

UMA ABORDAGEM VISUAL PARA A EXPLORAÇÃO DOS DADOS EXTRAGALÁCTICOS DA MISSÃO GAIA

A VISUAL APPROACH TO EXTRAGALACTIC GAIA DATA EXPLOITATION

Maria Eduarda Manfrinato Pimentel

UNESP Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Ciências,
Departamento de Computação, Bacharelado em Sistemas de Informação - Universidade de Lisboa,
Faculdade de Ciências, Bacharelado e Mestrado em Física, eduarda.pimentel@unesp.br

Resumo: *A missão Gaia, da Agência Espacial Europeia, foi um marco na astrofísica galáctica, uma vez que catalogou mais de dois bilhões de objetos com precisão nunca antes obtida, criando um mapa completo da Via Láctea. Apesar de originalmente ter sido desenvolvida para objetos galácticos, seus instrumentos são sensíveis o suficiente para detectar fontes extragalácticas brilhantes, abrindo as portas para uma nova gama de estudos extragalácticos. Para explorar todo o potencial da missão Gaia, porém, é necessário processar um grande volume de dados, tarefa que as ferramentas astronômicas tradicionais não conseguem realizar. Este processamento de dados é um dos grandes desafios da Astronomia moderna. Neste trabalho, traz-se um resumo da dissertação de mestrado da autora, onde é desenvolvido um sistema para o processamento de dados do terceiro Data Release da missão Gaia, com o objetivo de criar visualizações focadas nas fontes extragalácticas que se encontram em meio aos dados galácticos.*

Palavras-chave: Astroinformática; Visualização de dados; Big Data; missão Gaia

Abstract: *The Gaia mission, launched by the European Space Agency, was a milestone in galactic astrophysics, as it cataloged over two billion objects with unprecedented precision, providing a comprehensive map of the Milky Way. Although it was originally designed for galactic objects, its instruments are sensitive enough to detect bright extragalactic sources, opening the door to a new range of extragalactic studies. However, to fully explore the mission's potential, it is necessary to process a large volume of data, a task that traditional astronomical tools are unable to handle. This data handling is one of the major challenges of modern Astronomy. The present work presents a summary of the author's Master's dissertation, in which a system is developed to process data from the third Data Release of the Gaia mission, with the aim of creating visualizations focused on extragalactic sources embedded within galactic data.*

Keywords: Astroinformatics; Data Visualization, Big Data, Gaia mission

INTRODUÇÃO

O potencial extragaláctico da Missão Gaia

A medição das posições e movimentos dos objetos no céu é um interesse para os astrônomos desde a Antiguidade, e ainda é parte essencial da Astronomia moderna. Os avanços na instrumentação permitiram que os dados astrométricos ficassem cada vez mais precisos, e a missão Gaia é um exemplo disto.



Lançada em 2013 pela Agência Espacial Europeia (ESA), o objetivo primário desta missão era de criar um mapa tridimensional completo da Via Láctea, permitindo estudos sobre a dinâmica, formação e estrutura da galáxia. Para isto, os instrumentos deste telescópio medem diversas propriedades das fontes detectadas, como as posições, paralaxes, velocidades radiais e movimentos próprios. Desde que começou a operar, já foram mais de dois bilhões de objetos catalogados. As resoluções atingidas pelo Gaia também são sem precedentes, com um erro padrão na ordem dos 10 micro-segundos de arco (mas), superando em dez vezes a resolução da missão que a antecedeu, a Hipparcos.

Os instrumentos da missão Gaia possuem um limite de brilho detectável, sendo calibrados para poder detectar fontes pontuais, priorizando observar objetos dentro da Via Láctea (Krone-Martins et al., 2013). Porém, estas características também permitiram que a missão detectasse as estrelas mais brilhantes das galáxias mais próximas; galáxias distantes não são resolvidas, mas ainda podem ser detectadas pelos seus bojos centrais mais brilhantes. Assim, objetos extragalácticos também fazem parte do catálogo do telescópio Gaia (Bailer-Jones, 2023).

Para filtrar os objetos extragalácticos entre os milhões de dados observados, é necessário definir características que possam ser únicas ou típicas destes objetos, permitindo separá-los dos objetos da Via Láctea, que são a maioria. O trabalho de isolar objetos extragalácticos individuais já foi feito antes para objetos bem conhecidos, como nos estudos da Nuvem de Magalhães produzidos por Jiménez-Arranz et al. (2023a, 2023b), e nos estudos de Helmi et al. (2018) que envolveram também a cinemática de nove galáxias anãs vizinhas da Via Láctea. Este processo de separação em geral envolve fazer cortes espaciais centrados numa região de interesse, e em seguida, realizar cortes astrométricos e fotométricos, segundo as propriedades de cada objeto, como a cinética e as cores.

Outra maneira de identificar objetos extragalácticos seria selecionando as paralaxes mais próximas de zero, uma vez que a paralaxe diminui com a distância. Apesar da alta precisão das medições do Gaia, para objetos distantes os erros destas observações fazem com que este filtro não seja, por si só, confiável (Gaia Collaboration et al., 2023; Jiménez-Arranz et al., 2023b).

Astronomia na era de *Big Data*

Uma vez estabelecida a possibilidade de estudos extragalácticos com os dados da missão Gaia, é necessário notar a dificuldade de explorar o potencial completo desta base de dados, devido ao seu tamanho. Isto está atrelado a dois fatores principais: a necessidade de representações visuais dos dados da missão, e à grande quantidade de dados produzida.

A criação de representações gráficas das bases de dados é frequentemente uma parte essencial do processo de descoberta científica. Estas visualizações facilitam o reconhecimento de padrões e comportamentos (Moitinho et al., 2017), que não são tão facilmente percebidos olhando apenas para números numa tabela. Além disso, visualizações servem também para auxiliar na filtragem e consulta dos dados, permitindo a validação de resultados científicos, interpretações quantitativas e qualitativas (Hansen et al., 2009).



