 \pm 0.022$	$-22.154 \pm 0.052$	0.9159
$I_{CE}$	$0.490 \pm 0.004$	$0.500 \pm 0.012$	0.3993
$I_{AC}$	$3.058 \pm 1.067$	$1.954 \pm 0.760$	0.6851
$D_{BCG}$	$0.481 \pm 0.041$	$0.525 \pm 0.080$	0.6905
$V_{BCG}$	$131.932 \pm 7.622$	$194.078 \pm 29.718$	0.0062
$HD$	$0.132 \pm 0.020$	$0.192 \pm 0.009$	< 0.0001
$\Delta_{DS}$	$1.332 \pm 0.016$	$2.205 \pm 0.061$	< 0.0001

Tabela 5.19 – Propriedades a partir do indicador AD

vem do mínimo de galáxias para que HD forneça resultados confiáveis (vide de Carvalho et al. 2017). Nenhuma outra regra de seleção foi imposta sobre os dados. A amostra final contém 595 aglomerados de galáxias com massas maiores que  $10^{14} M_{\odot}$ . Em seguida, calculamos os indicadores fotométricos  $\Delta M_{12}$ ,  $\Delta M_{14}$ ,  $M_1$ ,  $D_{BCG}$ ,  $V_{BCG}$ ,  $I_{CE}$ ,  $I_{AC}$  e  $N_{picos}$  para os mesmos 595 aglomerados.

As tabelas 5.4.1, 5.4.1 e 5.4.1 sintetizam os resultados. Nelas apresentamos, para cada um dos indicadores espectroscópicos, os valores médios dos indicadores fotométricos assim como das estatísticas dos testes espectroscópicos complementares. Os valores médios são calculados para subamostras gaussiana (G) e não-gaussiana (NG) (no caso dos indicadores HD e AD) e subamostras com subestruturas (CS) e sem subestruturas (SS) (no caso do indicador DS). Na última coluna da tabela consta o resultado do teste de duas amostras com permutação (função *permTS* do R), onde as subamostras comparadas são  $G \times NG$  e  $CS \times SS$ . Por essa coluna, vemos quais indicadores fotométricos apresentam sensibilidade ao indicador espectroscópico particular.

A análise dessas tabelas revela de imediato que os valores das estatísticas dos testes espectroscópicos são sensíveis, cada uma delas, aos demais indicadores. Ou seja HD, AD e DS mostram algum grau de consistência interna. Aglomerados com valores menores de HD,  $A^2$  e  $\Delta_{DS}$  provavelmente são sistemas mais regulares, o que significa que possuem distribuição



Propriedade	SS	CS	valor-p
$\Delta M_{12}$	$0.488 \pm 0.021$	$0.472 \pm 0.024$	0.6082
$\Delta M_{14}$	$0.909 \pm 0.024$	$0.860 \pm 0.028$	0.1979
$M_1$	$-22.069 \pm 0.025$	$-22.265 \pm 0.034$	$< 0.0001$
$I_{CE}$	$0.475 \pm 0.005$	$0.515 \pm 0.006$	$< 0.0001$
$I_{AC}$	$1.813 \pm 0.341$	$4.498 \pm 2.220$	0.1556
$D_{BCG}$	$0.491 \pm 0.049$	$0.482 \pm 0.058$	0.9116
$V_{BCG}$	$131.781 \pm 8.498$	$152.640 \pm 14.427$	0.1864
$HD$	$0.148 \pm 0.003$	$0.192 \pm 0.009$	$< 0.0001$
$A^2$	$0.379 \pm 0.010$	$0.452 \pm 0.023$	$< 0.0011$

Tabela 5.20 – Propriedades a partir do indicador DS

de velocidades gaussiana e não apresentam subestruturas. Além disso, as tabelas 5.4.1, 5.4.1 e 5.4.1 mostram que o indicador espectroscópico HD foi aquele para o qual um maior número de indicadores fotométricos apresentaram diferenças significativas entre subamostras G e NG. Resumidamente, vemos a sensibilidade dos indicadores (fotométricos e espectroscópicos) para com os testes estatísticos (HD, AD e DS):

HD  $\rightarrow \Delta M_{12}, \Delta M_{14}, M_1, ICE, IAC$

AD  $\rightarrow \Delta M_{12}, \Delta M_{14}, VBCG$

DS  $\rightarrow M_1, ICE$

Isso sugere que HD pode ser um bom calibrador para o uso dos indicadores fotométricos na análise subsequente. Note que HD contém todos os indicadores fotométricos sensíveis à classificação dos sistemas obtida via DS ( $M_1$  e  $I_{CE}$ ), além de conter dois dos três indicadores fotométricos sensíveis a AD ( $\Delta M_{12}$  e  $\Delta M_{14}$ ). E ainda contabiliza o indicador  $I_{AC}$ . Essa constatação nos faz escolher HD como a base da calibrador a ser feita. No final da calibração, como uma regra adicional de decisão, podemos usar o indicador  $N_{picos}$ , calculado com o *pdfCluster* (Seção 4.7.4).

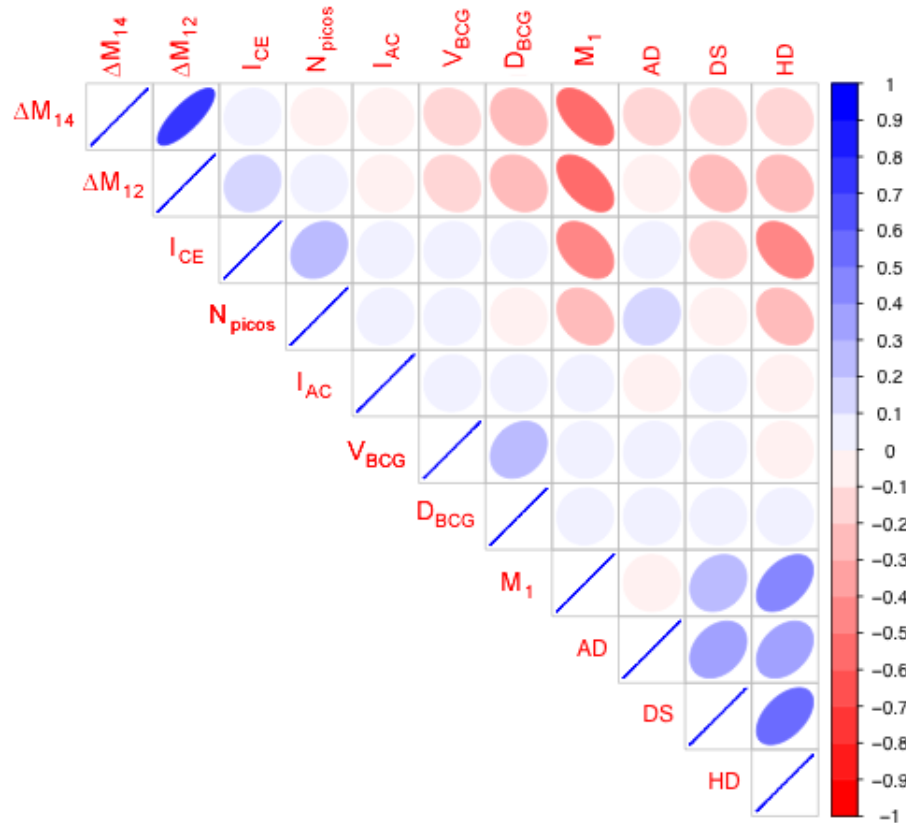


Figura 5.26 – Matriz de correlação entre os indicadores de subestruturas.

A figura 5.26 mostra a correlação entre todos os indicadores de subestruturas para os aglomerados do catálogo simulado. As elipses mais alongadas e as cores mais intensas indicam correlações (em azul) ou anticorrelações (em vermelho) mais fortes. Note que o indicador  $M_1$  é aquele que mais se anticorrelaciona com os demais indicadores fotométricos ( $\Delta M_{12}$ ,  $\Delta M_{14}$ ,  $I_{CE}$  e  $N_{picos}$ ), enquanto HD é o indicador espectroscópico com mais correlações ou anticorrelações (fortes ou moderadas) possui com os indicadores fotométricos ( $\Delta M_{12}$ ,  $\Delta M_{14}$ ,  $M_1$ ,  $I_{CE}$  e  $N_{picos}$ ). A calibração entre indicadores espectroscópicos e fotométricos deve começar com o estudo da relação entre HD e  $M_1$ , que será feita a seguir.

#### 5.4.2 Calibração do método fotométrico

A figura 5.27 mostra a relação entre HD e  $M_1$ . As cores indicam as classificações feitas a partir de HD de acordo com de Carvalho et al. (2017): G (vermelho), NG (azul) e IND – INDETERMINADO (verde). Esta última categoria refere-se aos sistemas cuja validação de diagnóstico através de reamostragens da distribuição de velocidades tem os diagnósticos G e NG com frequência menor que 70%. É importante notar que tanto sistemas G como NG possuem uma anticorrelação com  $M_1$ , ou seja, quanto mais luminosa a BCG do sistema, menor o valor de

HD. Notemos ainda que as anticorrelações possuem inclinações diferentes, sendo mais íngreme para sistemas NG, para os quais, se  $M_1 > -22.5$ , os valores de HD sobem significativamente, enquanto para sistemas G, o crescimento é mais moderado. Essa diferença de comportamento permite identificar uma separação em torno do valor  $M_1 = -22.5$ , abaixo do qual contabilizamos uma fração de 74% de sistemas G, e acima do qual temos uma queda que chega a 32% de sistema G.

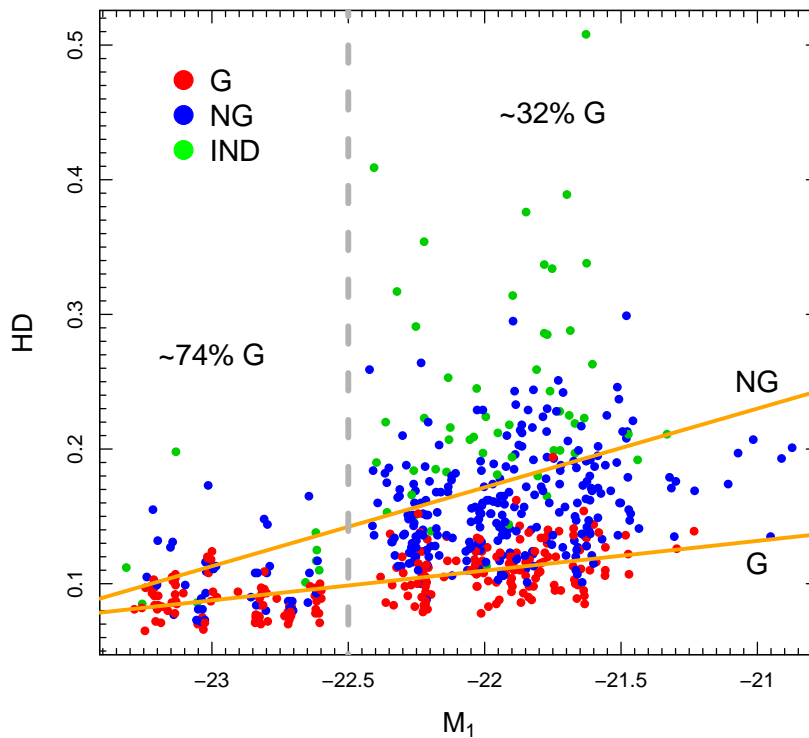


Figura 5.27 – Relação entre  $M_1$  e HD. G – gaussianos (vermelho), NG – não-gaussianos (azul) e IND (verde). As regressões para sistemas G e NG são mostradas em amarelo.

A tabela 5.21 mostra como variam os indicadores usando-se como divisor  $M_1 = -22.5$ . Notemos que diferenças significativas entre as médias das subamostras acima e abaixo desse valor ocorrem para  $\Delta M_{12}$ ,  $\Delta M_{14}$  e  $I_{CE}$  (entre os indicadores fotométricos) e HD e DS (entre os indicadores espectroscópicos). Este resultado reforça a adoção de  $M_1 > -22.5$  como um critério para indicar maior probabilidade de um sistema possuir subestruturas.

Critérios adicionais podem ser impostos através dos demais indicadores fotométricos. Na figura 5.28 vemos as relações entre  $M_1$  e  $\Delta M_{12}$ ,  $\Delta M_{14}$ ,  $I_{CE}$  e  $N_{picos}$ . A linha tracejada vertical indica o separador  $M_1 = -22.5$ . As linhas horizontais indicam as regiões a partir de onde a probabilidade de uma classificação gaussiana de HD é maior ou igual a 70%. Para o indicador  $\Delta M_{12}$ , o corte fica em  $\Delta M_{12} > 0.7$  (que corresponde a 75% de sistemas G); para

Tabela 5.21 – Propriedades a partir do indicador  $M_1$ 

propriedade	$< -22.5$	$> -22.5$	valor-p
$\Delta M_{12}$	$0.871 \pm 0.025$	$0.365 \pm 0.016$	$< 0.0001$
$\Delta M_{14}$	$1.290 \pm 0.026$	$0.769 \pm 0.019$	$< 0.0001$
ICE	$0.563 \pm 0.006$	$0.470 \pm 0.005$	$< 0.0001$
IAC	$1.467 \pm 0.307$	$3.341 \pm 1.203$	0.3959
DBCG	$0.385 \pm 0.066$	$0.518 \pm 0.044$	0.1331
VBCG	$139.885 \pm 14.726$	$140.408 \pm 9.061$	0.9773
HD	$0.095 \pm 0.002$	$0.154 \pm 0.003$	$< 0.0001$
$A^2$	$0.405 \pm 0.018$	$0.410 \pm 0.013$	0.8366
$\Delta_{DS}$	$1.312 \pm 0.024$	$1.943 \pm 0.020$	0.0021

os demais indicadores, temos  $\Delta M_{14} > 1.0$  (77% G),  $I_{CE} > 0.5$  (71%) e  $N_{picos} = 1$  (92% G). Deve-se notar que os indicadores  $\Delta M_{14}$  e  $N_{picos}$  são aqueles com maiores frações de sistemas G com os cortes adotados. Esses serão os nossos critérios para definir um sistema com maior probabilidade de ser regular (ou seja,  $\Delta M_{12}$  e  $I_{CE}$  não serão utilizados). Por oposição, um sistema será dito irregular quando pelo menos um desses critérios não for satisfeito. Este método de classificação será empregado na análise da amostra de Wen e Han (2013). Sua aplicação ao catálogo MOCK, usando apenas os dados fotométricos, leva a  $\sim 15\%$  de sistemas regulares, um valor consideravelmente mais baixo do que os 39% de aglomerados G obtidos por HD (contra 11% NG e 50% IND). Isso se deve ao fato de as restrições que maximizam a probabilidade de sistemas regulares vistas na figura ?? correspondem a regiões com baixa população de objetos. Uma quantidade significativa de sistemas G pode ser vista nos demais espaços da figura, mas são descartados por habitarem regiões dos gráficos com muitos sistemas G e IND. Isso é uma limitação do método e corresponde à dificuldade em caracterizar sistemas regulares com informação fotométrica somente.

