A new method for unveiling Open Clusters in Gaia:

New nearby Open Clusters confirmed by DR2

Alfred Castro-Ginard

May, 29 2018

Motivation

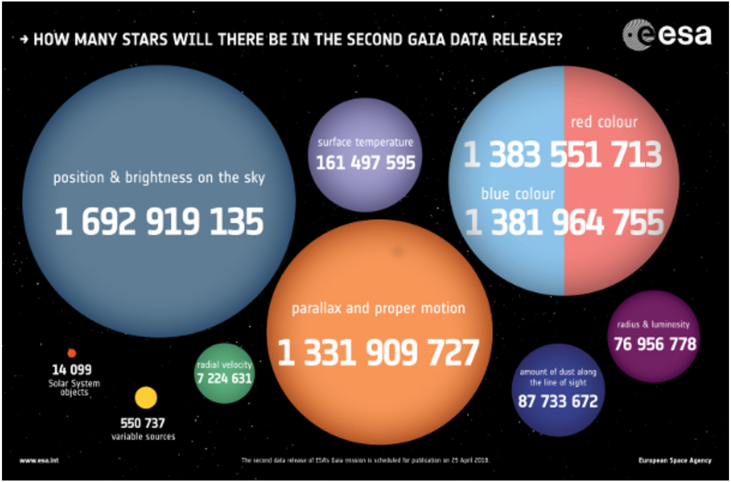
Why Open Clusters?

Open Clusters are fundamental building blocks of galaxies, and key objects for several astrophysical aspects:

- very young OCs are informative of the star formation mechanism
- young OCs trace the star forming regions
- itermediate and old OCs allow studying the chemical enrichment of the galactic disk
- stellar structure and evolution
- more...

Motivation

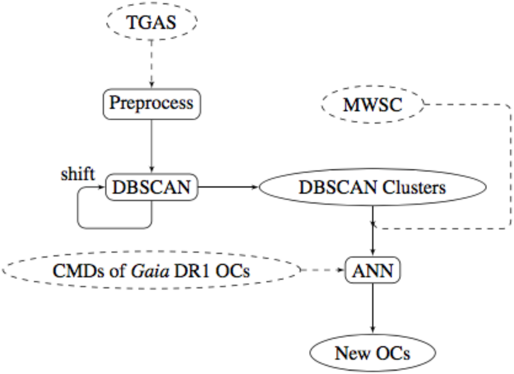
Why data mining / machine learning?



Method

Initially developed using astrometric data from TGAS and photometric data from 2MASS, aiming for its application to *Gaia* DR2

- Overdensities in astrometrical parameters $(l, b, \varpi, \mu_{\alpha^*}, \mu_\delta)$
- Located in the Galactic disk $|b| < 20$
- Not close to MWSC catalogue OCs
- CMDs similar to OCs from DR1

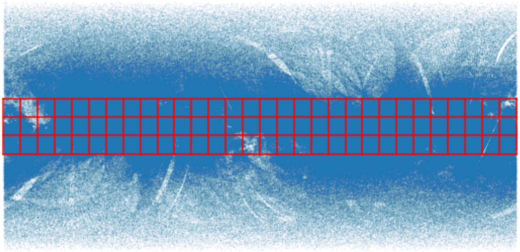


Preprocessing

Consider only the Galactic disk $|b| > 20$. Around 95% of the clusters catalogued in Dias[Dias+ 2002] and MWSC[Kharchenko+ 2013] are in this region.

Reject stars with $|\mu_{\alpha*}|, |\mu_\delta| > 30$mas·yr^{-1} and 0mas$< \varpi <$ 7mas.

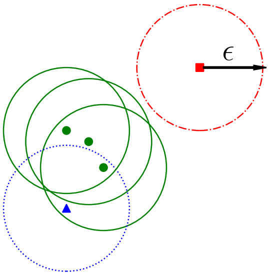
Divide the area of study in rectangles of size L deg.



DBSCAN

Use a density based unsupervised algorithm to search for overdensities in the parameter space.