Em relação à quantidade de dados sendo produzida, a missão Gaia é um exemplo típico do que acontece com as missões espaciais atuais, onde os avanços tecnológicos permitiram não só um aumento na qualidade dos dados, mas também um aumento na quantidade de dados sendo produzida. No caso do Gaia, a base de dados completa é da ordem de *petabytes*. Isto coloca estas novas missões, como a missão Gaia, o telescópio Euclid e o observatório Vera Rubin, por exemplo, no campo da *Big Data*. Esta característica é simultaneamente uma das maiores vantagens e maiores desafios quando se trata da missão Gaia.

Primeiramente, não só os dados do Gaia contêm bilhões de objetos, mas estas medições também são multidimensionais, com as tabelas contendo mais de cem colunas, entre medições diretas e propriedades derivadas destas. Representações 2D tradicionais, como o diagramas cor-cor ou diagramas cor-magnitude podem ser enganadores neste caso: estes gráficos de dispersão, quando chegamos aos milhões de pontos, sofrem com um efeito de excesso de elementos, onde padrões interessantes acabam sendo escondidos em meio à enorme quantidade de informação sendo exibida de uma só vez (Ellis; Dix, 2007).

No campo de *Big Data*, outro problema é que os dados não cabem na memória de um computador comum. Deste modo, operações simples como compartilhar arquivos e aplicar transformações aos seus dados exigem mais recursos computacionais; isto, por sua vez, implica uma maior demora no processamento. É necessário então aplicar estratégias para que a exploração destes dados ocorra de maneira segura e o mais rápido possível.

Dentro deste contexto, fica clara a necessidade de abordar a Astronomia sob a ótica da *Big Data*, uma vez que esta é uma tendência global através das diversas áreas científicas. Apesar disso, os *softwares* e módulos atuais utilizados na Astronomia atualmente não estão preparados para estas quantidades de dados (Hassan et al., 2013).

O presente trabalho sintetiza os resultados de uma dissertação de mestrado (Pimentel, 2024) que propôs um *pipeline* para processar dados do terceiro *data release* (DR3) da missão Gaia, e identificar galáxias dentro desta amostra. O sistema possibilita a produção de imagens customizáveis através de filtragens e rotações. As seções a seguir detalham o percurso metodológico e discutem as conclusões obtidas no estudo.

PANORAMA DO CÓDIGO

Para manipular a totalidade dos dados do terceiro *data release* (DR3) da missão Gaia, a estratégia adotada foi a de *split - apply - combine* (ou *separar - aplicar - combinar*, em tradução livre). Ela possui três etapas gerais (Wickham, 2011):

1. Dividir os dados em pedaços menores, de mais fácil manuseio
2. Manipular estes pedaços individualmente
3. Recombinar os pedaços de modo a obter uma estrutura de dados maior

Este método foi escolhido por ser um padrão dentro da ciência de dados, e de ser suportado por bibliotecas em diversas linguagens de programação, como é o caso da linguagem Python, que foi utilizada para o desenvolvimento do código.



A Figura 1 traz um fluxograma geral das quatro etapas do código. O código ainda conta com um arquivo de configuração, onde o usuário pode customizar parâmetros para utilizar este fluxo como for melhor para seus objetivos científicos.

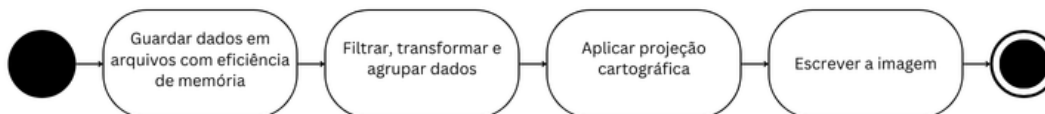


Figura 1: Fluxograma das quatro etapas principais do código desenvolvido para obtenção de uma visualização dos dados da missão Gaia.

Guardar os dados num formato eficiente

Os dados da missão Gaia são guardados num formato de arquivo desenvolvido especialmente para a missão, o *gbin* (*Gaia binary*). Para extrair os dados destes arquivos e acessá-los através do código, foi necessário utilizar um programa fornecido pela equipe da missão Gaia.

Como mencionado anteriormente, é impossível ler todos os dados para a memória do computador. Por isso, o usuário pode informar quais são as colunas da tabela que lhe interessam, e então os dados são lidos em blocos de linhas (*split*). Não são feitas manipulações nesta etapa, e então os dados são salvos num outro arquivo, desta vez no formato HDF5 (*combine*). Este é um formato de arquivo criado especificamente para armazenar grandes quantidades de dados.

O propósito desta etapa é eliminar informações irrelevantes para o usuário (através da pré-seleção de colunas), e tornar a leitura dos dados mais fácil, através da sua replicação no arquivo HDF5, que servirá como a nova fonte de dados para o resto do processamento.

Filtrar, transformar e agrupar os dados

Com os dados já acessíveis, inicia-se a etapa de manipulação e agrupamento. Mesmo após a seleção inicial de colunas, os dados da missão Gaia ainda contêm bilhões de fontes, o que torna pouco eficiente calcular estatísticas ou gerar visualizações diretamente sobre todos os pontos.

Por esse motivo, esta etapa tem como objetivo agrupar os dados armazenados no arquivo HDF5. O agrupamento é realizado por meio do esquema HEALPIX (Hierarchical Equal Area isoLatitude Pixelization), um método de pixelização da esfera que a divide em células de igual área (Górski et al., 2005).

Utilizando a biblioteca *healpy*, que implementa o HEALPIX em Python, cada fonte do Gaia é associada a uma célula com base em suas coordenadas. Dessa forma, é possível calcular estatísticas resumidas por célula, como a média do fluxo em uma banda fotométrica ou o movimento próprio médio.

Nessa etapa também são aplicados os filtros definidos pelo usuário, que seleciona as fontes de acordo com os valores das colunas escolhidas anteriormente. Isso reduz ainda mais a quantidade de dados processados.

