P2 - e-Science em Astronomia

Kethelin Parra Ramos - 9898349

QUESTÕES

1. O aglomerado de galáxias de Coma (*Coma Cluster*) é provavelmente o aglomerado mais popular e mais estudado. Ele foi identificado inicialmente por Willian Herschel um século antes da verdadeira natureza das galáxias ser compreendida.

(I) Reproduza a Fig. 1 (enunciado) do aglomerado de Coma usando três bandas no óptico/infra-vermelho próximo.

Resposta:

(I) Para criar a imagem RGB, fiz *download* das imagens *fits* do aglomerado nas bandas g, r e i do SDSS (óptico). No caso dos contornos de raios-X, usei "PSPC 0.6 Deg" do ROSAT. Estes dados foram retirados do *SkyView*.

- Imagem RGB:
  Criando um *frame* RGB no DS9, adicionando uma imagem por vez de cada banda (banda g para o B, r para o G e i para o R). A imagem final "colorida" foi ajustada em escala logarítmica (`log`) e cada banda teve seu contraste (em *scale parameters*) editado conforme a necessidade.

- Contornos de raios-X:
  Um novo *frame* foi criado no DS9 e então abri a imagem do ROSAT, coloquei na escala `sqrt` e acertei os limites do contraste. Clicando em *Analysis*→*Contour Parameters* para gerar os contornos, após ajustar a quantidade de níveis e a suavização do mesmo, o arquivo dos contornos foi salvo (.ctr) e inserido na imagem RGB.

- Grades das coordenadas:
  As grades foram adicionadas em *Analysis*→*Coordinate Grid Parameters*. Nesta mesma janela foi possível editar as cores, linha, tamanhos, sistema de coordenadas (FK5) e etc.

- Marcação das galáxias do catálogo NGC:

  Para isto pesquisei por "NGC" em *Analysis*→*Catalogs*→*Search for Catalogs* e selecionei os dados mais recentes (Corwin, 2004). Na janela que abriu editei o tamanho/cor das *regions*.

- Barra de escala:

  Foi adicionado em *Regions*→*Shape*→*Line*. Dando dois cliques na linha, abriu uma janela onde foi possível editar a cor/tamanho e adicionar texto.

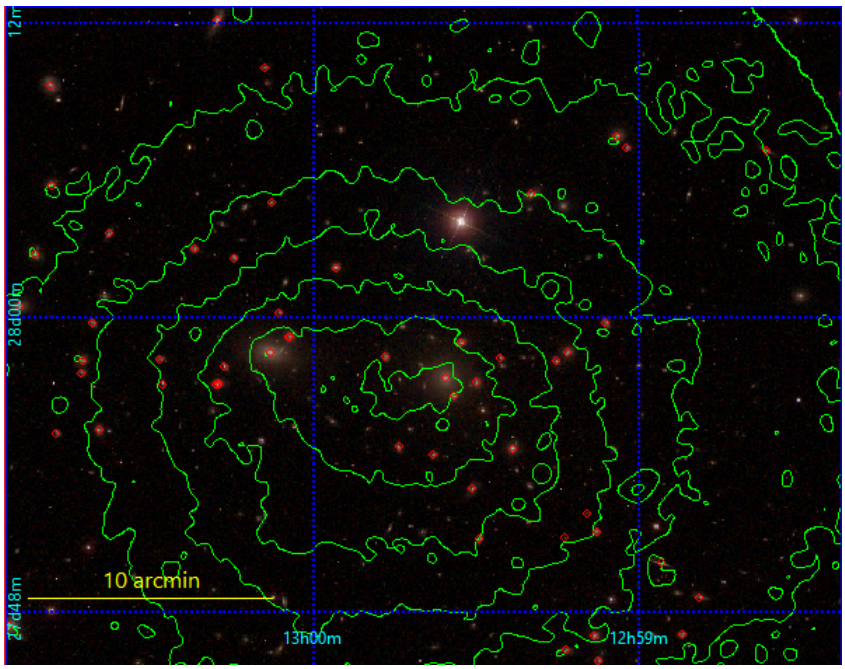

Figura 1: Imagem RGB e contornos de raios-X de Abell 1656.

2. **(I)** Faça uma seleção de objetos usando os seguintes critérios:

- Ser uma par de galáxias do catálogo H. Arp (1966);
- Redshift no intervalo $0.01 \leqslant z \leqslant 0.03$;
- Coordenadas: $120° \leqslant \ell \leqslant 240°$ e $b \geqslant 45°$ (ℓ e b são coordenadas galácticas).

(II) Faça um mosaico destes objetos usando figuras JPG do SDSS. Consulte a seguinte página, `http://skyserver.sdss.org/dr12/en/help/docs/api.aspximgcutout` para ver como é possível obter uma "*Image Cutout*" no formato JPG. O comando `wget` é útil para usar em um script e obter automaticamente um conjunto de imagens.

