Universidade de São Paulo
Instituto de Astronomia, Geofísica e Ciências Atmosféricas
Departamento de Astronomia

Nathan Vieira de Oliveira

Agrupamento de Quasares no Catálogo QuCatS/S-PLUS

São Paulo
2025

Nathan Vieira de Oliveira

Agrupamento de Quasares no Catálogo QuCatS/S-PLUS

Trabalho de Conclusão de Curso apresentado ao Instituto de Astronomia, Geofísica e Ciências Atmosféricas da Universidade de São Paulo como requisito parcial para a obtenção do título de Bacharel em Astronomia.

Vertente: Pesquisa Básica
Orientador: Prof. Dr. Laerte Sodré Junior

São Paulo

2025

Aos meus pais, Elaine e Gilson, minha irmã, Beatriz, as minhas avós Nadir e Zilda e à memória do meu vô Paulo.

Agradecimentos

Agradeço à minha família que apoiou meus estudos durante toda a minha vida. Em especial, à minha irmã que me incentivou a, de fato, optar por esse curso. Também agradeço ao projeto Acaia Sagarana e a todos seus professores, monitores e coordenadores, que, mesmo durante uma pandemia, me forneceram conhecimentos importantes para o meu ingresso e formação nesta universidade.

Agradeço aos meus amigos Gabriel Capellini, Gustavo Chinen, Vitor Saito, Lucas Barão e Rafael Gregório pelas trocas de conhecimento e por terem tornado essa jornada muito mais agradável. Também agradeço à minha namorada Thamires Lopes Ferreira, que me apoiou e incentivou durante todos os momentos.

Por fim, também agradeço à Profª. Dra. Tatiana Alexandrovna Michtchenko, que me ensinou e orientou durante dois projetos de Iniciação Científica e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pela fomentação desses projetos. Agradeço também ao Prof. Dr. Laerte Sodré Júnior pelas orientações e direcionamentos durante todo esse trabalho de graduação.

Esta tese/dissertação foi escrita em LaTeX com a classe IAGTESE, para teses e dissertações do IAG.

"*Todo dia é dia de viver.*"

Milton Nascimento

"*[...] E filhos netos bisnetos tataranetos despojados dos bens mais sólidos e rutilantes*
[portanto os mais completos
irão tomando a pouco e pouco desapego de toda fortuna
e concentrando seu fervor numa riqueza só, abstrata e una [...]"

Carlos Drummond de Andrade

Resumo

Neste trabalho, estudamos os agrupamentos de quasares brilhantes ($r < 20$) do catálogo *Quasar Candidates at S-PLUS* (QuCatS; Nakazono et al. 2024). Para isso, cruzamos os dados com o catálogo de candidatos a quasares do GAIA e geramos, além do QuCatS, mais duas amostras - resultantes de comparações diretas entre as probabilidades de serem quasares e dos *redshifts* fotométricos. Para uma análise global, calculamos funções de correlação para nossas amostras e as comparamos com os resultados obtidos do catálogo de quasares (candidatos) do SDSS (GAIA), obtendo que elas se comportam de acordo com o esperado de dados fotométricos. Para encontrar as estruturas, utilizamos um algoritmo de agrupamento cuja base é o DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*; Ester et al. 1996), e as analisamos através das estatísticas básicas de grafos e de funções de multiplicidade. Com isso, concluímos que não existem diferenças significativas entre nossas amostras e que suas abundâncias, em relação a recortes semelhantes do catálogo SDSS, estão de acordo com o esperado. Ademais, através do DBSCAN e do cruzamento de dados, identificamos 31 candidatos promissores a pares de quasares próximos (distâncias angulares menores que 10 segundos de arco). Desses, um caso binário e um caso projetado já foram previamente confirmados por espectro na literatura. Por fim, através de um algoritmo baseado no método CHMS (*Convex Hull of Members Spheres*; Clowes et al. 2011), identificamos e apresentamos dois candidatos promissores a Large Quasar Group (LQG).

Abstract

In this work, we study clusterings of bright quasars ($r < 20$) from the *Quasar Candidates at S-PLUS* (QuCatS; Nakazono et al. 2024) catalog. For this purpose, we cross-matched the data with the GAIA quasar candidate catalog and generated, in addition to QuCatS, two additional samples - derived from direct comparisons between quasar probabilities and photometric redshifts. For a global analysis, we computed correlation functions for our samples and compared them with results from the SDSS (GAIA) quasar (candidate) catalog, finding that they behave as expected for photometric data. To identify structures, we employed a clustering algorithm based on DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*; Ester et al. 1996) and analyzed them using basic graph statistics and multiplicity functions. From this, we conclude that there are no significant differences between our samples and that their abundances, relative to similar subsets of the SDSS catalog, are consistent with expectations. Furthermore, using DBSCAN and data cross-matching, we identified 31 promising candidates to quasar pairs with angular separations smaller than 10 arcseconds. Among these, one binary system and one projected pair were spectroscopically confirmed in previous studies. Finally, using an algorithm based on the CHMS method (*Convex Hull of Members Spheres*; Clowes et al. 2011), we identified and present two promising Large Quasar Group (LQG) candidates.

Lista de Figuras

2.1 Comparação direta das probabilidades 22
2.2 Comparação das distribuições de probabilidades: QuCatS×GAIA 23
2.3 Comparação dos *redshifts*: QuCatS×GAIA 24

3.1 Regiões filtradas para o cálculo da função de correlação 26
3.2 Visualização das intersecções das cascas esféricas adotadas 28
3.3 Percolação e Saturação do parâmetro L 30

4.1 Funções de correlação 36
4.2 Funções de multiplicidade 39
4.3 Estruturas típicas encontradas na amostra QuCatS. 40
4.4 Distribuições estatísticas de grafos utilizadas para descrever as estruturas. . 41
4.5 Candidatos a pares de quasares detectados pelo DBSCAN. 43
4.6 Candidatos promissores a LQGs. 47

A.1 Comparação entre K-means e DBSCAN 59

B.1 Exemplo da redefinição dos grupos 61

C.1 Distribuição dos candidatos do QuCatS na esfera celeste 63

D.1 Regiões selecionadas do catálogo de quasares do SDSS-DR16 65
D.2 Região selecionada do catálogo de candidatas a quasares do GAIA 66

E.1 Funções de correlação: Prob_QuCatS × Prob_GAIA e z_QuCatS×z_GAIA 67

G.1 Candidatos a pares de quasares descartados. 72

H.1 Candidatos advindos dos cruzamentos encontrados na literatura. 75

Lista de Tabelas

4.1 Distribuição dos agrupamentos resultantes da amostra QuCatS 37
4.2 Características dos candidatos a pares de quasares inicialmente encontrados 42
4.3 Pares de objetos próximos com espectro. 44
4.4 Candidatos a LQGs. 46

F.1 Distribuição dos agrupamentos resultantes da amostra Prob_QuCatS×Prob_GAIA. 69
F.2 Distribuição dos agrupamentos resultantes da amostra z_QuCatS×z_GAIA. 69

H.1 Pares de objetos encontrados através do cruzamento de dados. 73

Sumário

1. Introdução 19

2. Tratamento de dados 21

2.1 Comparação das probabilidades 22

2.2 Comparação dos *redshifts* 23

3. Metodologia 25

3.1 Definição da função de correlação 25

3.2 Descrição da aplicação do algoritmo DBSCAN 27

3.3 Redefinição dos grupos 28

3.4 Seleção do parâmetro L 29

3.5 Estatísticas básicas dos grafos 31

3.6 Função de multiplicidade 32

3.7 Descrição da aplicação do método CHMS 33

4. Resultados 35

4.1 Função de correlação 35

4.2 Distribuição dos agrupamentos 37

4.3 Função de multiplicidade 38

4.4 Estudo das estruturas 39

4.4.1 Aplicação das estatísticas de grafos 40

4.5 Pares de quasares 42

4.5.1 Pares detectados pelo DBSCAN 42

4.5.2 Pares detectados pelo cruzamento com o GAIA 44

4.6 Candidatos a LQGs 45

5. Conclusões 49

Referências 51

Apêndice 57

A. Comparação entre DBSCAN e K-means 59

B. Exemplo da redefinição dos grupos 61

C. Distribuição dos candidatos do QuCatS na esfera celeste 63

D. Regiões selecionadas dos catálogos SDSS e GAIA 65

E. Função de correlação das amostras restantes 67

F. Tabelas de distribuição dos agrupamentos 69

G. Candidatos a pares de quasares descartados 71

H. Candidatos a pares de quasares detectados pelo cruzamento 73

Capítulo 1

Introdução

Descobertos na década de 1950, os quasares, simplificação do termo *quasi-stellar object*, foram descobertos por suas emissões em rádio. Com tamanhos físicos maiores do que estrelas, porém menores do que galáxias, esses objetos possuem origens ligadas aos seus núcleos ativos. Esses, por sua vez, são formados por buracos negros supermassivos que, através de mecanismos de acreção, geram emissões de altíssimas energias (Jatenco-Pereira et al., 2006).

Diante disso, por serem fortes emissores e serem relativamente comuns nos primórdios do Universo, esses objetos podem ser observados em grandes *redshifts* (z) - recentemente, Wang et al. (2021) reportou um quasar com z igual a 7.642, ou seja, quando esses fótons foram emitidos, o Universo possuía menos de um bilhão de anos. Por isso, o estudo desses corpos e de suas distribuições espaciais é de extremo interesse para a cosmologia.

Em especial, destacamos o estudo dos Grandes Grupos de Quasares (*Large Quasar Group* - LQG). Esses grupos, que variam de algumas unidades até dezenas de componentes, são considerados precursores das maiores estruturas conhecidas como superaglomerados e Estruturas em Grande Escala (*Large Scale Structure* - LSS) (Clowes, 2001). Essa conexão é fundamentada no atual modelo padrão, Λ-CDM (*Cold Dark Matter*), que descreve a formação das estruturas de modo hierárquico, ou seja, das menores para as maiores.

O maior grupo conhecido até o momento, Huge-LQG (Clowes et al., 2013), possui 73 componentes e está distribuído ao redor de $z \sim 1.27$. Ele e outras LSS, como o *Giant Arch* (Lopez et al., 2022) e a *Sloan Great Wall* (Gott III et al., 2005), são tidos por alguns autores como estruturas que violam o modelo Λ-CDM. Isso ocorre, pois o Princípio Cosmológico, que fornece a base para o seu desenvolvimento teórico, assume que, em grandes escalas, o Universo deve ser isotrópico e homogêneo (Ryden, 2016) e, por isso, a presença dessas

grandes estruturas pode violar ou, ao menos, propor um novo limite para que ele possa ser válido. Ressaltamos que essas violações não são um consenso - mais detalhes dessa discussão podem ser encontrados em Abdalla et al. (2022).

Destacamos, também, a procura por pares de quasares próximos entre si, ou seja, com distâncias angulares menores do que 10 segundos de arco. Sendo eles binários ou resultantes de lentes gravitacionais, exemplares desses objetos são raríssimos de serem encontrados. Apesar disso, essas duplas podem ajudar em estudos de diversos fenômenos. Por exemplo, através deles, podemos restringir a constante de Hubble e a Função de Massa Inicial (lentes; Lemon et al. 2022) ou limitar a abundância de Buracos Negros Primordiais e caracterizar a formação e a população de Buracos Negros Supermassivos (binários; Shen et al. 2023).

Com isso em mente, propomos o estudo do agrupamento dos candidatos a quasares no catálogo *Quasar Candidates at S-PLUS* (QuCatS). Mais especificamente, os objetivos deste projeto são: identificar objetos isolados, pares e estruturas maiores em função do *redshift*; visualizar objetos de interesse, como pares de quasares próximos, e, por fim, descrever estatisticamente as estruturas encontradas.

Desse modo, utilizaremos o algoritmo DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*; Ester et al. 1996) para encontrar as estruturas e as ferramentas disponibilizadas por Cardoso (2022) para visualizar os objetos de interesse. Além disso, neste trabalho, adotaremos os seguintes parâmetros cosmológicos para um modelo Λ-CDM: $\Omega = 1$, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ e $H_0 = 70\ kpc/km/s$.

Capítulo 2

Tratamento de dados

Descrito em Nakazono et al. (2024), o catálogo QuCatS fornece candidatos a quasares presentes no projeto *Southern Photometric Local Universe Survey* (S-PLUS; Mendes de Oliveira et al. 2019). Contendo 645980, 244912 e 144991 fontes com probabilidades de serem quasares maiores que 80%, 90% e 95%, respectivamente, ele também fornece *redshifts* fotométricos, z_{fot}, magnitudes em diversas bandas e as posições das fontes no sistema internacional de referência celeste. Destacamos que, por uma questão de tratabilidade, uma vez que as análises feitas são, em geral, custosas computacionalmente, utilizaremos uma amostra das fontes mais brilhantes desse catálogo. Ou seja, baseando-se na magnitude da banda r, filtraremos os dados para $r \leq 20$, totalizando 54977 candidatos.

Antes da aplicação dos algoritmos pretendidos, realizamos o cruzamento dos dados da amostra brilhante com os do telescópio espacial GAIA (*Global Astrometric Interferometer for Astrophysics*), mais especificamente dos seus catálogos de parâmetros astrofísicos (Creevey, L. et al., 2023a) e extragalácticos (Gaia Collaboration et al., 2023). Ressaltamos que, por possuírem medidas astrométricas das fontes, os modelos de classificação desses catálogos conseguem descartar candidatos com grandes movimentos próprios, os quais, em geral, são estrelas próximas. Porém, conforme descrito em Creevey, L. et al. (2023b), por uma questão de completude da amostra, nenhum limite geral foi adotado.

Nesse sentido, dos 54977 objetos da amostra inicial, apenas 104 não foram identificados nos catálogos do GAIA. Ou seja, temos uma alta taxa de correspondência (99.81%) que sugere que o método utilizado para identificação das fontes é adequado. Além disso, em 33 casos, obtivemos duas correspondências dentro de um segundo de arco. Desse modo, apesar de não haver uma correspondência direta dessas fontes no QuCatS, por se tratarem de objetos próximos aos candidatos, também os examinamos em busca de pares de quasares

próximos.

Assim, nas Seções 2.1 e 2.2 discutiremos, respectivamente, as diferenças encontradas em relação as probabilidades e aos *redshifts* desses candidatos em comum. Por uma questão de completude, aplicaremos o DBSCAN para três amostras, sendo elas, pré e pós comparação de probabilidades e pós comparação de *redshift*. Essas amostras serão denotadas, respectivamente, como QuCatS, Prob_QuCatS× Prob_GAIA e z_QuCatS×z_GAIA.

2.1 Comparação das probabilidades

Dos 54873 candidatos em comum, todos possuem suas probabilidades de serem quasares estimadas pelo GAIA. Na Figura 2.1, apresentamos uma comparação direta dos valores estimados pelos dois catálogos. Nela, percebe-se que não existe uma correlação entre essas estimativas; um possível motivo para isso ocorrer está no fato de que os modelos classificadores são treinados com medidas independentes e de diferentes origens físicas - por exemplo, as medidas astrométricas do GAIA.