Ao final do processo, o arquivo HDF5 passa a armazenar apenas as estatísticas de cada célula HEALPIX, e não mais os dados individuais das fontes. Assim, o volume de dados é reduzido de bilhões de pontos para milhões de células, número que depende da resolução adotada: resoluções mais altas produzem mapas mais detalhados, mas geram um maior número de células (Figura 2).

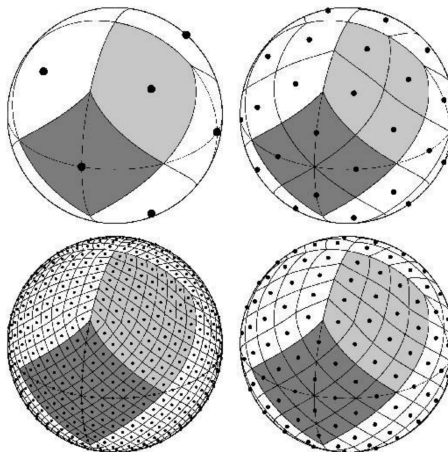


Figura 2: Partição HEALPIX demonstrada numa esfera em diferentes resoluções. Os pontos pretos representam o centro de cada célula. (fonte: Górski, 2005)

Aplicar uma projeção cartográfica e criação da matriz imagem

Nesta etapa, os dados já agrupados em células HEALPIX são utilizados para gerar uma imagem por meio de uma projeção cartográfica. O usuário escolhe qual grandeza será representada, como o brilho nas bandas fotométricas G, RP ou BP da missão *Gaia*.

É possível definir a resolução da imagem final, e selecionar uma entre três projeções (Hammer, cartesiana ou ortográfica). Cada célula HEALPIX possui um ponto central, que é projetado no plano da imagem de acordo com a projeção escolhida. Dessa forma, define-se a qual pixel da imagem cada célula corresponde.

O resultado deste processo é uma matriz, a chamada matriz imagem. Ela possui as mesmas dimensões da imagem final, que guarda em cada pixel o valor da propriedade selecionada, já sendo assim uma espécie de mapa do céu. Porém, a escala destes valores ainda é a escala da grandeza a visualizar. Para poder de fato produzir uma imagem, é preciso transformar esta escala numa escala de base 8, para se obter a imagem em tons de cinza.

Escrever a imagem

Aqui, o valor guardado em cada pixel da matriz imagem passará por uma normalização, de modo a passar este valor para uma escala de 8-bits ou 16-bits, dependendo da escolha do usuário.

A escala 8-bit permite guardar $2^8 = 256$ valores diferentes, enquanto a escala 16-bit permite $2^{16} = 65536$ valores diferentes. A escolha de escala dependerá da amplitude de valores presentes na grandeza selecionada, para as fontes selecionadas; a escala 16-bit permite traduzir uma maior sensibilidade nas medições.

Por fim, estes valores normalizados são utilizados para escrever a imagem PNG final, trazendo a visualização escolhida numa escala cinzenta, como será visto em breve.

METODOLOGIA

Com o código desenvolvido, torna-se possível fazer o estudo das fontes extragalácticas do catálogo da missão Gaia. Para identificá-los, primeiramente foi necessário buscar características comuns entre galáxias vizinhas bem conhecidas, compará-las com dados da região do disco da Via Láctea e usar essa análise para definir quais parâmetros utilizar para filtrar o conjunto de dados completo.

Para realizar a análise das galáxias vizinhas, utilizou-se as mesmas galáxias anãs selecionadas em Helmi et al. (2018), acrescentando também as galáxias M33, Andrômeda e as Nuvens de Magalhães, escolhidas por serem também objetos bem conhecidos e facilmente identificáveis nas visualizações já existentes do Gaia. Amostras de dados ao redor destas fontes extragalácticas foram obtidas com o serviço *online* do Gaia Archive.

A amostra de dados do disco Via Láctea foi obtida a partir de uma subamostra de 1% do DR3, onde foram selecionadas apenas as fontes com latitudes galácticas entre -12° e 26° . Assim, estas fontes pertencem de maneira geral ao disco e ao centro da galáxia, que são as principais fontes de contaminação para objetos extragalácticos.

Como a base de dados do Gaia traz múltiplas medidas astrométricas para suas fontes, foram escolhidas algumas destas medidas para definir os critérios de separação entre fontes galácticas e extragalácticas.

Movimento próprio

O movimento próprio (μ) é uma mudança angular na posição observada de um objeto astronômico, que resulta da projeção do movimento do objeto na esfera celeste. De maneira geral, objetos distantes possuem movimentos próprios menores em comparação ao movimento dos objetos na vizinhança solar.

Assim, é esperado que as galáxias possuam movimentos próprios próximos de zero. Porém, o mesmo pode ocorrer para fontes distantes do Sol, dentro da Via Láctea, de modo que este não é um critério definitivo. Para refinar esta análise, serão analisados os movimentos próprios nas coordenadas galácticas de latitude (μ_{pmb}) e longitude (μ_{pml}), medidos em milissegundos de arco por ano (mas/ano), uma vez que estas irão refletir o movimento das fontes ao redor da Via Láctea.

Deste modo, uma análise das fontes em cada um dos quatro quadrantes galácticos poderá revelar tendências de movimento para os objetos galácticos, que não aparecerão para os objetos extragalácticos. Também aplica-se um filtro para controle de qualidade de cor, seguindo os critérios definidos em Helmi et al. (2018).

Paralaxe

A paralaxe (ϖ) é também uma mudança angular da posição de um objeto em comparação com os objetos de fundo no céu. Este efeito é causado não só pela mudança de posição do observador, mas também pelo movimento do objeto através da galáxia, isto é, seu movimento próprio, como descrito anteriormente. A paralaxe

é de grande importância para a Astronomia, pois pode ser utilizada para calcular a distância até um objeto:

$$d = 1/\varpi$$

onde d é a distância em parsecs, e ϖ é a paralaxe em segundos de arco.