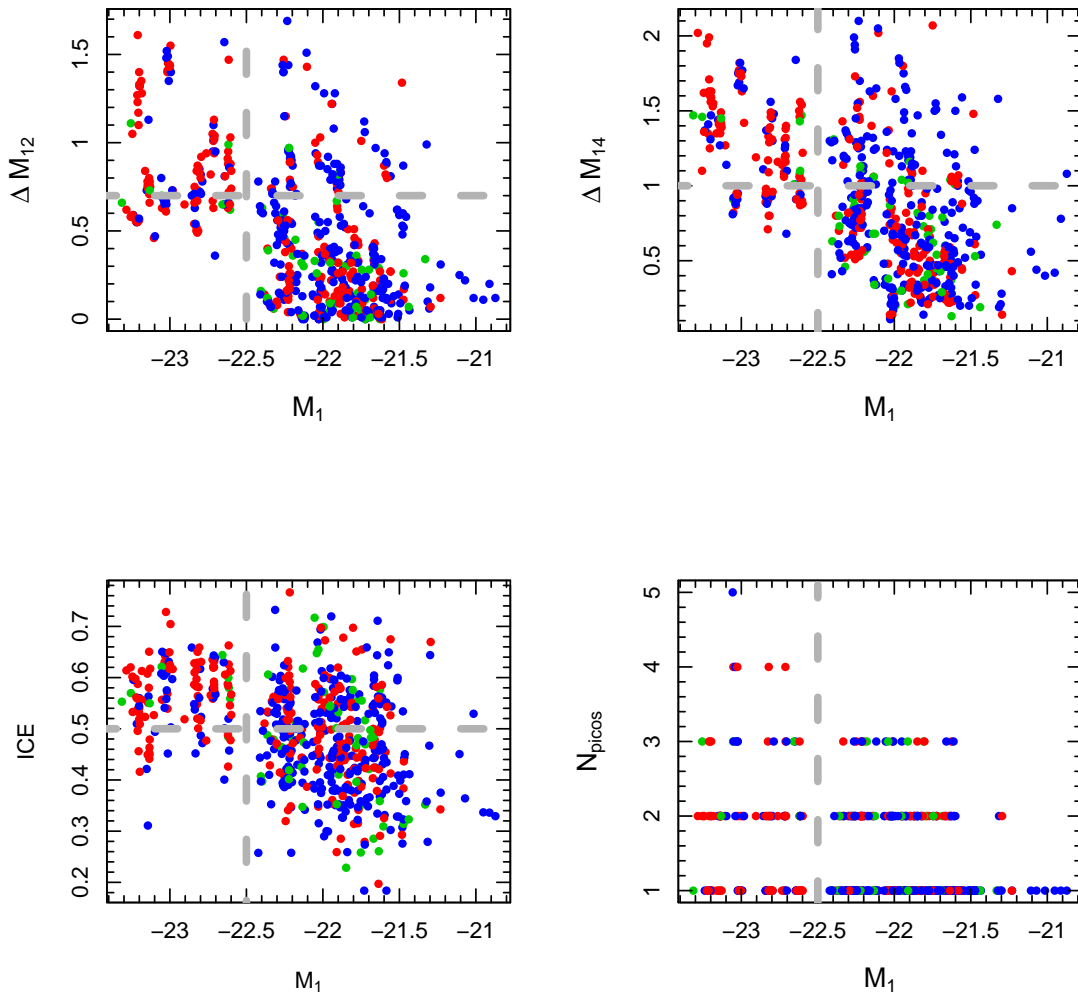


Figura 5.28 – Relação entre  $M_1$  e HD. G – gaussianos (vermelho), NG – não-gaussianos (azul) e IND (verde). As regressões para sistemas G e NG são mostradas em amarelo.

### 5.4.3 Aplicação do método $M_1 - \Delta_{M14} - N_{picos}$ aos dados observados

Finalmente, aplicamos este método ao conjunto de 689 aglomerados para os quais foi possível estimar  $R_{dens}$  na etapa anterior. Deve-se notar que o indicador de subestruturas opera somente dentro deste raio, ou seja, subestruturas periféricas e possíveis grupos de acreção não serão considerados como indicadores de perturbações ao sistema. Dado que o método apoia-se no parâmetro HD, o qual fornece diagnósticos mais confiáveis a partir de um mínimo de 20 galáxias em  $R_{200}$ , adotamos aqui critério semelhante: pelo menos 20 galáxias dentro de  $R_{dens}$ . Esse critério adicional reduz a amostra para 650 aglomerados.

A aplicação do método revelou um total de 502 sistemas (78%) com maior probabilidade de serem irregulares – IRR, e 148 (22%) com maior probabilidade de serem regulares – R. As

distribuições de  $M_1$  e  $\Delta M_{14}$  são apresentadas na figura 5.29. Os resultados são aproximadamente consistentes (MOCK: 15% R e 85% IRR; OBS: 22% R e 78% IRR). Ou seja, a tendência do método é subestimar a quantidade de sistemas regulares e superestimar a quantidade de sistemas irregulares. No estudo de Wen e Han (2013), sobre determinação de subestruturas em aglomerados com dados fotométricos, os autores encontram também um número modesto de sistemas relaxados, 28%, usando uma metodologia totalmente diferente da nossa. Isso reforça a conclusão de que a identificação de subestruturas sobre dados fotométricos é uma tarefa com significativo grau de dificuldade.

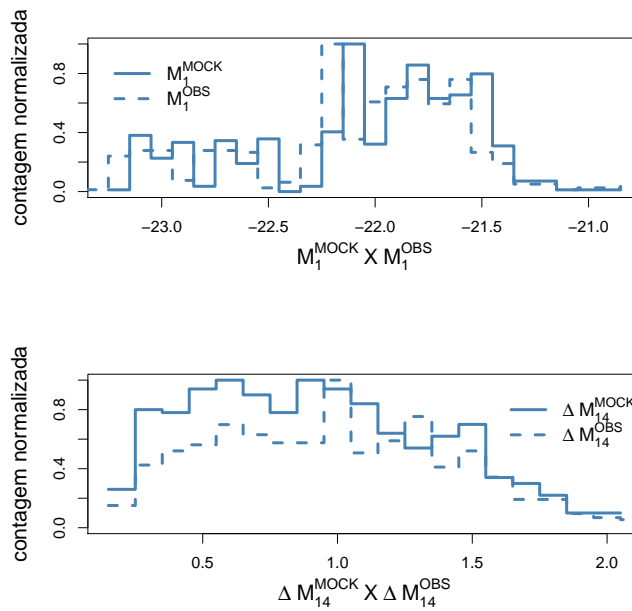


Figura 5.29 – Histogramas das distribuições de  $M_1$  e  $\Delta M_{14}$  para as amostras MOCK e observada.

Finalmente, apresentamos dois exemplos de nossa amostra: um caso com classificação regular e outro com classificação irregular.

#### 5.4.3.1 Aglomerado 39

O aglomerado 39 já foi apresentado na figura 5.25. Já sabemos que  $R_{dens} = 0.28$  graus e que, usando a informação do *redshift* fotométrico do aglomerado,  $z = 0.13$ , calculamos  $R_{dens} = 1.84$  Mpc. Dentro desse raio encontram-se 353 galáxias, correspondendo a  $N_{dens}$ . Vemos agora na figura 5.30 o resultado da combinação do uso das ferramentas e análise parcelas+ $R_{dens}$ +*pdfCluster*, sobre os dados fotométricos, indo até o limite de magnitude  $m_r = 21$ . A seguir, ordenamos as etapas do *pipeline* de análise.

1. Inicialmente, o método das parcelas verifica que este campo é bem centrado (o que constatamos visualmente) e que podemos rejeitar a hipótese de que seja indistinguível

estatisticamente de um processo espacial de Poisson.

2. Então, o pico máximo de densidade do campo é determinado através de *pdfCluster*. O centro é indicado pelo X central da figura 5.30.
3. Um perfil de densidade superficial de galáxias é calculado e comparado com o perfil de calibração definido com o catálogo MOCK. A densidade  $\Sigma_{200}$  é determinada e uma estimativa de  $R_{200}$ , que chamamos de  $R_{dens}$ , é obtida também com o perfil de calibração. Este raio permite definir a região central do aglomerado, indicada na figura 5.30 por um círculo.
4. Então usamos a saída do *pdfCluster* do item (2) para obter todos os picos de densidade (além do pico central) que correspondam a  $3\sigma$  acima da densidade média do campo. Os picos são marcados na figura 5.30 também com a letra “X”.
5. Distinguimos então picos centrais (dentro do círculo) de picos externos. Somente picos centrais são contabilizados para  $N_{picos}$ . Neste caso, portanto,  $N_{picos} = 1$ .<sup>4</sup>
6. Finalmente, obtemos os valores  $M_1 = -23.40$  e  $\Delta M_{14} = 1.1$ . Completamos assim a informação necessária para classificar o sistema do ponto de vista dinâmico. Neste caso, o sistema é classificado como regular:  $M_1 < -22.5$ ,  $\Delta M_{14} > 1.0$  e  $N_{picos} = 1$ .

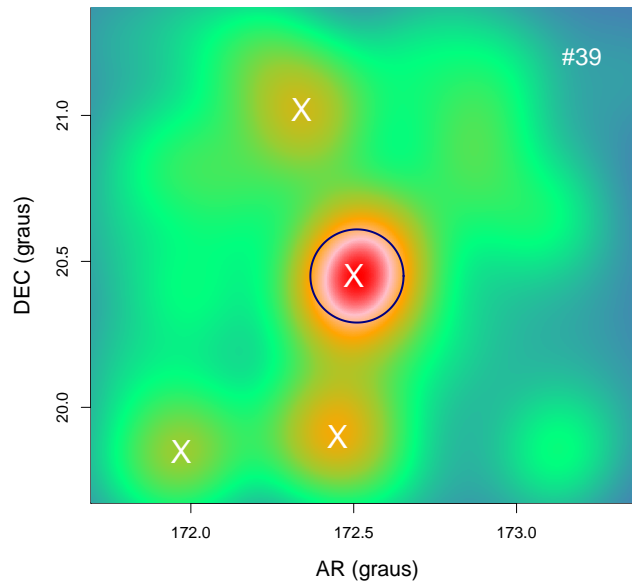


Figura 5.30 – Campo de densidade do aglomerado 39 obtido por kernel adaptativo. Centro e picos de densidade estão indicados pela letra “X”. Um círculo contendo o raio de densidade é superposto à imagem.

<sup>4</sup>Os demais picos identificados podem ser utilizados para investigar o campo de *infall* em torno do aglomerado, um tema a ser estudado em trabalhos futuros.

## 5.4.3.2 Aglomerado 53

Apresentamos agora um campo com características diferentes. O procedimento de análise é o mesmo que o descrito para o aglomerado 39. Seu mapa de densidade é apresentado na figura 5.31. O campo é bem centrado e distingue-se de um processo de Poisson. O raio de densidade estimado corresponde a 0,6 graus, que para o *redshift* fotométrico do campo,  $z = 0.089$ , corresponde a  $R_{dens} = 2.73$  Mpc. Um raio bem maior que no caso anterior, possivelmente superestimado pelo aumento de densidade a sudoeste do pico central. Na direção sudoeste, *pdfCluster* identificou mais dois picos de densidade significativos, um dentro de  $R_{dens}$ , e outro fora. O alongamento do sistema nesta direção pode estar associado ao processo de *infall* através de um filamento. Seja como for, desta vez temos que  $N_{picos} = 2$ , o que já caracterizaria o sistema como não-relaxado. Completando a análise, calculamos  $M_1 = -22.3$  e  $\Delta M_{14} = 1.8$ . Note que o indicador  $M_1$  reforça o caráter não relaxado do sistema, enquanto  $\Delta M_{14}$ , isoladamente, indicaria um sistema relaxado. Como definimos que basta um dos critérios não ser respeitado (e aqui temos dois), a classificação final do sistema é não-relaxado.

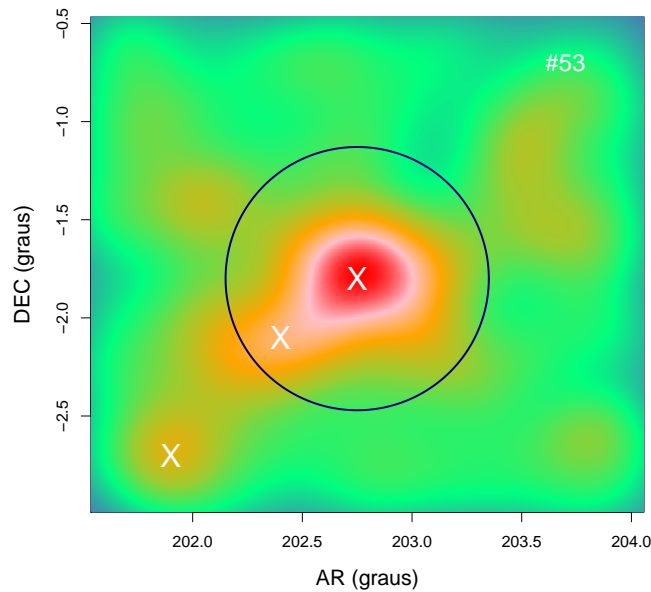


Figura 5.31 – Campo de densidade do aglomerado 53 obtido por kernel adaptativo. Centro e picos de densidade estão indicados pela letra X. Um círculo contendo o raio de densidade é superposto à imagem.

