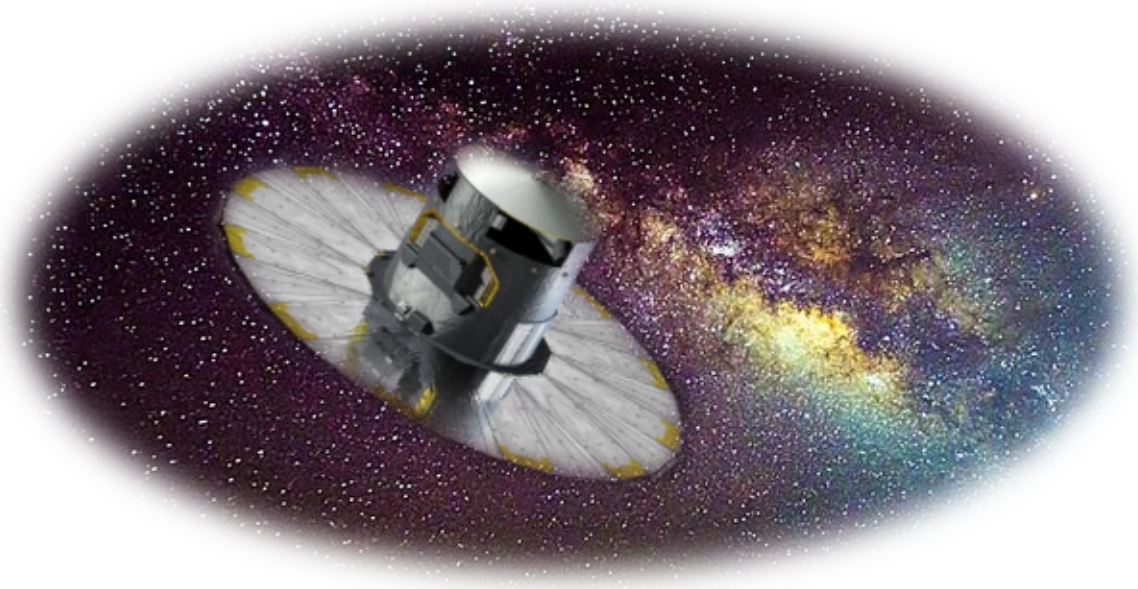
university of groningen / faculty of science and engineering / kapteyn astronomical institute

Discovery and characterization of substructures in TGAS and RAVE

Author
Titania Virginiflosia

Supervisors
Dr. Jovan Veljanoski
Dr. Lorenzo Posti
Prof. Dr. Amina Helmi



University of Groningen
Kapteyn Astronomical Institute
The Netherlands
December 2017

Abstract

We exploit the powerful combination of astrometric data from Gaia and radial velocity data from RAVE to find substructures in the solar neighborhood, using a friends-of-friends algorithm in six phase space dimensions. We show that this algorithm is successful in recovering the known substructures, as well as finding new substructures. A significance test reveals that the newly discovered substructures are not likely to be found by chance. Their physical sizes and the velocity dispersions are typically larger than Galactic open clusters, yet visual analyses of the color magnitude diagrams reveals the stars are likely to have the same age and origin. We conclude that some of these substructures are candidates of dissolving open clusters.

Contents

1 Introduction **3**

2 Data **5**

2.1 TGAS catalogue 5

2.2 RAVE DR5 5

2.3 Cross-matched data set 6

2.4 6-dimensional phase space 8

3 Methods **11**

3.1 Friends-of-friends algorithm 11

3.2 ROCKSTAR 12

3.3 Configuration and Input Files 13

3.4 Choice of Parameters Values 13

4 Results **17**

4.1 Comparison between different sets of parameters 17

4.2 Properties of the substructures 18

4.3 Commonality 22

4.4 Cross-match to catalogues of open clusters and OB-associations 23

4.5 Combining the 4 experiments 25

4.6 Significance test 28

4.7 CMD analysis 29

4.8 Finding more members in the TGAS data set 31

5 Conclusion **36**

5.1 Conclusion 36

5.2 Future prospects 36

Bibliography **38**

Appendix A Plots of positions and velocities **40**

Appendix B Plots of proper motions, color magnitude diagrams, and metallicities **41**

The cover image is a self-made composite of two pictures. The starry background was downloaded in December 2017 from *https://pngtree.com/freebackground/hd-vast-starry-background-image_719161.html*.
The Gaia spacecraft is an artist's impression by D. Ducros from ESA, downloaded in December 2017 from *http://blogs.esa.int/rocketscience/2015/01/14/a-year-on-station-for-gaia/*

Chapter 1

Introduction

The disk of the Milky Way contains many groups of stars that share the same kinematic properties, known as moving groups (Proctor, 1869). Among these groups are associations and star clusters that have been dissolved over time by the action of both internal forces (mass loss through the dynamical evolution and stellar evolution) and external ones, such as interactions with the Galactic tidal field, collisions with molecular clouds, and the Galactic differential rotation (Zhai et al., 2017). The origin of moving groups is still under active debate (Bovy and Hogg, 2010): are they remnants of a coeval star formation event with similar chemical composition? Or are they formed by dynamical effects of nonaxisymmetric features of the Galaxy such as spirals and bars? The dynamics of cluster dissolution provides important clues to understanding the stars formation history and the dynamical evolution of the Milky Way. The shape of these clusters can even shed light on the disruption process directly (Zhai et al., 2017).

As the velocity dispersions in moving groups are small, typically a few km/s or less (Tian et al., 1996), proper motions and radial velocities can be used to detect the common space motion and thus determine membership. For the majority of the moving group candidate members, only proper motions *or* radial velocities are available (Hoogerwerf and Aguilar, 1999). In the past decades, several methods have been developed to disentangle stellar systems from the field star population based on proper motion data alone. One is the convergent-point method, which uses the perspective effect that makes the proper motions of the stars point towards a convergent point in the sky. The vector-point diagram method also uses proper motions for membership selection, where member stars show a concentrated distribution compared to that of field stars. These traditional methods have several shortcomings. The convergent point method is useful when there is a significant sample of stars in a region of the sky. The presence of more than one group within the sample affects the performance of this method. The vector-point diagram method is constrained to small regions of the sky. This method is especially suited for member selection in open clusters. Neither of these methods use parallax information.

In the Hipparcos era, high quality measurements of positions, parallaxes, and proper motions have triggered the search for new, better methods to identify moving groups. Hoogerwerf and Aguilar (1999) introduced a new method using these five astrometric parameters measured by Hipparcos (ESA 1997), called the Spaghetti method. No information of radial velocities is assumed to be available. The membership selection is based on a combination of the classical convergent point method and a new selection method which makes use of the parallaxes as well as the proper motions, and searches for members in velocity space. The basic difficulty of this method is that the two measured velocity components are not the same for different stars, as each measured pair lies on a plane orthogonal to the unique line of sight to the corresponding star. Brown et al. (2016) also used the same method to select members of OB-associations. They mentioned that this method is not sufficient to study OB-associations which are located close to the Solar Antapex, as their space motions are purely radial with respect to the Sun and hence are traced by their radial velocities.

About 20 years after Hipparcos, the first data release of Gaia (Gaia Collaboration: Brown et al. 2016) has provided highly accurate positions, parallaxes, and proper motions for about 2 million stars in common

with Hipparcos and Tycho-2 catalogues (Høg et al. 2000). In addition to that, it has more than two hundred thousand stars in common with the ground-based survey for radial velocity (Radial Velocity Experiment, RAVE, Kunder et al 2017). These stars now have 6-dimensional phase space information, which is useful to study the moving groups in the Galaxy. Several authors have used this phase space information to find open cluster groupings (Conrad et al., 2017) and binary pairs (Oh et al., 2017).

In this work, we exploit the powerful combination of astrometric data from Gaia and radial velocity data from RAVE to find substructures in the solar neighborhood. The approach is based on adaptive hierarchical refinement of the friends-of-friends algorithm in six phase space dimensions, which allows for robust tracking of substructures (Behroozi et al., 2013). An attractive feature of the friends-of-friends algorithm is its simplicity: the results depend solely on the linking length in units of the mean interparticle separation. The algorithm does not assume any particular shape and therefore it is optimal to study non-axisymmetric mass distributions (More et al., 2011). Our goal is to run the friends-of-friends algorithm on the data set containing 6-dimensional information of the positions and velocities. Some substructures are matched to known stellar systems, for others no clear match is found in existing catalogues. All substructures are characterized in their spatial and kinematic distribution membership as well as color magnitude diagram.

This thesis proceeds as follows: in Chapter 2, we describe the data set used in this work. In Chapter 3, we explain the algorithm to identify the substructures. In Chapter 4, we present the results and discuss their significance. Finally, we give a summary in Chapter 5.

Chapter 2

Data

The primary data set used in this thesis is the cross-match between the Tycho-Gaia Astrometric Solution (TGAS, Gaia Collaboration: Brown et al. 2016) catalogue and the Radial Velocity Experiment Data Release 5 (RAVE DR5, Kunder et al., 2017). In this section, we describe each data set and how the combination of both can provide full phase space information of stars needed for this work.

2.1 TGAS catalogue

The European Space Agency satellite Gaia was launched in December 2013 to collect astrometric and photometric data for more than 1 billion sources brighter than magnitude 20.7 with an accuracy level of $5 - 25$ μmas. After 14 months from the start of nominal operations, the first data release (Gaia Data Release 1, Gaia DR1) was made available to the public. The two main components of Gaia DR1 are: 1) The astrometric data set which consists of two subsets: (a) The primary astrometric data set containing positions, parallaxes, and mean proper motions for about 2 million sources in common between the Gaia DR1, Hipparcos and Tycho-2 catalogues, (b) The secondary astrometric data set containing positions in the sky for an additional 1 billion sources, 2) The photometric data set containing the mean G-band magnitudes for all the sources in Gaia DR1.

The determination of proper motions and parallaxes in the primary astrometric data set benefits from the different epochs of observations between the Hipparcos and Tycho-2 catalogues (J1991.25) and Gaia (J2015.0), since the data from only 1 year of Gaia observations alone may not be reliable enough (Michalik et al., 2015). The realization of the joint-solution is called the Tycho-Gaia astrometric solution (TGAS). The typical uncertainty of TGAS sources is 0.3 mas for the positions, 1 mas/yr for the proper motions, and 0.3 mas for the parallaxes. For about a hundred thousand sources in common with Hipparcos, the proper motions are considerably more precise, with typical uncertainty of about 0.06 mas/yr.

2.2 RAVE DR5

The Radial Velocity Experiment (RAVE) is a ground-based spectroscopic survey of stars selected in the Southern Hemisphere within the magnitude of $9 \leq I \leq 12$ (Kunder et al., 2017). The RAVE spectra were taken using the multi-object spectrograph 6 degree field of view, mounted on the 1.2 meter UK Schmidt Telescope of the Australian Astronomical Observatory. The fifth data release of RAVE (RAVE DR5) contains about half a million stars, in which radial velocities are derived from the spectra are available for all of the stars. Stellar atmospheric parameters such as temperature, surface gravity, and metallicity are determined using a pipeline that is based on the combination of a decision tree: one to renormalize iteratively the spectra and obtain stellar parameter estimations for the low signal-to-noise spectra, and another one to derive the parameters for stars having high signal-to-noise (see the summary in Kordopatis

et al, 2011). The stellar atmospheric parameters are then used to estimate parallaxes using the Bayesian method (Binney et al., 2014).

2.3 Cross-matched data set

A cross-matched data set between TGAS and RAVE DR5 was made by Helmi et al., (2017) and contains 210,263 stars in common. These stars now have positions, proper motions, and parallaxes from TGAS, in addition to radial velocities, stellar atmospheric parameters, and spectro-photometric parallaxes from RAVE DR5. Distances are derived from either the TGAS or RAVE DR5 parallaxes ($d = 1/\bar{\omega}$) that have a smaller relative parallax error. This derivation of distances based on parallaxes ($d = 1/\bar{\omega}$) is not always optimal, but it is found to be the best distance estimator in the case of a small relative parallax error (Binney et al., 2014). Thus, we made a quality cut on the parallax by taking stars that have positive parallaxes and relative parallax errors smaller than 20%. If the parallax comes from RAVE, an additional cut was made following the selection criteria described by Kunder et al., (2017), i.e., we select stars with radial velocity errors $\sigma \leq 8$ km/s, correlation coefficient ≥ 10, signal-to-noise ratio SNR ≥ 20, and AlgoConv $\neq 1$. Combining these criteria, we have the TGAS and RAVE data set containing 108,916 stars (hereafter the TGAS×RAVE data set, unless stated otherwise). We show the comparison of the distribution of stars in the sky before and after applying the quality cut criteria in Figure 2.1.