A paralaxe não é medida diretamente pelos instrumentos da missão Gaia, mas sim calculada a partir de uma modelagem aplicada às observações da mudança de posição nos sentidos da ascensão reta (α) e declinação (δ). Esta modelagem revela um padrão de hélice, de onde se extraem os valores da paralaxe para aquele objeto, bem como a sua incerteza (Luri et al., 2018).

Pela equação apresentada, objetos muito distantes apresentam paralaxes muito pequenas, próximas de zero. Análises estatísticas (Bailer-Jones, 2023; Luri et al., 2018) mostram que, nesses casos, as incertezas do modelo fazem com que a paralaxe medida tenha probabilidade semelhante de assumir valores positivos ou negativos.

Tradicionalmente, valores negativos de paralaxe seriam descartados, pois levariam a uma “distância negativa”, que não possui significado físico, e são um efeito da incerteza na medição. Porém, para obter apenas as fontes extragalácticas da missão Gaia, serão selecionadas exclusivamente fontes com paralaxe negativa, uma vez que isto é um forte indicador de objetos longínquos.

Cor BP-RP

A fotometria traz muitas informações sobre uma fonte astronômica. A missão Gaia possui três bandas fotométricas, G (350–1000 nm), que cobre grande parte do espectro visível, BP (330–680 nm), associada à região azul, e RP (640–1000 nm), associada à região vermelha do espectro. A cor BP–RP é definida como a diferença entre as magnitudes medidas nas bandas BP e RP para um mesmo objeto.

A cor está relacionada principalmente à temperatura da fonte, mas também pode refletir o avermelhamento causado pelo meio interestelar. No caso das estrelas, ela pode ainda fornecer informações sobre idade e massa. Assim, a cor de uma galáxia é um indicador de sua população estelar.

Como a escala de magnitude é invertida — valores maiores correspondem a objetos menos brilhantes — valores baixos de BP–RP indicam objetos mais azuis, geralmente mais jovens e massivos, enquanto valores altos indicam objetos mais vermelhos, tipicamente mais velhos e menos massivos.

As galáxias apresentam diferentes distribuições de cor, dependendo de suas populações estelares e de sua história evolutiva, sem um padrão único. No entanto, as estrelas do disco e do centro da Via Láctea tendem a apresentar distribuições de cor distintas das fontes extragalácticas, o que torna essa medida útil também como critério de seleção.

Análise dos quadrantes

Com base na discussão apresentada, são construídos histogramas de movimento próprio, tanto nas duas direções das coordenadas galácticas (pml e pmb) quanto no seu valor total (pm), e de cor para os objetos extragalácticos em cada quadrante galáctico, bem como para as possíveis fontes de contaminação

provenientes do disco da Via Láctea. Em todos os casos, são escolhidas apenas fontes com paralaxe negativa.

Esses histogramas são utilizados para definir valores de corte específicos para cada quadrante. A análise detalhada, assim como os histogramas correspondentes, encontra-se na dissertação de Pimentel (2024); apresenta-se aqui apenas um quadro-resumo com os cortes adotados em cada quadrante.

Quadro 1: Resumo das fontes utilizadas como referência para os objetos extragalácticos em cada quadrante, e também dos cortes adicionais de movimento próprio e cor aplicados a cada quadrante.

Quadrante	Fontes extragalácticas selecionadas	<i>pml</i> (mas/ano)	<i>pmb</i> (mas/ano)	BP-RP
1°	Galáxia anã de Draco (Draco dSph), Galáxia anã de Sagitário (Sgr dSph)	-3 < pml < 3	-1 < pmb < 1	BP-RP < 1.3
2°	Andrômeda, Ursa Menor (Ursa Min), M33	-1 < pml < 1	-1 < pmb < 1	Nenhum
3°	Galáxia anã de Fornax (Fornax dSph), Galáxia anã de Carina (Carina dSph), Leo-I, Leo-II	-1 < pml < 1	-1 < pmb < 1	BP-RP < 1.7
4°	Grande Nuvem de Magalhães (LMC), Pequena Nuvem de Magalhães (SMC), Bôötes-I	-1.5 < pml < 1.5	-2 < pmb < 2	BP-RP < 1.4

RESULTADOS

Aqui são apresentadas as visualizações de cada quadrante, com base nas filtragens descritas anteriormente, destacando os objetos identificados. Todas as imagens foram feitas com uma divisão HEALPix de 805.306.368 células e mostram mapas de densidade de fontes: as áreas mais claras indicam maior concentração de objetos, conforme o filtro aplicado.

Cada quadrante foi sobreposto ao mapa de densidade sem filtros da missão Gaia, permitindo comparar visualmente o efeito dos filtros. Em todas as imagens foi usada a projeção ortográfica, mostrando o mesmo quadrante visto a partir de dois hemisférios diferentes. Para identificar as fontes encontradas nas visualizações, utilizou-se o programa Aladin Sky Atlas.

Primeiro quadrante

Na Figura 3 vê-se uma nuvem escura na região central da galáxia, consistente com as faixas de poeira que existem na Via Láctea. Estas faixas são regiões de formação estelar, mas a poeira interestelar causa a extinção da luz de



frequências mais altas. Desta forma, as regiões mais poeirentas da galáxia tendem a ser vermelhas.