Resposta:

(I) - Baixei a tabela "*Atlas of Peculiar Galaxies (Arp 1966)*" diretamente no site do Vizier, adicionando uma coluna extra com coordenadas galácticas.

- No TOPCAT criei um *subset* para filtrar os dados de acordo com a região solicitada em coordenadas galácticas.

- Primeiramente, tentei fazer *cross-match* destes dados com o SIMBAD (para obter os *redshifts* e classificações) mas no final sobraria apenas 9 objetos na tabela. Por causa disso, baixei a tabela desse catálogo no NED e então fiz o *cross-match*.

- Criei um *subset* para filtrar os *redshifts*. Esta tabela final possui 24 objetos (arquivo Q2_P2.csv). Não foi possível filtrar as classificações (a maioria constava "G" e não "GPair"), para o item **(II)** a seleção dos objetos corretos foi "a olho".

(II) Escrevi o programa abaixo (em *Python*) para coletar as imagens dos objetos selecionados no item anterior. Para fazer o mosaico faltaram 2 galáxias (da imagem exemplo), então coloquei outras duas da minha amostra no lugar.

```
# Bibliotecas usadas
import numpy as np
from google.colab import files
#Leitura da tabela
dados = np.loadtxt("Questao2.csv", skiprows=1,delimiter=",")
RA=dados[:,0]
DEC=dados[:,1]
#Download das imagens com wget
for i in range(len(RA)):
  url_gal="'http://skyserver.sdss.org/dr12/SkyserverWS/ImgCutout/getjpeg?ra="
+str(RA[i])+"&dec="+str(DEC[i])+"&width=800&height=500'"
  nome="galaxia_"+str(i)+".jpeg"
  !wget -P /content/ -O $nome -A jpeg $url_gal
#Para não ter que baixar um arquivo por vez no google colab
for i in range(len(RA)):
  files.download('galaxia_'+str(i)+'.jpeg')
```

Figura 2: Mosaico dos pares de galáxias do catálogo Arp (1966).

3. O arquivo "`clusterdata2020.csv`" contém uma lista de de coordenadas (X e Y, unidades arbitrárias). Determine quantas subestruturas e em qual subestrutura cada ponto pertence. Mostre o seu resultado na forma de um gráfico, onde cada subestrutura tem uma cor. Quais são as coordenadas centrais (X_c,Y_c) de cada subestrutura? Descreva detalhadamente como você procedeu e qual algoritmo/biblioteca usou.

Resposta:

Utilizei o método de agrupamento "*K-means clustering*" para a identificação de subestruturas, que foi implementado utilizando a função *KMeans()* da biblioteca *Scikit-learn*. A distribuição de dados sem as subestruturas identificadas se encontra na Fig. 3.

Este algoritmo exige que o número de agrupamentos seja pré-definido, então foi testado $N = \{3, 4, 5, 6\}$ *clusters*. Por causa disto, a quantidade "real" de subestruturas depende da minha análise visual dos resultados de cada simulação.

Na Fig. 4 temos os resultados para cada N testado, portanto, o caso que me parece ser o mais razoável é o $N = 3$ subestruturas. As coordenadas centrais (X_c,Y_c) de cada subestrutura se encontram na Tab. I.

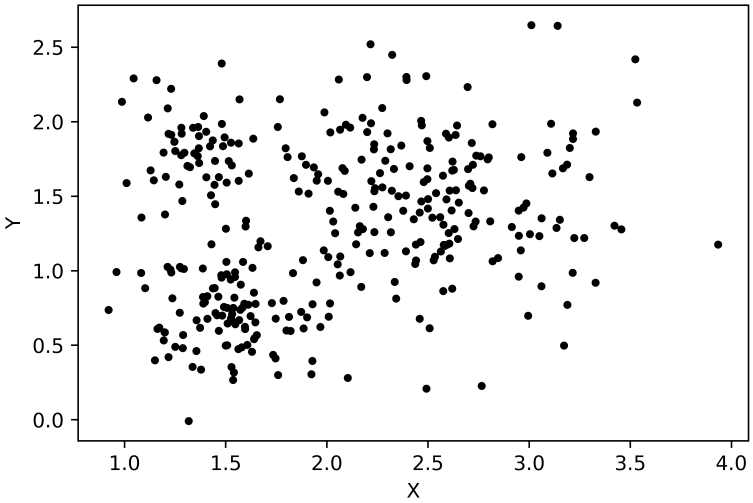

Figura 3: Distribuição das coordenadas (X,Y) do arquivo sem as subestruturas identificadas.