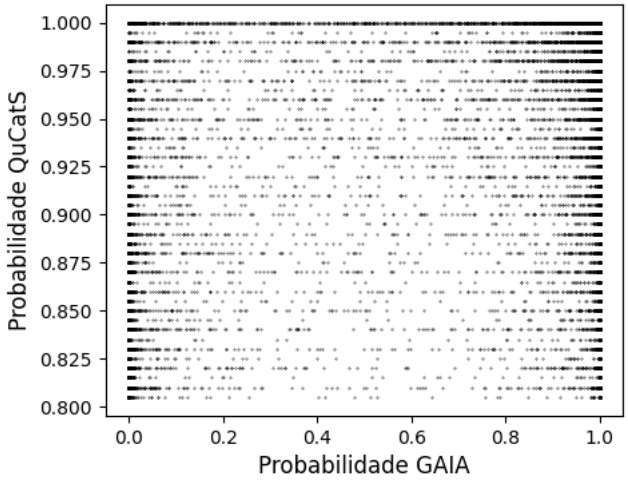
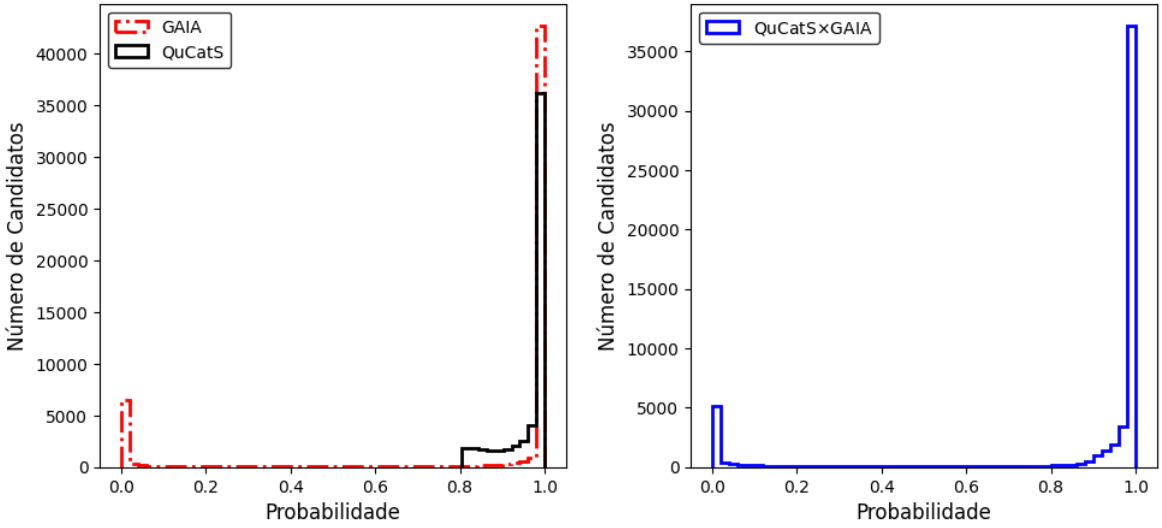


Figura 2.1: Comparação direta das probabilidades dos 54873 candidatos em comum. Note que as probabilidades estimadas pelo QuCatS estão, de fato, limitadas a serem maiores que 0.8 e possuem uma distribuição discreta, enquanto as do GAIA variam continuamente em todos os valores possíveis.

Assim, para esses objetos, podemos obter uma nova estimativa através da média geométrica das duas probabilidades (Equação 2.1). Escolhemos essa média por possibilitar que a multiplicação por um valor quase nulo também gere um novo valor quase nulo. Isso

pode ser visto na comparação das distribuições na Figura 2.2.

$$P_{novo} = \sqrt{P_{GAIA} \times P_{QuCatS}} \tag{2.1}$$



(a) (b)

Figura 2.2: Comparação das distribuições de probabilidades. Apresentamos, em (a), numa linha preta contínua e numa linha vermelha pontilhada as distribuições do GAIA e do QuCatS. Já em (b) apresentamos, em azul, a nova estimativa média QuCatS×GAIA. Nota-se que, conforme o esperado, para baixos valores de probabilidade, as distribuições do GAIA e da média seguem próximas ao nulo.

Aplicando, assim como feito inicialmente no QuCatS, um filtro para probabilidades maiores que 80% na nova estimativa média, obtivemos um total de 8912 candidatos descartados por serem pouco prováveis. Dentre esses objetos, através das outras estimativas do GAIA, concluímos que 8591 são estrelas e 321 são galáxias. Assim, conforme o esperado, descartamos uma grande quantidade de estrelas, resultando numa amostra com 46065 candidatos restantes denominada Prob_QuCatS× Prob_GAIA. Por fim, destacamos que o GAIA nem sempre possui inferências de todos os parâmetros, principalmente para as fontes menos brilhantes ($r > 18$), por isso, ainda é possível que hajam estrelas na amostra, uma vez que nossas fontes vão até $r \approx 20$.

2.2 *Comparação dos redshifts*

Por outro lado, dos 46065 candidatos restantes, 45960 possuem estimativas de *redshift* disponíveis no GAIA. Nesse caso, apresentamos uma comparação direta desses valores na

Figura 2.3. Nela, observamos que a região linear de inclinação igual a um é a mais densa, indicando que, para a maior parte dos dados, os catálogos estão de acordo. Nesse sentido, definimos as retas $y(x) = x \pm 0.2$ como os limites dessa região central. Logo, aplicando esses limites, obtemos uma amostra com 30141 candidatos em concordância que, por sua vez, formam z_QuCatS$\times z$_GAIA. Destacamos que diferentes inferências de *redshifts* não desqualificam os dados como possíveis quasares e, por isso, essa amostra representa apenas uma versão mais robusta dos dados.

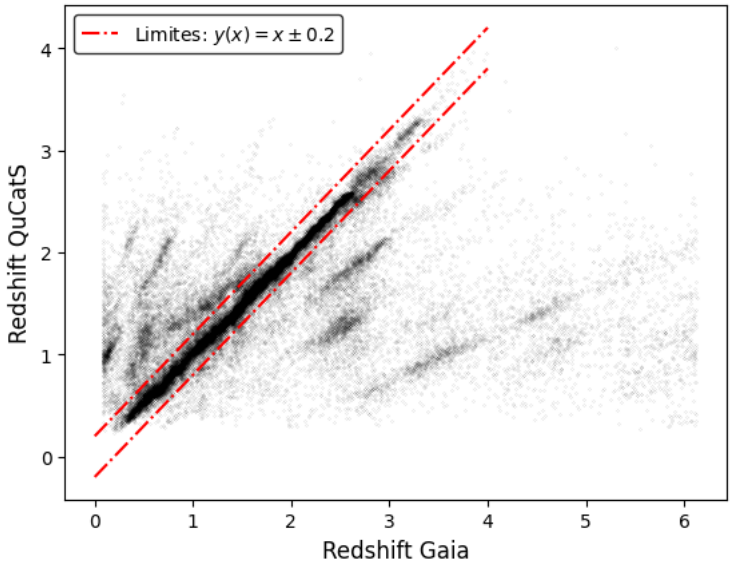


Figura 2.3: Comparação das estimativas de *redshifts* entre GAIA e QuCatS. Em preto temos o espalhamento dos dados e em vermelho temos os limites aplicados para os valores serem considerados concordantes.

Ressaltamos que também é possível observar outras estruturas de alta densidade na Figura 2.3, algumas, inclusive, com características lineares. Ou seja, possivelmente, as bases de dados apresentam diferenças sistemáticas em suas estimativas de *redshift*. Encontrar a origem dessas diferenças exigiria uma análise mais aprofundada dos métodos empregados em ambos os catálogos, não sendo compatível com os objetivos específicos deste projeto.

Por fim, as três amostras resultantes dessa Seção - QuCatS, Prob_QuCatS$\times$ Prob_GAIA e z_QuCatS$\times z$_GAIA constituídas, respectivamente, por 54977, 46065 e 30141 candidatos - formam a base de dados deste projeto e serão submetidas à metodologia descrita no Capítulo 3.

Capítulo 3

Metodologia

Neste capítulo, apresentaremos as metodologias aplicadas neste trabalho. Na Seção 3.1, definiremos a função de correlação e de que modo a calculamos. Essa estimativa nos permite testar a distribuição de quasares da nossa base de dados por meio de suas estatísticas globais, ou seja, sem entrar em detalhes dos agrupamentos posteriores.

Na Seções 3.2 e 3.3, descrevemos, respectivamente, de que forma aplicamos o algoritmo DBSCAN e como redefinimos os agrupamentos obtidos por ele. Na Seção 3.4, discutimos as escolhas de valores do seu parâmetro livre L.

Já nas Seções 3.5 e 3.6, definiremos, respectivamente, alguns conceitos de estatística de grafos e a função de multiplicidade - ambas serão utilizadas para analisar as características gerais das estruturas encontradas. Por fim, na Seção 3.7, discutiremos a aplicação do método *Convex Hull of Member Spheres*, CHMS, na análise da significância das estruturas de interesse.

3.1 Definição da função de correlação

As funções de correlação são ferramentas fundamentais na quantificação de estruturas do Universo. Nesse sentido, elas descrevem o excesso (ou a falta) de probabilidade, em relação a uma distribuição espacialmente uniforme e independente, de se encontrar dois objetos separados por uma distância r num dado volume (Peebles, 1980).

Seu estimador mais comum, desenvolvido por Landy e Szalay (1993), é dado pela Equação 3.1. Nela, as funções $DD(r_p)$, $RR(r_p)$ e $DR(r_p)$ representam, respectivamente, contagens de pares de dados reais, pares de dados randômicos e pares cruzados (reais $\times$ randômicos) dentro de uma distância r_p.

$$\xi(r_p) = \frac{DD(r_p) - 2DR(r_p) + RR(r_p)}{RR(r_p)} \tag{3.1}$$

Ou seja, para sua aplicação, devemos gerar pontos aleatórios, no mesmo espaço em que os dados estão distribuídos. Assim, por conta da inomogeneidade espacial da nossa base de dados, algumas precauções devem ser tomadas para evitarmos vieses no cálculo dessa função. Em especial, destacamos a necessidade da filtragem de regiões angularmente pequenas que, por efeitos de borda, criam erros intrínsecos nas estimativas.

Para realizar essa filtragem, dividimos o espaço angular total da nossa base de dados em malhas de 2 graus quadrados (mesma ordem do campo de visão do S-PLUS) e atribuímos os valores 1 e 0, respectivamente, aos campos significativos (mais do que 5 candidatas) e não significativos. Com isso, reduzimos nosso problema ao tratamento de uma imagem binária e, utilizando um conhecido algoritmo de visão computacional denominado Union-Find, filtramos as regiões que possuíam menos do que 30 componentes (60 graus quadrados). Na Figura 3.1 apresentamos as regiões selecionadas para o cálculo dessa função.

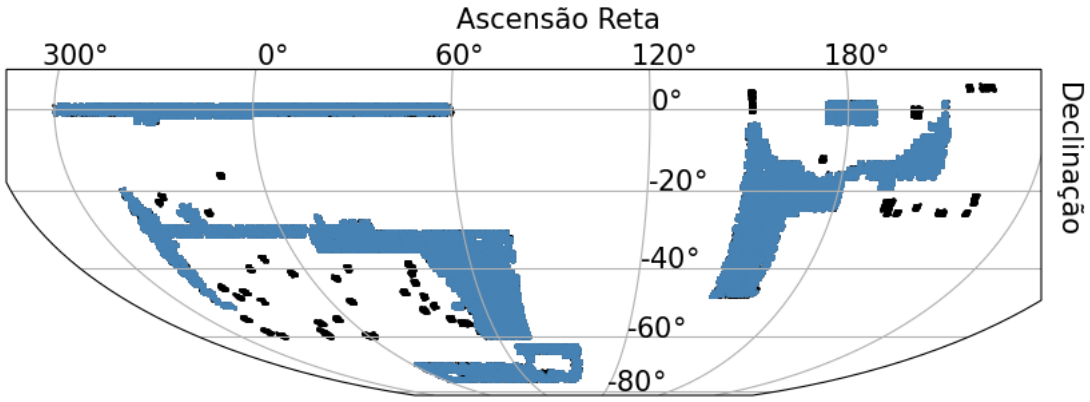


Figura 3.1: Regiões filtradas para o cálculo da função de correlação. Espalhadas, respectivamente, em preto e em azul, temos os dados da amostra QuCatS e as regiões filtradas pelo algoritmo Union-Find.

Assim, destacamos que, com o intuito de reduzir os erros associados a flutuações estatísticas, geramos duzentos dados randômicos por dado real. Esses dados tiveram suas coordenadas angulares amostradas de forma uniforme e seus *redshifts* amostrados de acordo com a distribuição do QuCatS. Além disso, por conta do alto tempo de computação exigido no cálculo das contagens, utilizamos os algoritmos já paralelizados da biblioteca *Corrfunc* (Sinha e Garrison, 2019).

Em conclusão, para uma análise mais rigorosa, outras correções que não foram consideradas neste trabalho podem ser feitas. Dentre elas, enfatizamos a aplicação de máscaras que levam em conta a presença de estrelas brilhantes e eventuais defeitos nas imagens que, em tese, tendem a aproximar nossas estimativas do valor esperado.

3.2 Descrição da aplicação do algoritmo DBSCAN

O processo de agrupamento do DBSCAN é baseado na distância de um ponto até seus pontos vizinhos. Por isso, distintivamente de algoritmos como o *K-means*, ele consegue, desde que as condições sejam satisfeitas, montar conjuntos de qualquer formato, enquanto ignora a presença de possíveis ruídos (Apêndice A). Ou seja, temos um algoritmo aplicável à procura de estruturas em todas as escalas: componentes isoladas, pares, superaglomerados e filamentos (Springel et al., 2005).

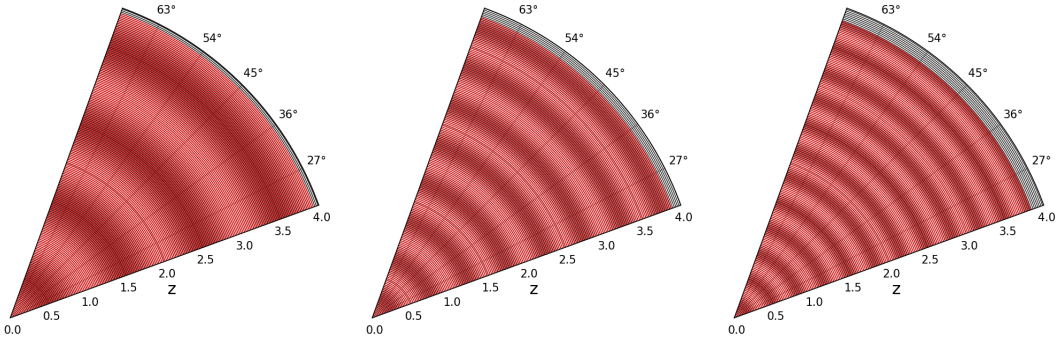
Nesse algoritmo, temos dois parâmetros livres, N_{min} e L, que se referem, respectivamente, ao número de componentes mínimo de um grupo e à distância máxima em que N_{min} pontos podem estar para formá-lo. Desse modo, como a menor estrutura de interesse são pares, aplicamos o algoritmo aos dados com N_{min} igual a dois e, para o parâmetro L, vamos considerar os valores de 1, 5 e 10 *Megaparsecs* ($Mpc\ h_{70}^{-1}$) - discutimos, na Seção 3.4, a qualidade dessas escolhas.

Por outro lado, para evitar a detecção de estruturas que dependem exclusivamente de projeções e a subestimação da quantidade de estruturas (discutida na Seção 4.1), separamos os dados em intervalos de *redshifts* de largura Δz. Esses intervalos são dados pela Equação 3.2. Nela, σ é o erro típico de z_{fot} (normalizado por $1 + z$) e α é um fator variável que controla as intersecções desses intervalos.