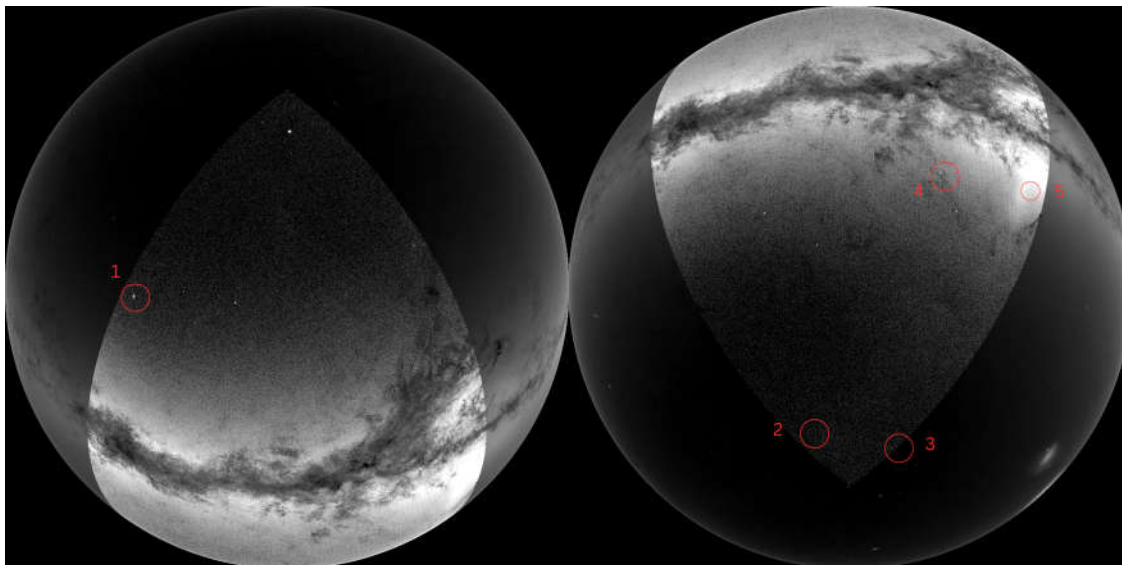


Figura 3: Visualizações anotadas do primeiro quadrante galáctico, centralizadas nos hemisférios norte e sul. Os objetos anotados são: 1 - Draco dSph, 2 - Galáxia Wolf-Lundmark-Melotte; 3 - NGC 7793; 4 - NGC 6822; 5 - Sgr dSph.

Como no corte realizado para este quadrante foram priorizadas as fontes azuis (BP-RP baixos), esta região de poeira foi majoritariamente excluída da imagem, dando origem à nuvem escura observada na visualização.

Quanto ao conteúdo extragaláctico encontrado para este quadrante, foi possível recuperar visualmente as duas galáxias usadas como referência para os parâmetros de corte (Draco dSph, Sgr dSph); bem como identificar três outras. Destas, uma é uma galáxia espiral (NGC 7793), e duas são galáxias irregulares (a galáxia Wolf-Lundmark-Melotte, WLM) e a galáxia NGC 6822. As duas últimas já foram estudadas anteriormente com dados do Gaia (Yang et al., 2022; Dmitrova et al., 2022).

Segundo quadrante

Das onze galáxias apontadas na Figura 4, sete não faziam parte da amostra utilizada para a decisão dos parâmetros da filtragem, apresentando uma diversidade de morfologias: M81, uma galáxia do tipo Seyfert-2; NGC 2403, que é um AGN; NGC 4228, uma galáxia *starburst*, com muita formação estelar; NGC 449, uma galáxia espiral de Magalhães (isto é, assemelha-se à Grande Nuvem de Magalhães, com uma estrutura de barra irregular); M101, a Galáxia do Catavento, que é uma galáxia espiral; e IC 1613, uma galáxia anã irregular.

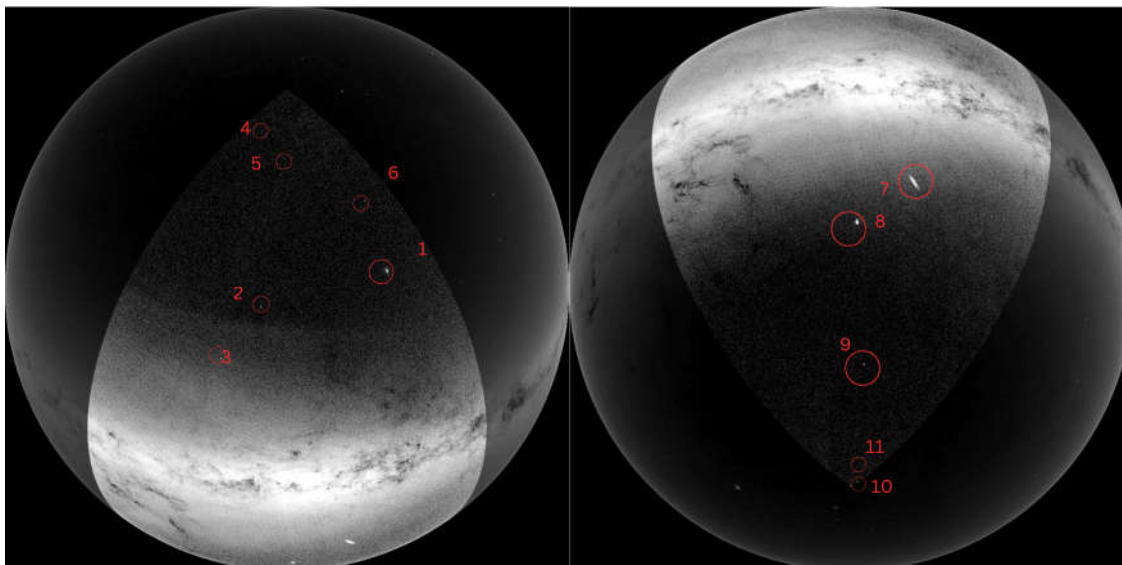


Figura 4: Visualizações anotadas do segundo quadrante galáctico, centralizadas nos hemisférios norte e sul. Os objetos anotados são: 1 - Ursa Minor dSph; 2 - M81; 3 - NGC 2403; 4 - NGC 4228; 5 - NGC 449; 6 - M101; 7 - Andrômeda; 8 - M33; 9 - IC 1613; 10 - NGC 253; 11 - NGC 247.

Destas, encontram-se estudos com dados do Gaia particularmente para M81 (Pan et al., 2022), NGC 2403 (Forbes et al., 2022) e IC 1613 (Taibi et al., 2024).