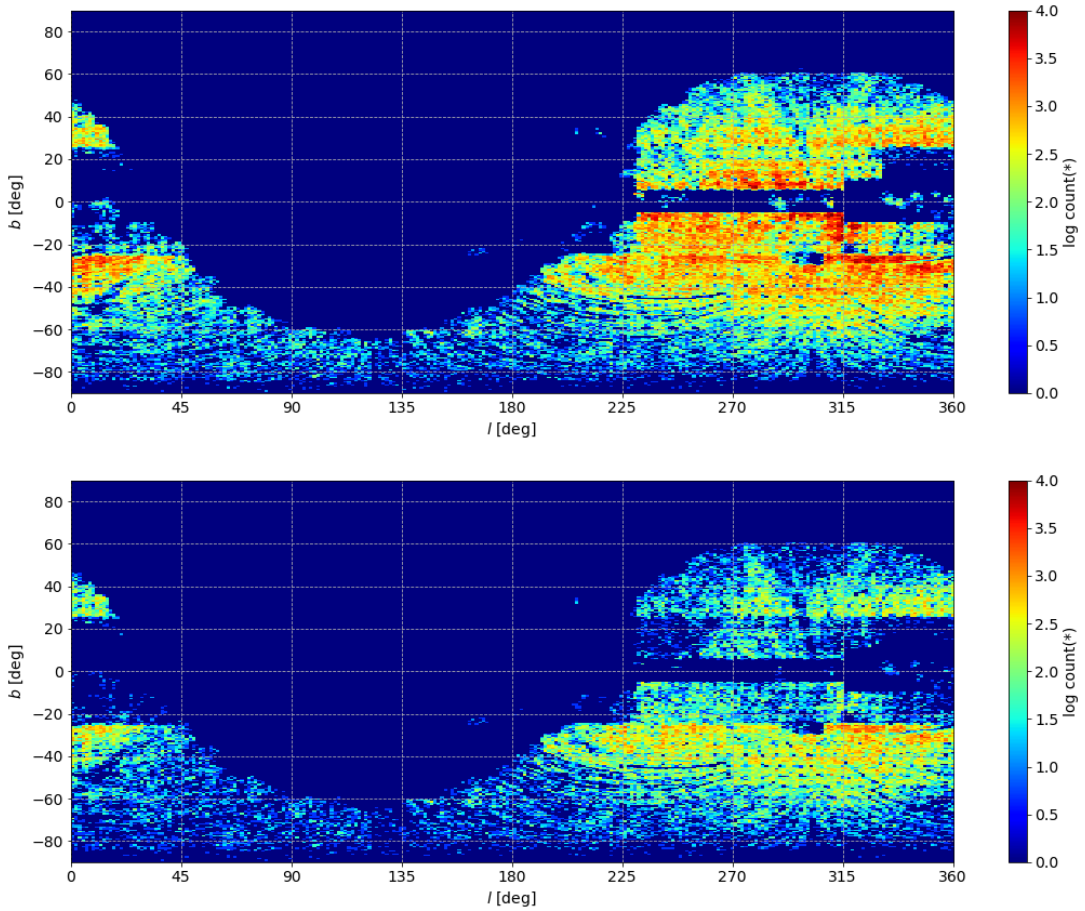


Figure 2.1: Visualizations of the distribution of stars in the sky in Galactic coordinates before (top panel) and after (bottom panel) applying the quality cut criteria (see text for details). The number of stars reduces by 52%: from 210,263 stars to 108,916 stars.

We show the distribution of parallaxes and distances from the TGAS×RAVE data set. Most of the stars

($\geq 97\%$) are located within 1 kpc away from the Sun. The closest star in the data set has a distance of 7 pc, while the farthest one has a distance of 8.4 kpc. The radial velocities and their uncertainties are shown in Figure 2.3. The stars have radial velocities that range between ± 300 km/s and their distribution peaks around 8 km/s. About 80% of the stars have radial velocity uncertainties below 2 km/s.

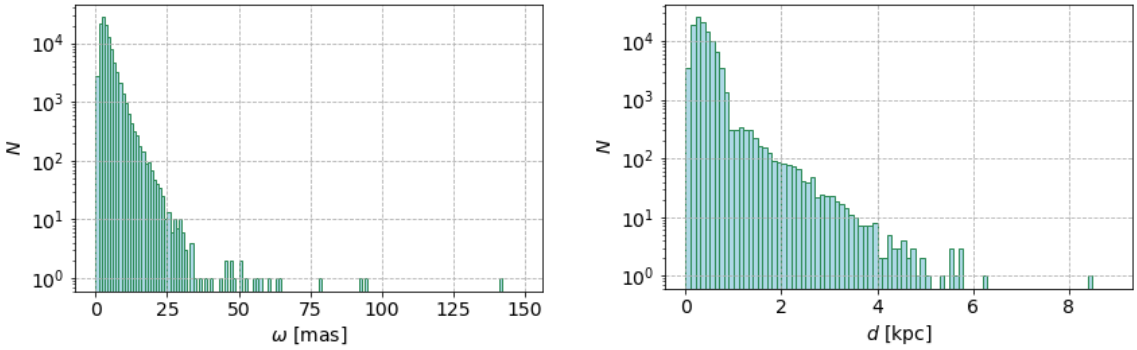


Figure 2.2: Distribution of parallaxes (left panel) and distances (right panel) of stars in the TGAS×RAVE data set. The bin width is 1 mas and 0.1 kpc for the parallaxes and the distances respectively.

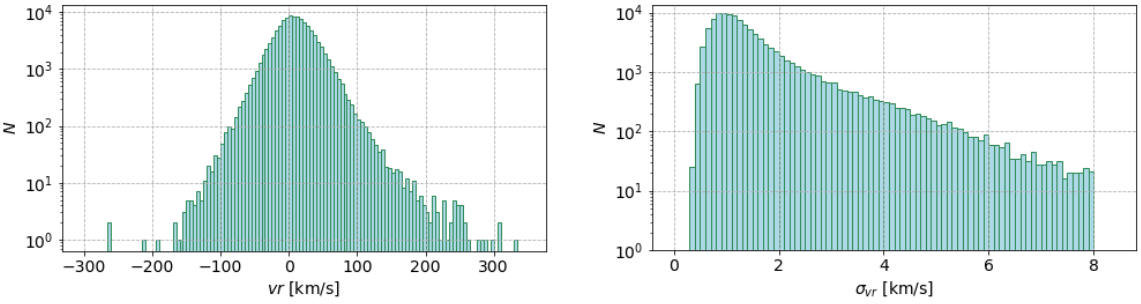


Figure 2.3: Distribution of radial velocities (left panel) and their corresponding uncertainties (right panel) of stars in the TGAS×RAVE data set, with a bin width of 5 km/s and 0.1 km/s respectively.

From the TGAS×RAVE data set we can construct a color magnitude diagram using the G-band magnitudes from TGAS and the K-band magnitudes from the Two Micron All-Sky Survey (2MASS, Skrutskie et al. 2006) included in RAVE DR5. The absolute magnitude M_G is calculated using the apparent magnitude G with the following equation,

$$M_G = G - 5 \log d + 5 \tag{2.1}$$

where d is distance in pc. The apparent magnitudes G and K_s in the data set range between $6 \leq G \leq 13$ and $2 \leq K_s \leq 13$ respectively, while the absolute magnitude M_G ranges between $-3 \leq M_G \leq 10$.

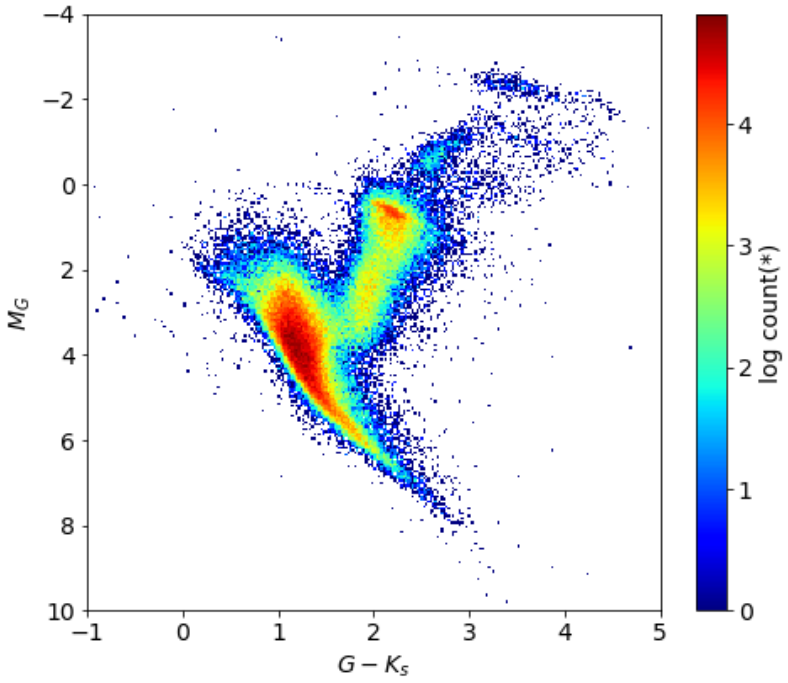


Figure 2.4: The color magnitude diagram for stars in the TGAS×RAVE data set, using G-band magnitudes observed by Gaia and K-band magnitude from 2MASS.

2.4 6-dimensional phase space

In this thesis we use 6-dimensional phase information in the Cartesian coordinate system to find substructures in the data set. The position is denoted as x, y, z with the Sun as the origin and axes pointing to the Galactic center (x), to the direction of the Galactic rotation (y), and to the north Galactic pole (z). The corresponding velocity components are v_x, v_y, v_z, respectively. The position and velocity components in Cartesian coordinates are calculated directly from the observed quantities in the TGAS×RAVE data set, using the distance (d [kpc]), position in the sky (l and b [rad]), radial velocity (vr [km/s]), and proper motions (μ_l and μ_b [mas/yr])[1],

$$x = d \cos b \cos l \tag{2.2}$$
$$y = d \cos b \sin l \tag{2.3}$$
$$z = d \sin b \tag{2.4}$$
$$v_x = v_r \cos l \cos b - v_l \sin l - v_b \cos l \sin b \tag{2.5}$$
$$v_y = v_r \sin l \cos b + v_l \cos l - v_b \sin l \sin b \tag{2.6}$$
$$v_z = v_r \sin b + v_b \cos b, \tag{2.7}$$

where v_l and v_b are the transverse velocities computed from the proper motion, with k being the conversion constant,

$$v_l = k \; d \; \mu_l \cos b \tag{2.8}$$
$$v_b = k \; d \; \mu_b. \tag{2.9}$$

The spatial distribution of stars in Cartesian coordinates is shown in Figure 2.5, with colors indicating the density on a logarithmic scale. The stars have peak density around the Sun and decline smoothly as the distance increases. This spatial distribution, however, is not isotropic with respect to the Sun. More than 80% stars are located in the third and fourth quadrant (that is, in the region of $y \leq 0$, left panel of Figure 2.5), and about 75% of stars are located below the Galactic plane ($z \leq 0$). The distributions of the relative uncertainty to the distance are shown in Figure 2.6.

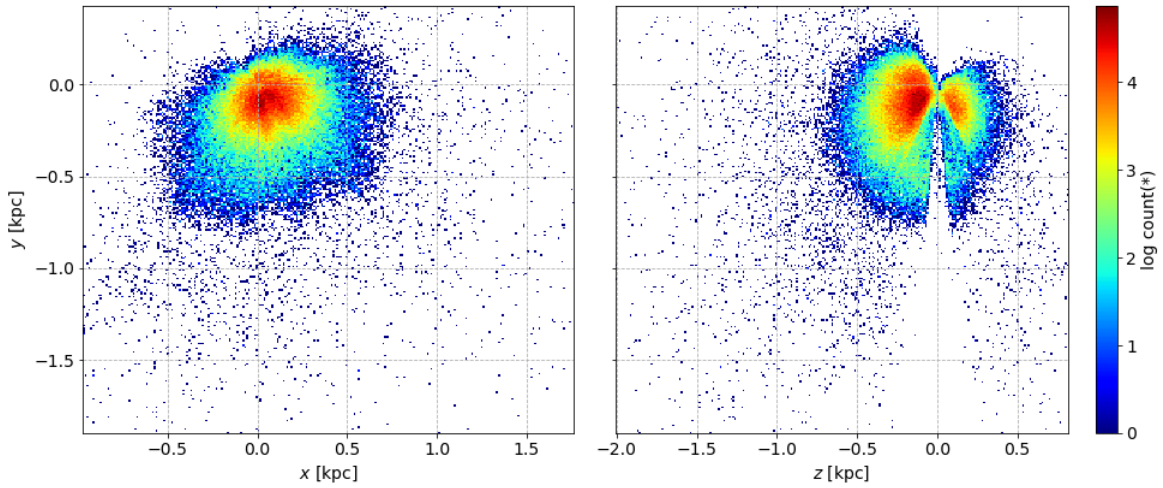


Figure 2.5: Spatial distribution of stars in Cartesian coordinates. The stellar density is shown in colors with logarithmic scale. Note that the spatial distribution of stars is not isotropic with respect to the Sun.

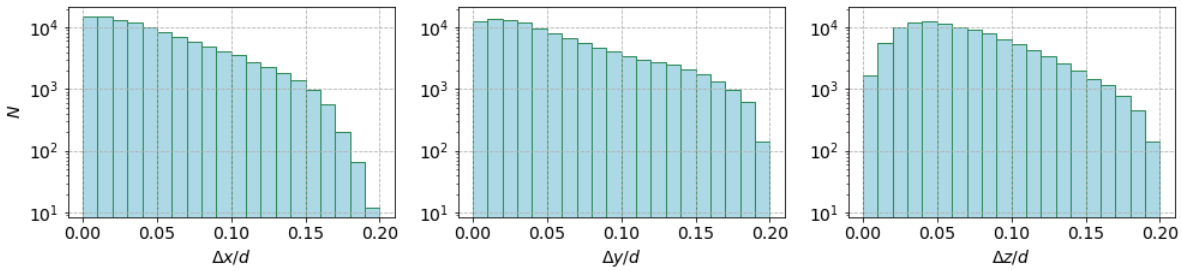


Figure 2.6: Distribution of the relative uncertainty of the distance in Cartesian coordinates x, y, and z.