Também destacamos a dependência de Δz com $z_{central}$, nesse sentido, quanto maior o *redshift*, mais os intervalos tendem a se intersectar, pois maior é o erro associado à estimativa - esse comportamento pode ser visualizado na Figura 3.2. Neste trabalho, adotamos σ igual a 0.02 e os seguintes valores de α: 0.5, 1 e 1.5. Além disso, $z_{central}$, inicializado como nulo, é incrementado em passos fixos de tamanho σ para cada intervalo.

$$\Delta z(z_{central}) = \pm\alpha\sigma(1 + z_{central}) \tag{3.2}$$

Assim, por termos os prováveis quasares distribuídos em cascas esféricas, utilizaremos



(a) $\alpha = 0.5$ (b) $\alpha = 1$ (c) $\alpha = 1.5$

Figura 3.2: Visualização das intersecções das cascas esféricas adotadas. Nessas figuras, dado um $z_{central}$, temos, em vermelho, o limite mínimo e, em preto, o limite máximo de cada casca. Assim, vemos que o efeito de franjas (tons mais escuros seguidos de tons mais claros de vermelho) indicam a passagem do máximo de uma casca pelo mínimo de outra, ou seja, suas intersecções. Note que, até a primeira intersecção ocorrer, essas cascas não varrem o espaço por completo.

a fórmula de Haversine para calcular as distâncias angulares entre eles. Após, multiplicaremos essa distância angular pela Distância de Diâmetro, D_A, que, para um dado $z_{central}$, fornece a proporção entre ângulo e comprimento. Logo, dessa multiplicação, vamos obter uma distância projetada, ou seja, sem considerarmos as diferenças de z_{fot} dentro da mesma casca. Apresentamos na Equação 3.3 a fórmula final da distância entre dois quasares de mesma casca esférica, em que α_i é a ascensão reta e δ_i é a declinação. Nesse sentido, essa será a distância considerada no algoritmo DBSCAN:

$$d(i,j) = 2D_A(z_{central}) \arcsin\left(\sqrt{\sin^2(\frac{\delta_j - \delta_i}{2}) + \cos\delta_i \cos\delta_j \sin^2(\frac{\alpha_j - \alpha_i}{2})}\right). \tag{3.3}$$

Por fim, destacamos que, por conta do erro típico, σ, ser relativamente alto, para muitos intervalos de Δz, a distância na direção da linha de visada desses objetos pode ultrapassar os valores estipulados de L.

3.3 Redefinição dos grupos

Conforme discutido na Seção 3.2, a separação dos candidatos a quasares é feita em cascas esféricas que se intersectam mais ou menos vezes a depender de α e $z_{central}$. Nesse sentido, ao aplicarmos o DBSCAN nessas diferentes cascas, podemos identificar que um

mesmo objeto faz parte de dois grupos distintos.

Logo, vamos transformar os grupos que possuem um objeto em comum num grupo único (Apêndice B). Salientamos que esse procedimento pode levar à construção de estruturas dependentes de projeções. Porém, acentuamos que essa dependência surge do próprio erro associado às estimativas de z_{fot}. Ou seja, esse procedimento é uma forma de contemplar a incerteza da posição dos objetos na direção da linha de visada.

3.4 Seleção do parâmetro L

Uma análise crucial para a aplicação do DBSCAN consiste numa seleção adequada de valores do seu parâmetro livre L. Para isso, propomos investigar os conceitos de saturação e percolação, que correspondem, respectivamente, à fração de indivíduos agrupados e ao número de estruturas geradas em função desse parâmetro.

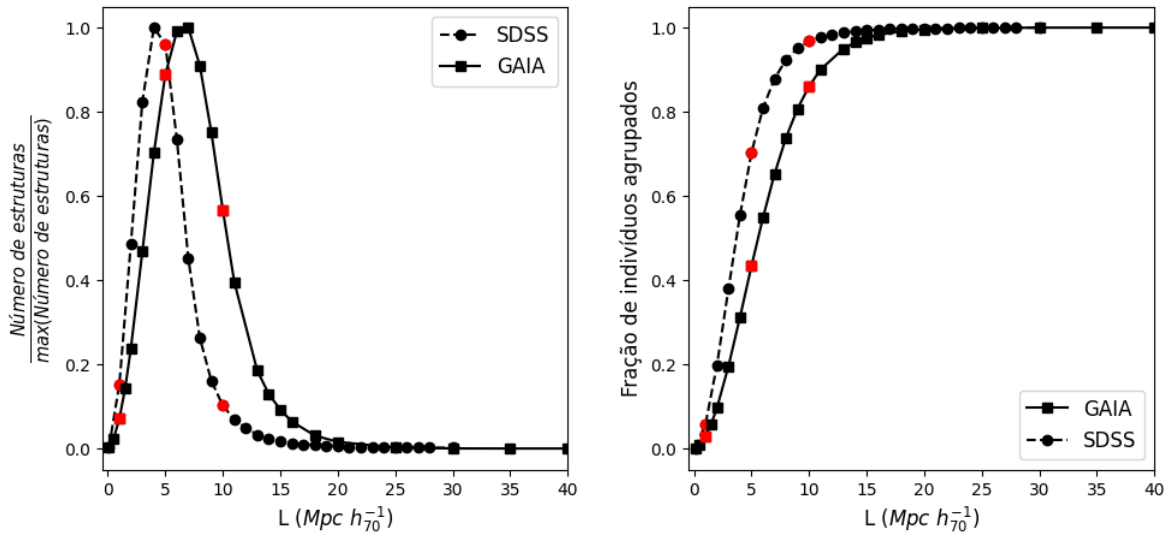
Quando L é superestimado, esperam-se casos de saturação e percolação completas. Ou seja, todos os objetos estão agrupados e formam um único grupo. Nesse caso, ou em casos próximos a esse, perdemos informações relevantes sobre os agrupamentos reais. Por outro lado, a perda de informações também ocorre para valores subestimados, uma vez que poucas estruturas serão formadas. Por isso, valores adequados são definidos, neste estudo, como aqueles distribuídos entre esses dois casos extremos.

Para se estimar esses valores com maior robustez, dois fatores são importantes: a completude da amostra e a homogeneidade espacial da região de estudo - uma vez que descontinuidades nas distribuições, como as presentes no QuCatS (Apêndice C), podem interferir nas estimativas. Por isso, aplicamos os algoritmos, descritos nas Seções 3.2 e 3.3, em regiões homogêneas de dois catálogos, sendo eles, o catálogo de candidatos a quasares do GAIA e o catálogo de quasares do SDSS-DR16 (*Sloan Digital Sky Survey - Data Release 16*; Lyke et al. (2020)) - as regiões selecionadas estão disponíveis no Apêndice D (primeira região para o SDSS).

Nesse sentido, optamos pelo catálogo SDSS por ser, com cerca de 750 mil quasares confirmados por espectro, o maior levantamento de seu tipo, conferindo, assim, uma alta confiança nas medidas de *redshift* e a maior completude possível entre catálogos espectroscópicos. Já o catálogo de candidatos do GAIA, além de possuir uma boa concordância com nossa base de dados (Capítulo 2), também é, com cerca de 6.6 milhões de candi-

datos, um dos maiores catálogos fotométricos disponíveis, possibilitando, portanto, uma comparação mais direta com nossas amostras. Ademais, para garantir a qualidade desses candidatos, estipulamos uma probabilidade mínima de ser quasar de 99%.

Apresentamos, na Figura 3.3, os resultados obtidos para suas regiões homogêneas. Nesses cálculos, utilizamos o parâmetro α igual a 1.5, ou seja, o caso com maior facilidade para formação de agrupamentos maiores. Logo, obtemos estimativas conservadoras da percolação e saturação, pois elas ocorrem para os menores valores de L possíveis. Destacamos que para valores próximos a 20 (15) $Mpc\ h_{70}^{-1}$ para o GAIA (SDSS) todos os candidatos passam a fazer parte de um único grupo; assim, temos uma distribuição saturada e percolada. Isto é, a partir desses valores críticos, mudanças em L não influenciam nas distribuições das estruturas, pois elas já formam a maior estrutura possível.



(a) Percolação (b) Saturação

Figura 3.3: Percolação (a) e Saturação (b) do parâmetro L para as regiões homogêneas do catálogo de candidatas a quasares do GAIA (quadrados) e do catálogo de quasares do SDSS-DR16 (círculos). Em vermelho apresentamos os valores utilizados em nossas amostras.

Por fim, vemos que, para o GAIA, nossas escolhas iniciais (1, 5 e 10 $Mpc\ h_{70}^{-1}$) sãos adequadas, pois estão distantes do valor crítico e bem distribuídas no intervalo disponível. Já para o SDSS, o valor de 10 $Mpc\ h_{70}^{-1}$ está mais próximo do limiar de se tornar percolado. Assim, concluímos que propor um valor maior que 10 $Mpc\ h_{70}^{-1}$ não seria, de fato, vantajoso para a busca de estruturas maiores, pois estaríamos superestimando seu valor.

3.5 Estatísticas básicas dos grafos

Nesta Seção, definimos conceitos básicos de grafos utilizados para analisar as estruturas encontradas. Primeiramente, um grafo, G, é definido como um conjunto de vértices e arestas, $G = \{V, A\}$. Esses conjuntos podem adotar diversas formas, porém, nossas estruturas serão tratadas como grafos simples e não direcionados: estruturas que não permitem que uma aresta tenha apenas um vértice (laço) ou que ela possua uma direção. Além disso, como estamos lidando com estruturas já formadas pelo DBSCAN, os vértices vão possuir uma distância projetada máxima, L, e assim, cada grafo pode ser tratado como totalmente conectado, ou seja, não há nenhum vértice sem conexão dentro dos limites de L.

Adotaremos algumas estatísticas básicas dos grafos para analisarmos nossas estruturas. A primeira delas, denominada de diâmetro, D(G), é a maior distância projetada entre todos os vértices (Equação 3.4, em que $d(i, j)$ vem da Equação 3.3).

$$D(G) \doteq \max_{i,j} d(i, j) \tag{3.4}$$

Ademais, temos que a densidade, também chamada de densidade de arestas, de um grafo simples e não direcionado, $\Gamma(G)$, é definida pela Equação 3.5. Nela, $|A|$ é o número de arestas e $|V|$ é o número de vértices - a notação $|A|$ representa a norma do conjunto A. Nesse sentido, essa medida verifica a proporção da quantidade de arestas em relação a quantidade máxima possível. Ou seja, em grafos em que existe a possibilidade de termos vértices sem arestas, esse valor varia de 0 a 1. No nosso caso, em que todos os vértices possuem ao menos uma aresta, a estrutura menos densa possível possui uma densidade igual a $2/|V|$. Logo, também podemos usar essa medida em sua forma renormalizada, dada pela Equação 3.6.

$$\Gamma(G) \doteq \frac{2|A|}{|V|(|V| - 1)} \tag{3.5}$$

$$\Gamma_{norm}(G) \doteq \frac{\left(\frac{2|A|}{|V|(|V|-1)} - \frac{2}{|V|}\right)}{1 - \frac{2}{|V|}} \tag{3.6}$$

Por fim, temos que o grau de um vértice, $\eta(i)$, apresentado na Equação 3.7, é definido como a soma do número de arestas que se conectam a ele. Assim, vértices centrais tendem a possuir graus de altos valores. As distribuições desses graus geralmente seguem formas

que, comumente, apresentam características de caudas longas, ou seja, poucos, mas não nulos, pontos de graus elevados. Ademais, é válido ressaltar que algoritmos geradores de grafos, normalmente, se baseiam nesse tipo de distribuição.

$$\eta(i) \doteq |A_i| \tag{3.7}$$

Portanto, a partir dessas quatro métricas, podemos caracterizar e estudar, de forma geral, a topologia das estruturas encontradas pelo DBSCAN para as nossas amostras. Ressaltamos que todas as definições apresentadas nesta Seção são baseadas nas formulações descritas em Diestel (2005).

3.6 Função de multiplicidade

Outra forma de estudar as estatísticas dos grupos obtidos pelos nossos algoritmos se dá pelas funções de multiplicidade $n_{grp}(R)$. Em linhas gerais, na sua forma diferencial, elas descrevem a quantidade de estruturas esperadas por volume efetivo do catálogo por intervalo (*bin*) de quantidade de componentes - também chamada de riqueza (R) nesse tipo de análise.

Logo, por se tratar de uma função que descreve a abundância absoluta de objetos (Puddu, E. et al., 2003), a completude do catálogo selecionado é um fator crucial que deve ser considerado no cálculo dessa função. Desse modo, supondo que os cortes realizados no Capítulo 2 - das magnitudes, $r < 20$, das probabilidades médias e dos *redshifts* discordantes - são os principais fatores de incompletude em nossas amostras, esperamos que as maiores diferenças entre elas e um catálogo completo, ou com alto grau de completude, possam ser corrigidas realizando os mesmos cortes, ou cortes de características semelhantes - mais detalhes dessa discussão são apresentados na Seção 4.3.

Por fim, por conta da baixa quantidade de grandes grupos, a função tende a ser ruidosa para altos valores de R. Para evitá-los, conforme descrito em Berlind et al. (2006), adotamos *bins* com tamanhos maiores do que o unitário, diminuindo, por consequência, o ruído dessas estimativas.

3.7 Descrição da aplicação do método CHMS

Proposto em Clowes et al. (2011), o método CHMS, *Convex Hull of Member Spheres*, possibilita a análise da significância de um grupo dentro de seu catálogo. Nele, expande-se uma esfera de raio $\frac{L}{2}$ ao redor dos N candidatos de um grupo e calcula-se seu volume convexo, ou seja, o volume do menor poliedro convexo que contém essas esferas. Em seguida, através da densidade média do catálogo naquele intervalo de *redshift*, $< \delta >$, geram-se N pontos randômicos em um cubo de volume $N < \delta >^{-1}$. Assim, calculando o volume convexo de K estruturas randômicas e comparando-os com o volume convexo da estrutura analisada, obtemos um modo de estimar a significância dessa estrutura.

Em suma, esse método estima a probabilidade de um grupo de candidatos ser formado aleatoriamente pela densidade média da região em que está distribuído, logo, grupos de alta significância (maior que 0.999) são improváveis de serem formados aleatoriamente.

Neste trabalho, também aplicamos uma versão modificada desse método em grupos que, inicialmente, não possuem alta significância. Nela, retiramos, ponto a ponto, os objetos mais distantes em relação ao centro e reaplicamos o método CHMS, encontrando, por consequência, a maior estrutura significativa possível para aquele grupo. Note que, conforme dito na Seção 3.3, a retirada dos pontos mais distantes, na maioria dos casos, ocorrerá na direção da linha de visada. Logo, essa versão atua, principalmente, na redução do viés induzido pelo código de redefinição.

Salientamos que, apesar de ser aplicável para todos os agrupamentos, neste trabalho, esses métodos serão aplicados apenas nas maiores estruturas encontradas ($R > 10$). Ou seja, assim como no trabalho que o origina (Clowes et al., 2011), apenas no estudo dos LQGs.

Em conclusão, apresentamos, neste capítulo, a função de correlação - utilizada para analisar as estatísticas globais de distribuição espacial das nossas amostras - o algoritmo utilizado para identificar os agrupamentos físicos e alguns métodos utilizados para estudá-los. No Capítulo 4, aplicaremos essa metodologia na base de dados definida no Capítulo 2 e, em alguns casos, para os catálogos GAIA e SDSS.

Capítulo 4

Resultados

Neste capítulo, apresentamos, na Seção 4.1, por meio das funções de correlação calculadas, uma análise global da distribuição dos quasares.