Utilizando o *software* Aladin, foi possível fazer uma visualização mais detalhada da galáxia de Andrômeda, onde foi possível inclusive distinguir melhor sua estrutura espiral.

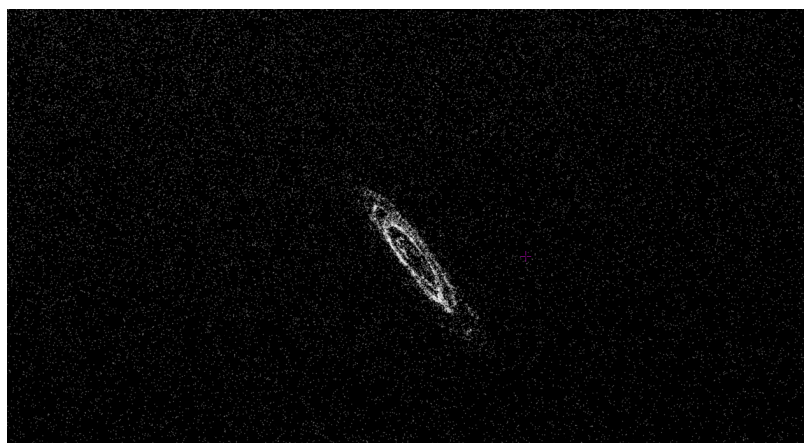


Figura 5: Mapa de densidade da região da galáxia de Andrômeda, visualizado utilizando o Aladin Sky Atlas.

Terceiro quadrante

Como explicado, o terceiro quadrante galáctico é uma região periférica da Via Láctea, menos populoso. Na Figura 6, vê-se uma estrutura de nuvens finas na

região central, e mais uma vez as galáxias são mais facilmente identificadas nas regiões polares.

Além das galáxias anãs utilizadas como amostra para a filtragem (Carina dSph, Fornax dSph, Leo-I e Leo-II), foram encontrados dois outros objetos: a galáxia anã do Sextante (Sextans dSph), que é uma galáxia satélite da Via Láctea; e NGC 3109, que é uma candidata a AGN. Ambas já foram estudadas anteriormente com dados da missão Gaia (Menzies et al., 2019; Cicúendez e Battaglia., 2018).

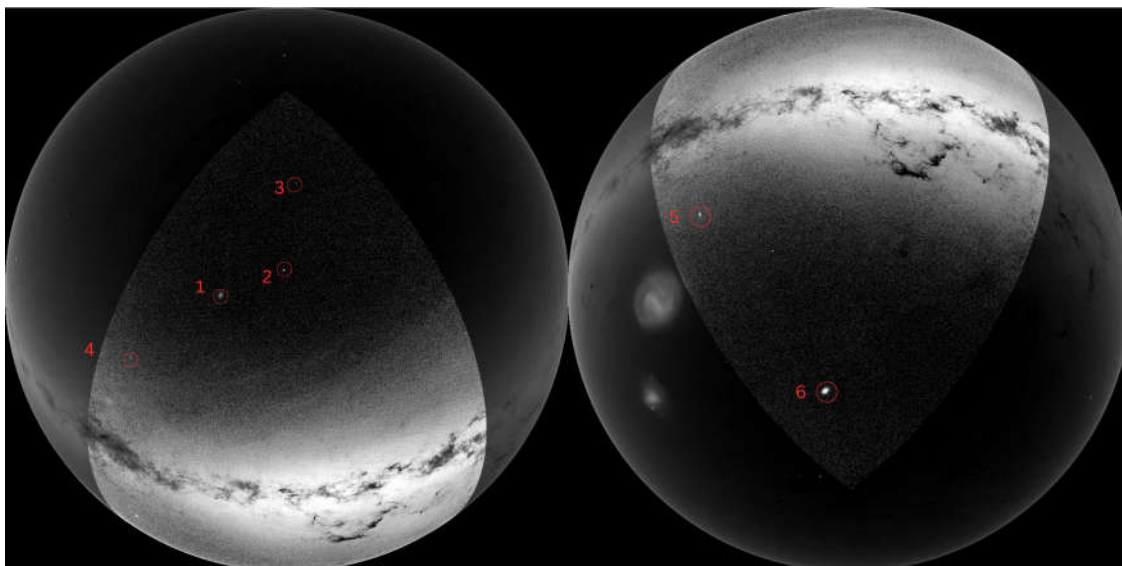


Figura 6: Visualizações anotadas do terceiro quadrante galáctico, centralizadas nos hemisférios norte e sul. Os objetos anotados são: 1 - Sextans dSph; 2 - Leo-I; 3 - Leo-II; 4 - NGC 3109; 5 - Carina dSph; 6 - Fornax dSph.

Quarto quadrante

Chegando no quarto quadrante, volta-se a observar a região central da galáxia, e vê-se na Figura 7 o reaparecimento das nuvens escuras no centro da galáxia, correspondentes às nuvens de poeira.

Além das nuvens de Magalhães, utilizadas como parâmetros na filtragem, foi possível encontrar outras cinco galáxias: a galáxia do Escultor (NGC 253), uma galáxia espiral; NGC 300 e NGC 55, galáxias do grupo do Escultor, sendo NGC 55 também uma galáxia irregular de Magalhães; e a galáxia anã irregular da Fênix (Phoenix dIrr). Destas, há estudos com dados do Gaia para NGC 253 (Lyu et al., 2023), e para Phoenix dIrr (Battaglia et al., 2022).

Nota-se que a galáxia Boötes-I, que era parte da amostra de fontes extragalácticas, não está presente na visualização resultante da filtragem. Isto pode ter sido causado por um viés na filtragem, uma vez que as Nuvens de Magalhães são galáxias maiores, com muito mais membros observáveis do que a galáxia de Boötes, de modo que elas influenciaram mais a filtragem.