We show the distribution of velocities of stars in Figure 2.7. More than 50% of the stars in the data set move towards the outer part of the Galaxy ($v_x \leq 0$) and about 80% of the stars trail behind the Sun ($v_y \leq 0$). Note that we do not apply a correction for the Sun's peculiar motion to the data set, since it will not affect the results significantly. The uncertainty distributions of the velocities are shown in Figure 2.8. The median uncertainties for v_x, v_y, and v_z are 2.68 km/s, 2.32 km/s, and 1.94 km/s respectively.

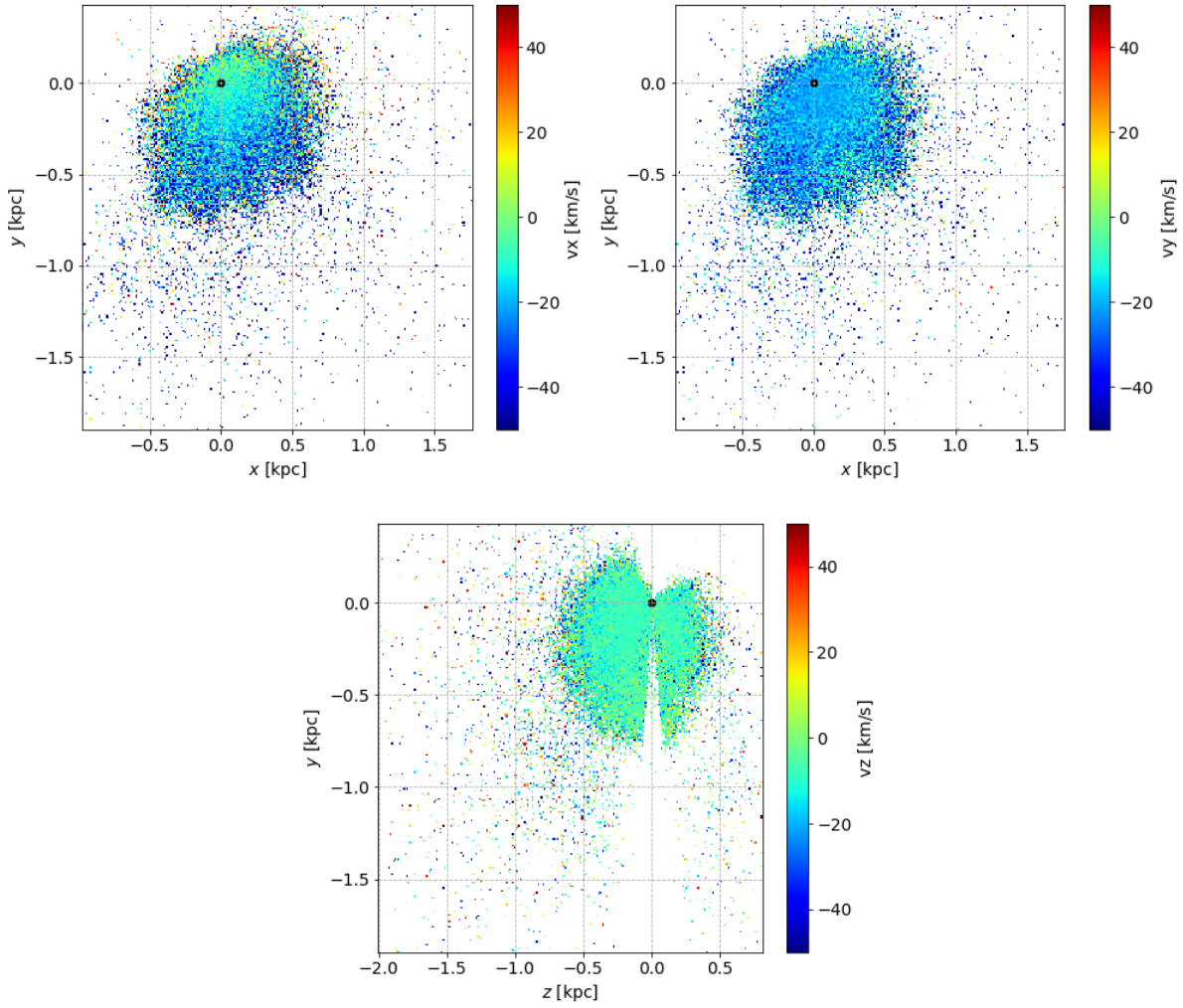


Figure 2.7: Colormaps of velocities in Cartesian coordinates, scale with colors. Top left: velocities in the x direction in the x-y plane. Most stars move towards the outer part of the Galaxy, as can be seen by the turquoise and blue colors. Top right: velocities in the y direction in the same plane as the top left panel. Most stars have a negative velocity in the y direction (indicated by the blue color), meaning that the stars are trailing behind the Sun. Bottom panel: velocities in the z direction in the z-y plane.

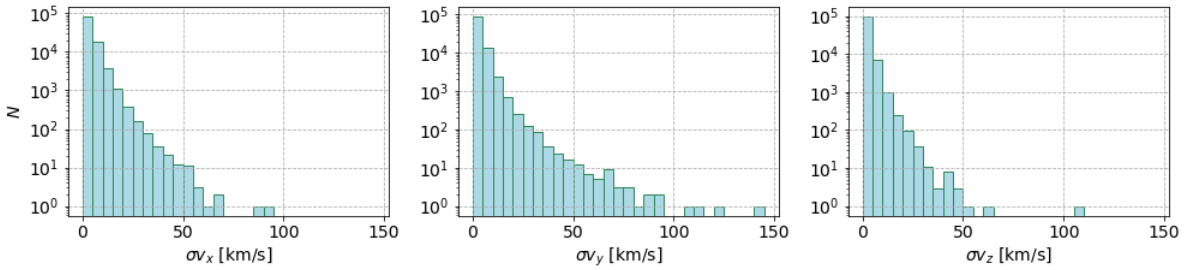


Figure 2.8: Distributions of the uncertainties of velocities v_x, v_y, and v_z in Cartesian coordinates.

Chapter 3

Methods

3.1 Friends-of-friends algorithm

The basic method used for identifying substructures is a "friends-of-friends" algorithm. This algorithm links together particles that fall within a specified distance in parameter space. Each set of joined particles constitutes a group. The parameter space can be defined by multiple physical parameters, such as position, velocity, metallicity, and redshift. An appropriate weighting of these parameters should be determined in advance in order to set the group linking criteria. Since each parameter has a different unit, each must be converted to an equivalent unitless quantity on the same scale before calculating distances in parameter space. For example, if the positions have a range of 5 kpc and the velocities range over 200 km/s, a metric must be made such that the group is not defined based on position only, because of its comparatively small numerical range. In principle, the metric puts all of the parameters to the same effective scale. In addition to the metric, the algorithm requires a linking length that defines the maximum distance allowed between group members in the parameter space.

For each particle, the algorithm links a nearby particle to it if

$$(\Delta s)^2 = \sum_i \Upsilon_i^2 (\Delta x_i)^2 < L, \tag{3.1}$$

where Δs is the total separation in parameter space of a neighbor to the particle in question, Υ_i is the scaling or weighting factor defined for parameter i, Δx_i is the separation between the particle and its neighbor in parameter i, and L is the linking length (Perrett et al., 2003). Traditionally, L is specified as a fraction Λ of the mean interparticle separation in parameter space for the entire system,

$$L = \Lambda \bar{\Delta} s. \tag{3.2}$$

Another criterion can be used to filter out groups which do not have a sufficient number of particles,

$$N_j \geq N_c, \tag{3.3}$$

where N_j is the number of members of group j and N_c is the minimum number of particles that is considered reliable (Liu et al., 2008).

In this thesis, we use the friends-of-friends algorithm called Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement (ROCKSTAR), originally made for identifying dark matter halos, substructures, and tidal features from N-body simulations (Behroozi et al., 2013). ROCKSTAR finds substructures based on six phase space dimensions of positions and velocities. We provide a more detailed explanation about the algorithm in the next section.

3.2 ROCKSTAR

As a first step, the algorithm finds overdense regions in the data set based on the pre-defined 3-dimensional (3D) linking length between the positions and divides them into 3D groups. The 3D linking length is defined as a fraction b of the mean interparticle distance $\bar{l}$ in the data, where $\bar{l}$ is related to the mean number density as $\bar{l} = \bar{n}^{-1/3}$. In ROCKSTAR, the 3D linking length is used to divide the data into manageable units. If a particle has more than a certain number of neighbors within the linking length, then the algorithm will start looking for neighbors within twice the original linking length. If any of those particles belong to another group, that corresponding group is joined to that of the original particle.

For each group, the positions and velocities of the particles are normalized by the position and velocity dispersion of the corresponding group. For two particles p_1 and p_2, the phase space density metric is defined as,

$$d(p_1, p_2) = \left(\frac{|\mathbf{x}_1 - \mathbf{x}_2|^2}{\sigma_x^2} + \frac{|\mathbf{v}_1 - \mathbf{v}_2|^2}{\sigma_v^2} \right)^{1/2}. \tag{3.4}$$

The 6-dimensional (6D) linking length, as opposed to the pre-defined 3D linking length, is adaptively chosen such that a constant fraction of f particles is linked together with another particle. There are two things to consider when choosing the value of this fraction. If the value is too large, the algorithm can find spurious subgroups. On the other hand, if the value is too small, the algorithm may not find small subgroups.

ROCKSTAR repeats the process of renormalization, choosing linking length, and calculating a new hierarchy level progressively for each subgroup. This repetition continues until the algorithm reaches its deepest level given the pre-specified minimum number of particles inside the 6D subgroups (see equation 3.3).

For each 6D subgroup at the deepest level of hierarchy, a seed is generated. The algorithm then recursively analyzes higher levels of the hierarchy to assign particles to the subgroup seed until all particles in the original 3D groups have been assigned. Two subgroups are merged if

$$\sqrt{(x_1 - x_2)^2 \mu_x^{-2} + (v_1 - v_2)^2 \mu_v^{-2}} < 10\sqrt{2}, \tag{3.5}$$

where σ_x and σ_v are the position and velocity dispersions, n is the number of particles, $\mu_x = \sigma_x/\sqrt{n}$ and $\mu_v = \sigma_v/\sqrt{n}$, all for the smaller subgroup between the two. Once particles have been assigned to 6D subgroups, unbound particles are removed and subgroup properties are calculated. A visual summary of the algorithm is given in Figure 3.1.

ROCKSTAR sets a default value of $b = 0.28$ and $f = 0.7$. We will see later that, despite these values of b and f in ROCKSTAR have been tested and used for finding dark matter halos, substructures, and tidal features (Behroozi et al., 2013), they have not yet been tuned for finding Galactic substructures such as open clusters, OB-associations, and moving groups. Thus, experiments need to be done to explore these values and to optimize the outputs (see section 3.4 for more details).

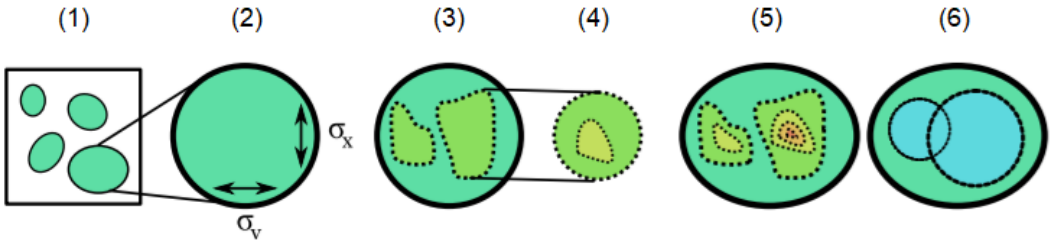


Figure 3.1: A summary of the ROCKSTAR algorithm. (1) The data set is divided into 3-dimensional groups. (2) For each group, the positions and velocities of the particles are normalized by the position and velocity dispersion of the group. (3) A phase space linking length is adaptively chosen such that a constant fraction f of particles is linked together in 6-dimensional subgroups. (4) The process repeats progressively for each subgroup: renormalization, a new linking length, and a new level of substructure are calculated. (5) Once the algorithm reaches its deepest level of hierarchy, particles are assigned to the closest subgroup in phase space. (6) Unbound particles are then removed and the properties of the substructures are calculated. This figure is adapted from Behroozi et al., (2013).

3.3 Configuration and Input Files

To run ROCKSTAR, a configuration file that lists numerous parameters is required. The user has to provide information for the linking length b, refinement constant f, minimum number of particles considered as members in a substructure, as well as the input and output directory. If not specified, the algorithm uses the default values. The input file for ROCKSTAR should include the information of positions, velocities and IDs of the particles. We use the positions and velocities in the Cartesian coordinate system, as described in Section 2.4.

