

Comparison of Density-Based Clustering Algorithms for Asteroid Family Identification

D. Vujičić¹ and R. Pavlović²

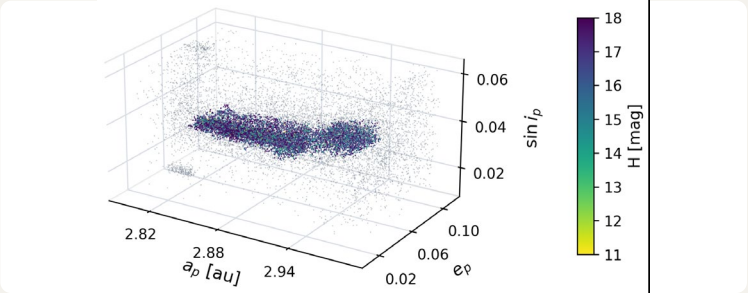
¹ Faculty of Technical Sciences Čačak, University of Kragujevac, Serbia ² Institute of Physics, University of Belgrade, Serbia

01 Motivation

The linking distance is tuned with care; the clustering algorithm is taken as given. Can density-adaptive **HDBSCAN** replace fixed-radius **DBSCAN** on a full catalog?

02 Data & pipeline

1,307,229 synthetic proper elements, AFP catalog (Jan 2025)
747,215 background left after quality cuts and Random Forest removal of 79 families
2 engines DBSCAN ($\epsilon = 55$ m/s) vs HDBSCAN (leaf), Zappalà metric, 3 zones
4 filters $N \geq 60$ · new · compact · V-shape in H



Known family (158) Koronis: 12,203 members coloured by H over the background (grey).

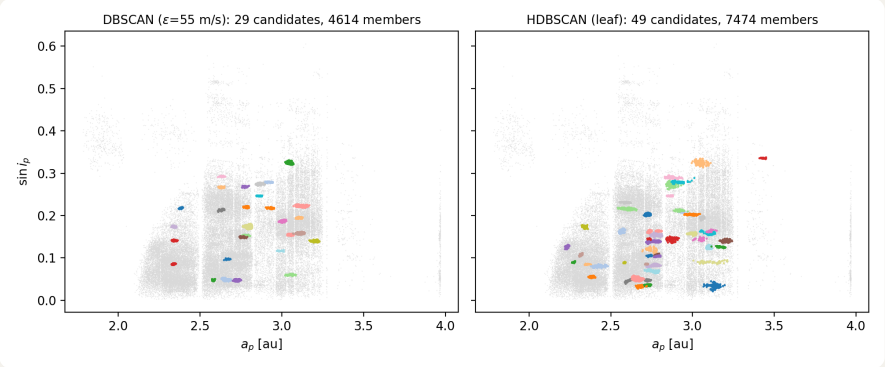
03 Default HDBSCAN fails

206,982 inner belt	286,179 middle belt	225,064 outer belt
------------------------------	-------------------------------	------------------------------

Excess-of-mass selection returns each zone as **one cluster**.
Leaf selection is mandatory.

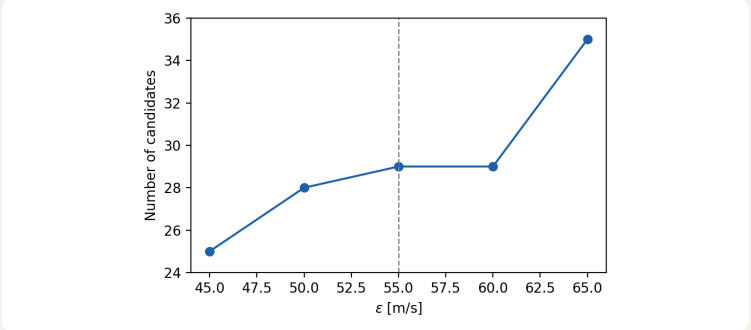
04 Head to head, identical vetting

Metric	DBSCAN	HDBSCAN
Candidates	29	49
Members	4,614	7,474
Nesvorný match	48.3%	42.9%



05 Where they agree

6/8 DBSCAN candidates stable over $\epsilon = 45\text{--}65$ m/s are recovered independently by HDBSCAN.



DBSCAN candidates vs ϵ (dashed: 55 m/s)

06 Take-home

- HDBSCAN needs leaf selection; defaults return the belt itself.
- Leaf HDBSCAN finds more candidates but matches Nesvorný less often.
- Run both:** candidates stable across ϵ and algorithms deserve follow-up.