Astronomical Data Analysis Software and Systems XXVII
ASP Conference Series, Vol. 522
Pascal Ballester, Jorge Ibsen, Mauricio Solar, and Keith Shortridge, eds.
©2019 Astronomical Society of the Pacific

Using the PLC and Modern Algorithms to Detect Clustering and Associations in Nearby Galaxies

C. Feinstein,[1,2] G. Baume,[1,2] M.J. Rodríguez,[1] and M.M. Vergne,[1,2]

[1] *Instituto de Astrofísica,(IALP), Conicet, Argentina;*
`cfeinstein@fcaglp.unlp.edu.ar`

[2] *Facultad de Ciencias Astronómicas y Geofísicas, UNLP, La Plata, Argentina*

Abstract. The Hubble Space Telescope (HST) is not affected by the atmospheric seeing, bringing observations that have an amazing spatial resolution that allows us to isolate the individual stars in nearby galaxies. In consequence, to have good photometry of large samples of stars in these objects. Therefore, high accuracy studies of extragalactic stellar associations and clusters could be done. One of most powerful algorithm for detecting clustering in a large amount of data is the Path Linkage Criterion (PLC), Battinelli (1991). We show in this work the results of our implementation of a high speed version of the PLC that was applied to HST data of two galaxies: NGC 300 and NGC 253. Also, we show the results obtained with PLC and others popular methods found in the literature of clustering, applied to the real data and to simulated data.

1. Observational Data

The data are from the Camera for Surveys of the Hubble Space Telescope (ACS/HST). Images and photometric data were taken from the archive of The Space Telescope Science Institute (STScI: MAST [1]).

These data were taken with the Wide Resolution Channel (WFC) of the ACS. This camera has two CCD detectors and the instrumental configuration give an amazing scale of 0".049 pixel^{-1} on an observed field of 3'.3 x 3'.3

These images cover part of NGC 300 and NGC 253 with the following characteristics: NGC 300 (1 field); exp. time = 360 seg Filters: F435W, F555W y F814W; Cycle 11; PI: Bresolin, Program: 9492, and

NGC 253 (5 fields) exp. time = 1500 seg; Filters: F475W, F606W y F814W; Cycle 15; PI: J. Dalcanton, Program 10915

1.1. Photometry

Photometric data were obtained from de STScI archive and correspond to the "star files" of ACS Nearby Galaxy Survey (ANGST). These files contain photometry of all objects classified as stars, having a good S/N (S/N > 4 and "dataflag" < 8). Photometry was obtained with the Dolphot code using the "point spread function" (PSF) adapted for the ACS/HST and the reduction process of Dalcanton et al. (2008).

[1] http://archive.stsci.edu/

2. Methodology

For each of the tree data bands, the following steps have been done:

- ☐ Selection of the data with low errors (e_{phot}, see Table 1)
- ☐ Selection of blue objects applying K-Means over the previously selected low error data using both color indices simultaneously
- ☐ Identification of stellar groups over the data, applying the following algorithms: DBSCAN, HDBSCAN and PLC
- ☐ Construction of the dendrograms for PLC case were built, where clearly is show the subdivision of the different clusters

The free parameters of each algorithm were adjusted in order to obtain the largest number of clusters. These free parameters and the obtained results are presented in Table 1.

The employed codes correspond to Python libraries developed by Pedregosa et al. (2011) Scikit-learn, McInnes et al. (2017) and Jones et al. (2001) Scipy. In the case of PLC, the code was developed by the authors.

3. Numerical Simulations

In order to test the detection efficiency of different methods, we made several simulations of stellar clusters submerged in galaxies fields with parameters (mag, colors, mass, etc) that mimic real galaxies.

A total of 48 clusters with a random number of stars following the Salpeter (1955) mass distribution were made. Colors of these star are from a population of stars with an age of $t \sim 10^7$ yrs. These groups were uniform and randomly distributed over a region with same size as the ACS CCD. Each group has a Gaussian spatial distribution that corresponds to real size clusters. Size, distance and extinction were chosen to mimic the average data found in NGC 300. Field stars are from a sample of NGC 300 but randomly uniform distributed over the simulated area. So, the density of clusters and the field stars were chosen to reproduce the real observed data.

The simulated observations are useful to find the number of clusters detected and lost for an stellar density with the different algorithms. Also, to check for spurious stochastic clusters that could be produce by random fluctuations of the number of the field stars and to test for problems in the algorithms.

4. Preliminary results

- ☐ The applied algorithms succeed to separated each populations of members (blue vs. red) and to identity several of the young clusters.
- ☐ NGC 300 does not have an important differential extinction while NGC 253 have a significant one. Due to this different extinction, the K-Means procedure have identified the blue population.

Table 1. Parameters for the different algorithms

Galaxy	Algorithm	K	DC	N_{MIN}	NG
NGC 300	K-Means	3	-	-	4104
(e_{phot} < 0.05)	DBSCAN	55	2.6"	10	1693
	HDSCAN	54	-	10	1784
	PLC	159	1.0" to 2.5"	10	2110
NGC 253	K-Means	2	-	-	22780
(e_{phot} < 0.03)	DBSCAN	340	2.0"	8	12814
	HDBSCAN	401	-	8	13237
	PLC	1266	0.3" to 2.0"	8	12790
Numeric	K-Means	3	-	-	11805
Simulation	DBSCAN	44	2.6"	10	11101
	HDBSCAN	48	-	10	11273
	PLC	44	1.0" to 2.5"	10	11607

K = quantity of groups found; DC = Characteristic distance; N_{MIN} = Minimum number of stars allowed to define a group; NG = quantity of blue members (K-Means) or total quantity of members of all groups (other algorithms)

- ☐ HDBSCAN allows to detect groups of variable density. In the other hand, the algorithm PLC applied with different values "linking length" allows to detect subgroups on the larger groups.
- ☐ Dendrograms are useful to understand the hierarchical structure of the young blue population.

References

Battinelli, P. 1991, A&A, 244, 69

Dalcanton, J., Williams, B., & ANGST Collaboration 2008, Astrophysics and Space Science Proceedings, 5, 115

Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open source scientific tools for Python. URL "`http://www.scipy.org/`"

McInnes, L., Healy, J., & Astels, S. 2017, The Journal of Open Source Software, 2. URL `http://dx.doi.org/10.21105/joss.00205`

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. 2011, Journal of Machine Learning Research, 12, 2825

Salpeter, E. E. 1955, ApJ, 121, 161