- No *a priori* knowledge of the number of clusters
- Finds arbitrary shaped clusters
- Need to define two parameters (ϵ, *minPts*)



Statistical distance between two stars:

$$d(i,j) = \sqrt{(l_i - l_j)^2 + (b_i - b_j)^2 + (\varpi_i - \varpi_j)^2 + (\mu_{\alpha^*,i} - \mu_{\alpha^*,j})^2 + (\mu_{\delta,i} - \mu_{\delta,j})^2}$$

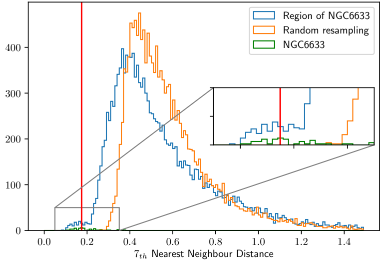
DBSCAN

Determination of $(\epsilon, minPts)$

Leave *minPts* to be optimized using simulated data.

For the determination of ϵ:

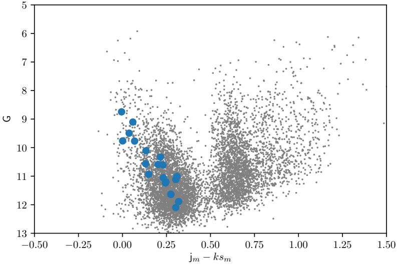
- Distance between the k_{th} nearest neighbours in a cluster should be smaller than distance between stars belonging to the field
- Compute ϵ as: $\epsilon = (\epsilon_{kNN} + \epsilon_{rand})/2$

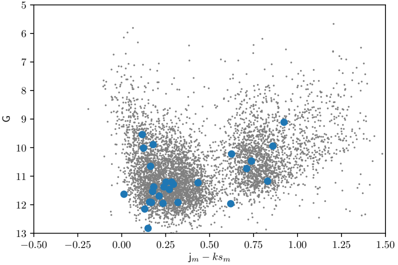


Artificial Neural Network

To distinguish between statistical clusters and Open Clusters

How to distinguish between all the clusters, real OCs or not, from DBSCAN?





Treat it as a pattern recognition problem to automatically classify in real OCs and statistical clusters, decide if stars are following an isochrone in a Color Magnitude Diagram.

Results

Validation of the candidates

Each of the 32 OCs candidates detected with TGAS were validated using *Gaia* DR2:

- Focus on a cone search of 2 deg centred in the centre of the candidate.
- Re-run the algorithm to re-detect the OC, now with more members due to the increase in the limiting magnitude ($G < 17$).
- Check if the mean parameters of the found OC are compatible with the previous ones.
- Check if the member stars are following an isochrone in the CMD, using *Gaia* photometry.

Results

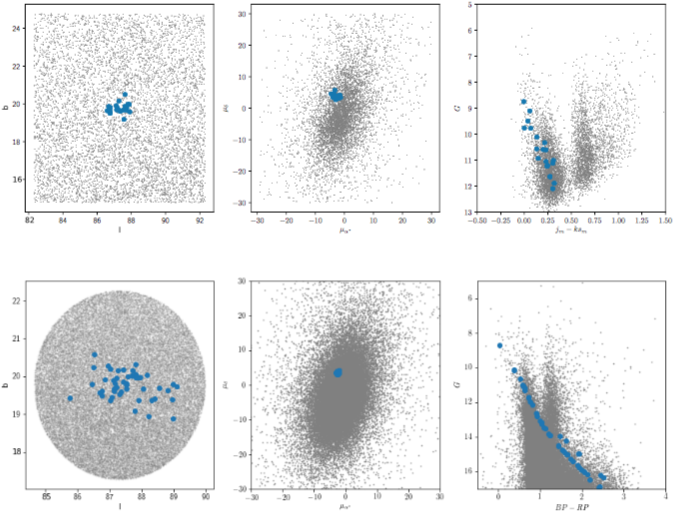
Proposed new OCs

After the validation of the TGAS candidates with *Gaia* DR2 data, the proposed new OCs are:

Name	α [deg]	δ [deg]	l [deg]	b [deg]	ϖ [mas]	μ_{α^*} [$\mathrm{mas \cdot yr^{-1}}$]	μ_δ [$\mathrm{mas \cdot yr^{-1}}$]	V_{rad} [$\mathrm{km \cdot s^{-1}}$]	N ($N_{V_{\mathrm{rad}}}$)
UBC1	288.00 (0.84)	56.83 (0.63)	87.55 (0.74)	19.76 (0.35)	3.05 (0.02)	−2.49 (0.25)	3.69 (0.24)	−21.46 (2.36)	47 (14)
UBC2	5.80 (0.84)	46.59 (0.34)	117.89 (0.62)	−15.99 (0.32)	1.74 (0.03)	−6.34 (0.12)	−5.03 (0.13)	−9.73 (2.22)	23 (4)
UBC3	283.77 (0.16)	12.34 (0.22)	44.35 (0.24)	4.79 (0.12)	0.58 (0.04)	−0.60 (0.08)	−1.36 (0.09)	−7.25 (13.54)	29 (2)
UBC4	60.96 (1.07)	35.35 (0.74)	161.42 (1.05)	−12.75 (0.50)	1.64 (0.05)	−0.75 (0.13)	−5.72 (0.13)	3.67 (1.65)	44 (3)
UBC5	238.42 (0.74)	−47.72 (0.41)	331.74 (0.56)	4.68 (0.32)	1.78 (0.01)	−6.69 (0.15)	−4.18 (0.09)	−14.91 (−)	29 (1)
UBC6	343.95 (0.48)	51.19 (0.19)	105.13 (0.29)	−7.63 (0.21)	0.67 (0.01)	−4.64 (0.06)	−4.90 (0.08)	−31.64 (1.51)	76 (3)
UBC7	106.92 (0.61)	−37.74 (0.65)	248.80 (0.71)	−13.25 (0.42)	3.56 (0.05)	−9.74 (0.19)	6.99 (0.20)	16.42 (4.71)	77 (21)
UBC8	84.36 (0.86)	57.16 (0.54)	154.83 (0.64)	13.30 (0.36)	2.05 (0.03)	−3.14 (0.17)	−3.99 (0.16)	−5.96 (3.94)	103 (21)
UBC9	276.64 (0.41)	26.40 (0.39)	54.48 (0.40)	16.80 (0.38)	2.87 (0.02)	0.60 (0.16)	−5.35 (0.18)	−17.98 (3.12)	25 (6)
UBC10a	324.46 (1.36)	61.75 (0.95)	102.03 (1.02)	7.02 (0.55)	1.07 (0.01)	−2.14 (0.11)	−3.03 (0.12)	−23.12 (−)	43 (1)
UBC10b	326.87 (0.96)	61.10 (0.47)	102.49 (0.36)	5.75 (0.55)	1.01 (0.01)	−3.46 (0.09)	−1.86 (0.10)	−46.90 (−)	40 (1)
UBC11	246.16 (1.91)	−59.94 (0.87)	326.81 (1.15)	−7.39 (0.61)	2.13 (0.04)	−0.30 (0.37)	−6.78 (0.28)	−18.18 (5.35)	44 (4)
UBC12	126.13 (0.65)	−8.56 (0.47)	231.81 (0.71)	16.24 (0.41)	2.21 (0.05)	−8.27 (0.20)	4.07 (0.28)	31.34 (−)	19 (1)
UBC13	120.90 (0.79)	3.60 (1.14)	218.04 (1.02)	17.68 (0.99)	1.60 (0.04)	−7.76 (0.19)	−1.16 (0.21)	22.91 (5.48)	36 (6)
UBC14	294.80 (0.58)	3.64 (1.01)	41.70 (1.06)	−8.91 (0.52)	1.30 (0.02)	0.14 (0.16)	−2.09 (0.20)	−9.85 (−)	46 (1)
UBC17a	83.38 (0.22)	−1.58 (0.86)	205.23 (1.04)	−18.01 (1.06)	2.74 (0.04)	1.59 (0.27)	−1.20 (0.35)	18.96 (7.64)	180 (18)
UBC17b	83.35 (0.76)	−1.54 (0.94)	205.18 (0.95)	−18.02 (0.79)	2.36 (0.04)	0.05 (0.17)	−0.16 (0.24)	33.19 (4.41)	103 (4)
UBC19	56.48 (0.37)	29.91 (0.22)	162.25 (0.24)	−19.32 (0.32)	2.39 (0.11)	2.71 (0.53)	−5.19 (0.27)	31.38 (3.46)	34 (2)
UBC21	130.35 (0.81)	−20.68 (0.94)	244.56 (1.10)	12.87 (0.55)	1.12 (0.02)	−6.51 (0.22)	2.48 (0.17)	− (−)	47 (0)
UBC26	285.24 (0.69)	21.92 (0.74)	53.61 (0.86)	7.80 (0.49)	1.66 (0.03)	2.01 (0.17)	−5.18 (0.21)	6.79 (17.43)	64 (2)
UBC27	294.31 (0.25)	15.58 (0.25)	52.00 (0.24)	−2.73 (0.25)	0.88 (0.03)	−0.82 (0.07)	−6.22 (0.08)	− (−)	65 (0)
UBC31	61.11 (1.21)	32.76 (1.13)	163.33 (1.04)	−14.55 (1.14)	2.70 (0.07)	3.77 (0.22)	−5.43 (0.24)	22.74 (5.73)	84 (12)
UBC32	279.43 (0.66)	−14.04 (0.93)	18.87 (0.96)	−3.38 (0.60)	3.56 (0.04)	−1.75 (0.26)	−9.26 (0.29)	−21.58 (7.24)	60 (14)