Já na Seção 4.2, delineamos os agrupamentos, obtidos da aplicação do DBSCAN e da redefinição dos grupos, das três amostras consideradas. Nas Seções 4.3 e 4.4, os descrevemos qualitativa e estatisticamente, por meio da função de multiplicidade e da teoria dos grafos.

Na Seção 4.5, apresentamos os pares de quasares próximos confirmados, descartados e promissores e suas características. Finalmente, na Seção 4.6, utilizamos o CHMS, em conjunto com outras estatísticas, para encontrarmos e descrevermos candidatos promissores a LQGs.

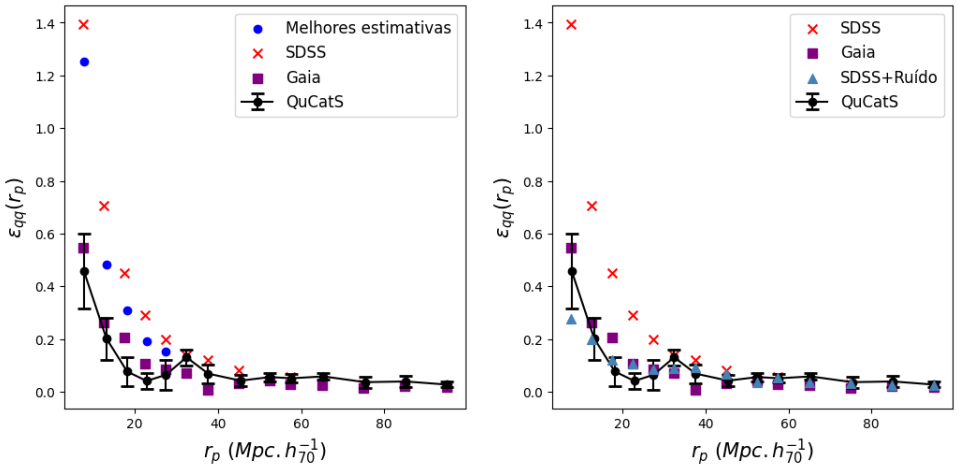
4.1 Função de correlação

Exibimos, na Figura 4.1, as funções de correlação das amostras QuCatS, GAIA e SDSS - utilizando as mesmas regiões da Seção 3.4 para os últimos dois casos. Nela, conforme mostrado em Fujii (2023), podemos tratar os dados do SDSS como excelentes aproximações da curva prevista pelo Λ-CDM.

Desse modo, é possível observar que, em todos os casos, temos um excesso de probabilidade para distâncias, r_p, suficientemente baixas (menores que 35 $Mpc\ h_{70}^{-1}$). Ou seja, para essas distâncias, os quasares, em comparação com uma distribuição aleatória, tendem a se agrupar mais do que o esperado por conta da interação gravitacional. No entanto, também é possível observar que os catálogos fotométricos - QuCatS e GAIA - subestimam esse excesso.

Uma possível explicação para esse comportamento está relacionada às incertezas das

medidas de *redshifts* - que são maiores nos catálogos fotométricos - e que não estão sendo consideradas nesse cálculo. Nesse sentido, os erros associados a essa grandeza tendem a aleatorizar a distribuição real dos quasares, gerando, por consequência, uma supressão na estimativa da função real (Chaves-Montero et al., 2018), ou seja, estamos subestimando, de modo global, a distância real dos quasares. Para verificar esse fenômeno, introduzimos erros gaussianos, de desvio padrão $0.01(1+z)$ (metade do estimado para o QuCatS), nos *redshifts* do SDSS e recalculamos sua função de correlação. Assim, apresentamos em (b), como SDSS+Ruído, o resultado dessa verificação e, vemos que, de fato, a introdução de erros sistemáticos nessa grandeza tende a suprimir as estimativas corretas.



(a) (b)

Figura 4.1: Funções de correlação. Em ambas, apresentamos como cruzes vermelhas, quadrados roxos e pontos pretos, respectivamente, os dados obtidos pelo SDSS, GAIA e QuCatS. Em (a), temos, como bolas azuis, as melhores estimativas atingidas pelas diferentes regiões do QuCatS. Em (b), apresentamos, como triângulos, os dados obtidos da amostra ruidosa do SDSS.

Também destacamos o uso do método *Jackknife* para se estimar os erros na função de correlação da amostra QuCatS. Nesse método, removemos, da nossa amostra total, uma subamostra por vez e recalculamos a função de correlação. Após isso, tomamos a variância dessas novas estimativas como incertezas. Em nosso caso, utilizamos as sete regiões conexas, apresentadas na Figura 3.1, como subamostras e ponderamos os resultados pelo número de quasares em cada região.

Além disso, em (a), apresentamos as melhores estimativas dessas diferentes regiões para a parte suprimida da função. Logo, nota-se que, ocasionalmente, dentro das variações de

cada região, os dados ainda podem se aproximar da estimativa mais precisa (SDSS).

Por último, disponibilizamos no Apêndice E as funções de correlação das amostras Prob_QuCatS× Prob_GAIA e z_QuCatS×z_GAIA que possuem comportamentos semelhantes ao da amostra original. Portanto, concluímos que nossas amostras se comportam como o esperado de dados fotométricos e que, por conta da supressão de pares mais próximos que 35 $Mpc\ h_{70}^{-1}$, as contagens de grupos, caso não levem em conta os erros nos *redshifts*, podem ser subestimadas. Assim, destacamos a importância dos intervalos de z_{fot}, utilizados na aplicação do algoritmo de agrupamento, serem proporcionais ao seu erro típico (Seção 3.2).

4.2 Distribuição dos agrupamentos

Passando, nesta seção, para as análises dos agrupamentos, relembramos que, conforme discutido nas Seções 3.2 e 3.3, aplicamos os algoritmos, em nossa base de dados, com nove diferentes conjuntos de parâmetros. Esses conjuntos são constituídos de variações de α - parâmetro que controla as quantidades de intersecções dos intervalos de *redshift* - e L - distância máxima para pares serem considerados parte de um mesmo grupo.

Tabela 4.1 - Distribuição dos agrupamentos resultantes da amostra QuCatS.

α	L(Mpc)	Duplas	Trios	Quartetos	Quintetos	Estruturas	Quantidade de Grupos	Objetos Agrupados	Fração do Agrupamento
0.5	1	106	0	0	0	0	106	212	0.004
0.5	5	1916	128	8	2	2	2056	4272	0.078
0.5	10	4852	885	213	68	50	6068	13934	0.253
1.0	1	230	0	0	0	0	230	460	0.008
1.0	5	3564	423	63	17	10	4077	8807	0.160
1.0	10	6089	1812	756	324	390	9371	25450	0.463
1.5	1	322	3	0	0	0	325	653	0.012
1.5	5	4644	785	150	49	27	5655	12703	0.231
1.5	10	5498	2199	994	535	1003	10229	33376	0.607

As distribuições desses agrupamentos para os diferentes conjuntos podem ser encontradas na Tabela 4.1 e nas Tabelas F.1 e F.2 do Apêndice F. Nesse caso, grupos maiores que quintetos são, neste trabalho, denominados estruturas. Além disso, a quantidade de grupos

corresponde ao número de agrupamentos obtidos, e a fração de agrupamento é obtida da razão do número de objetos agrupados pelo total.

Nessas tabelas, de acordo com o esperado, é possível observar um aumento do número de objetos agrupados conforme α e L crescem. Nota-se, também, que as maiores frações de agrupamento diminuem em proporção ao número de objetos disponíveis por amostra. Isso ocorre porque os cortes de candidatos não são baseados nas distribuições espaciais. Ou seja, estamos diminuindo a densidade numa mesma região fixa do espaço; por isso, atingimos menores frações de objetos agrupados.

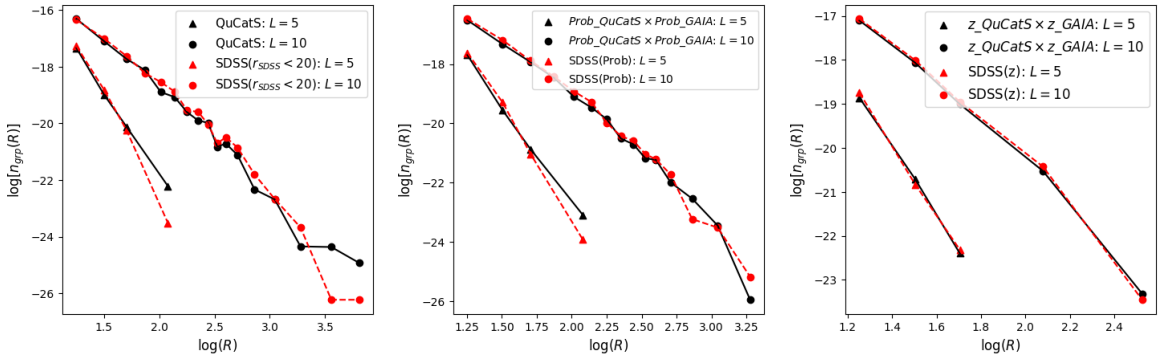
Além disso, conforme discutido na Seção 3.4, pela falta de completude e homogeneidade espacial das amostras, os valores dessas frações, em comparação com os apresentados na Figura 3.3, são, de fato, subestimados. Por exemplo, com $L = 10\ Mpc\ h_{70}^{-1}$, temos, respectivamente, para o SDSS, GAIA e QuCatS, os valores: $0.97 > 0.86 > 0.61$.

Por último, verifica-se que, para $L = 1\ Mpc\ h_{70}^{-1}$, com poucas exceções, grupos maiores que pares não são formados. Portanto, esse caso não é considerado na análise da função de multiplicidade, pois não há dados disponíveis para tal.

4.3 Função de multiplicidade

Exibimos, na Figura 4.2, a comparação entre as funções de multiplicidade calculadas para nossa base de dados e para o SDSS. Desse modo, conforme discutido na Seção 3.6, com o objetivo de compararmos os dados de modo direto, nós selecionamos, no SDSS, a segunda região do Apêndice D. Ou seja, para estimarmos abundâncias de mesma ordem, selecionamos, com cerca de 3000 graus quadrados, uma região de mesma dimensão que o S-PLUS. Além disso, para as comparações específicas de cada amostra, aplicamos uma filtragem correspondente.

Nesse sentido, para a comparação com a amostra QuCatS (a) realizamos o mesmo corte de magnitudes: $r_{SDSS} < 20$. Já para as amostras Prob_QuCatS× Prob_GAIA (b) e z_QuCatS×z_GAIA (c) consideramos que seus cortes ocorreram de modo homogêneo. Assim, amostramos os dados do SDSS de modo aleatório, mantendo a condição de que as densidades, quantidade de quasares por grau quadrado, das duas amostras em comparação sejam da mesma ordem. Por sua vez, essas amostras, denominadas SDSS($r_{SDSS} < 20$), SDSS(Prob) e SDSS(z), possuem, respectivamente, 48698, 42226 e 27629 quasares.



(a) (b) (c)

Figura 4.2: Funções de multiplicidade calculadas para $\alpha = 1.5$. Apresentamos em linhas pretas contínuas e em linhas vermelhas tracejadas, respectivamente, as funções de multiplicidade calculadas para as amostras resultantes do QuCatS e do SDSS. Em círculos e triângulos, vemos, respectivamente, os valores de L iguais a 5 e 10 $Mpc\ h_{70}^{-1}$.

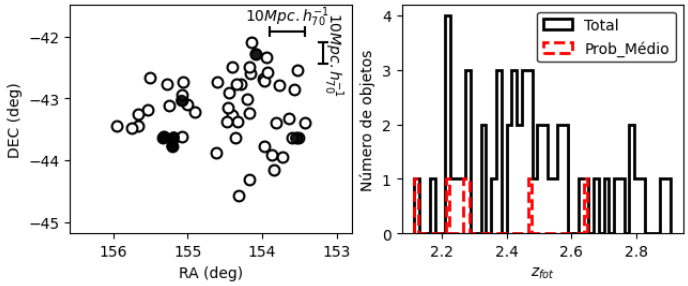
Finalmente, destacamos a boa concordância entre as amostras, principalmente fora da ruidosa região das altas riquezas, o que indica, por consequência, que os intervalos de *redshifts* serem proporcionais ao erro típico, de fato, impedem a atenuação do número de estruturas encontradas (Seção 4.1).

4.4 Estudo das estruturas

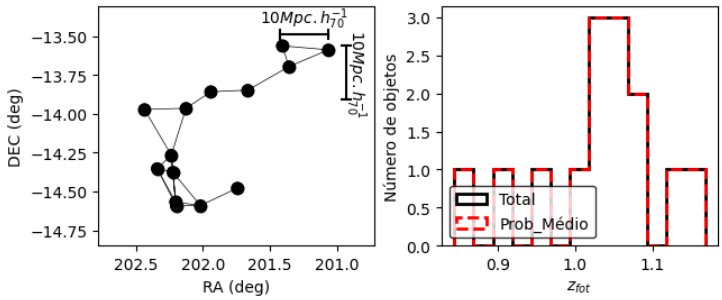
Passando, nesta seção, para a análise das características das estruturas, apresentamos, na Figura 4.3, exemplos de grafos típicos encontrados na amostra QuCatS, bem como suas distribuições de *redshift* - mais especificamente, resultantes dos parâmetros α e L iguais a 1.5 e 10 $Mpc\ h_{70}^{-1}$. No canto superior direito, indicamos escalas de distância de dimensão $Mpc\ h_{70}^{-1}$ em termos da fórmula de Haversine.

Ainda nessa Figura, as probabilidades controlam as cores das arestas e do interior dos pontos. Ou seja, pontos e arestas brancas indicam que o objeto foi descartado através das comparações com as probabilidades do GAIA (indicadas de forma pontilhada e em vermelho nas distribuições de z_{fot}). Nesse caso, estruturas formadas quase completamente por objetos descartados, como em (a), são encontradas com frequência. Logo, o cruzamento de dados é, de fato, essencial para esse tipo de análise. Ademais, estruturas de baixas densidades e projeções (b), assim como seu caso oposto, altas densidades e projeções (c), também são comuns. As interpretações para esses casos são, respectivamente, pequenos

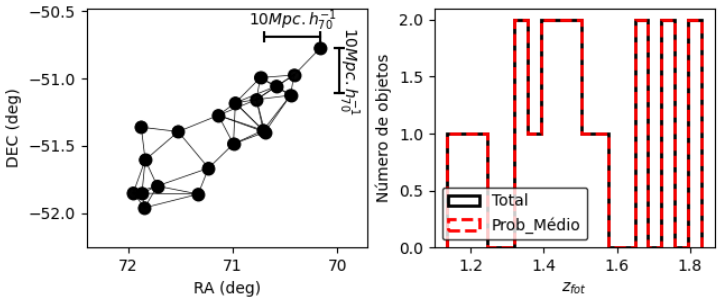
grupos ou filamentos próximos da perspectiva de seu comprimento e grandes filamentos vistos mais próximos da perspectiva de sua seção reta.



(a) Muitos candidatos descartados



(b) Baixas densidades e projeções



(c) Altas densidades e projeções

Figura 4.3: Estruturas típicas encontradas na amostra QuCatS.

Uma boa forma de se parametrizar a discussão acima se dá na descrição estatística desses grupos. Dessa forma, na Seção 4.4.1, utilizaremos as estatísticas básicas de grafos com esse intuito.