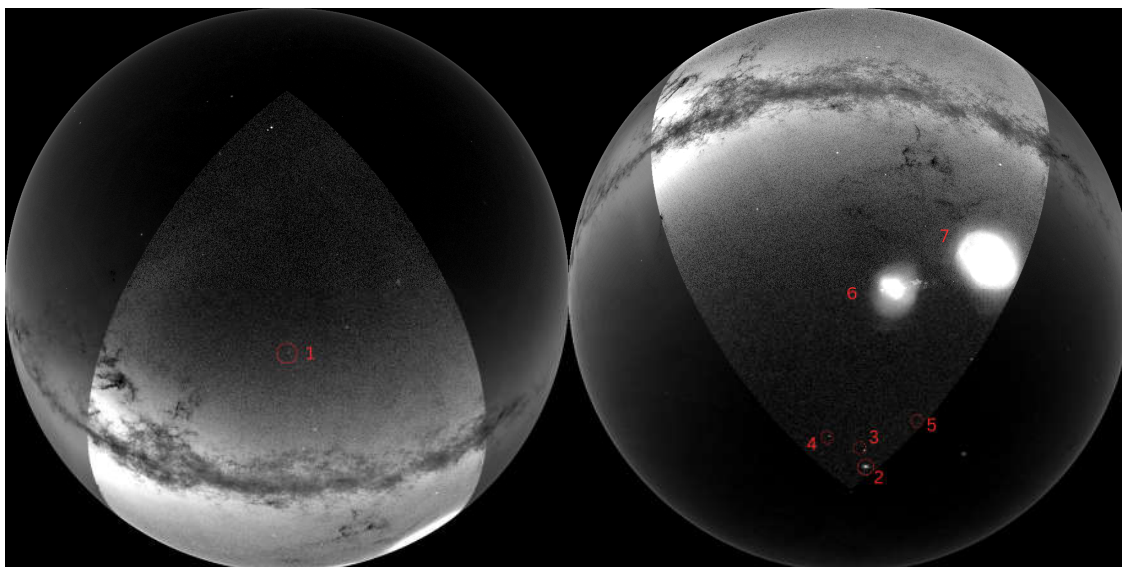


Figura 7: Visualizações anotadas do quarto quadrante galáctico, centralizadas nos hemisférios norte e sul. Os objetos anotados são: 1 - M83; 2 - NGC 253; 3 - NGC 300; 4 - NGC 55; 5 - Phoenix dlrr; 6 - Pequena Nuvem de Magalhães; 7 - Grande Nuvem de Magalhães.

As nuvens de Magalhães (Figura 8) são outro caso de estudo interessante, pois através da filtragem aplicada é possível distinguir melhor suas estruturas mais sutis, como a barra distorcida da Grande Nuvem, e a Ponte de Magalhães, uma ponte de gás que conecta as duas galáxias, provavelmente originária de uma interação entre as duas nuvens (Wang et al., 2022).

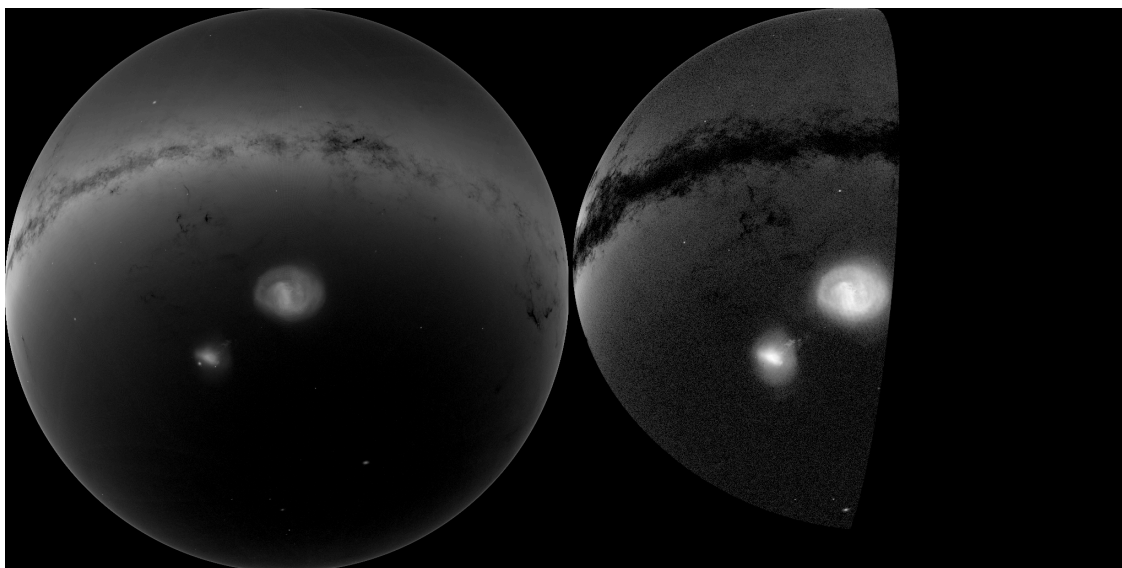


Figura 8: Comparação do mapa de densidade centralizado na Grande Nuvem de Magalhães. À esquerda, está a projeção ortogonal sem filtros. À direita está o resultado da filtragem e do recorte do quarto quadrante.

Conclusões

Com este projeto, foi possível produzir uma ferramenta robusta para a criação de visualizações da base de dados gerada pela missão Gaia, tendo sido especialmente construída para ser capaz de processar este grande volume de dados.

Dentro do contexto de uma Astronomia cada vez mais voltada para o ambiente de *Big Data*, é essencial que as ferramentas de trabalho dos astrônomos se adaptem para receber esta quantidade cada vez maior de dados, e permita que estes dados sejam facilmente manipulados, para que a extração de informações relevantes seja possível, colaborando para o processo da descoberta científica.

Ainda mais, mostrou-se que a missão Gaia tem um enorme potencial para estudos extragalácticos, devido à sensibilidade e qualidade dos seus instrumentos. Aliando estes dados de alto calibre ao código desenvolvido, foi possível criar visualizações que evidenciam fontes extragalácticas, incluindo galáxias que potencialmente ainda não foram estudadas utilizando dados da missão.