A análise individual dos campos pode revelar nuances dos métodos utilizados e ajudar a implementar melhorias no procedimento. A definição de uma amostra de controle (contendo tanto fotometria como espectroscopia) será determinante para um refinamento da técnica. Tal amostra deverá ser analisada individualmente para que identifiquemos os possíveis vieses astrofísicos e estatísticos envolvidos. Este será um desdobramento necessário do presente trabalho.



## 6 DISCUSSÃO FINAL

### 6.1 Contexto da tese

A abundância de aglomerados de galáxias, por intervalo de massa, pode ser utilizada como uma probabilidade de flutuação de densidade linear e, uma vez que o seu crescimento é determinado diretamente a partir das propriedades da energia escura, pode-se medir seu número, seguindo o intervalo de massa em função do *redshift* (FRIEMAN et al., 2008). Um dos parâmetros cosmológicos que é bastante sensível a abundância de aglomerados é o parâmetro da equação de estado  $\omega$ . Por outro lado, estimativas diretas de massa de halos em aglomerados não são possíveis. As massas de aglomerados de galáxias podem ser determinadas por uma variedade de métodos. Contudo, estes estimadores são sempre indiretos e inferidos a partir de certos observáveis que se correlacionam com a massa, com algum grau de incerteza. Nesse sentido, faz uso dos *mass proxies* que são observáveis que melhor traduzem as medidas de massa de aglomerados. Neste contexto, as relações de escala de aglomerados de galáxias são altamente utilizadas. Um exemplo, é a relação massa-observável (MO), bastante investigada no que diz respeito à evolução dessas grandes estruturas em função de *redshift*, tanto no óptico quanto em raio-X ((VIKHLININ et al., 2002); (LEVINE et al., 2002); (ETTORI et al., 2004); (STANEK et al., 2006) e (ROZO et al., 2010)). A calibração da relação MO pode ser feita por meio da contagem de aglomerados bem como de sua variância. Essa análise pode ser extrapolada através do estudo de simulações cosmológicas que, por sua vez, determina não só a normalização da relação MO, mas a escolha do modelo cosmológico que melhor se adapte aos dados (vide Apêndices A1 e A2).

As relações de escala envolvendo observáveis, tais como a riqueza ou luminosidade, com uma propriedade fundamental, como a massa, podem revelar importantes informações que remontam ao universo atual no que se refere à formação e à evolução das grandes estruturas. Por exemplo, Lopes et al. (2009) compara, utilizando 127 objetos do *Northern Sky Optical Cluster Survey* (NoSOCS), a massa ( $M_{200}$  e  $M_{500}$ ) com a riqueza ( $N_{gal}$ ), estimada dentro dos raios  $R_{200}$  e  $R_{500}$ , respectivamente (ver figura 5.1).

Naquele trabalho, os resíduos foram obtidos por  $\log(M_{obs}/M_{fit})$  onde  $M_{obs}$  representa a massa observada dentro do raio determinado, já  $M_{fit}$  é a solução da regressão linear.

As relações de escala foram obtidas segundo a forma

$$\ln(Y) = A + B \times \ln\left(\frac{X}{C}\right), \quad (6.1)$$

onde os parâmetros  $X$  e  $Y$  definem as relações entre as quantidades correlacionadas.  $A$  e  $B$  representam, respectivamente, a interceptação e a inclinação da regressão linear e  $C$  representa o

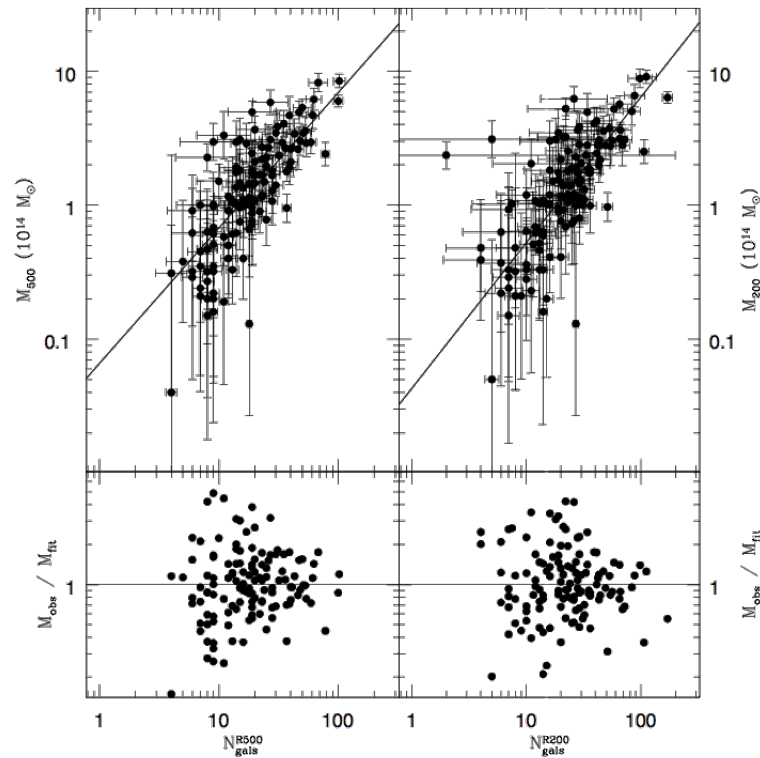


Figura 6.1 – A conexão entre a massa e riqueza, calculada com duas aberturas de raios:  $R_{500}$  (à esquerda) e  $R_{200}$  (à direita). Nos painéis inferiores os resíduos  $\log(M_{obs}/M_{fit})$  são mostrados.  $M_{obs}$  é o valor observado da massa, enquanto  $M_{fit}$  é o resultado de regressão linear.

valor médio do parâmetro  $X$  dentro de um determinado raio da amostra. Segundo Popesso et al. (2004), relações de escala que fazem uso de um raio de abertura fixo, por exemplo  $0,50h^{-1}Mpc$ , produz um menor erro nas estimativas de riqueza ou luminosidade (óptica) e minimiza a dispersão no óptico contra as relações de escala obtidas a partir de raios- $X$ . De fato uma vantagem óptica. Uma outra informação é quando os raios de abertura são comparados com raios dimensionados com a massa. Os primeiros podem ser grandes para sistemas de baixa massa e muito pequenos para os mais massivos. Esse fato tem influência na inclinação da regressão linear das relações de escala, tornando-as mais íngremes.

Um tema bastante discutido ao se calibrar uma relação MO trata-se da redução do espalhamento entre *mass proxies* e a massa dos aglomerados. Para tal Rykoff et al. (2012) apresenta um estimador de riqueza óptico capaz de resgatar a riqueza das galáxias de fundo (*background*). Já Andreon e Hurn (2010) utiliza a metodologia Bayesiana<sup>1</sup> contando os termos usualmente negligenciados em outras técnicas (contagens de galáxias via Poisson e dispersão intrínseca).

<sup>1</sup>Inferência estatística que descreve, probabilisticamente, as incertezas de quantidades cujos dados são difíceis de se obter. Assim, utiliza-se da operação bayesiana a fim de calibrar a medida das incertezas.

Por outro lado, Andreon e Bergé (2012) utilizaram uma amostra controlada não aleatória para calibrar a relação massa-riqueza extraída do *PanStarrs 1+Euclid-like survey*. Tal método é impulsionado pela ideia de que esse tipo de seleção consegue restituir os parâmetros da regressão entre as quantidades correlacionadas. Essa regressão podemos ver na figura 5.2, a qual compara o ajuste do modelo (linha cheia) com a relação real de entrada composta por 201 aglomerados. (ver por exemplo Gelman et al. (2014); Stanek et al. (2006); Pacaud et al. (2007) e Andreon e Moretti (2011)).

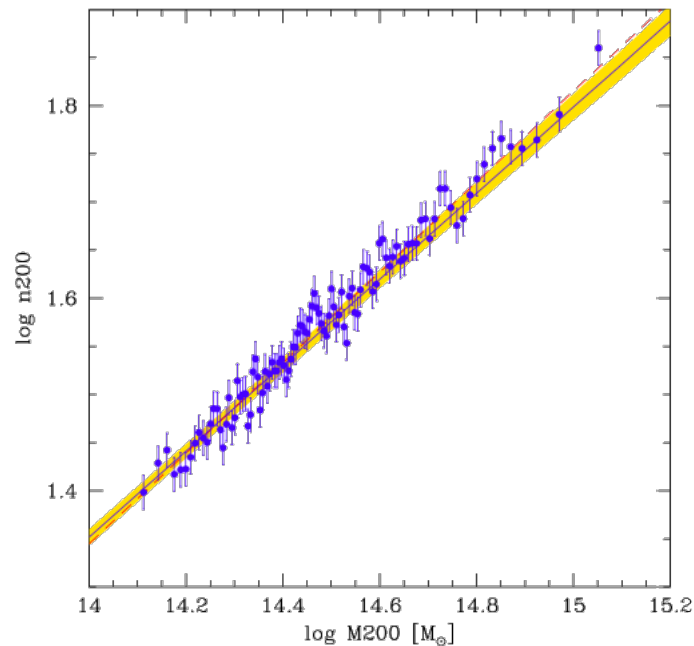


Figura 6.2 – Relação de escala entre a riqueza e a massa extraída do *PanStarrs 1+Euclid-like survey*. A linha sólida marca a linha de regressão ajustada aos dados observados. A região sombreada marca o 68% de confiabilidade para a regressão. A linha pontilhada vermelha indica a relação de entrada dos dados.

Dado que o *proxy* de massa a ser utilizado seja a riqueza, isto significa que simultaneamente estamos buscando uma abertura em torno do aglomerado dentro da qual a contagem de galáxias será feita. Determinar tal abertura em levantamentos fotométricos é uma etapa decisiva para se estabelecer e estudar a relação massa-riqueza e o seu respectivo espalhamento. Neste trabalho, introduzimos uma técnica que visa inferir um raio baseado no perfil de densidade superficial de galáxias. Este raio, chamado aqui de  $R_{dens}$  é definido para ser o valor esperado de  $R_{200}$  através de um perfil de densidade calibrado previamente. É neste contexto, portanto, que se insere esta tese: uma etapa anterior à determinação da riqueza e de seu espalhamento na relação com a massa. Determinamos a riqueza em  $R_{dens}$ , chamada  $N_{dens}$ , e estimamos ainda se os sistemas são ou não regulares dentro dessa região. Uma síntese do trabalho e nossas conclusões estão reunidas a seguir.

## 6.2 Síntese

### 6.2.1 Estimando riqueza

O trabalho segue duas sequências de passos, cada uma corresponde a um *pipeline*. No primeiro deles, limpamos a amostra de “impurezas” definidas por aglomerados mal centrados e aqueles campos cujos aglomerados+*background* não se distinguem suficientemente de um processo espacial de Poisson. Naturalmente, nada impede que um campo mal centrado venha a ser redefinido e o campo resgatado para a análise. Este tipo de recuperação de dados, contudo, não foi realizada neste trabalho.

Após esta etapa, o perfil de densidade superficial do aglomerado é traçado e comparado com um perfil calibrado previamente. No presente trabalho, o perfil usado como modelo é obtido a partir de um catálogo MOCK para o qual dispomos da lista de membros e do campo total de galáxias. O perfil-modelo é definido como a média ponderada dos diversos perfis neste catálogo, usando-se  $\Sigma_{200}^i$  (a densidade que associada a  $R_{200}$  de cada aglomerado) como peso. Um perfil observado pode ser comparado com o perfil-modelo, buscando-se o ponto onde  $\Sigma_r = \Sigma_{200}^{mod}$ , ou seja, em que raio  $r$  do perfil observado é igual ao  $\Sigma_{200}$  do perfil-modelo. Esse raio, se existe, é definido como  $R_{dens}$ , uma abertura que deve corresponder aproximadamente a  $R_{200}$ <sup>2</sup>. Dado que o programa tenha identificado  $R_{dens}$ , uma contagem de galáxias é feita na região, gerando  $N_{dens}$ , uma estimativa da riqueza que deve corresponder a  $N_{200}$ . O diagrama na figura 6.3 ilustra esta etapa do trabalho.

O procedimento é extremamente flexível e permite diversas seleções no campo de galáxias que será utilizado. Da mesma forma, o catálogo MOCK usado para calibrar o perfil de densidade superficial é uma escolha do usuário do programa.

### 6.2.2 Estimando o estado dinâmico

Após essa etapa, nosso objetivo passa a ser determinar a presença ou não de subestruturas nos aglomerados em dados fotométricos. O estado dinâmico dos aglomerados pode afetar suas relações de escala (CHEN et al. 2007; RIBEIRO et al. 2010; CASSANO et al. 2010; ANDRADE-SANTOS et al. 2012). Isso significa que um indicador do estado desses sistemas é importante, se desejamos encontrar uma relação massa-riqueza com o menor espalhamento possível. Encontrar subestruturas ou estimar se um sistema é relaxado (ou não) já é uma tarefa difícil quando dispomos de *redshifts* espectroscópicos (vide PINKNEY et al. 1996; RIBEIRO et al. 2013). Usando dados fotométricos, a tarefa é muito pior. Como o objetivo é identificar amostras com maior probabilidade de conter sistemas relaxados, o ideal é definir um critério em que diversos indicadores concordem que o sistema deve ser regular.

O procedimento adotado aqui tem uma etapa preliminar em que usamos, mais uma

<sup>2</sup>Para todos os campos analisados neste trabalho, o raio de correspondência sempre foi identificado.