4.4.1 Aplicação das estatísticas de grafos

Utilizando os conceitos da Seção 3.5, chegamos às distribuições apresentadas na Figura 4.4. Nota-se, através delas, que as estruturas encontradas por essas diferentes amostras

possuem descrições estatísticas semelhantes para as estruturas mais prováveis. Nesse sentido, para os diâmetros em (a), temos picos centrados em 20 $Mpc\ h_{70}^{-1}$ e distribuições com caudas longas. Destacamos que o diâmetro mais provável ser o dobro do valor de L também ocorre em outros casos. Assim, esse comportamento parece ser intrínseco ao DBSCAN.

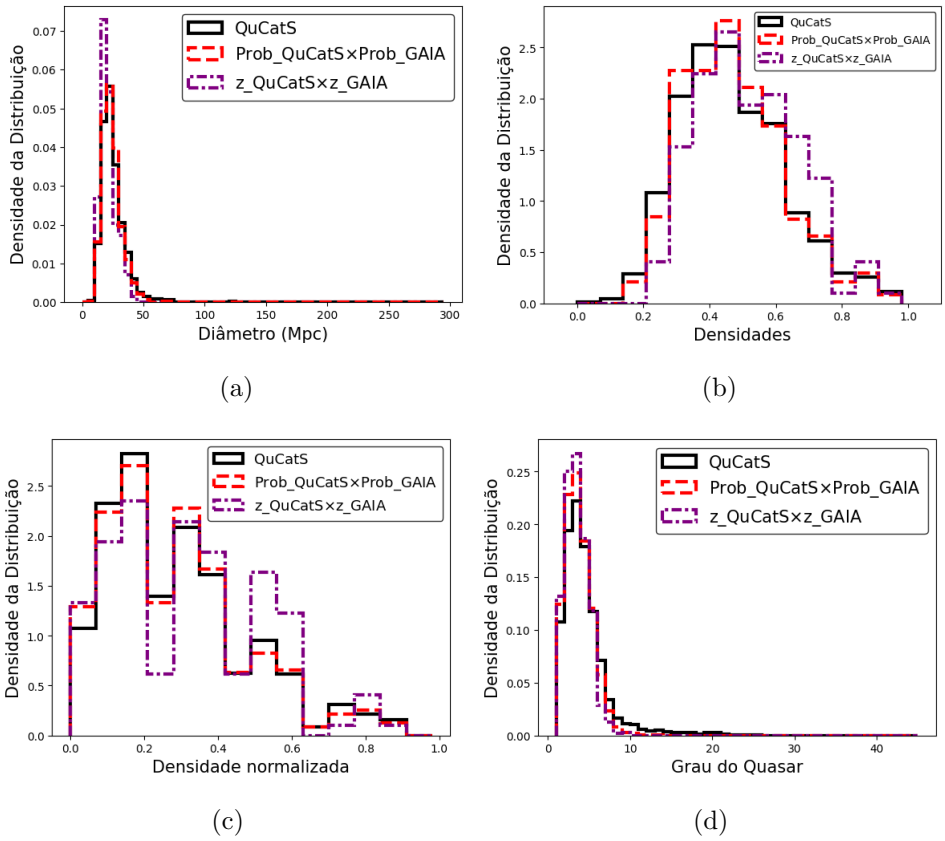


Figura 4.4: Distribuições estatísticas utilizadas para descrever as estruturas. Em (a) temos a distribuição de diâmetro para L igual a 10 $Mpc\ h_{70}^{-1}$. Em (b) e (c) temos as distribuições da densidades não normalizadas e normalizadas. Em (d) temos as distribuições de grau dos quasares.

Para as distribuições de graus, atingimos, conforme o esperado, algo próximo de uma lei de potência. Note que esse tipo de distribuição é esperado, pois, caso não tivéssemos caudas longas, estruturas densas não seriam possíveis e isso indicaria que estaríamos subestimando a quantidade de ligações plausíveis.

Por fim, obtemos para as densidades não normalizadas (b) um pico centrado em 0.4 e, para o caso normalizado (c), temos o maior pico próximo a 0.1, ou seja, em geral, as estruturas não são densas. Assim, conforme discutido na Seção 4.4, é válido argumentar que o DBSCAN consegue identificar filamentos, principalmente na perspectiva de seu com-

primento. Pois, junto ao fato da baixa densidade média, também temos a possibilidade de gerar grandes diâmetros.

4.5 Pares de quasares

Passamos a discutir, então, a procura por estruturas de interesse. Nesta seção, descreveremos a procura por pares próximos na amostra brilhante do QuCatS por meio da aplicação direta do DBSCAN (Seção 4.5.1) e pelo cruzamento de dados com o catálogo GAIA (Seção 4.5.2).

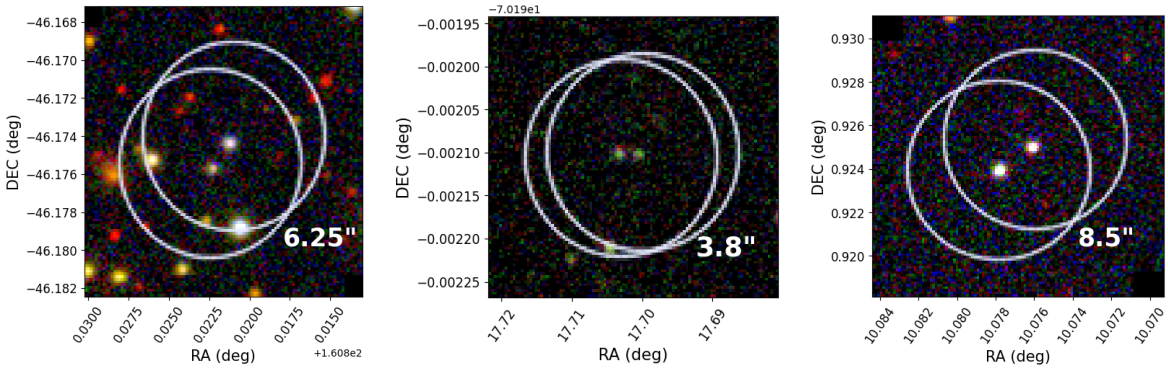
4.5.1 Pares detectados pelo DBSCAN

Inicialmente, destacamos que os sete pares encontrados pelo DBSCAN são menos dependentes de projeções do que estruturas maiores. Isso ocorre, pois o algoritmo de redefinição de grupos não é capaz de formá-los; ou seja, a projeção máxima está limitada a uma casca esférica.

Tabela 4.2 - Características dos candidatos a pares de quasares inicialmente encontrados.

Candidatos	RA	DEC	*Prob_GAIA*	*Prob_QuCatS*	*z_GAIA*	*z_QuCatS*	Tipo
(a)	160.820897	-46.174092	0.979187	0.990	1.678151	1.711574	Objetos
	160.822346	-46.175508	0.998813	0.990	1.670684	1.623595	resolvidos
(b)	17.699768	-70.192102	0.999442	0.965	1.776860	1.719651	Objetos
	17.702883	-70.192108	0.996044	0.855	2.442272	1.738883*	resolvidos
(c)	10.075894	0.925285	1.000000	1.000	2.030010	2.008920	Objetos
	10.077789	0.923876	1.000000	1.000	2.098253	2.089496	resolvidos
(d)	17.117453	0.003904	1.000000	1.000	1.190434	1.166972	Não resolvidos
	17.117541	0.003629		0.880		1.144902	ou únicos
(e)	36.034455	0.050089	1.000000	0.970	0.839071	1.584508	Não resolvidos
	36.034507	0.050454		1.000		1.571547	ou únicos
(f)	356.397217	0.028210	1.000000	1.000	2.775743	1.001162	Não resolvidos
	356.397249	0.027932		1.000		1.051522	ou únicos
(g)	67.607806	-44.658091	1.000000	1.000	1.075777	1.141577	Não resolvidos
	67.607335	-44.658167	-	0.875	-	1.231317	ou únicos

Informações desses pares são exibidas na Tabela 4.2. Nela, um dos objetos da dupla (g) não possui correspondência no GAIA e, por isso, não possui valores atribuídos. Além disso, informações dispostas em linhas únicas indicam que apenas um valor foi encontrado para ambas as coordenadas, ou seja, possivelmente, temos um objeto duplicado ou não resolvido. Esses objetos, através da ferramenta Aladin (Bonnarel et al., 2000), também foram visualmente verificados em outros telescópios e, em nenhum caso, a visualização de um companheiro foi possível. Também indicamos, através de um asterisco, uma estimativa de *redshift* fotométrico com uma grande incerteza associada. Logo, a grande distância desse valor para o do GAIA pode estar ligada ao seu desvio padrão de 1.07. Note que esse fato não se repete para os casos (e) e (f).



(a) (b) (c)

Figura 4.5: Candidatos a pares de quasares detectados pelo DBSCAN na amostra QuCatS. As imagens foram obtidas através do pacote do S-PLUS, em *Python*, mais especificamente de ferramentas de imageamento descritas em Coe et al. (2012) e em Lupton et al. (2004). No canto inferior direito das imagens são exibidas as distâncias angulares das fontes em segundos de arco. Os círculos brancos ao redor dos candidatos estão centralizados nas suas coordenadas.

Assim, por conta dos objetos possivelmente duplicados, reduzimos as duplas iniciais aos casos (a), (b) e (c) apresentados na Figura 4.5 - os casos restantes estão disponíveis no Apêndice G. Dessas, a dupla (c) é a única dupla que possui confirmações através de espectros (Meusinger, H. et al., 2011). Por meio deles, seus *redshifts* também foram medidos, nesse caso, os valores 2.83857 e 2.08719 foram, respectivamente, encontrados. Assim, esse é um caso de par projetado. Por outro lado, a dupla (b), por conta dos diferentes valores de *redshifts* encontrados no GAIA e pela alta incerteza em um dos valores do QuCatS, também aparenta ser uma projeção. Porém, a confirmação por espectro ainda se faz

necessária. Por conta disso, o caso (a) aparenta ser o mais promissor de ser comprovado numa eventual medição. Destacamos que esse, provavelmente, é um caso binário, pois, conforme descrito em Kochanek et al. (1997), pares de quasares com distâncias angulares maiores que três segundos de arco são, com grande probabilidade, pares físicos.

4.5.2 Pares detectados pelo cruzamento com o GAIA

Por outro lado, também verificamos, na literatura, os 33 pares de objetos encontrados durante o cruzamento de dados entre os catálogos QuCatS e GAIA (Capítulo 2). Na Tabela 4.3, apresentamos os pares de objetos que possuem sua natureza estudada por espectros. Nesse sentido, o par de coordenadas (a) é tido como uma fonte única com $z = 1.4349$ em Croom et al. (2004). O par (b) é tido como um par binário de quasares com *redshifts* iguais a 1.455 e 1.445 por Scialpi, M. et al. (2024) e o par (c) é formado por uma estrela do tipo M1 e um quasar ($z = 0.665$), respectivamente, em West et al. (2008) e Paris, Isabelle et al. (2018).

Tabela 4.3 - Pares de objetos próximos, encontrados no cruzamento entre os catálogos GAIA e QuCatS, que possuem medidas de espectro.

Dupla	RA (deg)	DEC (deg)	$\delta\theta$ $(arcseg)$	$Prob_G$	$Prob_Q$	z_G	z_Q	G (mag)	r (mag)	Natureza dos objetos
(a)	355.964579	-29.177059	0.96	1	0.84	3.44	1.33	20.19	19.74	Fonte única
	355.964759	-29.177276		0.69		2.33		-		
(b)	13.158248	-30.764511	0.39	0.01	0.98	-	1.39	19.78	19.09	Par de quasar em subarcosegundo
	13.158340	-30.764437		0.01		-		19.76		
(c)	330.619685	0.817290	1.22	0	1	-	1.21	20.24	19.92	Estrela M1
	330.619367	0.817172		0		-		20.56		Quasar

Logo, com desvios menores ou iguais a 0.01, é possível notar uma boa concordância entre as estimativas de *redshifts* do QuCatS e das fontes nos casos (a) e (b). Por outro lado, para o caso (a) no GAIA e para o caso (c) no QuCatS, obtemos desvios maiores ou iguais a 0.6, ou seja, eles apresentam uma péssima concordância.

Por outro lado, na Tabela 4.3, também apresentamos: a distância angular, $\delta\theta$, entre os objetos; as probabilidades de serem quasares para os catálogos GAIA ($Prob_G$) e QuCatS ($Prob_Q$); as estimativas de *redshifts* (z_G e z_Q) e suas magnitudes (G e r). Assim, destacamos

que o único par confirmado (b) possui, para seus dois quasares, baixas probabilidades associadas no GAIA. Logo, é possível que fontes próximas impeçam que essa estimativa seja feita de forma correta. Por isso, também disponibilizamos, na Tabela H.1 do Apêndice H, informações obtidas de todos os 33 pares encontrados, pois o descarte através das probabilidades pode incorrer em erros.

Por fim, exibimos, na Figura H.1 do Apêndice H, o imageamento dos casos (a), (b) e (c). Nota-se que os pares considerados não únicos, (b) e (c), são quase pontuais e, portanto, concluímos que o QuCatS não os identificou, principalmente, pela falta de resolução angular.

4.6 Candidatos a LQGs

Nesta seção, discutiremos os resultados da busca por outro tipo de estrutura de interesse, nesse caso, pelos grupos candidatos a LQGs. Desse modo, aplicamos a metodologia descrita na Seção 3.7 nas estruturas com mais de dez componentes da amostra Prob_QuCatS×Prob_GAIA ($\alpha = 1.5$ e $L = 10\ Mpc\ h_{70}^{-1}$). Escolhemos essa amostra com o intuito de evitarmos as grandes estruturas de estrelas presentes no QuCatS (Seção 4.4), enquanto ainda preservamos uma boa quantidade de grupos em relação ao caso mais restrito (z_QuCatS×z_GAIA).

Exibimos na Tabela 4.4 nossos candidatos a LQG, ou seja, grupos de alta significância (maior que 0.999). Nela, apresentamos, das estruturas iniciais e finais, as grandezas: riqueza (R_0 e R_f), significância (sig_0 e sig_f) e o intervalo de *redshift* (Δz_0 e Δz_f) em que os dados estão distribuídos. Ademais, considerando apenas a estrutura final, apresentamos o *redshift* médio ($< z >$), a maior distância entre seus pontos ($\max dist$) e sua razão pela riqueza ($\frac{\max dist}{R_f}$).

Note que os termos inicial e final se referem, respectivamente, a antes e após a retirada dos pontos distantes, isto é, aqueles que não fazem parte da região significativa. Desse modo, para nos mantermos no regime dos LQGs, o algoritmo que os retira foi limitado a parar quando a riqueza mínima, $R = 10$, é atingida. Ou seja, caso a maior parte significativa da estrutura tenha menos do que 10 componentes, o candidato a LQG é descartado.