Results

UBC1



Results

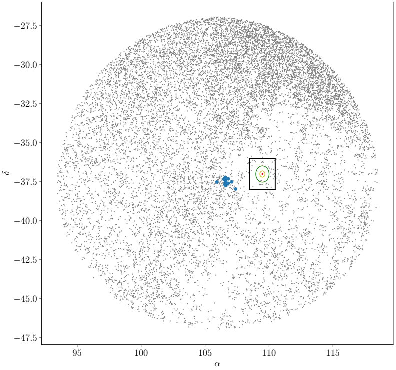
UBC7 vs Collinder135

Cone search of 10 deg centred in UBC7 with photometric observations > 120.

UBC7 is found in the vicinity of Collinder135, separated enough to be considered as a different cluster.

The common proper motions with Collinder135 suggests that UBC7 is probably a part of it.

More spurious clusters may be detected due to the scanning law of *Gaia*.



Conclusions and future work

- After the method is applied to TGAS, around 70% of the candidates were confirmed with *Gaia* DR2.
- Explore how the definition of $d(i,j)$ and the determination of ϵ adapts to the different OCs.
- Need to build a wider and more realistic training set for the ANN to use with *Gaia* DR2.
- The method is devised to work with large datasets such as *Gaia* DR2, but its implementation into a big data enviroment is work in progress.

More on...

Astronomy & Astrophysics manuscript no. main
May 8, 2018

A new method for unveiling Open Clusters in *Gaia*:

new nearby Open Clusters confirmed by DR2

A. Castro-Ginard[1], C. Jordi[1], X. Luri[1], F. Julbe[1], M. Morvan[1,2], L. Balaguer-Núñez[1], and T. Cantat-Gaudin[1]

[1] Dept. Fisica Quantica i Astrofisica, Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí Franquès 1, E08028 Barcelona, Spain
e-mail: acastro@fqa.ub.edu
[2] Mines Saint-Etienne, Institut Henri Fayol, F - 42023 Saint-Etienne France

Received date / Accepted date

ABSTRACT

Context. The publication of the *Gaia* Data Release 2 (*Gaia* DR2) opens a new era in Astronomy. It includes precise astrometric data (positions, proper motions and parallaxes) for more than 1.3 billion sources, mostly stars. To analyse such a vast amount of new data, the use of data mining techniques and machine learning algorithms are mandatory.
Aims. The search for Open Clusters, groups of stars that were born and move together, located in the disk, is a great example for the application of these techniques. Our aim is to develop a method to automatically explore the data space, requiring minimal manual intervention.
Methods. We explore the performance of a density based clustering algorithm, DBSCAN, to find clusters in the data together with a supervised learning method such as an Artificial Neural Network (ANN) to automatically distinguish between real Open Clusters and statistical clusters.