3.4 Choice of Parameters Values

In this section, we explore the effect of different parameters on the properties of the substructures identified by ROCKSTAR. We are particularly interested to know whether ROCKSTAR can find and characterize substructures that are known in the TGAS×RAVE data set. For this purpose, we run ROCKSTAR with all possible combinations of parameters and compare the results to the literature. We demand for each substructure to have at least 10 stars. As a first step, we focus the comparison on the Pleiades, since it is a very well-known open cluster. It is also located in a region which was specifically targeted by RAVE and where the survey footprint does not continue (see Figure 3.2), thus it is easy to know if there is contamination in the obtained results.

We compare four physical quantities between the Pleiades and the substructures found by ROCKSTAR: median position in the Cartesian coordinates, number of stars, distance, and velocity dispersion. Previous studies reported various distances to the Pleiades (see the summary on Figure 8 in Gaia Collaboration: Brown et al. 2016), ranging from 115 to 150 pc taking into account the measurement uncertainties. Given this range in distance, we expect to have the shift of median position to be less than 35 parsecs. There are 31 stars in the TGAS×RAVE data set that are located in the Pleiades area, thus we use this number as our limit. The internal velocity dispersion of Pleiades reported using Hipparcos measurement is $\sigma_v \lesssim 1$ km/s (Narayanan and Gould, 1999).

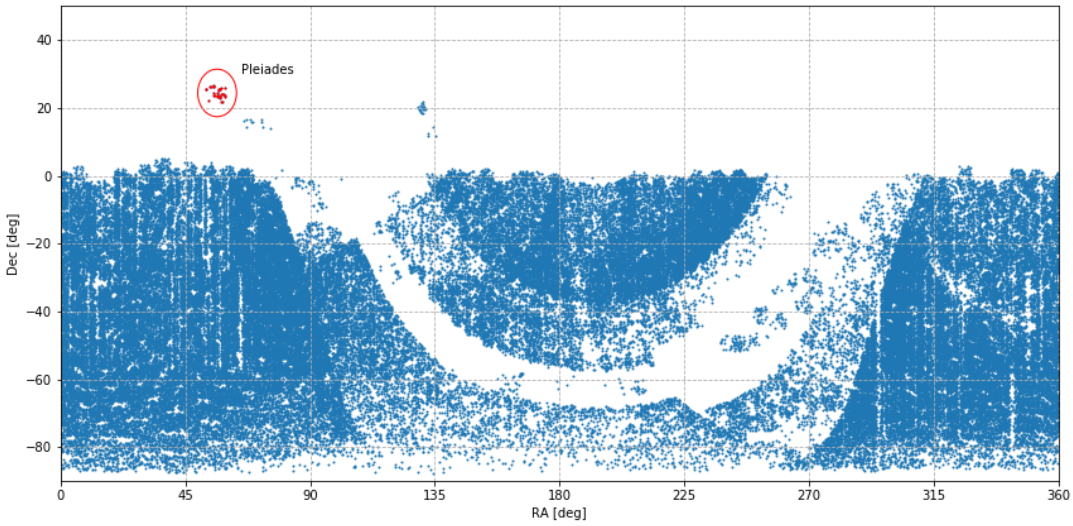


Figure 3.2: The distribution of stars in the TGAS×RAVE data set in the equatorial coordinate system. The Pleiades open cluster is shown in red dots. This open cluster is located in an isolated area in the sky with respect to the RAVE survey footprint. There are 31 stars located in this area.

First, we ran ROCKSTAR with all possible combinations of b and f. If we set the minimum number of stars within a substructure to be 10, we found that the possible values for both parameters are $0.05 \leq b, f \leq 0.90$. Any value of b that is less than 0.05 will give an error in which ROCKSTAR dumps the subgroup seed, while any value bigger than 0.90 will take a much longer time and possibly give substructures that are not statistically significant (Behroozi, Wechsler & Wu, 2013). We chose the step in each parameter search to be 0.01, resulting in 7396 combinations from 86 different values of each b and f. Each combination of parameters produces a number of substructures that ranges from 8 to 251. We then choose only one substructure from each run that might represent Pleiades, the one that has the closest median position to the real Pleiades in Cartesian coordinates. Here we use the median instead of the mean position in order to minimize the effect of outliers. The shift of median position r is defined as

$$r = \sqrt{(x_{\rm sub} - x_{\rm plei})^2 + (y_{\rm sub} - y_{\rm plei})^2 + (z_{\rm sub} - z_{\rm plei})^2}, \tag{3.6}$$

where $(x, y, z)_{\rm sub}$ are the median positions of the substructure and $(x, y, z)_{\rm plei}$ are the mean positions of the real Pleiades taken from Van Leeuwen (2009):

$$\begin{aligned} x_{\rm plei} &= -107 \text{ pc} \\ y_{\rm plei} &= \ \ 25.9 \text{ pc} \\ z_{\rm plei} &= -48.3 \text{ pc}. \end{aligned}$$

Figure 3.3 shows the properties of all substructures that might represent Pleiades. The first panel shows the shift of the median position of the substructures found by ROCKSTAR to the mean position of the real Pleiades. The values range from 13 to 143 pc, with most substructures ($\sim$ 97%) within a radial distance from the known Pleiades members that is smaller than 22 pc. The second panel shows the number of stars in the substructures, where the values range from 14 to 25000. The third panel shows the mean distance of the substructures in parsecs, ranging from 50 pc to 650 pc away from the Sun. The last panel shows the velocity dispersions in one dimension. The values range from 1.5 km/s to 52 km/s, with most substructures ($\sim$ 93%) having velocity dispersions less than 20 km/s.

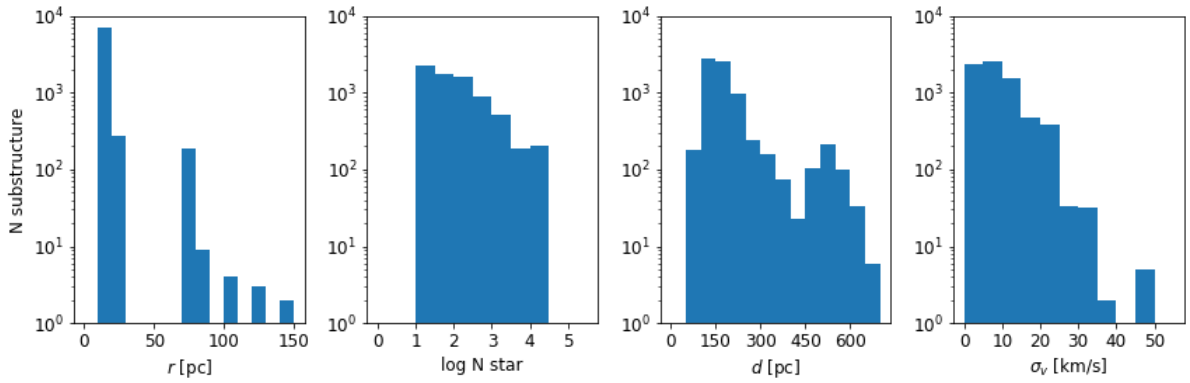


Figure 3.3: Properties of the Pleiades-like substructures from the 7396 runs of all possible combinations of f and b. From left to right panel: (1) Shift of the median position of the substructures relative to the real Pleiades in Cartesian coordinates, (2) Number of stars, (3) Distance (pc), (4) Velocity dispersion (km/s). Note that the y axis is a logarithmic scale.

Figure 3.4 shows the properties of these substructures in a colormap as a function of b and f. Now we need to examine this colormap in order to learn how different parameters contribute to the type of substructures produced by ROCKSTAR. On the first panel (top-left), we show the shift of median positions. It is obvious that almost all combinations of parameters b and f can produce a substructure that is relatively close to the real Pleiades. There is a slight dependence on the choice of f, as can be seen from the different horizontal layers.

On the second panel (top right), we show the number of stars in the substructures. Since the values range widely from 14 to 25000 stars, it may not be useful to show the overall distribution, especially since there are only 31 stars in the data set which are located in the Pleiades area. Thus, we focus on showing only the substructures that have members equal to or less than the number of Pleiades stars in the data set. The third panel (bottom-left) of Figure 3.4 shows the distance of the substructures, ranging from 115 to 150 pc. The second and third panels show a very strong dependence on the choice of both b and f. There are two different regions where one parameter plays a more dominant role than the other. For a low value of b ($b \leq 0.2$), the results of ROCKSTAR depend solely on the parameter b, in which the number of stars and distance changes as b increases, regardless of the value of f. For a higher value of b, the results of ROCKSTAR depend more strongly on the choice of parameter f, while b only contributes slightly to the results.

The last panel (bottom-right) of Figure 3.4 shows the velocity dispersion. In this case we show only the substructures that have velocity dispersions less than 20 km/s to see the colormap in more details.

The colormap indicates that, in order to reproduce a substructure such as the Pleiades open cluster, we have to choose a specific set of b and f. Some combinations can reproduce the Pleiades with reasonable physical properties, but some can not. These combinations, however, only apply to the Pleiades-like substructures. We still have to figure out how ROCKSTAR works for other substructures, for example those that are located in non-isolated areas, or the ones that have different morphology than Pleiades. We will explore this in more detail by running ROCKSTAR with a few sets of b and f to explore the results. Nonetheless, Figure 3.4 shows that values of $b \leq 0.2$ (all panels) are preferred, as well as values of $f \geq 0.4$ (first panel).

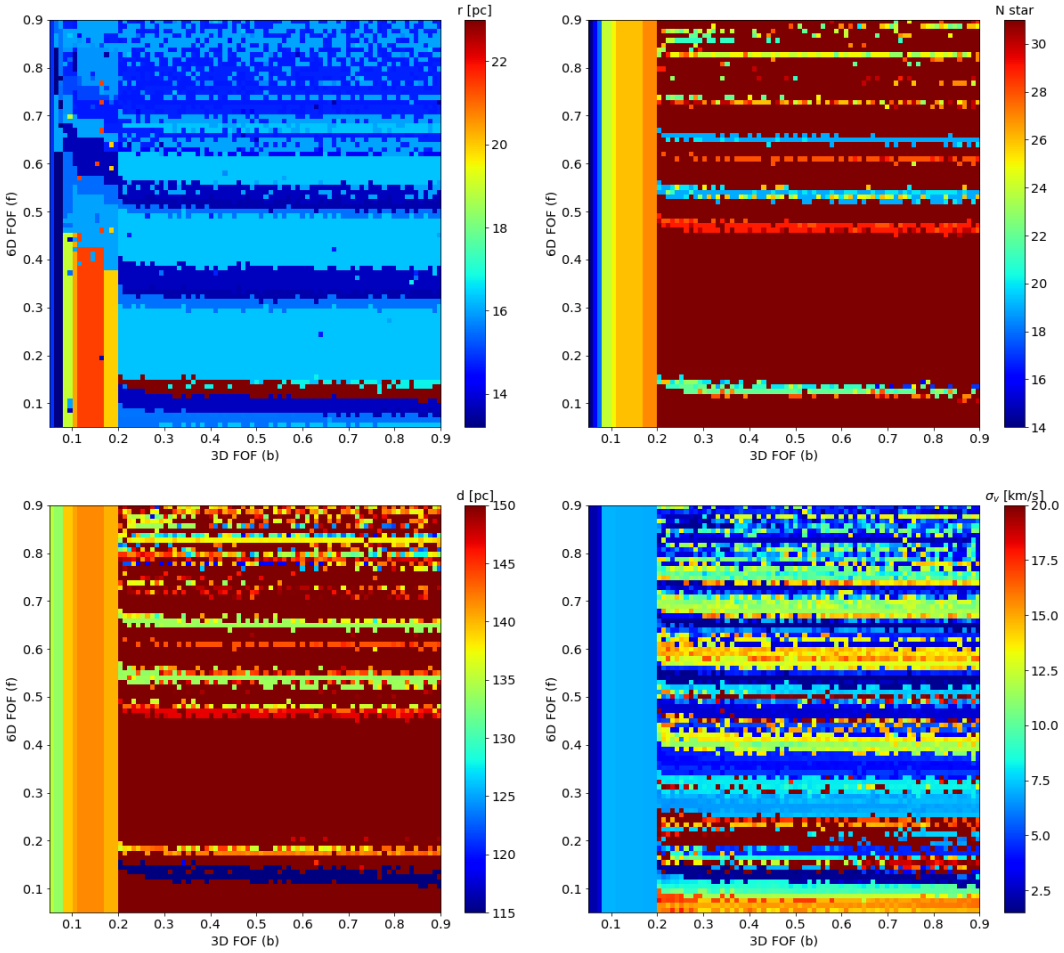


Figure 3.4: Properties of the Pleiades-like substructures in the $b - f$ plane. Top left: the shift of median positions in parsecs, for substructures with $r \leq 25$ pc. Top right: the number of stars in the substructures, ranging from the lowest number to the upper limit described in text. The dark red area represents substructures that have more than 31 stars. Bottom-left: distance of substructures in parsecs between the limit of $115 \leq d \leq 150$. Other areas shown by dark blue or dark red colors are for substructures with $d < 115$ pc or $d > 150$ pc respectively. Bottom-right: velocity dispersions of the substructures. The colorbar ranges from the minimum dispersion to 20 km/s.

Chapter 4

Results

4.1 Comparison between different sets of parameters

We run ROCKSTAR using 4 different sets of parameters covering both small and large values of b and f. We choose these values to be $b_1 = 0.05$, $b_2 = 0.20$, $f_1 = 0.40$, and $f_2 = 0.80$. Together they create 4 combinations (see Table 4.1), namely Experiment 1 (b_1, f_1), Experiment 2 (b_1, f_2), Experiment 3 (b_2, f_1), and Experiment 4 (b_2, f_2).