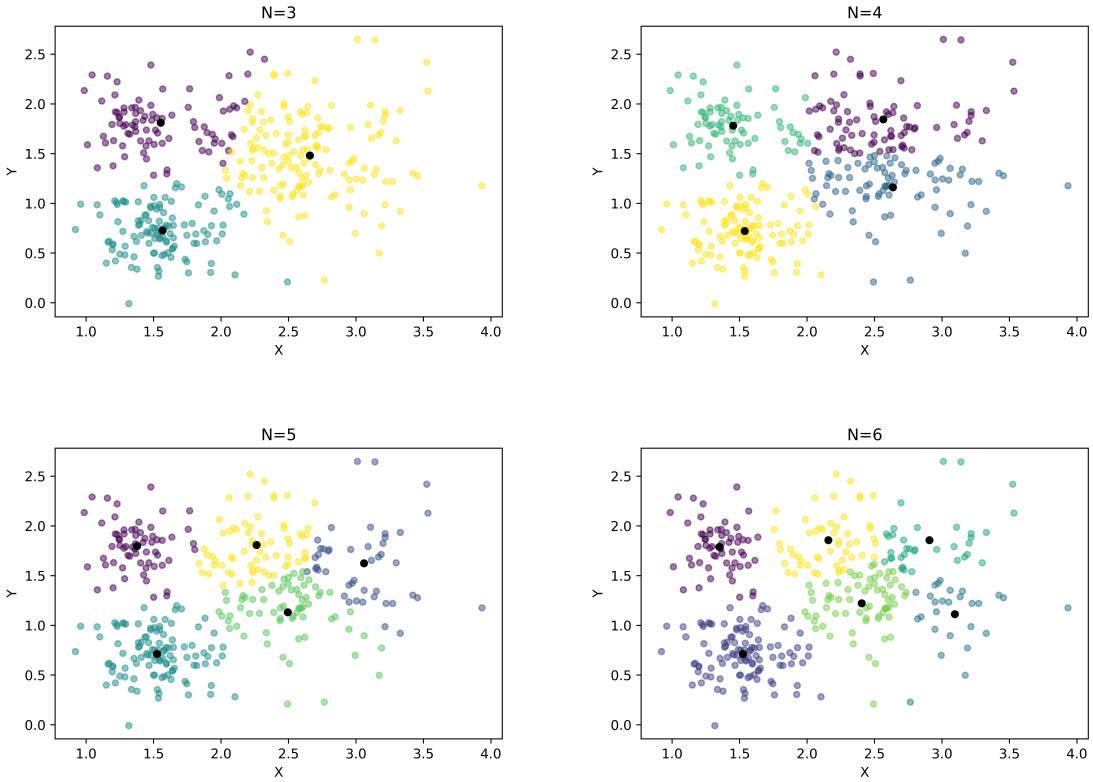

Figura 4: Identificação de subestruturas usando o método "*K-means clustering*" para $N = \{3, 4, 5, 6\}$ *clusters*. Os pontos pretos representam as coordenadas centrais de cada subestrutura.

Tabela I: Coordenadas centrais (X_c,Y_c) de cada subestrutura para $N = \{3, 4, 5, 6\}$.

$N_{clusters}$	$\mathbf{X}_c$	$\mathbf{Y}_c$	$N_{clusters}$	$\mathbf{X}_c$	$\mathbf{Y}_c$
3	1.55243846	1.81051742	4	2.56483389	1.84468006
	1.56528159	0.72768475		2.63619839	1.1609164
	2.65800134	1.48095154		1.45222521	1.78103933
				1.53786219	0.722391
5	1.37776777	1.79761075	6	1.35041466	1.78893917
	3.05999285	1.62408444		1.52464596	0.71293047
	1.52464596	0.71293047		3.09548266	1.11286503
	2.4941457	1.13211456		2.90696949	1.85620121
	2.262869	1.80795386		2.40453372	1.22180545
				2.15731314	1.85682871

Abaixo se encontra parte do código utilizado (apenas pra N=3):

```
# Bibliotecas usadas
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

#Leitura dos dados
dados = np.loadtxt("clusterdata2020.csv", skiprows=0,delimiter=",")
X=dados[:,0]
Y=dados[:,1]

#Gráfico para visualizar os dados sem a classificação
plt.figure()
plt.xlabel("X")
plt.ylabel("Y")
plt.plot(X,Y, 'k.')
plt.savefig("grafico_q3.png", dpi=900, quality = 85)
```

```
plt.show()

#Identificação de subestruturas
Ncluster=3  #Número de clusters
kmeans = KMeans(n_clusters=Ncluster).fit(dados) #Algoritmo de classificação
                                                #K-Means Clustering
cood_central = kmeans.cluster_centers_ #Dá os pontos centrais de cada cluster
print("- Coordenadas centrais de cada subestrutura:")
print(cood_central)

#Gráfico com as subestruturas
plt.scatter(X, Y, c= kmeans.labels_.astype(float), s=20, alpha=0.5)
plt.scatter(cood_central[:, 0], cood_central[:, 1], c='black', s=25)
plt.xlabel("X")
plt.ylabel("Y")
plt.title("N=3")
plt.savefig("Q3_KmeansN3.png", dpi=900, quality = 85)
plt.show()
```