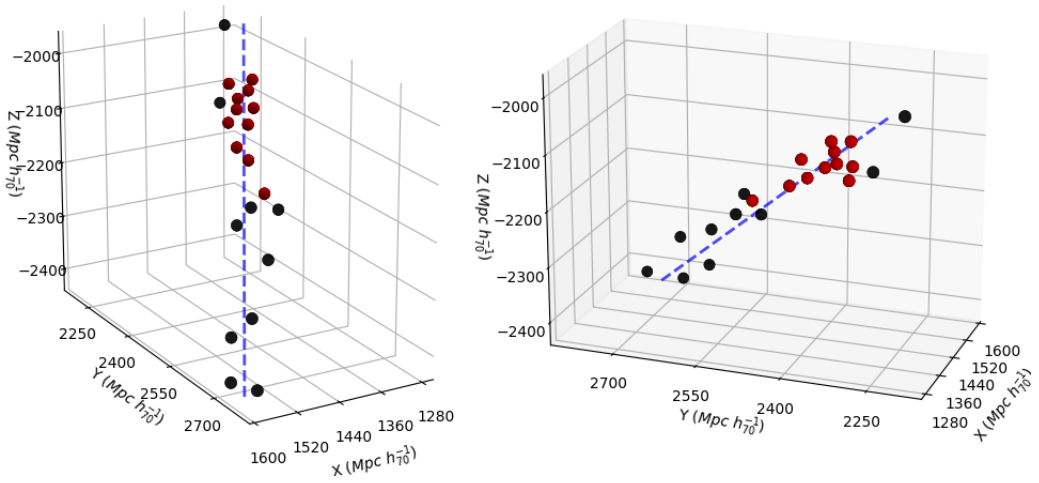
Tabela 4.4 - Candidatos a LQGs.

Candidato	R_0	sig_0	Δz_0	R_f	sig_f	Δz_f	$< z >$	$\max dist$ ($Mpc\ h_{70}^{-1}$)	$\frac{\max dist}{R_f}$ ($Mpc\ h_{70}^{-1}$/quasar)
(a)	22	0.00	0.47	10	1.00	0.14	1.31	305.73	30.57
(b)	20	0.00	0.70	11	1.00	0.25	1.44	479.65	43.60
(c)	16	0.64	0.27	11	1.00	0.15	1.64	264.81	24.07
(d)	18	0.15	0.40	12	1.00	0.22	1.55	405.93	33.83
(e)	19	0.00	0.34	10	1.00	0.12	1.59	226.16	22.62
(f)	23	0.00	0.55	13	1.00	0.20	1.65	335.49	25.81
(g)	22	0.00	0.55	11	1.00	0.27	1.38	534.51	48.59
(h)	18	0.04	0.35	13	1.00	0.22	1.09	508.08	39.08
(i)	21	0.02	0.33	11	1.00	0.10	1.06	233.58	21.23
(j)	18	0.02	0.38	10	1.00	0.15	1.29	316.10	31.61
(k)	18	0.35	0.35	11	1.00	0.12	1.77	204.96	18.63

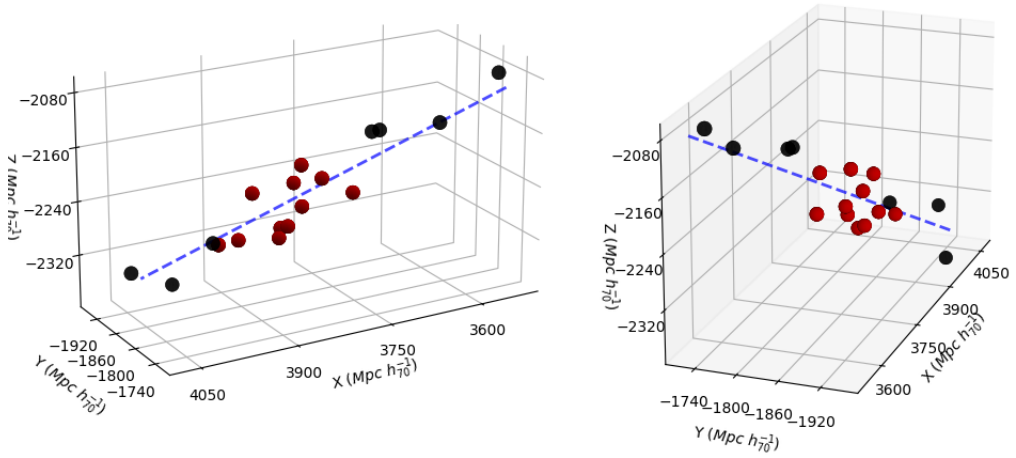
Dos resultados, destacamos que, inicialmente, nenhuma estrutura possui alta significância e, por isso, todas tiveram quasares distantes retirados. Além disso, conforme discutido na Seção 3.7, do caso inicial para o final, observamos uma grande redução do intervalo de *redshift* que atenua, por consequência, o viés introduzido pelo algoritmo de redefinição de grupos. Ademais, relembramos que o CHMS é um método que depende de variáveis aleatórias e que, por isso, as estruturas finais podem apresentar leves desvios entre diferentes aplicações.

Por outro lado, com o objetivo de compararmos nossos candidatos a outros da literatura, introduzimos a métrica $\frac{\max dist}{R_f}$ que nos dá, de acordo com o Λ-CDM, a dispersão do grupo ao longo do seu maior eixo. Nesse sentido, obtemos os valores 16.99, 20.53 e 18.53 $Mpc/quasar$, respectivamente, para os grupos *Huge-LQG*, U1.11 e U1.28 - maiores grupos com essa medida disponível na literatura (Lopez et al., 2022). Assim, para nossos dados, obtemos que (i) e (k) estão muito próximos dos valores da literatura e, por isso, eles são tidos como os candidatos mais promissores. Além deles, também destacamos os casos (c) e (e) que estão um pouco acima do esperado.

Por fim, apresentamos, na Figura 4.6, mantendo a proporção real das coordenadas cartesianas, a distribuição espacial dos casos (i) e (k) - mais especificamente, de suas esferas de raio 5 $Mpc\ h_{70}^{-1}$.



Candidato (i)



Candidato (k)

Figura 4.6: Candidatos promissores a LQGs. Em preto e em vermelho, estão, respectivamente, os quasares retirados e os que fazem parte da região significativa. Em ambos os casos, denotamos, como uma linha pontilhada em azul, o eixo da linha de visada. Logo, vemos que, de fato, a maiores variações ocorrem nesse eixo e que o método CHMS diminui esse viés.

Finalizamos, portanto, o capítulo de resultados deste trabalho. Nele, estudamos a distribuição dos quasares das nossas amostras desde uma perspectiva global, por funções de multiplicidade e correlação, até o estudo de casos de interesse, como pares e LQGs, encontrando, em muitos casos, boas concordâncias entre nossos resultados e os esperados.

Capítulo 5

Conclusões

Neste trabalho, caracterizamos o agrupamento de quasares brilhantes do catálogo QuCatS. Para isso, tratamos os dados através de um cruzamento com o catálogo do GAIA, encontrando, por consequência, uma boa concordância entre eles. Nessa etapa, separamos os candidatos em três amostras distintas, mais especificamente, pré e pós-comparação de probabilidades, e pós-comparação de *redshifts*.

Para essas amostras, calculamos suas funções de correlação e as comparamos com a função resultante do catálogo SDSS, concluindo, em sequência, que nossos resultados se comportam de acordo com o esperado de amostras fotométricas. Após, aplicamos nas amostras o algoritmo de agrupamento com diferentes parâmetros. Desses resultados, através da função de multiplicidade e das estatísticas de grafos, analisamos seu comportamento de maneira global, concluindo, respectivamente, que suas abundâncias estão de acordo com o esperado e que as diferentes amostras geram distribuições com características semelhantes.

Como estudo de caso, verificamos, na literatura, 40 candidatos a pares de quasares próximos, encontrando, por consequência, uma confirmação de par binário e três candidatos que podem ser descartados. Além disso, também aplicamos um método estatístico, CHMS, na procura de candidatos promissores a LQG. Disso, encontramos dois grupos, de alta significância, que aparentam estar em concordância com outros objetos da literatura.

Por fim, consideramos as observações desses quasares, a introdução das estimativas de erro e das máscaras de correção no cálculo da função de correlação e o aprofundamento do uso das estatísticas de grafos como possíveis continuações para este projeto.

Referências Bibliográficas

Abdalla E., Abellán G. F., Aboubrahim A., Agnello A., Akarsu O., Akrami Y., Alestas G., Aloni D., Amendola L., Anchordoqui L. A., Anderson R. I., Arendse N., Asgari M., Ballardini M., Barger V., Basilakos S., Batista R. C., Battistelli E. S., Battye R., Benetti M., Benisty D., Berlin A., de Bernardis P., Berti E., Bidenko B., Birrer S., Blakeslee J. P., Boddy K. K., Bom C. R., Bonilla A., Borghi N., Bouchet F. R., Braglia M., Buchert T., Buckley-Geer E., Calabrese E., Caldwell R. R., Camarena D., Capozziello S., Casertano S., Chen G. C.-F., Chluba J., Chen A., Chen H.-Y., Chudaykin A., Cicoli M., ... Zumalacárregui M., Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies, Journal of High Energy Astrophysics, 2022, vol. 34, p. 49–211

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., Brinkmann J., Gott III J. R., Kleinman S. J., Krzesinski J., Lee B. C., Miller C. J., Nitta A., Schneider D. P., Tucker D. L., Zehavi I., for The SDSS Collaboration Percolation Galaxy Groups and Clusters in the SDSS Redshift Survey: Identification, Catalogs, and the Multiplicity Function, The Astrophysical Journal Supplement Series, 2006, vol. 167, p. 1

Bonnarel F., Fernique P., Bienaymé O., Egret D., Genova F., Louys M., Ochsenbein F., Wenger M., Bartlett J. G., The ALADIN interactive sky atlas. A reference tool for identification of astronomical sources, A&AS, 2000, vol. 143, p. 33

Cardoso N. M., , 2022 Astrotools: Web-based astronomical tools

Chaves-Montero J., Angulo R. E., Hernández-Monteagudo C., The effect of photometric

redshift uncertainties on galaxy clustering and baryonic acoustic oscillations, Monthly Notices of the Royal Astronomical Society, 2018, vol. 477, p. 3892–3909

Clowes R. G., Large Quasar Groups - A Short Review.. In The New Era of Wide Field Astronomy , vol. 232 of Astronomical Society of the Pacific Conference Series, 2001, p. 108

Clowes R. G., Campusano L. E., Graham M. J., Söchting I. K., Two close large quasar groups of size 350Mpc at: Two close large quasar groups, Monthly Notices of the Royal Astronomical Society, 2011, vol. 419, p. 556–565

Clowes R. G., Harris K. A., Raghunathan S., Campusano L. E., Söchting I. K., Graham M. J., A structure in the early Universe at z 1.3 that exceeds the homogeneity scale of the R-W concordance cosmology, Monthly Notices of the Royal Astronomical Society, 2013, vol. 429, p. 2910

Coe D., Umetsu K., Zitrin A., Donahue M., Medezinski E., Postman M., Carrasco M., Anguita T., Geller M. J., Rines K. J., Diaferio A., Kurtz M. J., Bradley L., Koekemoer A., Zheng W., Nonino M., Molino A., Mahdavi A., Lemze D., Infante L., Ogaz S., Melchior P., Host O., Ford H., Grillo C., Rosati P., Jiménez-Teja Y., Moustakas J., Broadhurst T., Ascaso B., Lahav O., Bartelmann M., Benítez N., Bouwens R., Graur O., Graves G., Jha S., Jouvel S., Kelson D., Moustakas L., Maoz D., Meneghetti M., Merten J., Riess A., Rodney S., Seitz S., CLASH: PRECISE NEW CONSTRAINTS ON THE MASS PROFILE OF THE GALAXY CLUSTER A2261, The Astrophysical Journal, 2012, vol. 757, p. 22

Creevey, L. Sordo, R. Pailler, F. Frémat, Y. Heiter, U. Thévenin, F. Andrae, R. Fouesneau, M. Lobel, A. Bailer-Jones, C. A. L. Garabato, D. Bellas-Velidis, I. Brugaletta, E. Lorca, A. Ordenovic, C. Palicio, P. A. Sarro, L. M. Delchambre, L. Drimmel, R. Rybizki, J. Torralba Elipe, G. Korn, A. J. Recio-Blanco, A. Schultheis, M. S. ... Silvester, J. Gaia Data Release 3 - Astrophysical parameters inference system (Apsis). I. Methods and content overview, AA, 2023a, vol. 674, p. A26

Creevey, L. Sordo, R. Pailler, F. Frémat, Y. Heiter, U. Thévenin, F. Andrae, R. Fouesneau, M. Lobel, A. Bailer-Jones, C. A. L. Garabato, D. Bellas-Velidis, I. Brugaletta, E. Lorca,

A. Ordenovic, C. Palicio, P. A. Sarro, L. M. Delchambre, L. Drimmel, R. Rybizki, J. Torralba Elipe, G. Korn, A. J. Recio-Blanco, A. Schultheis, M. S. ... Silvester, J. Gaia Data Release 3 - Astrophysical parameters inference system (Apsis). I. Methods and content overview, AA, 2023b, vol. 674, p. A26

Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., The 2dF QSO Redshift Survey - XII. The spectroscopic catalogue and luminosity function, MNRAS, 2004, vol. 349, p. 1397

Diestel R., Graph Theory. Terceira Edição. Heidelberg, Springer-Verlag, 2005, 422 p.

Ester M., Kriegel H.-P., Sander J., Xu X., A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining , KDD'96, AAAI Press, 1996, p. 226–231

Fujii H., Critical assessment of the recent report on the gigaparsec-scale correlation of the orientations of large quasar groups, Monthly Notices of the Royal Astronomical Society, 2023, vol. 527, p. 1982

Gaia Collaboration Bailer-Jones, C. A. L. Teyssier, D. Delchambre, L. Ducourant, C. Garabato, D. Hatzidimitriou, D. Klioner, S. A. Rimoldini, L. Bellas-Velidis, I. Carballo, R. Carnerero, M. I. Diener, C. Fouesneau, M. Galluccio, L. Gavras, P. Krone-Martins, A. Raiteri, C. M. Teixeira, R. Brown, A. G. A. Vallenari, A. Prusti, T. ... Zwitter, T. Gaia Data Release 3 - The extragalactic content, AA, 2023, vol. 674, p. A41

Gott III J. R., Juri M., Schlegel D., Hoyle F., Vogeley M., Tegmark M., Bahcall N., Brinkmann J., A Map of the Universe, The Astrophysical Journal, 2005, vol. 624, p. 463–484

Jatenco-Pereira V., Friaça A. C. S., Pino E. D., Junior L. S., Astronomia: Uma Visão Geral Do Universo. Segunda Edição. São Paulo, Editora da Universidade de São Paulo, 2006

Kochanek C. S., Falco E. E., Munoz J. A., , 1997 Why Quasar Pairs Are Binary Quasars And Not Gravitational Lenses

Landy S. D., Szalay A. S., Bias and Variance of Angular Correlation Functions, ApJ, 1993, vol. 412, p. 64

Lemon C., Anguita T., Auger-Williams M. W., Courbin F., Galan A., McMahon R., Neira F., Oguri M., Schechter P., Shajib A., Treu T., Agnello A., Spiniello C., Gravitationally lensed quasars in Gaia – IV. 150 new lenses, quasar pairs, and projected quasars, Monthly Notices of the Royal Astronomical Society, 2022, vol. 520, p. 3305

Lopez A. M., Clowes R. G., Williger G. M., A Giant Arc on the Sky, Monthly Notices of the Royal Astronomical Society, 2022, vol. 516, p. 1557–1572

Lupton R., Blanton M., Fekete G., Hogg D., O'Mullane W., Szalay A., Wherry N., Preparing Red-Green-Blue Images from CCD Data, Publications of the Astronomical Society of the Pacific, 2004, vol. 116, p. 133–137