Table 4.1: Parameters for the 4 experiments

Name	3D linking parameter (b)	6D linking parameter (f)
Experiment 1	0.05	0.40
Experiment 2	0.05	0.80
Experiment 3	0.20	0.40
Experiment 4	0.20	0.80

There is a trend in which the smaller b produces more substructures than the larger b. The opposite happens for f, where the smaller f produces fewer substructures than the larger f (see the left panel of Figure 4.1). The number of stars assigned to substructures also strongly depends on the choice of b. Experiments with the smaller b have fewer stars assigned to substructures (see the right panel of Figure 4.1), even though the number of substructures itself is higher than the ones from experiments with the larger b. This overall trend can be understood by looking at how each parameter works in the ROCKSTAR algorithm (see section 3.2). The parameter b is only used to divide the data volume into manageable units. A smaller b will divide the data set into smaller volumes, and therefore increase the total number of 3D groups for the substructures while having a small number of stars in each group (hence, fewer stars assigned to a substructure). On the other hand, the 6D linking parameter works such that a larger f will find smaller substructures, while a smaller f will more likely merge distinct substructures together (hence, a smaller number of substructures in total).

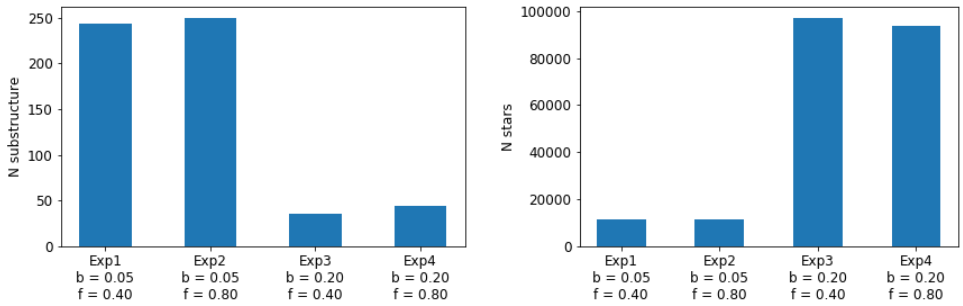


Figure 4.1: Left panel: The total number of substructures produced by ROCKSTAR for different sets of parameters. Right panel: The number of stars that belong to the substructures in each experiment. The exact values are given in Table 4.2.

Table 4.2: Number of substructures and stars for each experiment

Name	Total number of substructures	Total number of stars
Experiment 1	244	11332
Experiment 2	249	11353
Experiment 3	36	96744
Experiment 4	44	93836

4.2 Properties of the substructures

For each experiment, we calculate the properties of each substructure, such as the number of stars comprising a substructure, the physical size, and the velocity dispersion. The physical size and the velocity dispersion are calculated using Principal Component Analysis (PCA), in which we determine the principal axes of the substructures based on the biggest variance of the data. Instead of calculating the physical size or velocity dispersion in Cartesian coordinates, we determine these quantities in a coordinate system spanned by the principal axes of each substructure. The physical size is calculated in the principal axes of position, while the velocity dispersion is calculated in the principal axes of velocity. This method is useful to give a more intuitive sense of the shape of the substructure. The PCA can be done as follows: for the stars in each substructure, we calculate the variance and the covariance of each axis. An example of the variance of x and covariance between x and y is given in the following equations,

$$c_{xx} = var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{4.1}$$

$$c_{xy} = cov(xy) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \tag{4.2}$$

where n is the number of stars in the substructure. By taking all the variances and covariances, we can construct a 3×3 covariance matrix,

$$C = \begin{pmatrix} c_{xx} & c_{xy} & c_{xz} \\ c_{xy} & c_{yy} & c_{yz} \\ c_{xz} & c_{yz} & c_{zz} \end{pmatrix}. \tag{4.3}$$

The eigenvalues of the covariance matrix correspond to the variance of the data, that is equal to the squared dispersion of the positions and velocities of the substructures. If we denote σ_1^2, σ_2^2, σ_3^2 as the

3 eigenvalues, the physical sizes and the velocity dispersions of the substructures can be calculated by taking the root mean square,

$$\sigma = \sqrt{\frac{(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)}{3}}. \tag{4.4}$$

Figure 4.2 shows the sizes of the substructures in terms of the number of stars. Most substructures have a small number of stars, except a few outliers from Experiment 1 and 2 that contain 1-2 thousand stars. Even more extreme are Experiment 3 and 4, which can have up to 85 thousand stars, or equal to $\sim 80\%$ of the whole data set.

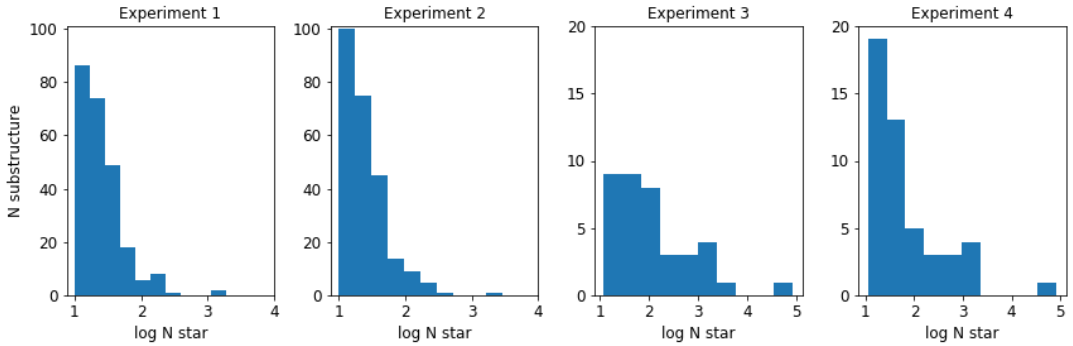


Figure 4.2: The number of stars in the substructures from each experiment.

Figure 4.3 shows the physical sizes and the velocity dispersions of the substructures calculated using equation 4.4. Experiments with the smaller b produce more substructures with smaller size compared to experiments with the larger b, but also with a larger velocity dispersion in general. For $b = 0.05$, the physical size ranges from 3 to 40 pc with a velocity dispersion ranging from 2 to 40 km/s. For experiments with $b = 0.20$, most substructures have physical sizes ranging from 8 to 60 pc and velocity dispersions ranging from 2 to 15 km/s, except for the biggest substructures that have physical sizes of around 170 pc and velocity dispersions of about 30 km/s.

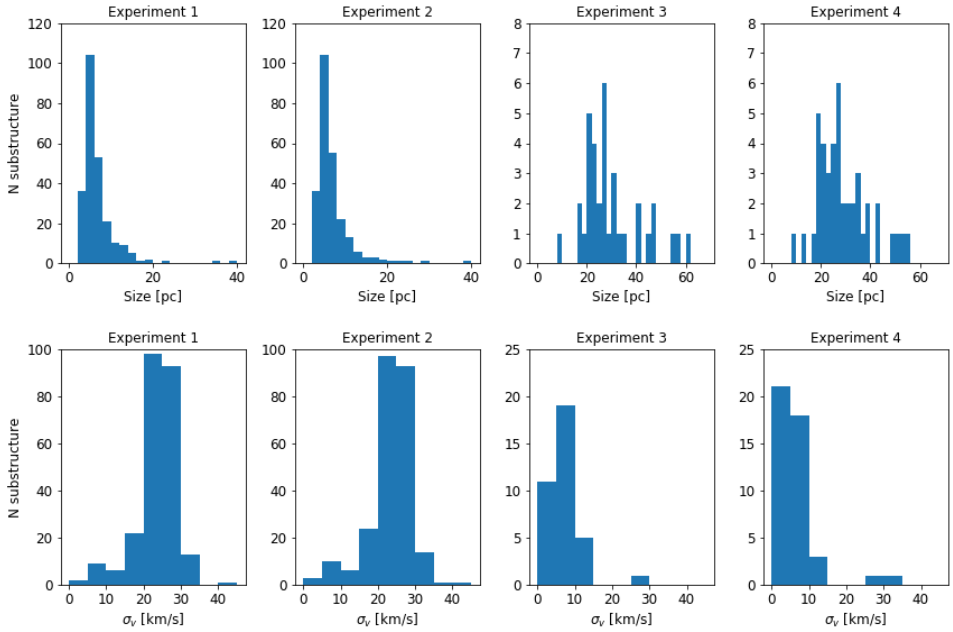


Figure 4.3: Top panel: distribution of the physical sizes of the substructures in each experiment with a bin size of 2 pc. The outliers from Experiment 3 and 4 are not included in the histogram. Bottom panel: distribution of the velocity dispersions with a bin size of 5 km/s. Experiment 1 and 2 produce more substructures with smaller sizes compared to Experiment 3 and 4, but also with larger velocity dispersions in general.

The distribution of the substructures and their velocities in Cartesian coordinates are shown in Figure 4.4. Each substructure is denoted by a blue dot and an arrow indicating its velocity. Most of the substructures are trailing behind the Sun (negative v_y) and moving towards the outside of the Galaxy (negative v_x). This is overall in good agreement with the velocity distribution of stars within a radius of $300 - 400$ pc around the Sun, as shown by the colormap in Figure 2.7. A more detailed distribution of the velocity and its comparison between the stars in the data set and stars in the substructures is shown by the histograms in Figure 4.5.

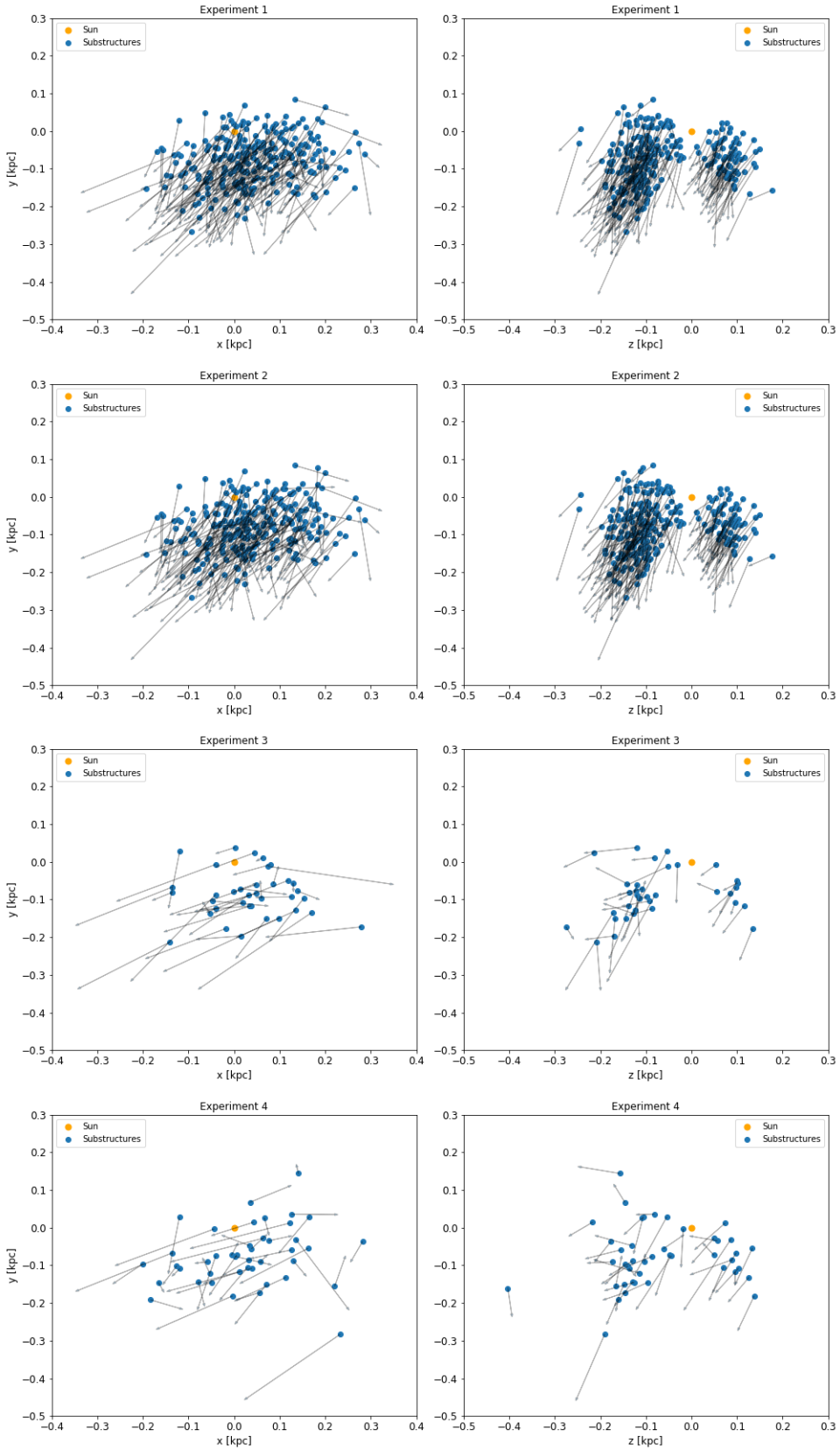


Figure 4.4: Velocity distributions of the substructures in heliocentric coordinates. Each blue dot indicates a substructure with an arrow showing its velocity. The yellow dots represent the Sun in the origin.