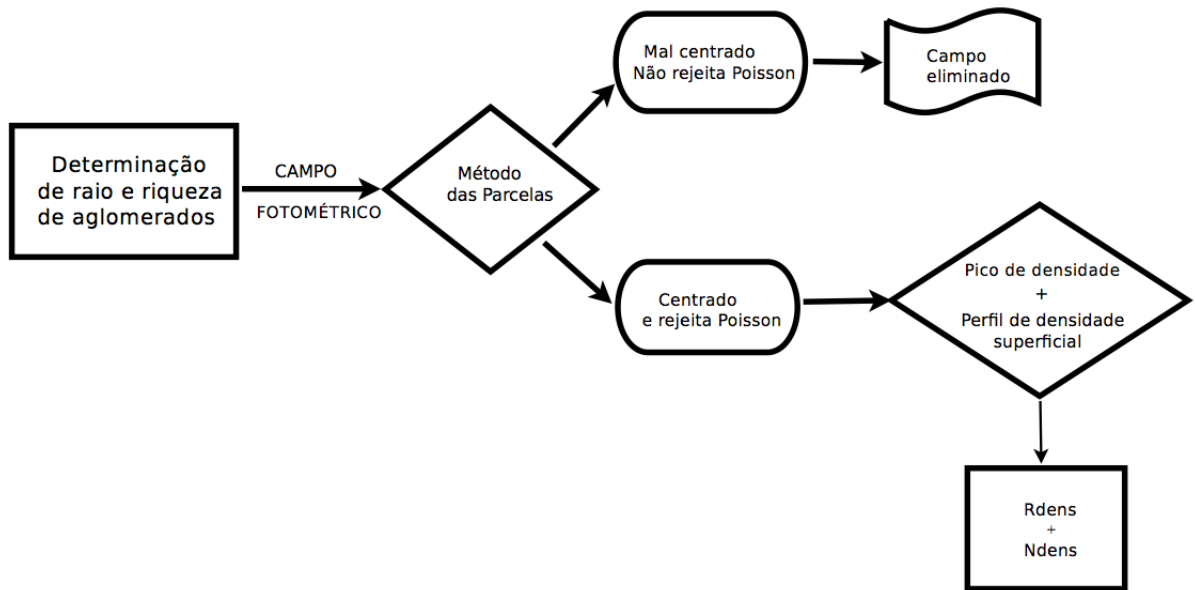


Figura 6.3 – Diagrama mostrando a sequência de passos para a determinação da riqueza de aglomerados.

vez, o catálogo MOCK para identificar as correlações entre os estimadores espectroscópicos e os estimadores fotométricos. Usamos os indicadores HD, AD e DS para determinar se os sistemas têm (ou não) distribuição de velocidades gaussianas (HD e AD) e se apresentam (ou não) desvios cinemáticos locais que possam ser interpretados como subestruturas (DS). Distribuições gaussianas e ausência de subestruturas indicam um sistema relaxado. Como indicadores fotométricos usamos:  $M_1$ ,  $\Delta M_{12}$ ,  $\Delta M_{14}$ , DBCG e VBCG, os índices ICE e IAC, assim como o número de picos no campo suavizado de densidade (todos descritos e discutidos nos Capítulos 4 e 5 desta tese). Verificamos que os índices com maiores números de correlações foram  $M_1$  e HD, e estes foram usados para determinar inicialmente um corte em  $M_1$  baseado em valores de HD. A figura 5.29 ilustra esta relação. O valor escolhido para restringir sistemas relaxados foi  $M_1 < -22.5$ . Em seguida, usamos as melhores relações entre  $M_1$  e os demais parâmetros fotométricos para definir os critérios  $\Delta M_{14} > 1.0$  e  $N_{picos}$ . Para ser relaxado, todos os critérios precisam ser satisfeitos simultaneamente. Isto envia os resultados para baixas frações de sistemas relaxados, o que é consistente com a ideia de definir amostras razoavelmente puras de sistemas relaxados. Nesse caso, maior pureza implica em reduzir a amostra.

Para esse método funcionar, precisamos usar o valor de  $R_{dens}$ , pois apenas dentro desta região vamos analisar o grau de relaxamento do sistema. A outra entrada são os *redshift* fotométricos ( $z_{phot}$ ) das galáxias 1 e 4 (a primeira e a quarta mais luminosas do sistema). O diagrama apresentado na figura 6.4 ilustra o procedimento.

O método é dependente da etapa anterior e, portanto, herda também a sua flexibilidade e

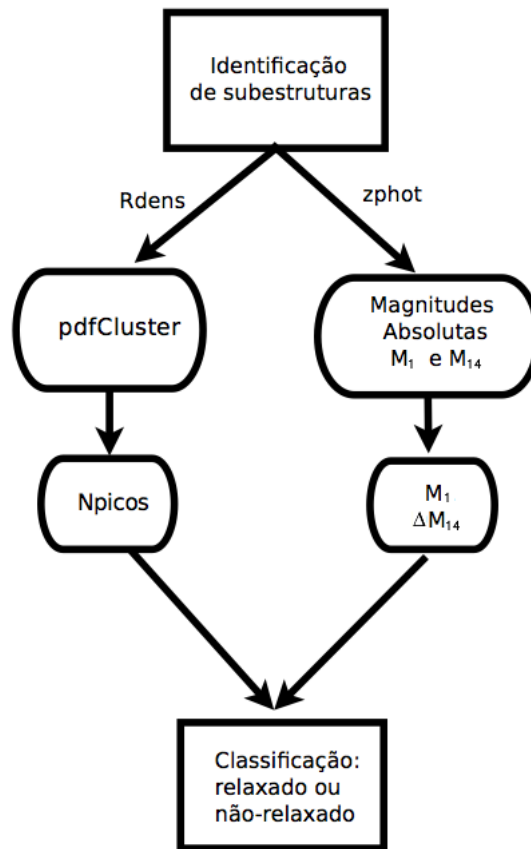


Figura 6.4 – Diagrama mostrando a sequência de passos para inferir o estado dinâmico dos aglomerados.

limitações. A seguir, destacamos os principais achados deste trabalho.

### 6.3 Conclusões e perspectivas

Resumimos alguns de nossos principais resultados, conclusões e perspectivas.

1. Nosso estudo revela que é possível determinar  $R_{dens}$  e  $N_{dens}$  de forma robusta, conseguindo relações lineares entre  $N_{dens}$  e  $N_{200}$  com coeficientes de correlação  $> 0,9$ , para diversas amostras de galáxias retiradas dos dados.
2. A robustez da relação  $N_{dens} - N_{200}$  é importante, pois as seleções, que visam separar o aglomerado do *background*, podem introduzir outros problemas e incertezas difíceis de quantificar. O fato de que o método funciona razoavelmente usando-se todos os dados fotométricos disponíveis até o limite de detecção diminui os efeitos de variabilidade dos dados. Apesar do funcionamento com os dados completos, pretendemos futuramente investigar seleções de galáxias que possam otimizar o funcionamento do programa e melhorar os resultados.

3. Na sequência deste trabalho, pretendemos melhorar e tornar mais universal o catálogo de calibração. Torná-lo mais universal significaria definir diferentes perfis-modelo, por exemplo, em diferentes intervalos de massa dos halos. Ou segundo algum outro parâmetro.
4. Obtivemos ainda uma estimativa conservadora da fração de sistemas relaxados na amostra do SDSS, 22%, aproximadamente consistente com a estimativa de Wen e Han (2013), que encontram 28% numa amostra maior, a partir da qual a nossa foi obtida.
5. O procedimento para inferir o estado dinâmico dos sistemas revelou uma relação que pode ter importância fundamental para uma série de trabalhos futuros: a relação  $HD-M_1$ . Sem dúvida um ponto a ser explorado brevemente.
6. O uso do *pdfCluster* também mostrou-se extremamente útil e importante neste trabalho. Notemos que definir uma abertura onde contar galáxias em um aglomerado depende de sabermos onde é o seu centro. Os centros obtidos pelo *pdfCluster* foram muito satisfatórios em todos os casos onde uma inspeção visual foi feita, e também quando outros métodos foram usados para simples conferência: baricentro ponderado por luminosidade (94% de concordância dentro de 100 kpc) e posição da BCG (78% de concordância dentro de 100 kpc).
7. A identificação dos picos de densidade secundários também permite uma série de estudos futuros, por exemplo explorando o padrão de acreção de galáxias em torno dos aglomerados.
8. Pretendemos fazer uma comparação mais efetiva com outros trabalhos e técnicas usadas em dados fotométricos. Por exemplo uma comparação direta com os resultados de Wen e Han (2013) pode ser um primeiro passo nesse sentido.

E assim chegamos ao fim.

## REFERÊNCIAS

- AGUERRI, J. A. L.; SANCHEZ-JANSSEN, R. A study of catalogued nearby galaxy clusters in the sdss-dr4. ii. cluster substructure. **arXiv preprint arXiv:1007.3497**, 2010.
- ALBARETI, F. D.; ALLENDE PRIETO, C.; ALMEIDA, A.; ANDERS, F.; ANDERSON, S.; ANDREWS, B. H.; ARAGÓN-SALAMANCA, A.; ARGUDO-FERNÁNDEZ, M.; ARMENGAUD, E.; AUBOURG, E.; AL. et. The 13th Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-IV Survey Mapping Nearby Galaxies at Apache Point Observatory. , v. 233, p. 25, dez. 2017.
- ANDERSON, T. W.; DARLING, D. A. A test of goodness of fit. **Journal of the American statistical association**, Taylor & Francis, v. 49, n. 268, p. 765–769, 1954.
- ANDREON, S.; BERGÉ, J. Richness-mass relation self-calibration for galaxy clusters. **Astronomy & Astrophysics**, EDP Sciences, v. 547, p. A117, 2012.
- ANDREON, S.; HURN, M. A. The scaling relation between richness and mass of galaxy clusters: a bayesian approach. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 404, n. 4, p. 1922–1937, 2010.
- ANDREON, S.; MORETTI, A. Do x-ray dark or underluminous galaxy clusters exist? **Astronomy & Astrophysics**, EDP Sciences, v. 536, p. A37, 2011.
- AZZALINI, A.; TORELLI, N. Clustering via nonparametric density estimation. **Statistics and Computing**, Springer, v. 17, n. 1, p. 71–80, 2007.
- BADDELEY A., M. J. P. A. **Properties of residuals for spatial point processes**. Vienna, Austria, 2008.
- BAHCALL, N. A. Core radii and central densities of 15 rich clusters of galaxies. **The Astrophysical Journal**, v. 198, p. 249–254, 1975.
- BAHCALL, N. A.; FAN, X. The most massive distant clusters: Determining  $\alpha$  and  $\sigma_8$ . **The Astrophysical Journal**, IOP Publishing, v. 504, n. 1, p. 1, 1998.
- BARTELMANN, M. Arcs from a universal dark-matter halo profile. **arXiv preprint astro-ph/9602053**, 1996.
- BAXTER, E. J.; ROZO, E.; JAIN, B.; RYKOFF, E.; WECHSLER, R. H. Constraining the mass–richness relationship of redmapper clusters with angular clustering. **Monthly Notices of the Royal Astronomical Society**, The Royal Astronomical Society, v. 463, n. 1, p. 205–221, 2016.
- BERLIND, A. A.; FRIEMAN, J.; WEINBERG, D. H.; BLANTON, M. R.; WARREN, M. S.; ABAZAJIAN, K.; SCRANTON, R.; HOGG, D. W.; SCOCCIMARRO, R.; BAHCALL, N. A. et al. Percolation galaxy groups and clusters in the sdss redshift survey: identification, catalogs, and the multiplicity function. **The Astrophysical Journal Supplement Series**, IOP Publishing, v. 167, n. 1, p. 1, 2006.



- BERLIND, A. A.; WEINBERG, D. H. The halo occupation distribution: toward an empirical determination of the relation between galaxies and mass. **The Astrophysical Journal**, IOP Publishing, v. 575, n. 2, p. 587, 2002.
- BODE, P.; OSTRIKER, J. P.; TUROK, N. Halo formation in warm dark matter models. **The Astrophysical Journal**, IOP Publishing, v. 556, n. 1, p. 93, 2001.
- BONFERRONI, C. Teoria statistica delle classi e calcolo delle probabilita. **Publicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze**, v. 8, p. 3–62, 1936.
- BOSCH, F. C. V. D.; YANG, X.; MO, H.; NORBERG, P. The abundance and radial distribution of satellite galaxies. **Monthly Notices of the Royal Astronomical Society**, Blackwell Science Ltd Oxford, UK, v. 356, n. 4, p. 1233–1248, 2005.
- BOYLAN-KOLCHIN, M.; SPRINGEL, V.; WHITE, S. D.; JENKINS, A.; LEMSON, G. Resolving cosmic structure formation with the millennium-ii simulation. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 398, n. 3, p. 1150–1164, 2009.
- BYTH, K. On robust distance-based intensity estimators. **Biometrics**, JSTOR, p. 127–135, 1982.
- CARETTA, C. A.; ROSA, R. R.; VELHO, H. de C.; RAMOS, F. M.; MAKLER, M. Evidence of turbulence-like universality in the formation of galaxy-sized dark matter haloes. **Astronomy & Astrophysics**, EDP Sciences, v. 487, n. 2, p. 445–451, 2008.
- CARLBERG, R. G.; YEE, H.; ELLINGSON, E. The average mass and light profiles of galaxy clusters. **The Astrophysical Journal**, IOP Publishing, v. 478, n. 2, p. 462, 1997.
- CARVALHO, R. de; RIBEIRO, A.; STALDER, D.; ROSA, R.; COSTA, A.; MOURA, T. Investigating the relation between galaxy properties and the gaussianity of the velocity distribution of groups and clusters. **The Astronomical Journal**, IOP Publishing, v. 154, n. 3, p. 96, 2017.
- CLARK, P. J.; EVANS, F. C. Distance to nearest neighbor as a measure of spatial relationships in populations. **Ecology**, JSTOR, p. 445–453, 1954.
- COLLESS, M.; DUNN, A. M. Structure and dynamics of the coma cluster. **arXiv preprint astro-ph/9508070**, 1995.
- COOK, R. D.; WEISBERG, S. **Residuals and influence in regression**. [S.l.]: New York: Chapman and Hall, 1982.
- COSTA, F. J. **DETERMINAÇÃO DE RAIOS E RIQUEZA DE AGLOMERADOS DE GALÁXIAS A PARTIR DE ANÁLISE ESPACIAL ESTATÍSTICA**. 2011. DISSERTAÇÃO (Licenciatura em Física), UESC (Universidade Estadual de Santa Cruz), Ilhéus, Brazil.
- CUNHA, C. E.; EVRARD, A. E. Sensitivity of galaxy cluster dark energy constraints to halo modeling uncertainties. **Physical Review D**, APS, v. 81, n. 8, p. 083509, 2010.
- DICKINSON, T. **NightWatch: A Practical Guide to Viewing the Universe**. [S.l.]: Firefly Books, 1998.
- DIGGLE, P. J. Statistical analysis of spatial point processes. **Academic, London**, 1983.
- DRESSLER, A.; SHECTMAN, S. A. Evidence for substructure in rich clusters of galaxies from radial-velocity measurements. **The Astronomical Journal**, v. 95, p. 985–995, 1988.

