The Predictive Power of Computational Astrophysics as a Discovery Tool
Proceedings IAU Symposium No. 362, 2023
D. Bisikalo, D. Wiebe & C. Boily, eds.
doi:10.1017/S1743921322001697

Clustering stellar pairs to detect extended stellar structures

Sergey Sapozhnikov and Dana Kovaleva

Institute of Astronomy RAS, Pyatnitskaya 48., Moscow, Russia

Abstract. Gaia data allows for search for extended stellar structures in phase (coordinates plus velocities) space. We describe a method of using DBSCAN clustering algorithm, which is used to group closely-packed-together data points, to a list of preliminary selected pairs of stars, with parameters expected to be found within stellar streams and comoving groups: loose structures in which stars are not gravitationally bound, but do share motion and evolutionary properties. To test our approach, we construct a model population of background stars, and use pair-constructing and clustering algorithms on it. Results show that transitioning to a list of pairs sharply reveals structures not presented in background model, which then become more apparent targets in coordinate-velocity phase space for DBSCAN algorithm thanks to now increased relative density of the extended stellar structure.

Keywords. stars: kinematics, galaxies: structure, galaxies: star clusters

---

Gaia data have many applications, in particular, they can be used to search for star clusters and extended stellar structures: loose groups of stars, which are not bound gravitationally, but do exhibit similarities in motion and age suggesting that they share a similar genesis, like shown in Harshil Kamdar *et al.* (1999). Approaches to find extended stellar structures include, for example, search in the vicinity of stellar clusters for kinematically similar stars (see Jerabkova, Tereza *et al.* (2021); Siegfried Roser *et al.* (2019)), or using clustering algorithms to detect new objects in the field Hunt, Emily L. *et al.* (2021). The latter approach is somewhat complicated by the fact that such structures are very stretched on the sky, with many unrelated stars in their vicinity. We aim to test the approach of using our algorithm initially developed to find ultra-wide binary stars to search for stellar pairs with common motion, and feed this catalogue of pairs to the DBSCAN clustering algorithm. By adjusting the restrictions on pairs properties, we aim to highlight the comoving groups of stars and make them stand out more clearly from the background and make it easier for clustering algorithm to recognize.

Creation of pairs method is applied to a large sample of stars from Gaia DR2 data, 30x30 deg on the sky and 100-1000 pc distance from Sun. Criteria for pairs are selected according to Harshil Kamdar *et al.* (1999) and are the following:

- projected separation $< 1\ pc$
- projected relative motion $< 3\ km/s$
- proper motion difference $< 6\ mas/yr$
- parallax consistency within $-0.1 < \pi_{cons} < 1$

Here, "parallax consistency" is

$$\pi_{cons} = 3(\sigma_{\pi_1} + \sigma_{\pi_2}) - |(\pi_1 - \pi_2)|. \tag{0.1}$$

Clustering is performed using DBSCAN clustering algorithm in four-parameter space (2 coordinates + 2 velocities). We use small sample of region in Coma Berenices to test

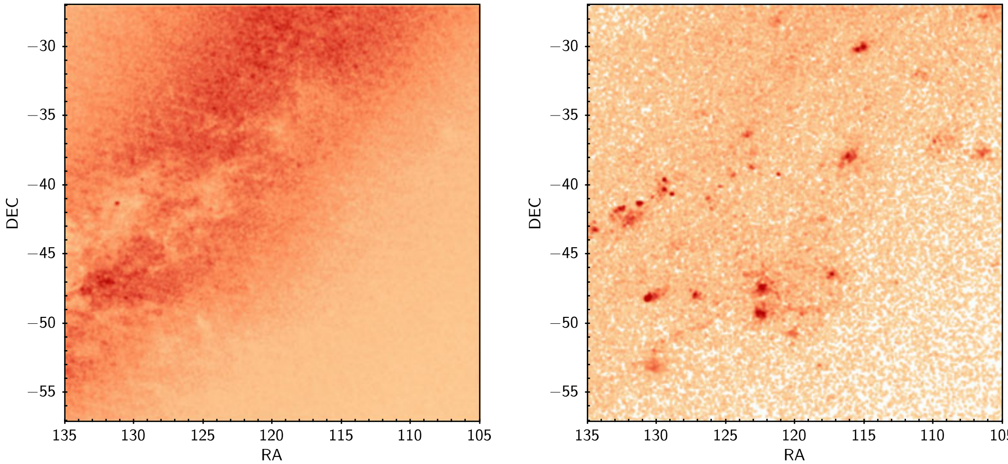

Figure 1. Scatter plots showing the sky positions of stars (left) and pairs (right) from the Gaia data in the selected region. More densely populated regions are closer to orange color. These scatter plots show how the translation from stars to pairs increases the contrast of the stellar structures.

for false negative result. Different parameters of DBSCAN clustering algorithm yield different amount of clusters when applied to the same set of pairs from Coma Ber. We compare our results with both clustering of pairs created from model uniform star distribution, and direct clustering of stars from that model, and find that clustering of stars from model devoid of clusters produce false positives much more easily compared to clustering of pairs created from the same sample. Also, clustering of pairs (compared to clustering of stars) for real data allow for easier recognition of clusters (in larger area of DBSCAN hyperparameters).

We conclude that:

• Transition from stars to pairs makes stellar structures stand out sharply against the stellar background. These structures are labelled by DBSCAN algorithm.

• We check for false positives using uniform stellar background model. No clusters are found in this sample, as expected.

• We applied our method to Coma Ber comoving group and verified that it correctly finds existing stellar structure there.

• Comparison with direct clustering of stars reveals that our method of pair clustering is less prone to false positives creation.

References

Harshil Kamdar, Charlie Conroy et al. 2019, *Astrophys. J*, 884(2):L42

Jerabkova, Tereza and Boffin, Henri M. J. et al. 2021, *Astronomy and Astrophysics*, 646:A104

Röser, Siegfried and Schilbach, Elena 2019, *Astronomy and Astrophysics*, 627:A4

Hunt, Emily L. and Reffert, Sabine 2021, *Astronomy and Astrophysics*, 646:A104

https://doi.org/10.1017/S1743921322001697 Published online by Cambridge University Press