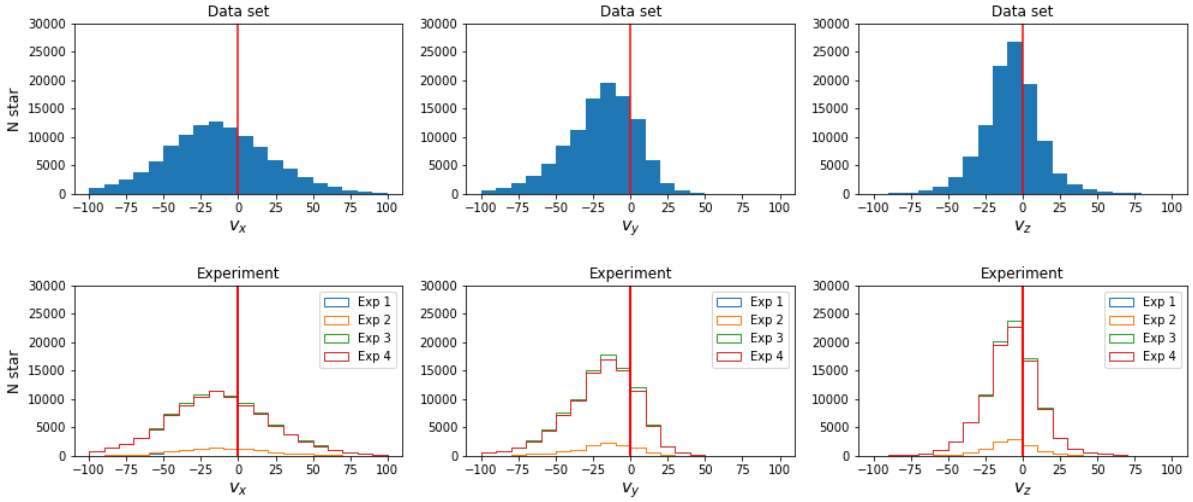


Figure 4.5: Top panel: Velocity distributions of stars from the TGAS×RAVE data set in Cartesian coordinates, centered at the Sun $(v_x, v_y, v_z = 0, 0, 0)$, which is marked by the red vertical lines. The peaks of the distributions in v_x and v_y are shifted from zero, meaning that more than half of the stars move outwards from the Galaxy (negative v_x) and trailing behind the sun (negative v_y). Bottom panel: Velocity distributions of stars that belong to the substructures in each experiment. The blue, orange, green, and red histograms represent the distributions from Experiment 1, 2, 3, and 4 respectively. These follow closely the distribution of all stars from the entire data set. The histogram from Experiment 1 might not be visible due to overlap with Experiment 2.

4.3 Commonality

In this section, we examine the substructures in terms of their star members. We do the comparison as follows: first we compare the substructures with the same b but different f, and then we compare the substructures with different b but the same f. This gives us 4 sets of comparisons as shown in Figure 4.6. After that we compare the results from 4 different sets of parameters in order to see if there are stars that always belong to the same substructure.

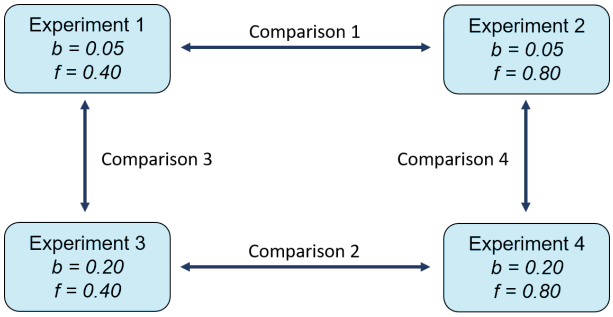


Figure 4.6: Diagram of the comparisons.

In Comparison 1, we found that around 90% of the substructures from Experiment 1 and 2 have exactly the same stars between each other (thus, exactly the same properties such as size and velocity), while the other 10% of the substructures consist of different stars. The difference is actually not because they have different stars, but because the experiments assign the stars to different substructures. For example, there is a group of stars that belong to one substructure in Experiment 1, but it breaks down into two smaller substructures in Experiment 2. On the other hand, there are some substructures from Experiment 2 that break down into smaller substructures in Experiment 1 (see Figure 4.7).

In Comparison 2, there is only one substructure that has completely identical stars between the two experiments. The rest of the stars are distributed differently in the substructures. The smaller substructures in Experiment 3 are merged into the biggest substructure in Experiment 4, while the smaller substructures in Experiment 4 are merged into the biggest substructure in Experiment 3.

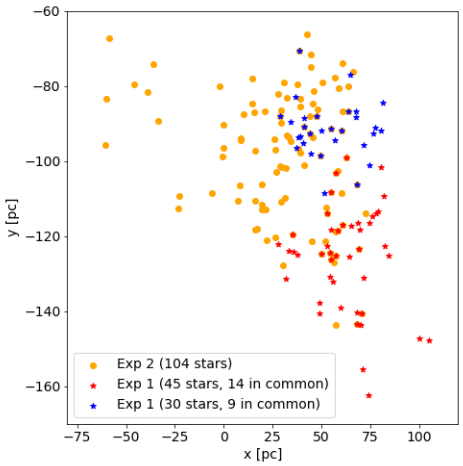


Figure 4.7: An example of a substructure that is considered as one group in an Experiment 2, but breaks down into two smaller substructures in Experiment 1.

In Comparison 3 and 4, we found that there are no substructures that share the same particle memberships at all. Some substructures are merged completely while some others are broken into smaller parts and join other stars to form bigger substructures (as seen on Figure 4.8).

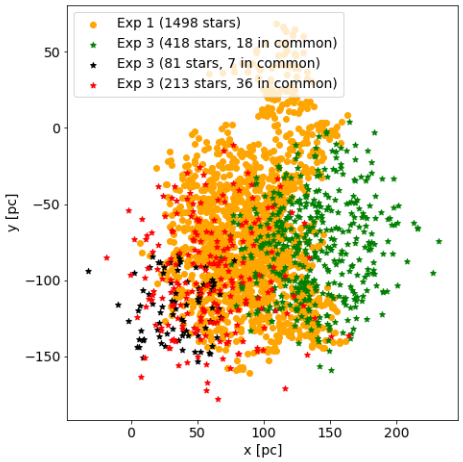


Figure 4.8: An example of a substructure that is considered as one group in Experiment 1, but breaks down into a few smaller substructures in Experiment 3.

4.4 Cross-match to catalogues of open clusters and OB-associations

We want to know if ROCKSTAR has reproduced known substructures in the Milky Way from the 4 different sets of parameters we used. To do this, we compared our results from the 4 experiments with a catalogue of open clusters and OB-associations. We used the Catalogue of Optically Visible Open Clusters and Candidates by Dias at al. (2002), containing positions, proper motions, distances and radii

for 2167 open clusters, including a few associations, up to 15 kpc away from the Sun. For the catalogue of OB-associations, we used the one listed in Melnik & Dambis (2017), which provides more accurate properties of 91 OB-associations taken from Gaia astrometric data.

We cross-matched the substructures from each experiment with both catalogues based on the positions in the sky. This can be done using TOPCAT (Tool for OPerations on Catalogues And Tables, Taylor, 2005), by comparing a table containing positions of the substructures with another table containing positions of known open clusters and OB-associations. For each row in the first table, TOPCAT finds the match in the second table by simply taking the smallest separation. Another value then needs to be set, namely the maximum separation. This value is a bit tricky to choose because: 1) the substructures found by ROCKSTAR might contain outliers that could shift the mean position in the sky, especially for big substructures, 2) the substructures might not recover the entire open cluster or OB-association given the limit of the magnitude from the data set and the particular area of sky being observed by RAVE, and 3) the projected size of the substructures in the sky may vary significantly, depending on the distance and their actual size. We can improve this by taking the median position of the ROCKSTAR substructures instead of the mean, but we need to be very generous when choosing the value for the maximum separation. For the small and very compact substructures, such as open clusters, we allow the separation to be a maximum of 2 degrees. For the more extended objects, such as OB-associations, this value is set to be a maximum of 10 degrees.

The results are given in Table 4.4. They include the information of the substructure ID given by ROCKSTAR for each experiment as well as their number of stars, position in the sky, and distance (N_{star}, l_{sub}, b_{sub}, d_{sub}), and the information of their matching counterpart from the catalogue in position, distance, and name (l_{cat}, b_{cat}, d_{cat}, Cluster/OB-A). The separation in the sky between the substructures from ROCKSTAR and from the catalogue is also given (Δ).

We found that doing a cross-match merely based on position in the sky is not enough to give reasonable matches between known substructures and those found by ROCKSTAR. Some substructures, despite appearing quite close in the sky from the known counterpart, are actually located very far away from each other, ranging from 1-4 kpc. Taking into account the distances, we confirmed that only two open clusters are properly recovered by ROCKSTAR: Melotte 22 (Pleiades) and NGC 2632 (Praesepe or Beehive cluster). Possibly there is an OB-association discovered, which is SCO OB2 from Experiment 1 and 2. Even though the ROCKSTAR substructure is located about 20 pc away from the reported location, SCO OB2 is actually an association consisting of at least 3 different regions in the sky with slightly different distances compared to each other. Brown et al. (1999) reported that the 3 subgroups of SCO OB2: Upper Scorpius, Upper Centaurus Lupus, and Lower Centaurus Crux, are located in 145 ± 2 pc, 140 ± 2 pc, and 118 ± 2 pc respectively. Knowing that the subgroups within SCO OB2 alone can be 20 pc away from each other, it is possible that ROCKSTAR might have recovered this association.

Table 4.3: Cross-match between ROCKSTAR substructures and open clusters

Exp	Sub ID	N	l_{sub} (deg)	b_{sub} (deg)	d_{sub} (pc)	Δ (degree)	l_{cat} (deg)	b_{cat} (deg)	d_{cat} (pc)	Cluster
1	100	14	166.63	-23.3	135.09	0.22	166.57	-23.52	133	Melotte 22
1	145	10	205.47	32.55	173.12	0.38	205.92	32.48	187	NGC 2632
2	102	14	166.63	-23.3	135.09	0.22	166.57	-23.52	133	Melotte 22
2	146	10	205.47	32.55	173.12	0.38	205.92	32.48	187	NGC 2632
3	6	64	166.75	-23.57	133.93	0.17	166.57	-23.52	133	Melotte 22
3	0	18	206.22	32.87	180.38	0.47	205.92	32.48	187	NGC 2632
4	35	29	166.62	-23.71	134.86	0.2	166.57	-23.52	133	Melotte 22
4	1	18	206.22	32.87	180.38	0.47	205.92	32.48	187	NGC 2632

Table 4.4: Cross-match between ROCKSTAR substructures and OB-associations

Exp	Sub ID	N	l_{sub} (deg)	b_{sub} (deg)	d_{sub} (pc)	Δ (degree)	l_{cat} (deg)	b_{cat} (deg)	d_{cat} (pc)	Cluster
1	228	47	352.42	27.96	153.07	9.0	351.31	19.02	130	SCO OB2
2	233	47	352.42	27.96	153.07	9.0	351.31	19.02	130	SCO OB2

4.5 Combining the 4 experiments

We combined the results from the 4 experiments to find substructures which are most compact in space and most coherent in velocity. We start by selecting substructures that have velocity dispersions along the principal axes less than 10 km/s. This reduces the number of substructures from 573 to 81. Among these substructures, there are some duplicates of the known open clusters of which ROCKSTAR found only the core in Experiment 1 and 2, but found more members as well as a few outliers in Experiment 3 and 4. In this case, we select one substructure that represents the core and remove those that contain outliers. We also found that many substructures are intersecting with each other, i.e. identical stars are present in 2 or more substructures, but those substructures also contain unique stars. There are at least 3 different cases which we treat differently in this selection:

1. In the case of 2 substructures intersecting in different experiments, but the intersection not being significant (less than 10%), we keep both substructures and treat them individually as separate groups. These substructures were included in the final sample.
2. In the case of 2 or more substructures intersecting in different experiments, and the intersection is significant (between 10% to 95% of the members), we select the most compact one in size, i.e. having the least number of stars, and remove the other substructures.
3. In the case of 2 substructures intersecting in different experiments, and both having the same number of stars, we select the one that has a smaller velocity dispersion along the principal axes and remove the other one.

Based on this selection, our final sample consists of 57 substructures. Figure 4.9 shows the distribution of the final substructures in the longitude-distance polar plane. Most substructures are located in the fourth quadrant, with distances ranging from as close as 100 pc from the Sun to about 400 pc.