- DUARTE, M.; MAMON, G. A. **How well does the Friends-of-Friends algorithm recover group properties from distance-and luminosity-limited galaxy catalogs?** [S.l.], 2014.
- DUONG, T. **Bandwidth selectors for multivariate kernel density estimation.** [S.l.]: University of Western Australia, 2004.
- DUPKE, R. A.; BREGMAN, J. N. Direct measurements of gas bulk flows in the intracluster medium of the centaurus cluster with the chandra satellite. **The Astrophysical Journal**, IOP Publishing, v. 639, n. 2, p. 781, 2006.
- EKE, V. R.; COLE, S.; FRENK, C. S. Cluster evolution as a diagnostic for ? **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 282, n. 1, p. 263–280, 1996.
- ETTORI, S.; TOZZI, P.; BORGANI, S.; ROSATI, P. Scaling laws in x-ray galaxy clusters at redshift between 0.4 and 1.3. **Astronomy & Astrophysics**, EDP Sciences, v. 417, n. 1, p. 13–27, 2004.
- EVARD, A. E. The intracluster gas fraction in x-ray clusters: Constraints on the clustered mass density. **Monthly Notices of the Royal Astronomical Society**, The Royal Astronomical Society, v. 292, n. 2, p. 289–297, 1997.
- EVARD, A. E.; MACFARLAND, T.; COUCHMAN, H.; COLBERG, J.; YOSHIDA, N.; WHITE, S.; JENKINS, A.; FRENK, C.; PEARCE, F.; PEACOCK, J. et al. Galaxy clusters in hubble volume simulations: cosmological constraints from sky survey populations. **The Astrophysical Journal**, IOP Publishing, v. 573, n. 1, p. 7, 2002.
- EVARD, A. E.; METZLER, C. A.; NAVARRO, J. F. Mass estimates of x-ray clusters. **arXiv preprint astro-ph/9510058**, 1995.
- FRIEMAN, J.; TURNER, M.; HUTERER, D. Dark energy and the accelerating universe. **arXiv preprint arXiv:0803.0982**, 2008.
- GARDNER, J. P.; CONNOLLY, A.; MCBRIDE, C. Enabling knowledge discovery in a virtual universe. In: **Proc. of the 2007 TeraGrid Symp.** [S.l.: s.n.], 2007.
- GARDNER, J. P.; CONNOLLY, A.; MCBRIDE, C. Enabling rapid development of parallel tree search applications. In: ACM. **Proceedings of the 5th IEEE workshop on Challenges of large applications in distributed environments.** [S.l.], 2007. p. 1–10.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis.** [S.l.]: Taylor & Francis, 2014. v. 2.
- GIRARDI, M.; BIVIANO, A.; GIURICIN, G.; MARDIROSSIAN, F.; MEZZETTI, M. Optical radii of galaxy clusters. **arXiv preprint astro-ph/9408090**, 1994.
- GLADDERS, M. D.; YEE, H. K. The red-sequence cluster survey. i. the survey and cluster catalogs for patches rcs 0926+ 37 and rcs 1327+ 29. **The Astrophysical Journal Supplement Series**, IOP Publishing, v. 157, n. 1, p. 1, 2005.
- GUNN, J. E.; SIEGMUND, W. A.; MANNERY, E. J.; OWEN, R. E.; HULL, C. L.; LEGER, R. F.; CAREY, L. N.; KNAPP, G. R.; YORK, D. G.; BOROSKI, W. N. et al. The 2.5 m telescope of the sloan digital sky survey. **The Astronomical Journal**, IOP Publishing, v. 131, n. 4, p. 2332, 2006.

- HAMILTON, A. Astrophysics and space science library series, vol. 231, selected topics on large-scale structure and on the properties of galaxies. **The Evolving Universe**, ed. D. Hamilton (Dordrecht: Kluwer), v. 185, 1998.
- HENRY, J. P.; EVRARD, A. E.; HOEKSTRA, H.; BABUL, A.; MAHDAVI, A. The x-ray cluster normalization of the matter power spectrum. **The Astrophysical Journal**, IOP Publishing, v. 691, n. 2, p. 1307, 2009.
- HOU, A.; PARKER, L. C.; BALOGH, M. L.; MCGEE, S. L.; WILMAN, D. J.; CONNELLY, J. L.; HARRIS, W. E.; MOK, A.; MULCHAEY, J. S.; BOWER, R. G. et al. Do group dynamics play a role in the evolution of member galaxies? **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 435, n. 2, p. 1715–1726, 2013.
- HOU, A.; PARKER, L. C.; HARRIS, W. E.; WILMAN, D. J. Statistical tools for classifying galaxy group dynamics. **The Astrophysical Journal**, IOP Publishing, v. 702, n. 2, p. 1199, 2009.
- HOU, A.; PARKER, L. C.; WILMAN, D. J.; MCGEE, S. L.; HARRIS, W. E.; CONNELLY, J. L.; BALOGH, M. L.; MULCHAEY, J. S.; BOWER, R. G. Substructure in the most massive geec groups: field-like populations in dynamically active groups. **Monthly Notices of the Royal Astronomical Society**, Blackwell Publishing Ltd Oxford, UK, v. 421, n. 4, p. 3594–3611, 2012.
- HUCHRA, J.; GELLER, M. Groups of galaxies. i-nearby groups. **The Astrophysical Journal**, v. 257, p. 423–437, 1982.
- JACKSON, J. A critique of rees’s theory of primordial gravitational radiation. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 156, n. 1, p. 1P–5P, 1972.
- JENKINS, A.; FRENK, C.; WHITE, S. D.; COLBERG, J.; COLE, S.; EVRARD, A. E.; COUCHMAN, H.; YOSHIDA, N. The mass function of dark matter haloes. **Monthly Notices of the Royal Astronomical Society**, Blackwell Science Ltd Oxford, UK, v. 321, n. 2, p. 372–384, 2001.
- JING, Y.; SUTO, Y. The density profiles of the dark matter halo are not universal. **The Astrophysical Journal Letters**, IOP Publishing, v. 529, n. 2, p. L69, 2000.
- KATZ, N.; WHITE, S. D. Hierarchical galaxy formation-overmerging and the formation of an x-ray cluster. **The Astrophysical Journal**, v. 412, p. 455–478, 1993.
- KAUFFMANN, G.; COLBERG, J. M.; DIAFERIO, A.; WHITE, S. D. Clustering of galaxies in a hierarchical universe?i. methods and results at  $z=0$ . **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 303, n. 1, p. 188–206, 1999.
- KING, I. R. Density data and emission measure for a model of the coma cluster. **The Astrophysical Journal**, v. 174, p. L123, 1972.
- KNEBE, A.; MÜLLER, V. Quantifying substructure in galaxy clusters. **arXiv preprint astro-ph/9912534**, 1999.
- LEVINE, E.; SCHULZ, A.; WHITE, M. Future galaxy cluster surveys: The effect of theory uncertainty on constraining cosmological parameters. **The Astrophysical Journal**, IOP Publishing, v. 577, n. 2, p. 569, 2002.

- LIDDLE, A. R.; LYTH, D. H.; VIANA, P. T.; WHITE, M. Cold dark matter models with a cosmological constant. **Monthly Notices of the Royal Astronomical Society**, The Royal Astronomical Society, v. 282, n. 1, p. 281–290, 1996.
- LOPES, P.; CARVALHO, R. D.; KOHL-MOREIRA, J.; JONES, C. Nosocs in sdss–ii. mass calibration of low redshift galaxy clusters with optical and x-ray properties. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 399, n. 4, p. 2201–2220, 2009.
- LOPES, P. A.; CARVALHO, R. D.; CAPELATO, H.; GAL, R.; DJORGOVSKI, S.; BRUNNER, R.; ODEWAHN, S.; MAHABAL, A. X-ray galaxy clusters in nosocs: Substructure and the correlation of optical and x-ray properties. **The Astrophysical Journal**, IOP Publishing, v. 648, n. 1, p. 209, 2006.
- LYNDEN-BELL, D. Statistical mechanics of violent relaxation in stellar systems. **Monthly Notices of the Royal Astronomical Society**, v. 136, p. 101, 1967.
- MARDIA, K. V.; JUPP, P. E. **Directional statistics**. [S.l.]: John Wiley & Sons, 2009. v. 494.
- MARKEVITCH, M.; GONZALEZ, A.; CLOWE, D.; VIKHLININ, A.; FORMAN, W.; JONES, C.; MURRAY, S.; TUCKER, W. Direct constraints on the dark matter self-interaction cross section from the merging galaxy cluster 1e 0657–56. **The Astrophysical Journal**, IOP Publishing, v. 606, n. 2, p. 819, 2004.
- MARTINEZ, E. S. V. J. Statistics of galaxy clustering. n. 1, 2002.
- MOLLET, J.; TRUMBLE, J.; SEVACHERIAN, V. Comparison of dispersion and regression indices for tetranychus cinnabarinus (boisduval)(acari: Tetranychidae) populations in cotton. **Environmental entomology**, Oxford University Press Oxford, UK, v. 13, n. 6, p. 1511–1514, 1984.
- NAVARRO, J. F.; FRENK, C. S.; WHITE, S. D. A universal density profile from hierarchical clustering. **The Astrophysical Journal**, IOP Publishing, v. 490, n. 2, p. 493, 1997.
- NOVAES, A. G. N. **Sistemas Logísticos: transportes, armazenagem e distribuição física de produtos**. [S.l.]: E. Blucher, 1989.
- NOVIKOV, D. I.; MELOTT, A. L.; WILHITE, B. C.; KAUFMAN, M.; BURNS, J. O.; MILLER, C. J.; BATUSKI, D. J. Cluster winds blow along supercluster axes. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 304, n. 1, p. L5–L9, 1999.
- PACAUD, F.; PIERRE, M.; ADAMI, C.; ALTIERI, B.; ANDREON, S.; CHIAPPETTI, L.; DETAL, A.; DUC, P.-A.; GALAZ, G.; GUEGUEN, A. et al. The xmm-iss survey: the class 1 cluster sample over the initial 5 deg<sup>2</sup> and its cosmological modelling. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 382, n. 3, p. 1289–1308, 2007.
- PEACOCK, J.; HEAVENS, A. Alternatives to the press–schechter cosmological mass function. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 243, n. 1, p. 133–143, 1990.
- PEEBLES, P. J. E. **Principles of physical cosmology**. [S.l.]: Princeton University Press, 1993.
- PINKNEY, J.; ROETTIGER, K.; BURNS, J. O.; BIRD, C. M. Evaluation of statistical tests for substructure in clusters of galaxies. **The Astrophysical Journal Supplement Series**, v. 104, p. 1, 1996.

- POPESSO, P.; BÖHRINGER, H.; BRINKMANN, J.; VOGES, W.; YORK, D. *Rass-sdss galaxy clusters survey-i. the catalog and the correlation of x-ray and optical properties.* **Astronomy & Astrophysics**, EDP Sciences, v. 423, n. 2, p. 449–467, 2004.
- PRESS, W. H.; SCHECHTER, P. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. **The Astrophysical Journal**, v. 187, p. 425–438, 1974.
- RAHMAN, N.; SHANDARIN, S. F. Constraining cosmological models by the cluster mass function. **The Astrophysical Journal Letters**, IOP Publishing, v. 550, n. 2, p. L121, 2001.
- RIBEIRO, A.; CARVALHO, R. de; TREVISAN, M.; CAPELATO, H.; BARBERA, F. L.; LOPES, P.; SCHILLING, A. Spider–ix. classifying galaxy groups according to their velocity distribution. **Monthly Notices of the Royal Astronomical Society**, The Royal Astronomical Society, v. 434, n. 1, p. 784–795, 2013.
- RIBEIRO, A. L.; LOPES, P. A.; TREVISAN, M. Segregation effects according to the evolutionary stage of galaxy groups. **Monthly Notices of the Royal Astronomical Society: Letters**, Blackwell Publishing Ltd Oxford, UK, v. 409, n. 1, p. L124–L127, 2010.
- RIBEIRO, A. L.; LOPES, P. A.; TREVISAN, M. Non-gaussian velocity distributions?the effect on virial mass estimates of galaxy groups. **Monthly Notices of the Royal Astronomical Society: Letters**, Blackwell Science Ltd Oxford, UK, v. 413, n. 1, p. L81–L85, 2011.
- RICKER, P. M.; SARAZIN, C. L. Off-axis cluster mergers: effects of a strongly peaked dark matter profile. **The Astrophysical Journal**, IOP Publishing, v. 561, n. 2, p. 621, 2001.
- ROBERTS, I. D.; PARKER, L. C. Evidence of pre-processing and a dependence on dynamical state for low-mass satellite galaxies. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 467, n. 3, p. 3268–3278, 2017.
- ROUKEMA, B. F.; PETERSON, B. A.; QUINN, P. J.; ROCCA-VOLMERANGE, B. Merging history trees of dark matter haloes: a tool for exploring galaxy formation models. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 292, n. 4, p. 835–852, 1997.
- ROZO, E.; RYKOFF, E.; KOESTER, B.; NORD, B.; WU, H.-Y.; EVRARD, A.; WECHSLER, R. Extrinsic sources of scatter in the richness-mass relation of galaxy clusters. **The Astrophysical Journal**, IOP Publishing, v. 740, n. 2, p. 53, 2011.
- ROZO, E.; RYKOFF, E. S.; EVRARD, A.; BECKER, M.; MCKAY, T.; WECHSLER, R. H.; KOESTER, B. P.; HAO, J.; HANSEN, S.; SHELDON, E. et al. Constraining the scatter in the mass-richness relation of maxbcg clusters with weak lensing and x-ray data. **The Astrophysical Journal**, IOP Publishing, v. 699, n. 1, p. 768, 2009.
- ROZO, E.; RYKOFF, E. S.; KOESTER, B. P.; MCKAY, T.; HAO, J.; EVRARD, A.; WECHSLER, R. H.; HANSEN, S.; SHELDON, E.; JOHNSTON, D. et al. Improvement of the richness estimates of maxbcg clusters. **The Astrophysical Journal**, IOP Publishing, v. 703, n. 1, p. 601, 2009.
- ROZO, E.; WECHSLER, R. H.; RYKOFF, E. S.; ANNIS, J. T.; BECKER, M. R.; EVRARD, A. E.; FRIEMAN, J. A.; HANSEN, S. M.; HAO, J.; JOHNSTON, D. E. et al. Cosmological constraints from the sloan digital sky survey maxbcg cluster catalog. **The Astrophysical Journal**, IOP Publishing, v. 708, n. 1, p. 645, 2009.