Em suma, este trabalho superou as dificuldades técnicas relacionadas ao processamento de grandes volumes de dados astronômicos, e com isto abriu um caminho para a exploração da base de dados da missão Gaia em todo o seu potencial. Outras implementações deste código poderão originar ainda mais visualizações inovadoras para a missão.

REFERÊNCIAS

- BAILER-JONES, Coryn AL. Estimating distances from parallaxes. **Publications of the Astronomical Society of the Pacific**, v. 127, n. 956, p. 994-1009, 2015.
- BAILER-JONES, Coryn AL et al. Gaia Data Release 3-The extragalactic content. **Astronomy & Astrophysics**, v. 674, p. A41, 2023.
- BATTAGLIA, Giuseppina et al. Gaia early DR3 systemic motions of Local Group dwarf galaxies and orbital properties with a massive Large Magellanic Cloud. **Astronomy & Astrophysics**, v. 657, p. A54, 2022.
- CICUENDEZ, Luis; BATTAGLIA, Giuseppina. Appearances can be deceiving: clear signs of accretion in the seemingly ordinary Sextans dSph. **Monthly Notices of the Royal Astronomical Society**, v. 480, n. 1, p. 251-260, 2018.
- DIMITROVA, Tzvetelina A. et al. Locating Red Supergiants in the Galaxy NGC 6822. **The Astronomical Journal**, v. 163, n. 2, p. 70, 2022.
- ELLIS, Geoffrey; DIX, Alan. A taxonomy of clutter reduction for information visualisation. **IEEE transactions on visualization and computer graphics**, v. 13, n. 6, p. 1216-1223, 2007.
- FORBES, Duncan A. et al. Low-metallicity globular clusters in the low-mass isolated spiral galaxy NGC 2403. **Monthly Notices of the Royal Astronomical Society**, v. 512, n. 1, p. 802-810, 2022.



GAIA COLLABORATION et al. The Gaia mission. **arXiv preprint arXiv:1609.04153**, 2016.

GORSKI, Krzysztof M. et al. HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. **The Astrophysical Journal**, v. 622, n. 2, p. 759, 2005.

HAMMER, F. et al. The Magellanic stream system. I. Ram-pressure tails and the relics of the collision between the Magellanic clouds. **The Astrophysical Journal**, v. 813, n. 2, p. 110, 2015.

HANSEN, Charles et al. Visualization for data-intensive science. In: TANSLEY, Stewart et al. **The fourth paradigm**: data-intensive scientific discovery. Redmond, WA: Microsoft Research, 2009. p. 153–163.

HASSAN, Amr H. et al. Tera-scale astronomical data analysis and visualization. **Monthly Notices of the Royal Astronomical Society**, v. 429, n. 3, p. 2442-2455, 2013.

HELMI, Amina et al. Gaia Data Release 2-Kinematics of globular clusters and dwarf galaxies around the Milky Way. **Astronomy & astrophysics**, v. 616, p. A12, 2018.

JIMÉNEZ-ARRANZ, Óscar et al. Application of a neural network classifier for the generation of clean Small Magellanic Cloud stellar samples. **Astronomy & Astrophysics**, v. 672, p. A65, 2023a.

JIMÉNEZ-ARRANZ, Óscar et al. Kinematic analysis of the Large Magellanic Cloud using Gaia DR3. **Astronomy & Astrophysics**, v. 669, p. A91, 2023b.

KRONE-MARTINS, A. et al. Pushing the limits of the Gaia space mission by analyzing galaxy morphology. **Astronomy & Astrophysics**, v. 556, p. A102, 2013.

LURI, Xavier et al. Gaia data release 2-using gaia parallaxes. **Astronomy & Astrophysics**, v. 616, p. A9, 2018.

LYU, Xuanyi et al. On the origin of the anomalous gas, non-declining rotation curve, and disc asymmetries in NGC 253. **Monthly Notices of the Royal Astronomical Society**, v. 524, n. 1, p. 1169-1190, 2023.

MENZIES, John W. et al. Luminous AGB variables in the dwarf irregular galaxy, NGC 3109. **Monthly Notices of the Royal Astronomical Society**, v. 483, n. 4, p. 5150-5165, 2019.

MOITINHO, André et al. Gaia data release 1-the archive visualisation service. **Astronomy & Astrophysics**, v. 605, p. A52, 2017.

PAN, Jiaming et al. New globular cluster candidates in the M81 group. **Monthly Notices of the Royal Astronomical Society**, v. 515, n. 1, p. 48-70, 2022.

PIMENTEL, Maria Eduarda Manfrinato. **Galactic and Extragalactic Insights: A Visualisation Approach to Gaia Data Exploitation**. Orientador: André Moitinho de Almeida. Dissertação (Mestrado em Física) - Faculdade de Ciências, Universidade de Lisboa, Lisboa, 2024.



TAIBI, S. et al. From gas to stars: MUSEings on the internal evolution of IC 1613. **Astronomy & Astrophysics**, v. 689, p. A88, 2024.

WANG, Jianling; HAMMER, Francois; YANG, Yanbin. Lessons from the Magellanic System and its modeling. **Monthly Notices of the Royal Astronomical Society**, v. 515, n. 1, p. 940-952, 2022.

WICKHAM, Hadley. The split-apply-combine strategy for data analysis. **Journal of statistical software**, v. 40, p. 1-29, 2011.

YANG, Yanbin et al. Evidence of ram-pressure stripping of WLM, a dwarf galaxy far away from any large host galaxy. **Astronomy & Astrophysics**, v. 660, p. L11, 2022.

Como citar este artigo (ABNT):

PIMENTEL, M. E. M. Uma abordagem visual para a exploração dos dados extragalácticos da missão Gaia. **Informativo do Observatório Didático de Astronomia (IODA)**, v. 6, n.1, p. 44-58, jun 2026.