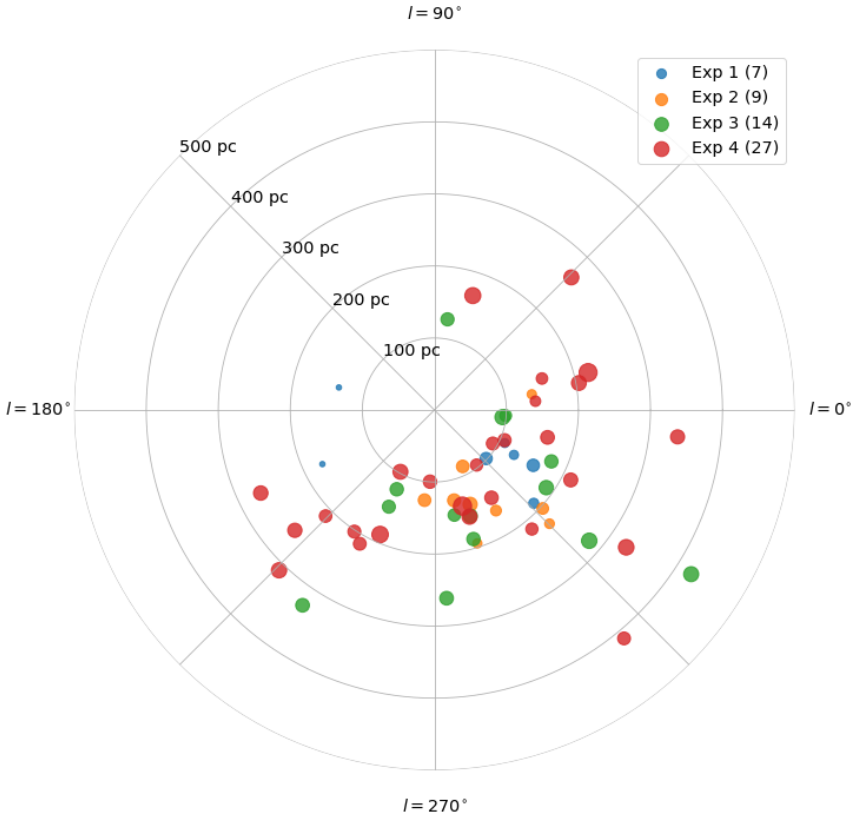


Figure 4.9: Distribution of the 57 substructures in the longitude-distance polar plane. The substructures are represented by circles, with radii equal to the physical sizes of the substructures. The colors indicate different experiments.

The distribution of the velocity dispersion along the principal axes is shown in Figure 4.10, ranked based on the smallest dispersion. Here the Experiment number is prefixed to the Substructure ID to create a more complete Substructure ID. For example, Substructure 145 in Experiment 1 will be referred to as 1.145. The first two substructures on the left are the Pleiades and Praesepe open clusters, while the rest are newly discovered by ROCKSTAR. From this distribution, we can see that there are some substructures that are very tight in 1 or 2 principal axes components (for example: Substructure 4.29, 4.23, and 4.3), while some others have isotropic dispersion (such as 4.11 or 4.34).

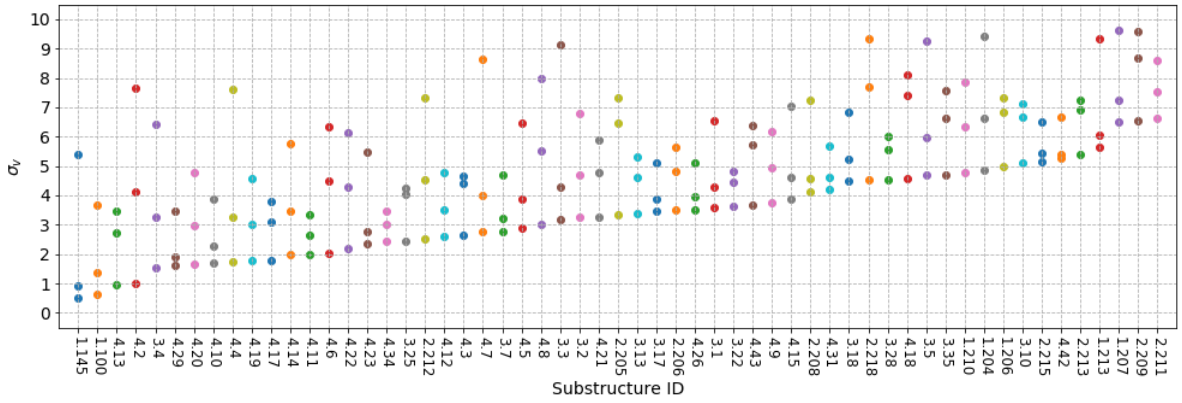


Figure 4.10: Distribution of the velocity dispersion along the principal axes. Each substructure has 3 dots representing the 3 components of velocity dispersion. The distribution is ranked from the smallest to the largest value among the 3 components. Colors are given randomly, to easily separate different substructures from one another.

We show two examples of the distribution of stars and their velocities in Cartesian coordinates in Figure 4.11. These substructures exhibit a coherent motion in space.

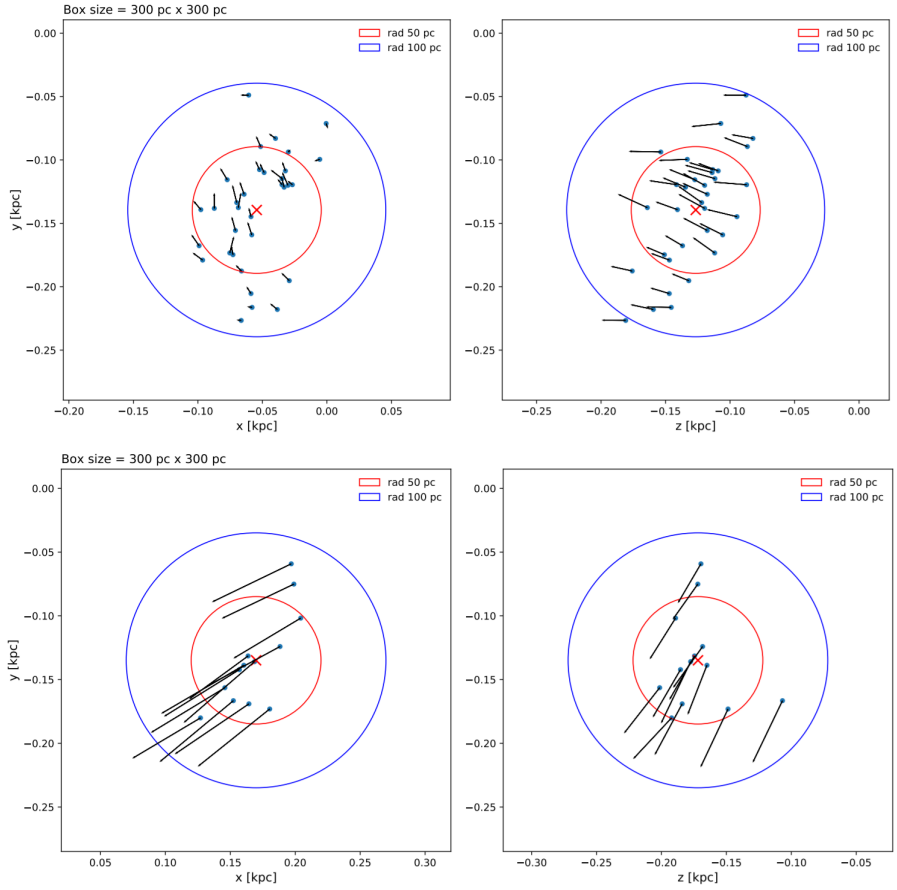


Figure 4.11: Distributions of stars in Cartesian coordinates. Top panels: an example from Experiment 4. Each star is denoted by a blue dot with its corresponding velocity denoted by an arrow. The distribution is shown in the $x-y$ plane (left) and $z-y$ plane (right). The red and blue circle has radius of 50 and 100 pc, respectively. The red cross indicates the median position of the substructure. Bottom panels: another example from Experiment 3.

We compiled a master table (see Table 4.5) for our final sample containing the properties of the substructures: positions in the sky, proper motions, physical sizes, velocity dispersions, and metallicities. The plots of the distributions and kinematics of the 57 substructures in the Cartesian coordinate system can be found in Appendix A. The plots of the observed properties (distributions of proper motions, color magnitude diagrams, metallicities and their relative errors) are available in Appendix B.

4.6 Significance test

To assess the probability of the detected substructures forming by chance, we did a significance test using a simple random sampling method. For each substructure, we defined a 3D box according to the size and the shape of the substructure in Cartesian coordinates. This box typically contains many more stars that are not members of the substructure. From this box, we drew a random sample with the same number of stars as in the substructure and calculated the velocity dispersion, both in Cartesian coordinates and the principal axes coordinates. This test was repeated ten thousand times without replacement, i.e. each star can only be selected once within a random sample. The probability is defined as $P = n/N$ where n is the number of samples having a velocity dispersion equal to that of the substructure, and N is the total number of realizations.

This significance test can be done for all except for 2 substructures (the Pleiades and Praesepe open clusters), where the stars are located in an isolated region, meaning that there are no field stars within the 3D box.

We found that all of the substructures are significant in terms of their kinematics, i.e. the probability of finding a corresponding velocity dispersion by chance is always less than one percent, if not zero (see examples in Figure 4.12).

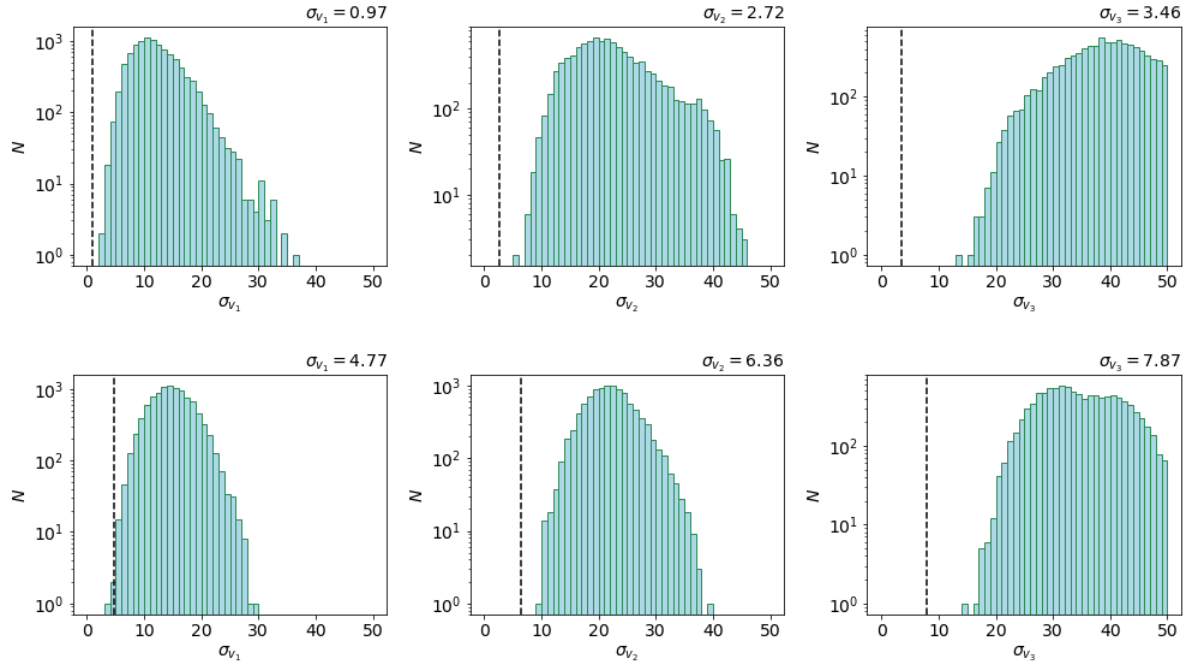


Figure 4.12: Comparison between the 3 components of velocity dispersion of a substructure along the principal axes (black dashed line) and the random sampling of field stars for 10,000 realizations (histogram). Top panel: An example of a substructure from Experiment 4, the random sampling has zero probability of finding such a small velocity dispersion. Bottom panel: An example of a substructure from Experiment 1, the random sampling has 0.04% probability of finding velocity dispersions as low as 4.77 km/s.

4.7 CMD analysis

We performed a visual analysis of the color magnitude diagrams of the substructures in our final sample. This is done by comparing the color magnitude diagrams of stars in the substructures with the field stars. First, we determine the median position of the 57 substructures in the sky. Then for each substructure, we calculate the distance of each member star to the median using the spherical law of cosine,

$$\Delta\sigma_{\rm star} = \arccos(\sin b_{\rm star} \cdot \sin b_{\rm med} + \cos b_{\rm star} \cdot \cos b_{\rm med} \cdot \cos(|l_{\rm star} - l_{\rm med}|)) \tag{4.5}$$

where $(l, b)_{\rm star}$ are the positions of each star and $(l, b)_{\rm med}$ are the median positions of the corresponding substructure in the sky. We then take the maximum value to be the radius of the field,

$$r_{\rm field} = \max(\Delta\sigma_{\rm star}). \tag{4.6}$$

By comparing the color magnitude diagram of the member stars to the field stars, we found that 27 substructures in our final sample have relatively tight distribution, as determined by our visual analysis. Figure 4.13 shows the color magnitude diagrams for substructures in Experiment 1 (Substructure 1.210, left) and Experiment 2 (Substructure 2.218, right). Both substructures have 15 and 13 stars, respectively. The stars from the substructures are denoted by blue dots, while the field stars are denoted by grey dots. We show two other examples (Substructure 4.21 and Substructure 4.34) of color magnitude diagrams of the substructures from Experiment 4 in Figure 4.14. These substructures have more members (44 and 54 stars, respectively) and also span to bigger area in the sky compare to those from Experiment 1 and 2, thus we see more stars within the field of these substructures.