- ROZO, E.; WECHSLER, R. H.; RYKOFF, E. S.; ANNIS, J. T.; BECKER, M. R.; EVRARD, A. E.; FRIEMAN, J. A.; HANSEN, S. M.; HAO, J.; JOHNSTON, D. E. et al. Cosmological constraints from the sloan digital sky survey maxbcg cluster catalog. **The Astrophysical Journal**, IOP Publishing, v. 708, n. 1, p. 645, 2010.
- RUIZ, R.; DEMÉTRIO, C. G.; ASSUNÇÃO, R. M.; LEANDRO, R. A. Modelos hierárquicos bayesianos para estudar a distribuição espacial da infestação da broca do café em nível local. **Revista Colombiana de Estadística**, v. 26, n. 1, p. 1–24, 2003.
- RYKOFF, E.; KOESTER, B.; ROZO, E.; ANNIS, J.; EVRARD, A.; HANSEN, S.; HAO, J.; JOHNSTON, D.; MCKAY, T.; WECHSLER, R. Robust optical richness estimation with reduced scatter. **The Astrophysical Journal**, IOP Publishing, v. 746, n. 2, p. 178, 2012.
- S., A. Notas de aula em estatística, métodos diferencial-geométricos em estatísticas. HeidelbergSpringer, v. 28, 1985.
- SKIBBA, R. A.; BOSCH, F. C. van den; YANG, X.; MORE, S.; MO, H.; FONTANOT, F. Are brightest halo galaxies central galaxies? **Monthly Notices of the Royal Astronomical Society**, The Royal Astronomical Society, v. 410, n. 1, p. 417–431, 2010.
- SOLANES, J. M.; SALVADOR-SOLÉ, E.; GONZÁLEZ-CASADO, G. Substructure in the enacs clusters. **arXiv preprint astro-ph/9812103**, 1998.
- STANEK, R.; EVRARD, A. E.; SCHUECKER, P.; NORD, B. et al. The x-ray luminosity-mass relation for local clusters of galaxies. **The Astrophysical Journal**, IOP Publishing, v. 648, n. 2, p. 956, 2006.
- STRATEVA, I.; IVEZIĆ, Ž.; KNAPP, G. R.; NARAYANAN, V. K.; STRAUSS, M. A.; GUNN, J. E.; LUPTON, R. H.; SCHLEGEL, D.; BAHCALL, N. A.; BRINKMANN, J. et al. Color separation of galaxy types in the sloan digital sky survey imaging data. **The Astronomical Journal**, IOP Publishing, v. 122, n. 4, p. 1861, 2001.
- UPTON, G.; FINGLETON, B. **Spatial Data Analysis by Example, Point Pattern and Quantitative Data**. 1985. [S.l.]: New York: Wiley, 1985.
- VIKHLININ, A.; VANSPEYBROECK, L.; MARKEVITCH, M.; FORMAN, W.; GREGO, L. Evolution of the cluster x-ray scaling relations since  $z > 0.4$ . **The Astrophysical Journal Letters**, IOP Publishing, v. 578, n. 2, p. L107, 2002.
- WARREN, M. S.; ABAZAJIAN, K.; HOLZ, D. E.; TEODORO, L. Precision determination of the mass function of dark matter halos. **The Astrophysical Journal**, IOP Publishing, v. 646, n. 2, p. 881, 2006.
- WEISBERG, S. **Applied linear regression**. [S.l.]: John Wiley & Sons, 1987.
- WEN, Z.; HAN, J.; LIU, F. A catalog of 132,684 clusters of galaxies identified from sloan digital sky survey iii. **The Astrophysical Journal Supplement Series**, IOP Publishing, v. 199, n. 2, p. 34, 2012.
- WEN, Z. á.; HAN, J. á. Substructure and dynamical state of 2092 rich clusters of galaxies derived from photometric data. **Monthly Notices of the Royal Astronomical Society**, The Royal Astronomical Society, v. 436, n. 1, p. 275–293, 2013.

WEST, M. J.; JR, A. O.; DEKEL, A. Subclustering in rich clusters of galaxies and their environs. **The Astrophysical Journal**, v. 327, p. 1–24, 1988.

WHITE, M.; COHN, J.; SMIT, R. Cluster galaxy dynamics and the effects of large-scale environment. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press, v. 408, n. 3, p. 1818–1834, 2010.

WHITE, S. D.; EFSTATHIOU, G.; FRENK, C. The amplitude of mass fluctuations in the universe. **Monthly Notices of the Royal Astronomical Society**, Oxford University Press Oxford, UK, v. 262, n. 4, p. 1023–1028, 1993.

WHITE, S. D.; NAVARRO, J. F.; EVRARD, A. E.; FRENK, C. S. The baryon content of galaxy clusters: a challenge to cosmological orthodoxy. **nature**, Nature Publishing Group, v. 366, n. 6454, p. 429, 1993.

YANG, X.; MO, H.; BOSCH, F. C. van den. Galaxy groups in the sdss dr4. ii. halo occupation statistics. **The Astrophysical Journal**, IOP Publishing, v. 676, n. 1, p. 248, 2008.

ZAROUBI, S.; SQUIRES, G.; HOFFMAN, Y.; SILK, J. Deprojection of rich cluster images. **The Astrophysical Journal Letters**, IOP Publishing, v. 500, n. 2, p. L87, 1998.

## Apêndices



## APÊNDICE A – APÊNDICE

### A.1 Extração de catálogos de grupos de galáxias

A extração de catálogos de grupos de galáxias a partir de dados do espaço de *redshift* (dos grupos verdadeiros) é difícil por várias razões, entre elas:

- é intrinsecamente difícil caracterizar os sistemas de poucos objetos (galáxias);
- o ambiente local requer uma definição precisa do centro do grupo, que também é abstruso para sistemas de baixa multiplicidade;
- o fluxo Hubble cria distorções nos *redshifts* (JACKSON, 1972), fazendo com que galáxias se estendam no espaço de *redshift*;

Há dois métodos básicos para construção de catálogos simulados de galáxias no espaço real:

- a) Utilizando o método de distribuição de ocupação de halo ((MARTINEZ, 2002); (BERLIND; WEINBERG, 2002)), cujo número de galáxias por halo é traçado a partir de uma função de distribuição de probabilidade dependente da massa, ou ainda, das luminosidades de galáxias ou massas estelares. Ainda supõe que a simetria da distribuição de galáxias é esférica, e que se segue o modelo de matéria escura nos halos expressa em função dos parâmetros cosmológicos.
- b) Utilizando modelos semi-analíticos (SAMs, por exemplo, Roukema et al. (1997) e Kauffmann et al. (1999)), cujas propriedades de galáxias (em particular, a massa estelar e luminosidade) são traçadas nos halos e sub-halos em simulações de  $N$  corpos ao longo do tempo cósmico. Este procedimento consegue produzir relativamente bem galáxias com propriedades e relações de escala observadas.

Neste trabalho, a segunda abordagem foi o método escolhido para utilização do catálogo simulado (BOYLAN-KOLCHIN et al., 2009). Esta simulação envolveu 21.603 partículas com massa de  $9,5 \times 10^6 M_{\odot}$ , em uma caixa de tamanho comóvel de 137 Mpc. Nas simulações se fez uso dos seguintes parâmetros cosmológicos:  $\Omega_m = 0,25$ ,  $\Omega_{\Lambda} = 0,75$ ,  $h = 0,73$ , e  $\sigma_8 = 0,9$ . A técnica FOF foi aplicada a estes dados, para mais de  $10^5$  partículas em grupos com massa superiores a  $10^{12} M_{\odot}$ . Com base neste método, a massa é incluída no interior da esfera de raio  $R_{200}$  centrada na partícula, cuja densidade média dentro da esfera é superior à densidade média do universo por um fator  $\Delta \sim 200$ . Este método considera, apenas, galáxias inseridas no raio  $R_{200}$ .

Agora descrevemos a construção do espaço de *redshift* de galáxias simulado do *Sloan Digital Sky Survey* (SDSS). A caixa de simulação não é grande suficiente para produzir um grupo de catálogo com *redshift* profundo. O tamanho da caixa de simulação é limitado de  $z = 0,034$ , de um canto para outro, ou  $z = 0,058$  ao longo da diagonal maior. Portanto, as caixas de simulação são replicadas ao longo das três coordenadas cartesianas afim de se chegar a um raio mínimo desejado, criando um superbox, cujo observador é colocado no meio de um dos lados. O superbox tem um total de 256 cubos de simulação (ver a figura A.1).

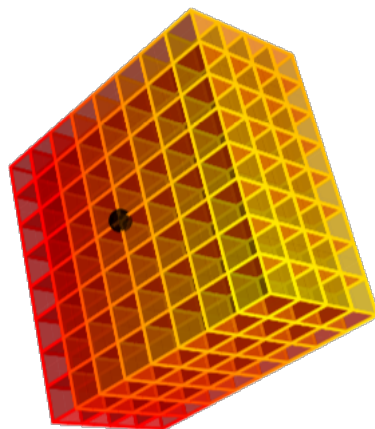


Figura A.1 – Representação do catálogo de galáxia simulado. Cada caixa tem 137 Mpc ao longo do cubo. O observador localiza-se no ponto preto, no meio de um dos lados do quadrado do superbox (DUARTE; MAMON, 2014).

### A.1.1 Friends-of-Friends (FoF)

O algoritmo Friends of Friends proposto por Huchra e Geller (1982), também conhecido por FoF, é considerado o método mais utilizado para identificar grandes estruturas no universo (Caretta et al. (2008) e Huchra e Geller (1982)). O método consiste na seguinte idéia: considera-se uma esfera de um dado raio, que engloba um objeto pertencente ao conjunto total e que está posicionado ao centro desta esfera. Todos os objetos que estiverem dentro da mesma esfera pertencem ao mesmo subconjunto, e são chamados de amigos. O procedimento se repete para cada objeto considerado amigo, e seguindo a regra “amigos dos amigos”, o grupo pode se expandir. O algoritmo se encerra quando não há mais “amigos” a incorporar.

O ponto fundamental do algoritmo é determinar o tamanho do raio de ligação. Huchra e Geller (1982) estabelecem dois parâmetros básicos que são  $D_L$  e  $V_L$ , onde  $D_L$  é a distância angular e  $V_L$  a velocidade de recessão. Dessa forma para verificar se um objeto é amigo, devemos comparar inicialmente, a sua distância angular e a sua velocidade de recessão. Ambas devem ser menores ou iguais aos valores comparados para serem considerados amigos. Duarte e Mamon (2014) fazem uma ampla discussão sobre a otimização na definição desses comprimentos de ligação na definição dos sistemas que farão parte do catálogo simulado.

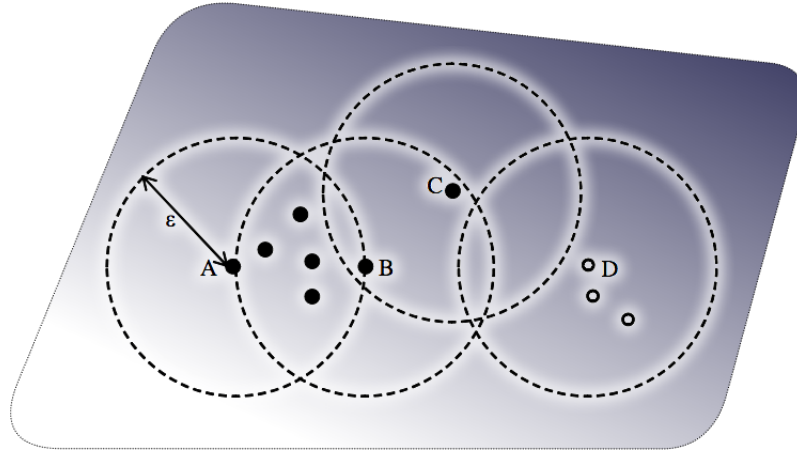


Figura A.2 – Duas partículas são consideradas amigas se a distância entre elas é inferior a uma distância limiar  $\varepsilon$ : A e B, B e C, são amigas, mas A e C não são. Amigos de amigos é definida entre dois pontos: se eles são amigos ou se eles estão contidos numa área de interesse (por exemplo, A e C é um amigo de amigo via B). Todos os pontos negros estão em um *cluster*, enquanto os pontos brancos caracterizam outro *cluster*.

assim, para o caso que a métrica torna-se diagonal

portanto,

### A.1.2 Evolução na Densidade Numérica de Aglomerados

Baseados em crescimento hierárquico, os modelos mais aceitos atualmente para descrever a formação de estruturas dispõem de um quadro em que pequenas flutuações de densidade no universo primordial aumentam com o passar do tempo. Os sistemas virializados observados atualmente (galáxias, aglomerados) são um produto final do crescimento dessas flutuações. É importante destacar a época de formação desses sistemas e, conseqüentemente, a distribuição das massas. No modelo que iremos apresentar, galáxias teriam sido formadas em *redshift* maior que 2, ao passo que aglomerados colapsariam após  $z \sim 1$ . Esse modelo prevê que estruturas virializadas com uma dada massa  $M$  são geradas quando flutuações de densidade atingem um valor superior a uma determinada densidade limite. Tal formalismo pode ser empregado para investigar a evolução da densidade numérica comóvel de halos de massa  $M$ , através de uma expressão  $n(z, M)$ , a qual representará a função de massa.