Lyke B. W., Higley A. N., McLane J. N., Schurhammer D. P., Myers A. D., Ross A. J., Dawson K., Chabanier S., Martini P., Busca N. G., Mas des Bourboux H. d., Salvato M., Streblyanska A., Zarrouk P., Burtin E., Anderson S. F., Bautista J., Bizyaev D., Brandt W. N., Brinkmann J., Brownstein J. R., Comparat J., Green P., de la Macorra A., Muñoz Gutiérrez A., Hou J., Newman J. A., Palanque-Delabrouille N., Pâris I., Percival W. J., Petitjean P., Rich J., Rossi G., Schneider D. P., Smith A., Vivek M., Weaver B. A., The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release, ApJS, 2020, vol. 250, p. 8

Mendes de Oliveira C., Ribeiro T., Schoenell W., Kanaan A., Overzier R. A., Molino A., Sampedro L., Coelho P., Barbosa C. E., Cortesi A., Costa-Duarte M. V., Herpich F. R., Hernandez-Jimenez J. A., Placco V. M., Xavier H. S., Abramo L. R., Saito R. K., Chies-Santos A. L., Ederoclite A., Lopes de Oliveira R., Gonçalves D. R., Akras S., Almeida L. A., Almeida-Fernandes F., Beers T. C., Bonatto C., Bonoli S., Cypriano E. S., Vinicius-Lima E., de Souza R. S., Fabiano de Souza G., Ferrari F., Gonçalves T. S., Gonzalez A. H., Gutiérrez-Soto L. A., Hartmann E. A., Jaffe Y., Kerber L. O., Lima-Dias C., The Southern Photometric Local Universe Survey (S-PLUS): improved SEDs, morphologies, and redshifts with 12 optical filters, Monthly Notices of the Royal Astronomical Society, 2019, vol. 489, p. 241–267

Meusinger, H. Hinze, A. de Hoon, A. Spectral variability of quasars from multi-epoch photometric data in the Sloan Digital Sky Survey Stripe 82, AA, 2011, vol. 525, p. A37

Nakazono L., R Valença R., Soares G., Izbicki R., Ivezic Z., R Lima E. V., T Hirata N. S., Sodré Jr L., Overzier R., Almeida-Fernandes F., Oliveira Schwarz G. B., Schoenell W., Kanaan A., Ribeiro T., Mendes de Oliveira C., The Quasar Catalogue for S-PLUS DR4 (QuCatS) and the estimation of photometric redshifts, Monthly Notices of the Royal Astronomical Society, 2024, vol. 531, p. 327–339

Paris, Isabelle Petitjean, Patrick Aubourg, Éric Myers, Adam D. Streblyanska, Alina Lyke, Brad W. Anderson, Scott F. ... Zarrouk, Pauline Zhao, Gong-Bo The Sloan Digital Sky Survey Quasar Catalog: Fourteenth data release, AA, 2018, vol. 613, p. A51

Peebles P. J. E., The large-scale structure of the universe, 1980

Puddu, E. De Filippis, E. Longo, G. Andreon, S. Gal, R. R. The multiplicity function of galaxies, AA, 2003, vol. 403, p. 73

Ryden B., Introduction to Cosmology. Segunda Edição, Cambridge University Press, 2016

Scialpi, M. Mannucci, F. Marconcini, C. Venturi, G. Pancino, E. Marconi, A. Cresci, G. Belfiore, F. Amiri, A. Bertola, E. Carniani, S. Cicone, C. Ciurlo, A. D'Amato, Q. Ginolfi, M. Lusso, E. Marasco, A. Nardini, E. Rubinur, K. Severgnini, P. Tozzi, G. Ulivi, L. Vignali, C. Volonteri, M. MUSE adaptive-optics spectroscopy confirms dual active galactic nuclei and strongly lensed systems at sub-arcsec separation, AA, 2024, vol. 690, p. A57

Shen Y., Hwang H.-C., Oguri M., Chen N., Matteo T. D., Ni Y., Bird S., Zakamska N., Liu X., Chen Y.-C., Kratter K. M., Statistics of Galactic-scale Quasar Pairs at Cosmic Noon, The Astrophysical Journal, 2023, vol. 943, p. 38

Sinha M., Garrison L. H., corrfunc – a suite of blazing fast correlation functions on the CPU, Monthly Notices of the Royal Astronomical Society, 2019, vol. 491, p. 3022–3041

Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H.,

Evrard A., Colberg J., Pearce F., Simulations of the formation, evolution and clustering of galaxies and quasars, Nature, 2005, vol. 435, p. 629–636

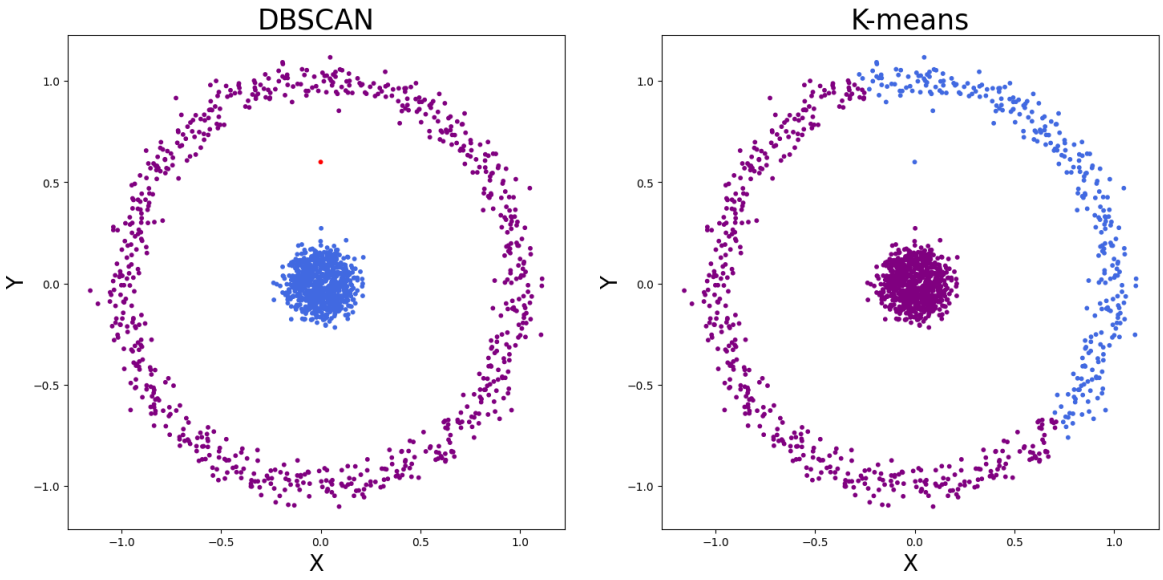
Wang F., Yang J., Fan X., Hennawi J. F., Barth A. J., Banados E., Bian F., Boutsia K., Connor T., Davies F. B., Decarli R., Eilers A.-C., Farina E. P., Green R., Jiang L., Li J.-T., Mazzucchelli C., Nanni R., Schindler J.-T., Venemans B., Walter F., Wu X.-B., Yue M., A Luminous Quasar at Redshift 7.642, The Astrophysical Journal Letters, 2021, vol. 907, p. L1

West A. A., Hawley S. L., Bochanski J. J., Covey K. R., Reid I. N., Dhital S., Hilton E. J., Masuda M., Constraining the Age-Activity Relation for Cool Stars: The Sloan Digital Sky Survey Data Release 5 Low-Mass Star Spectroscopic Sample, AJ, 2008, vol. 135, p. 785

Apêndice

Apêndice A

Comparação entre DBSCAN e K-means

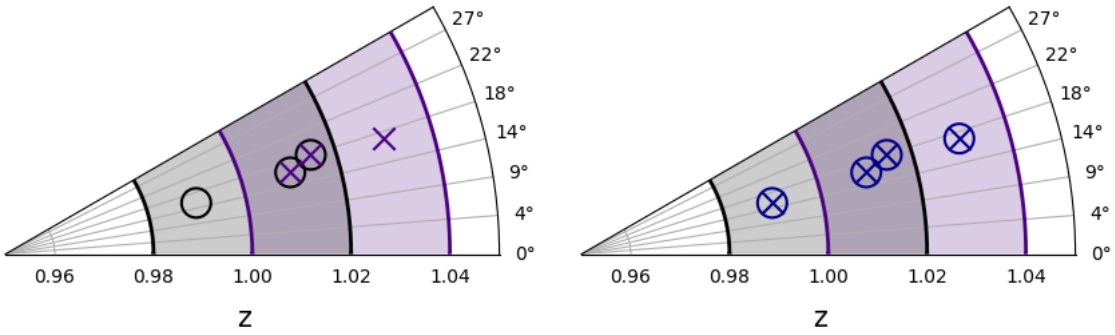


(a) (b)

Figura A.1: Comparação entre K-means e DBSCAN. Em (a) observamos que o algoritmo do DBSCAN consegue separar de forma satisfatória as duas estruturas (em roxo e em azul) enquanto ignora o ruído (ponto vermelho). Por outro lado, em (b), vemos que o K-means não é capaz de distinguir as duas estruturas de forma satisfatória e, além disso, não consegue ignorar o ruído entre elas.

Apêndice B

Exemplo da redefinição dos grupos



(a) Pré-redefinição

(b) Pós-redefinição

Figura B.1: Exemplo da redefinição dos grupos. Em (a) e em (b) temos, em preto e em roxo, duas cascas esféricas que se intersectam entre os valores 1.00 e 1.02. Em (a) temos os dados pré-redefinição, ou seja, dois grupos distintos que possuem dois objetos em comum, os elementos de cada grupo estão identificados como círculos e cruzes. Em (b) temos os dados pós-redefinição, em azul escuro, como parte de um mesmo grupo e não limitados em suas cascas originais.

Apêndice C

Distribuição dos candidatos do QuCatS na esfera celeste

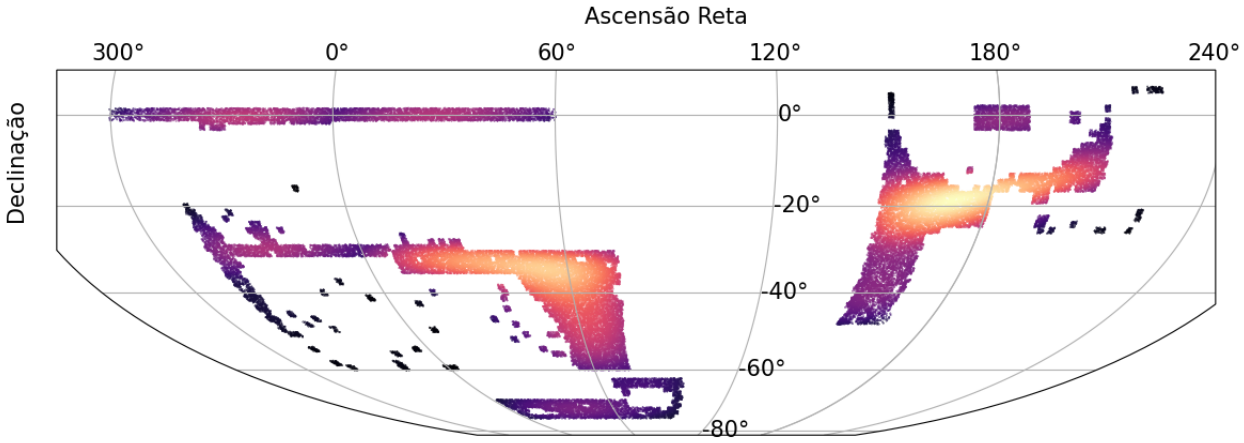


Figura C.1: Distribuição dos candidatos do QuCatS na esfera celeste. Os pontos mais claros indicam uma maior concentração de dados. Note que os objetos estão distribuídos em regiões específicas da esfera. Por isso, o DBSCAN não consegue formar um único grupo mesmo que possua um parâmetro saturado.

Apêndice D

Regiões selecionadas dos catálogos SDSS e GAIA

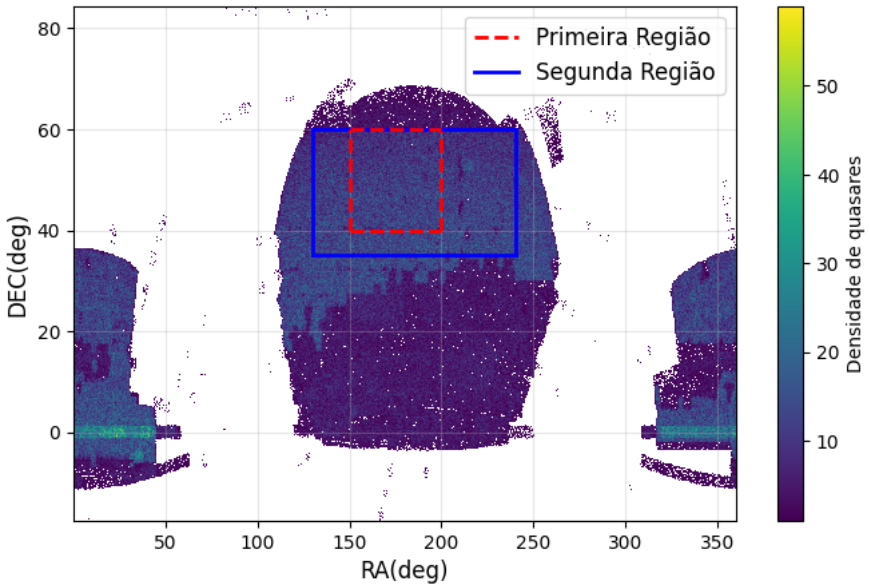


Figura D.1: Regiões selecionadas do catálogo de quasares do SDSS-DR16. A primeira contém 75197 quasares com valores de ascensão reta que variam de 150 a 200 graus e de declinação que variam de 40 a 60 graus. A segunda contém 215414 quasares com valores de ascensão reta que variam de 130 a 240 graus e de declinação que variam de 35 a 60 graus.

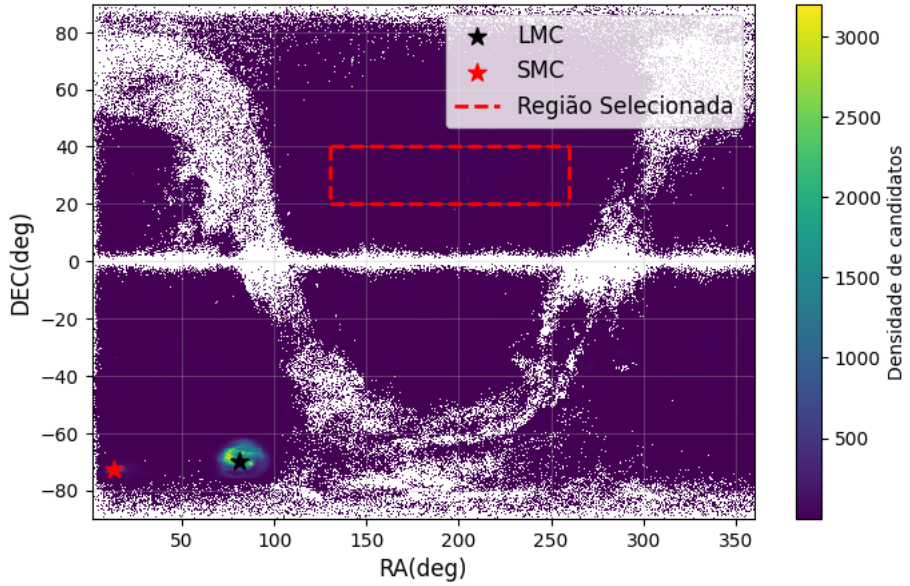


Figura D.2: Região selecionada do catálogo de candidatas a quasares do GAIA. Contendo 143609 candidatos a quasares, os valores de ascensão reta dessa região variam de 130 a 260 graus e os valores de declinação variam de 20 a 40 graus.