4. Baseado em um problema do GAVO, enunciado copiado *ipsis litteris*:

Our Galaxy is rotating. You don't have to take our word for it - ask Gaia. If you average the proper motions in the TGAS catalog over areas of 5 to 10 degrees in linear size and **plot the arrows**, you'll see a very distinctive pattern.

Since the Sun is on a largely circular orbit around the centre of the Galaxy, it should move perpendicular to the connecting line to the Galaxy's center. Hence, you'd expect that the "source"of the pattern (which is where we're moving) to be on the Galactic equator at Galactic longitude 90 degrees. Where do you see it (give both equatorial and Galactic coordinates)?

Resposta:

As coordenadas da região para onde o Sol estaria se movimentando parecem ser:

- Coordenadas Equatoriais: $\alpha = +90.0°$ $\delta = -40.0°$

- Coordenadas Galácticas: $\ell = +225.9°$ $b = -20.0°$

As setas parecem estar "entrando" no gráfico (ponto vermelho da Fig. 5) nesta região.

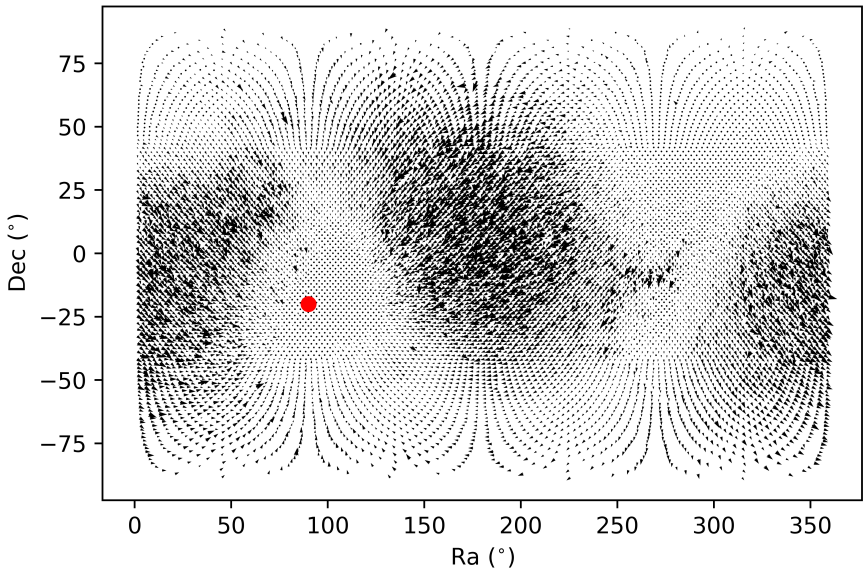

Figura 5: Gráfico da distribuição espacial do movimento próprio dos dados coletados do TGAS. O ponto vermelho representa a direção para onde o Sol está indo.

O código usado na atividade:

```
# Bibliotecas usadas
import pyvo
import matplotlib.pyplot as plt
from astropy.coordinates import SkyCoord
import astropy.units as u

#Acessando os dados
access_url = "http://gaia.ari.uni-heidelberg.de/tap" #URL do tap do gaia
service = pyvo.dal.TAPService(access_url) #Acesso a url

#Query utilizada no link (acessando os dados do TGAS)
result = service.run_sync(
"""SELECT gaia_healpix_index(5, source_id) AS healpix_5,avg(tgas_source.ra)
```

```
AS ra,avg(tgas_source.dec) AS dec,avg(tgas_source.pmra) AS pmra,
avg(tgas_source.pmdec) AS pmdec
FROM gaiadr1.tgas_source
GROUP BY healpix_5 """)

#Dados
RA = result["ra"]
DEC = result["dec"]
pmra=result["pmra"]
pmdec=result["pmdec"]

#Gráfico
fig, ax = plt.subplots()
q = ax.quiver(RA,DEC,pmra,pmdec) #cria os vetores
plt.plot(90.,-20.,'ro')  #ponto vermelho
plt.xlabel("Ra ($^{\circ}$)")
plt.ylabel("Dec ($^{\circ}$)")
plt.savefig("Q4_propermotion.png", dpi=900, quality = 85)
plt.show()

#Conversão de coordenadas (do ponto vermelho)
cood=SkyCoord(90.*u.deg, -20.*u.deg)
cood2=cood.galactic
print(cood2)
```