Uma simples estimativa de  $n(0, M)$  no universo local é suficiente para vincular  $\sigma_8$  e  $\Omega_m$  por  $\sigma_8 \Omega_m^{0,5} \sim 0,5$  (White et al. (1993a); Liddle et al. (1996)). Apesar de extremamente poderosa, esta relação é degenerada (a determinação unívoca de um parâmetro depende de uma determinação precisa do outro). Uma forma de quebrar essa degenerescência, até altos valores de desvio para o vermelho, é através da determinação da função de massa de aglomerados. A evolução da abundância de aglomerados é extremamente sensível a como esses parâmetros são

combinados. Em um universo com  $\Omega_m = 1$  (com baixo  $\sigma_8$ ) esperamos encontrar uma densidade de aglomerados ricos muito pequena em  $z \sim 1$  quando comparada com  $z \sim 0$ , o que corresponde a alta taxa de evolução na densidade de aglomerados. Por outro lado, num universo representado por um baixo valor de  $\Omega_m$  (com alto valor de  $\sigma_8$ ) a formação é quase interrompida em  $z \sim 1$ , representando uma pequena evolução na abundância de aglomerados (ver a figura A.3).

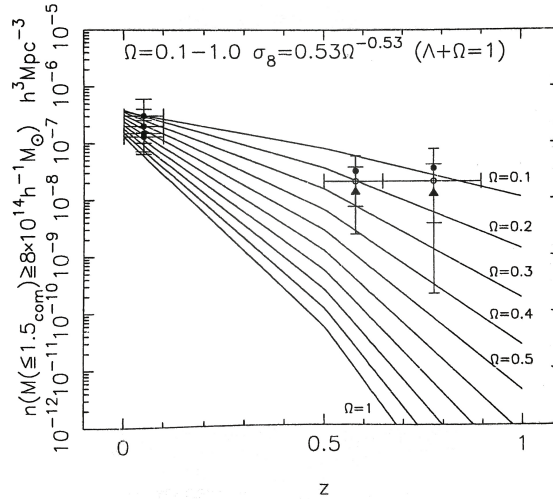


Figura A.3 – Evolução da abundância de aglomerados com desvio para o vermelho. As previsões são exibidas como curvas contínuas como função de  $\Omega_m(\sigma_8)$  (BAHCALL; FAN, 1998).

Portanto, a evolução da densidade numérica de aglomerados de alta massa representa um poderoso teste para determinação unívoca de  $\sigma_8$  e, conseqüentemente,  $\Omega_m$ . A função  $n(z, M)$  é válida para sistemas de qualquer massa, porém aglomerados ricos são os únicos sistemas com desvio para o vermelho de formação<sup>1</sup> não muito alto ( $z < 1$ ). Assim, catálogos de aglomerados ricos no universo local são geralmente completos e em mais alto  $z$  essas estruturas ainda são recuperadas sem grande custo observacional. E, dessa forma, podem ter a massa do virial determinada de forma precisa através de diferentes técnicas.

### A.1.3 Função de massa

As galáxias são objetos gravitacionalmente ligados no universo, abrangem uma grande dinâmica com relação a massa. Onde,  $F(M)dM$ , é a densidade do número de objetos no intervalo de massa  $(M, M + dM)$  (geralmente chamado de *função de massa*, também discutida no próximo tópico) e  $F(M)$  é a densidade de número de objetos com massas maiores que  $M$ . A formação destes objetos correspondem a um processo intrinsecamente não-linear, uma vez que a teoria linear desenvolvida até o momento não pode ser usada para determinar  $F(M)$ . Tal fato, no entanto, não é inteiramente verdade. Em qualquer uma realização da densidade linear

<sup>1</sup>Época em que o sistema tornou-se virializado.

de campo  $\delta_R(x)$ , (filtrado e usando uma função de janela de escala  $R$ ), haverá regiões com alta densidade (isto é, regiões com  $\delta_r > \delta_c$  onde  $\delta_c$  é o contraste de densidade crítico necessário para um colapso). Parece razoável supor que essas regiões eventualmente condensam como objetos ligados. Embora a dinâmica da região seja não-linear, o processo de condensação parece ser improvável, uma vez que muda a massa contida nessa região de forma significativa. Portanto, pode-se estimar o número médio de regiões com  $\delta_R > \delta_c$  em um campo Gaussiano aleatório, destes modo, seremos capazes de determinar  $F(M)$ . Uma maneira aproximada de conseguir tal descrição é fazer considerar um campo de densidade  $\delta_R(x)$ , suavizado por uma função de janela  $W_R$  com raio de escala  $R$ . A probabilidade de que esse campo tenha um valor  $\delta$ , em qualquer ponto escolhido é

$$P(\delta, t) = \left[ \frac{1}{2\pi\sigma^2(R, t)} \right]^{1/2} \exp\left(-\frac{\delta^2}{2\sigma^2(R, t)}\right), \quad (\text{A.1})$$

sendo

$$\sigma^2(R, t) = \int \frac{d^3k}{(2\pi)^3} |\delta_k(t)|^2 W_k^2(R), \quad (\text{A.2})$$

onde  $\sigma(R, t)$  é a variância do raio extrapolado no instante  $t$ . Como uma primeira aproximação, podemos supor que a região com  $\delta > \delta_c(t, t_i)$  irá formar um objeto gravitacionalmente ligada com massa  $M \propto \bar{\rho}R^3$  pelo tempo  $t$ . Aqui  $\delta_c \simeq 1$  no tempo  $t$ . Portanto, a fração de objetos ligados com massa superior a  $M$  será

$$F(M) = \int_{\delta_c(t, t_i)}^{\infty} P(\delta, R, t_i) d\delta. \quad (\text{A.3})$$

De A.1 têm

$$P(\delta, R, t_i) = \left[ \frac{1}{2\pi\sigma^2(R, t_i)} \right]^{1/2} e^{-\frac{\delta^2}{2\sigma^2(R, t_i)}}, \quad (\text{A.4})$$

substituindo em A.3

$$F(M) = \int_{\delta_c(t, t_i)}^{\infty} \left[ \frac{1}{2\pi\sigma^2(R, t_i)} \right]^{1/2} e^{-\frac{\delta^2}{2\sigma^2(R, t_i)}} d\delta \quad (\text{A.5})$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma(R, t_i)} \int_{\delta_c(t, t_i)}^{\infty} e^{-\frac{\delta^2}{2\sigma^2(R, t_i)}} d\delta, \quad (\text{A.6})$$

e sendo a **função complementar de erro**

$$\int_x^{\infty} e^{-a\lambda^2} d\lambda = \frac{\sqrt{\pi}}{2\sqrt{a}} \operatorname{erfc}(x\sqrt{a}), \quad (\text{A.7})$$

tem  $a = \frac{1}{2\sigma^2 R(R, t_i)}$  e  $x = \delta_c(t, t_i)$ .

É fácil mostrar que

$$F(M) = \frac{1}{2} \operatorname{erfc}\left(\frac{\delta_c(t, t_i)}{\sqrt{2}\sigma(R, t_i)}\right). \quad (\text{A.8})$$

A função de massa é apenas  $\frac{\partial F}{\partial M}$ ; a densidade de número,  $N(M, t)$ , pode ser encontrada dividindo essa expressão por  $\left(\frac{M}{\rho}\right)$ . Realização essas operações obtemos

$$N(M, t)dM = - \left(\frac{\bar{\rho}}{M}\right) \left(\frac{1}{2\pi}\right)^{1/2} \left(\frac{\delta_c}{\sigma}\right) \left(\frac{1}{\sigma} \frac{d\sigma}{dM}\right) \exp\left(-\frac{\delta_c^2}{2\sigma^2}\right) dM. \quad (\text{A.9})$$

Dado o espectro de potência  $|\delta_k(t)|^2$  e uma função de janela  $W(k, R)$  pode-se calcular explicitamente o lado direito dessa expressão. Há, contudo uma dificuldade fundamental com a expressão A.8. A integral de  $F(M)$ , em todo  $M$  deve ser a unidade, mas é fácil ver que, para a expressão em A.8

$$\int_0^\infty F(M)dM = \int_0^\infty dF = \frac{1}{2}. \quad (\text{A.10})$$

Isto acontece porque não levamos em conta as regiões sub-densas corretamente.

A probabilidade que um ponto satisfaça a condição  $\delta < \delta_c$  na escala  $R$ , representada pelos filtros  $R_1, R_2, \dots, R_n$ , com dispersões  $\Delta_1, \Delta_2, \dots, \Delta_n$  é dada por

$$P_{survive} \equiv P_s = \int_{-\infty}^{\delta_c} d\delta_1 \int_{-\infty}^{\delta_c} d\delta_2 \dots \int_{-\infty}^{\delta_c} d\delta_n P_J(\delta_1, \delta_2 \dots \delta_n), \quad (\text{A.11})$$

onde  $P_J[\delta_i]$  é a distribuição de probabilidade conjunta que as variáveis gaussianas  $\delta_i$  tomam um conjunto de valores simultâneos. Obviamente,  $(1 - P_s)$  dá a probabilidade de que um ponto se torna sobredenso em algum lugar ao longo da sequência de filtragens  $(R_1, \dots, R_n)$ . As variáveis gaussianas obtidas por diferentes escalas de filtragem não são independentes. Podemos ver que

$$\begin{aligned} \langle \delta_a \delta_b \rangle &= \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{d^3 \mathbf{p}}{(2\pi)^3} W_k(R_a) W_p^*(R_b) \langle \delta_k \delta_p^* \rangle e^{i(k-p).x} \\ &= \int \frac{d^3 \mathbf{k}}{(2\pi)^3} W_k(R_a) W_k^*(R_b) \sigma_k^2, \end{aligned} \quad (\text{A.12})$$

é, em geral, diferente de zero. Assim, o cálculo de A.11 é uma tarefa não trivial. Podemos olhar para este processo de uma forma diferente, porém, equivalente. Considere qualquer local fixo no espaço. Quando a escala de filtragem é um valor grande  $R_1$  (com uma dispersão  $\Delta_1$ ), podemos supor que este ponto tem um contraste de densidade  $\delta_1$ . Quando reduzimos a escala para  $R_2$ , teremos uma nova probabilidade de distribuição para que  $\delta$  deixe o valor do contraste de densidade cujo ponto escolhido passa a ser  $\delta_2$ . De forma sucessiva teremos uma sequência de escalas de filtragem,  $R_1, R_2, \dots, R_n$  (em ordem decrescente), o contraste de densidade realiza um passeio aleatório através dos pontos  $(\delta_1, \delta_2 \dots \delta_n)$ . Então atribuindo o ponto escolhido a uma massa  $M_k \propto \bar{\rho} R_k^3$ . Observa-se que, uma vez que  $\delta < \delta_c$  para todas as escalas maiores de filtragem, ou seja, para todos  $(R_1, R_2 \dots R_{k-1})$ , este ponto não pertence a nenhuma alta massa. O conceito de passeio aleatório simples se traduz em uma forma pictórica de seu conteúdo em A.11. Esse problema é igualmente difícil de resolver porque as etapas não são independentes. Na verdade, a resposta dependerá claramente da correlação entre as etapas, e de A.12, segue-se que

a resposta vai depender essencialmente da forma da função de janela. A partir de agora vamos considerar uma análise simples da função de janela através de seu truncamento no espaço  $\mathbf{k}$ , que é  $W_k(R) = (R^{-1} - k)$ . De A.12 segue-se que, (para  $a \leq b$ )

$$\langle \delta_a \delta_{a+1} \rangle = \sigma_a^2; \langle \delta_a \delta_b \rangle = \sigma_a^2. \quad (\text{A.13})$$

O comprimento dos passos dos caminhos aleatórios são  $l_1 \equiv (\delta_2 - \delta_1)$ ,  $l_2 = (\delta_3 - \delta_2)$ , ...,  $l_a = (\delta_{a+1} - \delta_a)$  etc. Cada um desses comprimentos é uma variável Gaussiana com a dispersão

$$\begin{aligned} \langle l_a^2 \rangle &= \langle (\delta_{a+1} - \delta_a)^2 \rangle \\ &= \langle \delta_{a+1} \delta_{a+1} - 2\delta_{a+1} \delta_a + \delta_a \delta_a \rangle \\ &= \langle \delta_{a+1} \delta_{a+1} \rangle - 2 \langle \delta_{a+1} \delta_a \rangle + \langle \delta_a \delta_a \rangle, \end{aligned}$$

e de A.13

$$\begin{aligned} &= \sigma_{a+1}^2 + \sigma_a^2 - 2 \langle \delta_{a+1} \delta_a \rangle \\ &= \sigma_{a+1}^2 - \sigma_a^2. \end{aligned} \quad (\text{A.14})$$

Já a correlação cruzada

$$\begin{aligned} \langle l_a l_b \rangle &= \langle (\delta_{a+1} - \delta_a)(\delta_{b+1} - \delta_b) \rangle \\ &= \langle \delta_{a+1} \delta_{b+1} - \delta_{a+1} \delta_b - \delta_a \delta_{b+1} + \delta_a \delta_b \rangle \\ &= \langle \delta_{a+1} \delta_{b+1} \rangle - \langle \delta_{a+1} \delta_b \rangle - \langle \delta_a \delta_{b+1} \rangle + \langle \delta_a \delta_b \rangle, \end{aligned}$$

recorrendo, novamente, a A.13, agora para  $(a+1) < b$ , temos

$$\langle l_a l_b \rangle = \sigma_{a+1}^2 - \sigma_{a+1}^2 - \sigma_a^2 + \sigma_a^2 = 0. \quad (\text{A.15})$$