Apêndice E

Função de correlação das amostras restantes

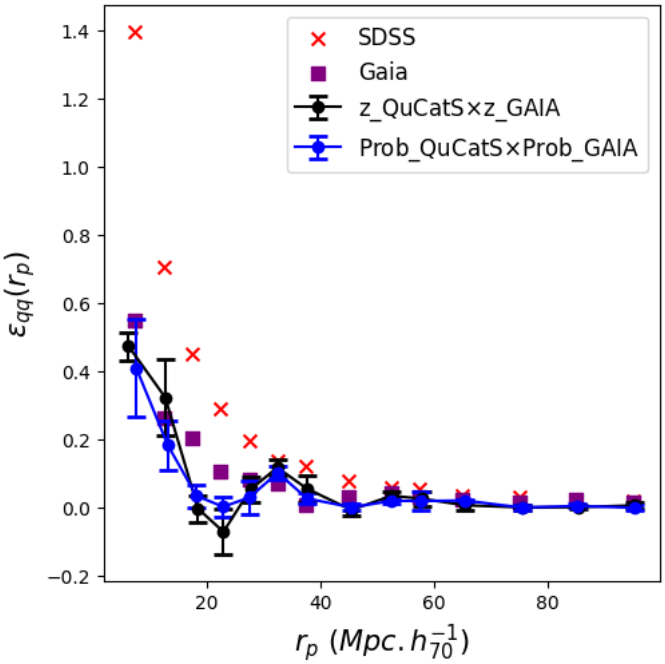


Figura E.1: Funções de correlação: Prob_QuCatS× Prob_GAIA e z_QuCatS×z_GAIA. Apresentamos como cruzes vermelhas, quadrados roxos e pontos azuis e pretos, respectivamente, os dados obtidos pelo SDSS, GAIA e Prob_QuCatS × Prob_GAIA e z_QuCatS×z_GAIA. Novamente, utilizamos o método *Jackknife* para estimarmos os erros da nossa base de dados.

Apêndice F

Tabelas de distribuição dos agrupamentos

Tabela F.1 - Distribuição dos agrupamentos resultantes da amostra Prob_QuCatS×Prob_GAIA.

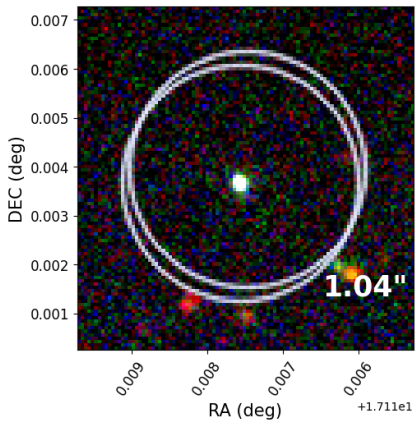
α	L(Mpc)	Duplas	Trios	Quartetos	Quintetos	Estruturas	Quantidade de Grupos	Objetos Agrupados	Fração do Agrupamento
0.5	1	76	0	0	0	0	76	152	0.003
0.5	5	1436	79	2	0	0	1517	3117	0.057
0.5	10	3793	646	128	35	17	4619	10323	0.188
1.0	1	165	0	0	0	0	165	330	0.006
1.0	5	2720	286	33	5	2	3046	6468	0.118
1.0	10	5118	1392	544	235	216	7505	19327	0.352
1.5	1	233	2	0	0	0	235	472	0.009
1.5	5	3605	552	86	23	10	4276	9388	0.171
1.5	10	4823	1788	797	435	678	8521	25822	0.470

Tabela F.2 - Distribuição dos agrupamentos resultantes da amostra z_QuCatS×z_GAIA.

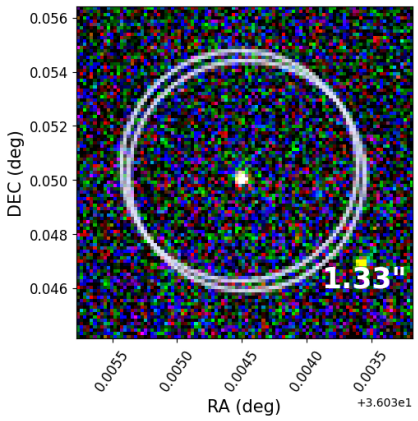
α	L(Mpc)	Duplas	Trios	Quartetos	Quintetos	Estruturas	Quantidade de Grupos	Objetos Agrupados	Fração do Agrupamento
0.5	1	39	0	0	0	0	39	78	0.003
0.5	5	659	21	0	0	0	680	1381	0.046
0.5	10	1893	230	42	8	2	2175	4698	0.156
1.0	1	75	0	0	0	0	75	150	0.005
1.0	5	1291	94	8	0	0	1393	2896	0.096
1.0	10	2936	626	201	73	31	3867	9136	0.303
1.5	1	105	1	0	0	0	106	213	0.007
1.5	5	1794	172	27	5	0	1998	4237	0.141
1.5	10	3196	998	381	149	140	4864	12621	0.419

Apêndice G

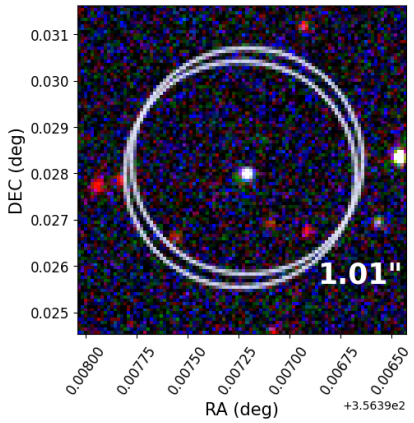
Candidatos a pares de quasares descartados



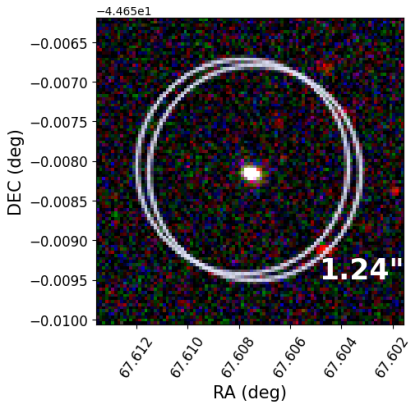
(d)



(e)



(f)



(g)

Figura G.1: Candidatos a pares de quasares detectados pelo DBSCAN na amostra QuCatS que foram descartados. As imagens foram obtidas através do pacote do S-PLUS, em *Python*, mais especificamente de ferramentas de imageamento descritas em Coe et al. (2012). No canto inferior direito das imagens são exibidas as distâncias angulares das fontes em segundos de arco. Os círculos brancos ao redor dos candidatos estão centralizados nas suas coordenadas.

Apêndice H

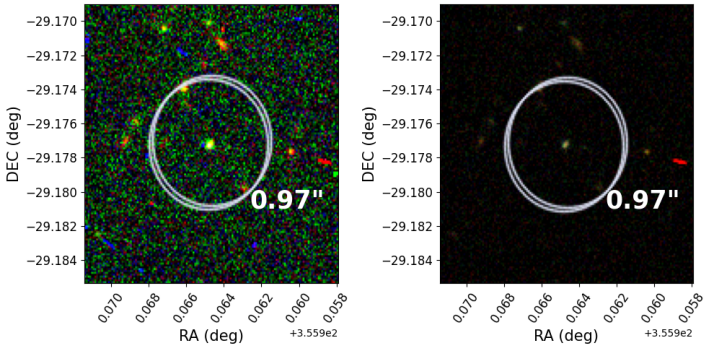
Candidatos a pares de quasares detectados pelo cruzamento

Tabela H.1 - Pares de objetos encontrados através do cruzamento de dados entre os catálogos QuCatS e GAIA.

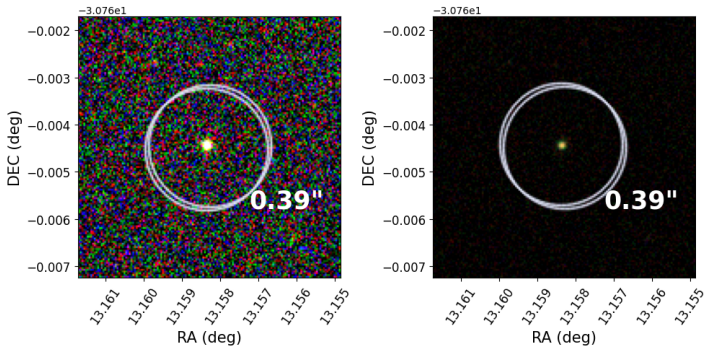
Dupla	RA (deg)	DEC (deg)	$\delta\theta$ $(arcosegundo)$	$Prob_{GAIA}$	$Prob_{QuCatS}$	z_{GAIA}	z_{QuCatS}	G (mag)	r (mag)
0	202.000725 202.001013	-7.803090 -7.802826	1.398865	0.998387 0.951235	0.860000	1.923974 5.411451	1.553151	20.266521 20.553738	19.759624
1	152.604709 152.604814	-17.439525 -17.439394	0.594222	0.021139 0.049911	0.930000	- -	0.674533	20.005096 20.088224	19.450663
(a)	355.964579 355.964759	-29.177059 -29.177276	0.965748	0.999904 0.691895	0.840000	3.441521 2.327521	1.330693	20.189411 -	19.740349
3	27.948835 27.948852	-31.203263 -31.203501	0.858748	0.996223 0.015480	0.820000	0.218687 -	1.012198	20.587486 20.713367	19.998478
4	316.581286 316.581255	-24.888283 -24.887913	1.336355	0.000000 0.983922	0.960000	- 1.884085	2.134444	19.088064 18.872353	18.236477
(b)	13.158248 13.158340	-30.764511 -30.764437	0.388286	0.014481 0.013156	0.980000	- -	1.388981	19.777527 19.763510	19.089005
6	49.565339 49.565445	-31.567447 -31.567323	0.549777	0.015954 0.054825	0.830000	- -	1.504530	20.637798 21.139286	18.641888
7	18.908960 18.908870	-34.842528 -34.842730	0.772377	0.021614 0.012248	0.810000	2.709864 -	2.013090	20.457579 20.796276	19.952677
(c)	330.619685 330.619367	0.817290 0.817172	1.223566	0.000000 0.001632	1.000000	- -	1.211785	20.235037 20.564340	19.920763
9	162.984102 162.984082	-33.483260 -33.483338	0.286107	0.009386 0.818664	1.000000	- 1.081647	2.088633	20.763678 20.366417	18.928844
10	5.171334 5.170084	-73.362941 -73.362961	1.290306	0.997775 0.108103	0.875000	0.743405 -	1.241490	20.666714 20.556868	19.904530

11	89.094125 89.094943	-74.941372 -74.941676	1.337214	0.999996 0.242846	0.830000	1.804671 -	0.966567	20.596806 20.580803	19.935202
12	73.740487 73.740248	-74.427902 -74.427736	0.640291	0.019342 0.002180	0.815000	- -	2.412553	20.651325 20.502863	19.896004
13	83.860470 83.860271	-72.410751 -72.410417	1.221014	0.015078 0.030252	0.825000	- -	2.435110	20.940657 20.705357	19.954224
14	83.002802 83.002627	-72.290458 -72.290627	0.639279	0.001202 0.013025	0.805000	- -	2.276964	20.937225 20.123259	19.714480
15	91.303374 91.303274	-72.570357 -72.570048	1.115705	0.001092 0.001429	0.810000	- -	1.186609	20.106482 20.042640	19.327740
16	84.732727 84.732046	-63.915033 -63.915244	1.317741	0.001020 0.997404	0.940000	- 0.559990	1.210902	20.778595 20.215340	19.817050
17	86.507572 86.508070	-64.821048 -64.821094	0.780655	0.054936 0.005607	0.840000	- -	2.336511	20.531076 20.744950	19.985540
18	85.440831 85.440675	-63.858806 -63.859072	0.990254	0.000276 0.062629	0.940000	- -	1.963278	20.598282 20.205988	19.714277
19	95.955600 95.956020	-67.062847 -67.062956	0.708838	0.647428 0.007048	0.925000	0.687301 -	1.740047	20.623426 20.649248	19.985947
20	69.338317 69.337509	-67.568387 -67.568483	1.162330	0.048401 0.999922	0.805000	- 2.080800	2.621506	20.802204 20.590878	19.940720
21	72.066077 72.065396	-66.834315 -66.834317	0.964093	0.039720 0.000001	0.815000	- -	1.781274	19.563662 20.347387	19.162731
22	77.984457 77.984699	-64.577059 -64.576916	0.637097	0.018685 0.002488	0.980000	- -	2.752522	20.651741 20.239151	19.735882
23	66.976560 66.976724	-73.512284 -73.512116	0.628807	0.041021 0.004676	0.850000	- -	2.310033	20.639292 20.522387	19.826744
24	88.728688 88.728829	-73.520369 -73.520216	0.568841	0.000713 0.181064	0.810000	- -	1.576865	20.678669 20.653960	19.984982
25	80.436732 80.437255	-74.193010 -74.192834	0.813572	0.000011 0.007100	0.855000	- -	2.408278	20.437315 20.620790	19.867002
26	7.822958 7.821603	-74.005027 -74.004883	1.440990	0.013479 0.007938	0.840000	- -	1.700836	20.303493 20.879816	19.978489
27	63.303789 63.303798	-72.769201 -72.768985	0.778559	0.001112 0.000000	0.940000	- -	1.684696	19.724585 19.773764	19.124752
28	79.636525 79.636347	-64.502399 -64.502635	0.893674	0.003730 0.006119	0.810000	- -	1.475155	20.114372 20.385946	19.547438
29	81.391169 81.391442	-63.974225 -63.974391	0.736784	0.053082 0.000190	0.880000	- -	2.233179	20.603570 20.502830	19.857445
30	18.658442 18.657491	-71.613840 -71.613539	1.529201	0.999999 0.190344	0.875000	2.500938 1.963115	2.322092	20.296722 20.658447	19.958628

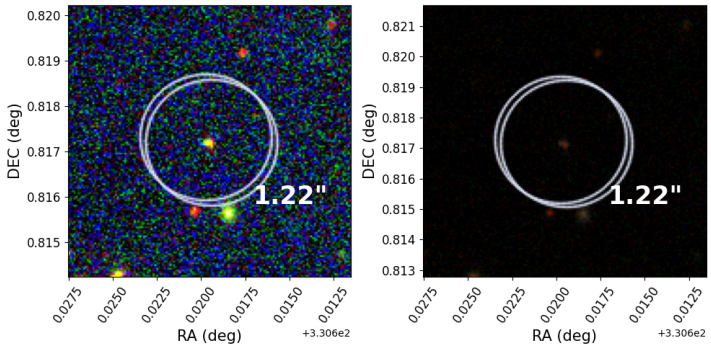
31	14.869538 14.868560	-71.612949 -71.612841	1.175779	0.000622 0.115397	0.885000	- -	2.554327	21.030695 20.443237	19.963366
32	16.573938 16.573733	-70.961177 -70.961122	0.311437	0.000000 0.000509	0.925000	- -	2.328518	20.244938 20.239832	19.837816



(a) Fonte única



(b) Par binário confirmado



(c) Estrela e quasar

Figura H.1: Candidatos advindos dos cruzamentos encontrados na literatura. No canto inferior direito das imagens são exibidas as distâncias angulares das fontes em segundos de arco. Os círculos brancos ao redor dos candidatos estão centralizados nas suas coordenadas. Note que no caso (a), apesar de ser considerado uma fonte única, pela separação angular, ainda é possível que seja um par de quasar.