Em outras palavras, um filtro apurado no espaço  $k$  produz um passeio aleatório, em que cada passo  $l_a$  é independente e extraído de uma variável gaussiana com dispersão  $(\sigma_{a+1}^2 - \sigma_a^2)$ . No limite contínuo, esse passeio aleatório é descrito por uma equação de difusão. A probabilidade  $P(\delta, \sigma^2)$  de uma partícula está em  $(\delta, \delta + d\delta)$ , quando a dispersão  $\delta^2$  obedece à equação de difusão

$$\frac{\partial P}{\partial \sigma^2} = \frac{1}{2} \frac{\partial^2 P}{\partial \delta^2}. \quad (\text{A.16})$$

Nós estamos interessados na probabilidade de que a trajetória chegue a  $(\delta, \sigma)$ , sem exceder  $\delta_c$  anterior, ou seja, menor que  $\sigma$ . Isso é equivalente a resolver A.16 com a condição de contorno  $\delta = \delta_c$ . Assim

$$P(\delta, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} \left[ \exp\left(-\frac{\delta^2}{2\sigma^2}\right) - \exp\left(-\frac{(\delta - 2\delta_c)^2}{2\sigma^2}\right) \right]. \quad (\text{A.17})$$

Integrando esta expressão de  $\delta_c$  a  $\infty$  e diferenciando com relação a  $M$ , tem

$$dF = \sqrt{\frac{2}{\pi}} \frac{\delta_c}{\sigma_2} \left( -\frac{\partial \sigma}{\partial M} \right) \exp \left( -\frac{\delta_c^2}{2\sigma^2} \right) dM, \quad (\text{A.18})$$

que é precisamente o dobro do valor obtido através de A.8. Este resultado pode ser expresso na forma explícita de espectros de potência com  $|\delta_c|^2 \propto k^n$ . Nesse caso  $\sigma^2 = c^2 M^{-(3+n)/3} t^{4/3}$ , onde  $c$  é uma constante. Uma vez que o limite da teoria linear ocorre quando  $\sigma \simeq 1$ , com escala característica  $M_{nl}$ , que por sua vez, obedece à relação de escala  $M_{nl} \propto t^{4/(3+n)}$ . Portanto,  $c^2 = M_{nl}^{(3+n)/3} (t_0) t_0^{-4/3}$ , sendo  $t_0$  a época atual. Assim:  $\delta_c(t_0, t_i) \cong (t_i/t_0)^2/3$ , temos

$$\begin{aligned} \sigma^2 &= M_{nl}^{(3+n)/3} (t_0) t_0^{-4/3} M^{-(3+n)/3} t_i^{4/3} \\ &= \frac{M_{nl}^{(3+n)/3} (t_0) t_i^{4/3}}{M^{(3+n)/3} t_0^{4/3}} \\ \sigma &= \left[ \frac{M_{nl}^{(3+n)/3} (t_0) t_i^{4/3}}{M^{(3+n)/3} t_0^{4/3}} \right]^{1/2} \\ \sigma(M, t_i) &= \frac{M_{nl}^{(3+n)/6} (t_0) t_i^{2/3}}{M^{(3+n)/6} t_0^{2/3}}, \end{aligned} \quad (\text{A.19})$$

portanto, a razão

$$\frac{\delta_c(t_0, t_i)}{\sigma(M, t_i)} = \left( \frac{t_i}{t_0} \right)^{2/3} \frac{M^{(3+n)/6}}{t_i^{2/3}} \frac{t_0^{2/3}}{M_{nl}^{(3+n)/6} (t_0)} = \left[ \frac{M}{M_{nl}(t_0)} \right]^{(3+n)/6}, \quad (\text{A.20})$$

onde  $M_{nl}(t_0)$  é a escala que está se tornando hoje (não-linear). A expressão para  $N(M, t_0)$  torna-se agora

$$N(M, t_0) dM = \frac{\bar{\rho}}{\sqrt{2\pi}} \left( 1 + \frac{n}{3} \right) \left( \frac{M}{M_{nl}} \right)^{(3+n)/6} \times \exp \left[ -\frac{1}{2} \left( \frac{M}{M_{nl}} \right)^{(3+n)/3} \right] \frac{dM}{M^2}, \quad (\text{A.21})$$

onde  $\bar{\rho}$  é a densidade média comóvel do universo. Uma escolha diferente da função de janela, em geral, dará também um diferente resultado, especialmente, no limite de baixa massa.

#### A.1.4 Função de massa e Aglomerados de galáxias

Tendo como base o crescimento de pequenas flutuações de densidade no universo primordial os modelos que melhor descrevem a formação de estruturas são baseados em crescimentos hierárquicos, isto é, galáxias, aglomerados de galáxias (sistemas virializados atualmente observados). Um modelo usualmente empregado para descrição (formação e evolução) de estruturas é o formalismo de Press-Schechter (PS). Este formalismo tem como ponto de partida a hipótese de que a estatística do campo primordial das perturbações de densidade é bem descrita por uma distribuição gaussiana. Talvez uma das dificuldades fundamentais encontrada pela tentativa de fazer previsões sobre a densidade de número de galáxias é que a escala galáctica é não-linear. Que



escala no universo (sem perturbações) encerra a massa de uma galáxia típica de  $M = 10^2 M_\odot$ ? Neste sentido, a densidade de uma região esférica de raio  $R$  é

$$\rho_m = \frac{M}{4\pi R^3/3}, \quad (\text{A.22})$$

sendo  $\rho_m = \Omega_m \rho_{cr}$ , encontramos

$$R = 0.951 h^{-1} Mpc \left( \frac{Mh}{10^{12} \Omega_m M_\odot} \right)^{1/3}. \quad (\text{A.23})$$

Então, uma galáxia vem da matéria dentro de um raio de cerca de  $1 Mpc$ , o que corresponde a flutuações em escalas da ordem  $k \sim 1 h Mpc^{-1}$ , dentro do regime não-linear. Isso responde a uma pergunta: Por que não tentar prever a densidade do número de galáxias? Se invertermos a equação A.23 para obter a massa contida dentro de uma esfera de raio  $R$ , temos

$$M = 1.16 \times 10^{15} \Omega_m h^{-1} M_\odot \left( \frac{R}{10 h^{-1} Mpc} \right)^3, \quad (\text{A.24})$$

então vemos que os aglomerados de galáxias, com massas até  $10^{15} M_\odot$  surgem de perturbações. Como então prever a abundância de aglomerados de galáxias? A ideia básica vem de um trabalho de Press e Schechter (1974). O formalismo de PS é um dos modelos mais empregados para descrever a formação e evolução de estruturas. Esta teoria prevê que estruturas virializadas com uma dada massa  $M$  são geradas quando flutuações de densidade atingem um valor superior a uma densidade limite ( $\delta_c$ ). Para entender o seu argumento, consideraremos um campo de densidade unidimensional, cuja inomogeneidade média é zero. O foco de nosso interesse concentra-se nas raras regiões de grande densidade cujas características entram em colapso, acumulam matéria (superior que a gravidade local) tornando-se mais importante do que o fluxo de Hubble. Partículas nesta região param de se afastar (longe um do outro) e estão presos no campo gravitacional local.

A Teoria Press-Schechter prevê a fração do volume que entrou em colapso como

$$f_{coll}(M(R), z) = \frac{2}{\sqrt{2\pi}\sigma^2(R, z)} \int_{\delta_c}^{\infty} d\delta e^{-\delta^2/2\sigma^2(R, z)}. \quad (\text{A.25})$$

Aqui,  $R$  é o raio sobre o qual o campo de densidade tem sido suavizado. Este raio é usado para calcular  $\sigma(R, z)$ , isto é, variância na escala  $R$  linearmente extrapolada no *redshift*  $z$ . Normalmente, inomogeneidades em grandes escalas são menores em magnitude quando comparadas em menores escalas. O lado direito da equação A.25 faz contagens de todas as partes da distribuição Gaussiana para as quais as maiores densidades superam a densidade crítica  $\delta_c$ .

Press-Schechter nos diz que a fração colapsada pode ser obtida usando  $\sigma$  (linear). Essas peculiaridades do formalismo Press-Schechter não diminuí sua eficácia. As simulações numéricas (por exemplo, White et al. (2010)) têm mostrado que funcionam muito bem. Além disso, alguns grupos têm justificado teoricamente os aspectos de sua fórmula (PEACOCK; HEAVENS, 1990).

Para obter a fração colapsada numa forma mais comparável com as observações, primeiro diferenciamos  $f_{coll}$  com relação a  $M$  e multiplicamos pelo seu intervalo infinitesimal  $dM$ . Isto dá a fração de volume colapsada de objetos com massa entre  $M$  e  $M + dM$ . Assim, multiplicamos este valor pela densidade média do número dos objetos  $\rho_m/M$  para obter a densidade do número de objetos colapsados com massa entre  $M$  e  $M + dM$ ,

$$dn(M, z) = -\frac{\rho_m}{M} \frac{df_{coll}(M(R), z)}{dM} dM. \quad (\text{A.26})$$

O sinal negativo aparece aqui desde  $f_{coll}$ , uma função decrescente da massa  $M$ . Assim, efetuamos a derivada usando o fato de que  $dM/dR = 3M/R$ , para obtermos a *função de massa*

$$\frac{dn(M, z)}{dM} = \sqrt{\frac{2}{\pi}} \frac{\rho_m \delta_c}{3M^2 \sigma} e^{-\delta_c^2/2\sigma^2} \left[ -\frac{R}{\sigma} \frac{d\sigma}{dR} \right], \quad (\text{A.27})$$

onde a densidade média de matéria pode ser escrita como  $\rho_m \equiv \Omega_m \rho_c$ ; e  $\delta_c(z) = \delta_c/D(z, \Omega_m, \Lambda)$ , com  $\delta_c \sim 1.68$  (contraste de densidade crítico necessário para um colapso), já  $D(z, \Omega_m, \Lambda)$  é conhecido como fator de crescimento de uma dada perturbação num regime linear. Se  $\sigma(z, M)$  (flutuação de massa para diferentes valores de  $M$ , num dado desvio para o vermelho) é pequena, em uma determinada escala, então a densidade do número de objetos em escala exponencial é suprimida. Há duas abordagens para contornar a questão do valor numérico de  $\delta_c$ . A primeira é contar com um modelo simples de colapso (modelo esférico). Estudos mostram que o colapso ocorre, quando  $\delta = 1,686$  (para  $\rho_m = \rho_{cr}$ ). Por outro lado, a outra forma é simplesmente tratar  $\delta_c$  como um parâmetro livre e calibrá-lo com simulações numéricas (ver (EKE et al., 1996)). O valor de  $\delta_c$  obtido desta forma é perto do valor obtido dentro de um modelo de colapso esférico adotado normalmente pela comunidade científica ( $\delta_c = 1,686$ ).

Medir a abundância de aglomerados e, portanto, testar teorias com a equação A.27 é um procedimento sutil. Uma das classes de dificuldades consiste na identificação dessas estruturas, onde sofisticados algoritmos vêm sendo desenvolvidos a fim de amenizar tamanho problema. Outra dificuldade em questão, gira em torno da determinação da massa de aglomerados de galáxias, sobretudo, em sua melhor calibração. Diante dessa dificuldade, é fundamental a obtenção de outras medidas globais que mantenham uma estreita relação com a massa. Essas quantidades são altamente desejáveis, por permitirem maior precisão ao mensurar a função de massa (ROZO et al., 2009b). Contudo, as quantidades consideradas observáveis diretas introduzem dispersões na calibração de massa.

Neste contexto, busca-se identificar e controlar as incertezas sistemáticas nas análises de dados sobre as diversas modelagens, como por exemplo, sondas de matéria escura, a fim de melhorar as estimativas. Uma das principais incertezas em contagem de aglomerados é a modelagem teórica da densidade espacial de halo em função da massa. Esta, por sua vez, incrementa incerteza na modelagem da função de seleção sob um levantamento específico. Estudos recentes mostram maiores consistências nas restrições cosmológicas, as quais vem

gerando maior volume e sensibilidade nos levantamentos a fim de melhorar a análise das fontes de erro. Estes se apresentam em duas categorias com relação massa-observável: o viés decorrente do conhecimento incerto da normalização e a variação de massa diante da dispersão na relação massa-observável. (CUNHA; EVRARD, 2010) assumem uma forma log-normal para a probabilidade de medir um sinal (observável), denotada  $M_{obs}$ , dada a massa  $M$ ,

$$p(M_{obs} | M) = \frac{1}{\sqrt{2\pi}\sigma_{\ln M}} \exp[-x^2(M_{obs})], \quad (\text{A.28})$$

onde

$$x(M_{obs}) \equiv \frac{\ln M_{obs} - \ln M - \ln M_{bias}(M_{obs}, z)}{\sqrt{2}\sigma_{\ln M}(M_{obs}, z)}. \quad (\text{A.29})$$

O erro sistemático em seu modelo (no *proxy* de massa) é introduzida de um viés (bias) dependente do *redshift* e da variância mediante a medição

$$\ln M_{bias}(z) = B_0 + B_1(1 + z), \quad (\text{A.30})$$

$$\sigma_{\ln M}^2(z) = \sigma_0^2 + \sum_{i=1}^3 S_i z^i, \quad (\text{A.31})$$

onde  $B_0$ ,  $B_1$ ,  $\sigma_0$  são coeficientes considerados independentes da massa, sendo também graus de liberdade na análise. Todavia, considerá-los independentes da massa constitui ainda assim uma atitude conservadora (CUNHA; EVRARD, 2010).

Contudo, a função de massa oferece ferramentas para o estudo da matéria (e que inclui a energia escura) e para a contagem de aglomerados de galáxias, bem como a análise da distribuição de aglomerados por massa (e como função do *redshift*), sendo sensível aos parâmetros cosmológicos (que contribuem para a função de massa dos halos de galáxias em aglomerados). Portanto, a função de massa estima, para um dado volume e *redshift*, a quantidade de objetos num certo intervalo de massa. Esta função tem importância central para a astrofísica e cosmologia em virtude de sua dependência com importantes ingredientes da teoria de formação de estruturas (espectro de potência de massa, fator de crescimento e o contraste de densidade crítico). Isto faz com que a contagem de aglomerados (determinação de abundância) seja um dos temas mais estudados atualmente.