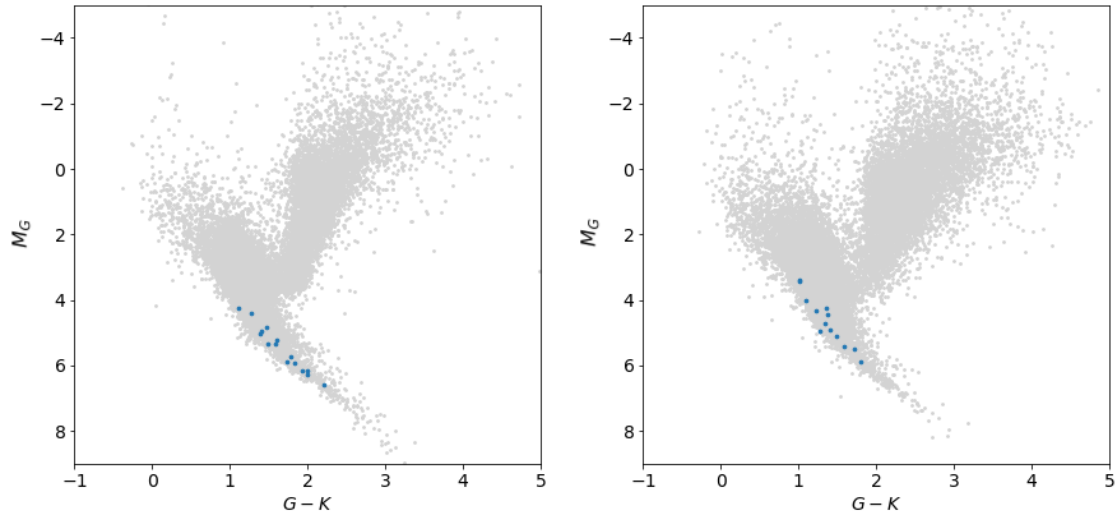


Figure 4.13: The color magnitude diagram of a substructure from Experiment 1 (Substructure 1.210, left) and Experiment 2 (Substructure 2.218, right). The stars in the substructures (blue dots) exhibit a relatively tight distribution compared to the field stars (grey dots).

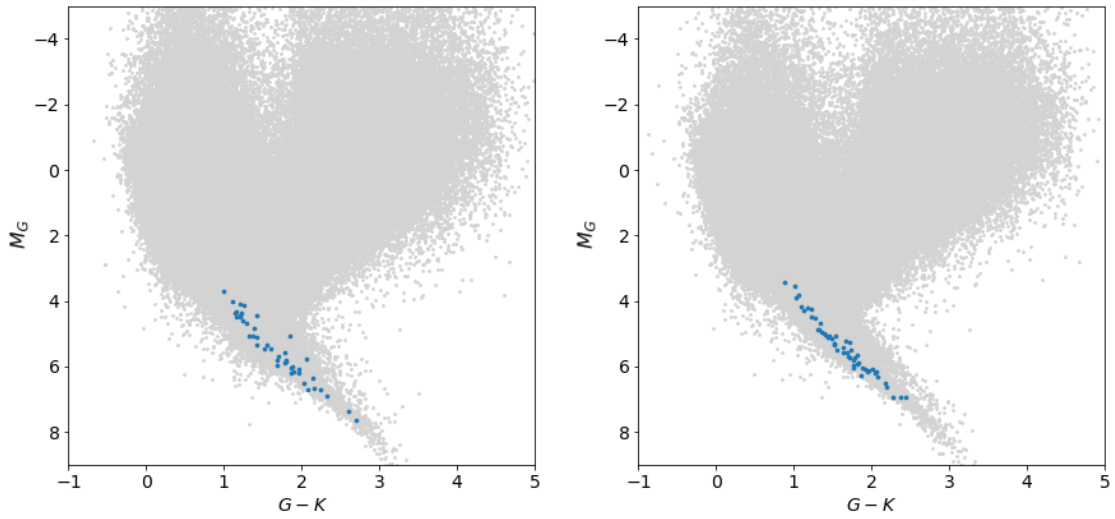


Figure 4.14: The color magnitude diagrams of two substructures (Substructure 4.21 on the left, Substructure 4.34 on the right) from Experiment 4. The colors are as described in Figure 4.13. These substructures also show tight distributions. Substructure 4.34 may contain members from the OB-association (SCO-OB2, (Brown et al., 1999)).

There are some substructures in our final sample that have very wide spread in the color magnitude diagram. We show two examples from Experiment 2 (Substructure 2.209) and Experiment 4 (Substructure 4.31) in Figure 4.15.

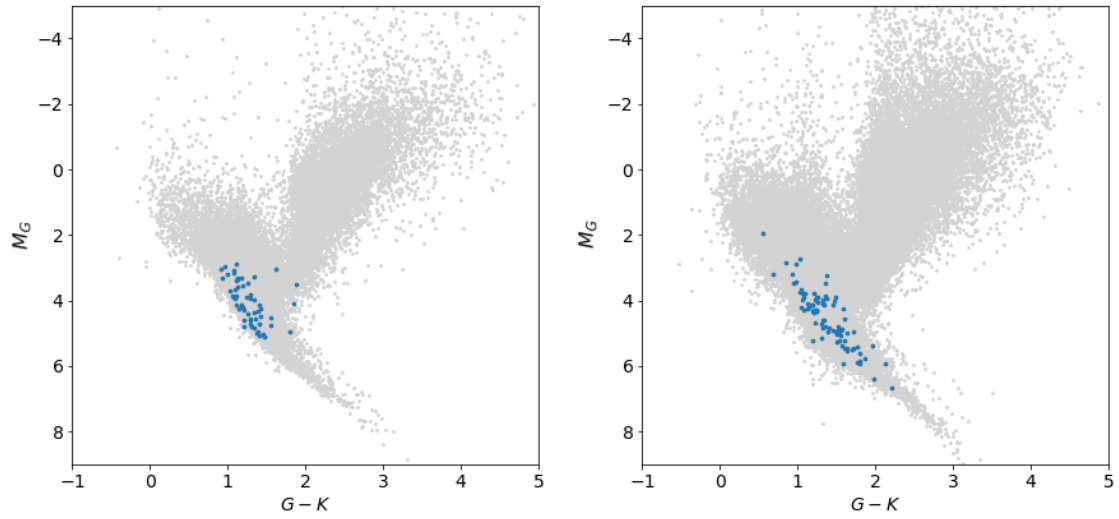


Figure 4.15: The color magnitude diagram of a substructure from Experiment 2 (Substructure 2.209, left) and Experiment 4 (Substructure 4.31, right), with the same color notations as Figure 4.13. These substructures contain 58 and 88 stars, respectively. The substructure in the left panel is located farther than the one in the right panel (200 pc versus 160 parsec), thus we can see the magnitude limitation of the stars in this substructure.

In summary, our final sample consist of substructures that have physical size ranging from 10 to 50 parsecs and 1-dimensional velocity dispersion ranging from 1 to 10 km/s. These substructures are more extended than open clusters, yet some appear to have reasonable distribution in the color magnitude

diagrams, meaning that they have the same origin and share the same age. These substructures are possibly the remnants of open clusters. Some other substructures, however, have very wide spread in the color magnitude diagram. This spread means that the stars in these substructures do not have the same origin, i.e., they were not born from the same giant molecular clouds. These substructure may be the results of dynamical effects of the Galaxy.

4.8 Finding more members in the TGAS data set

For the substructures in our final sample, we tried to find more members in the TGAS data set, which contains five astrometric quantities (positions, parallax and proper motions) without the radial velocities) for about two million stars. First, we removed stars that have negative parallaxes and relative parallax errors larger than 30% from this catalogue, resulting in one million stars with reliable quantities. A search radius is determined for each substructure based on equations 4.5 and 4.6. Then we make a selection based on the proper motions. For the TGAS stars in the search area, both proper motions in l and b should satisfy a condition in which the value is within the range of minimum and maximum values for proper motions of stars of the substructures.

$$\min(\mu_l \cos b) \leq \mu_{l,*} \cos b \leq \max(\mu_l \cos b), \tag{4.7}$$

$$\min(\mu_b) \leq \mu_{b,*} \leq \max(\mu_b), \tag{4.8}$$

where μ_l and μ_b are the proper motions of stars in the substructures, and $(\mu_l, \mu_b)_*$ are the proper motions of TGAS stars. The third selection is based on the parallax. We first calculated the median and standard deviation of the parallaxes of the stars in the substructures, and then used these values to select TGAS stars within 1σ median parallax, also taking into account the individual parallax errors of TGAS stars,

$$\frac{|\omega_{*,\mathrm{tgas}} - \omega_{\mathrm{sub,med}}|}{\sqrt{\sigma^2_{\omega_{*,\mathrm{tgas}}} + \sigma^2_{\omega_{\mathrm{sub}}}}} \leq 1. \tag{4.9}$$

We found this method works well for finding more members of the Pleiades and Praesepe open clusters. We show an example from the Pleiades in Figure 4.16. ROCKSTAR recovered 14 stars as members of the Pleiades in Experiment 1. By using the above criteria for the search radius, we found more than 1000 stars within the Pleiades field. Further selection based on proper motions and parallaxes gives 64 TGAS stars as additional members of the Pleiades. The stars in the color magnitude diagram also follow a very thin line, a characteristic of an open cluster.

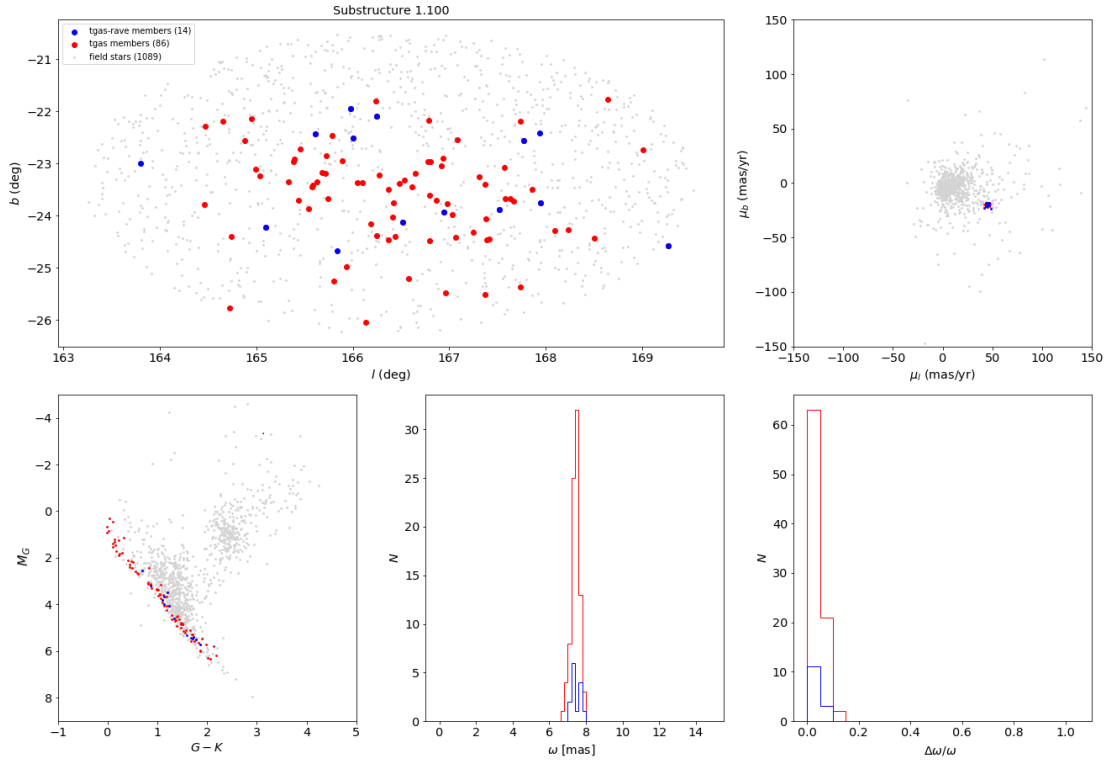


Figure 4.16: The properties of stars in the Pleiades open clusters. Top left: distribution of stars found by ROCKSTAR as the Pleiades members (blue dots), additional members based on the selection using the five astrometric quantities from TGAS (red dots), and field stars within the search radius (grey dots). Top right: distribution of the proper motions, showing a small, concentrated area in the vector point diagram. Bottom panels, from left to right: color magnitude diagram, distribution of parallax, distribution of relative parallax error.

We also found more members for other substructures. An example is shown in Figure 4.17. ROCKSTAR found 13 stars for Substructure 4.23 in Experiment 4, which span to almost a hundred degrees in the sky and more than 30 mas/yr in the proper motions. The search radius contains more than a hundred thousand stars. Further selection based on proper motions and parallaxes give about 150 stars in the TGAS data set that may belong to this substructure. The stars in the color magnitude diagram show a relatively tight distribution compared to the field stars.

We show another example in Figure 4.18, of Substructure 2.211 in which this method did not work very well. This substructure has 93 stars found by ROCKSTAR. Its position in the sky ranges over 70 degrees. The proper motions and the parallaxes of the stars have wide distributions, giving a large number of stars based on these cuts. We found more than 1800 stars classified as additional members from the TGAS data set. However, the results are not very convincing, as we can see that the color magnitude diagram shows a very wide scatter.

We found that our selections based on positions, proper motions, and parallaxes are not always reliable for finding more members in the TGAS data set. This method is difficult to apply when the substructures have a very wide span over the sky, or have a wide scatter in the vector point diagram and its parallax distribution.

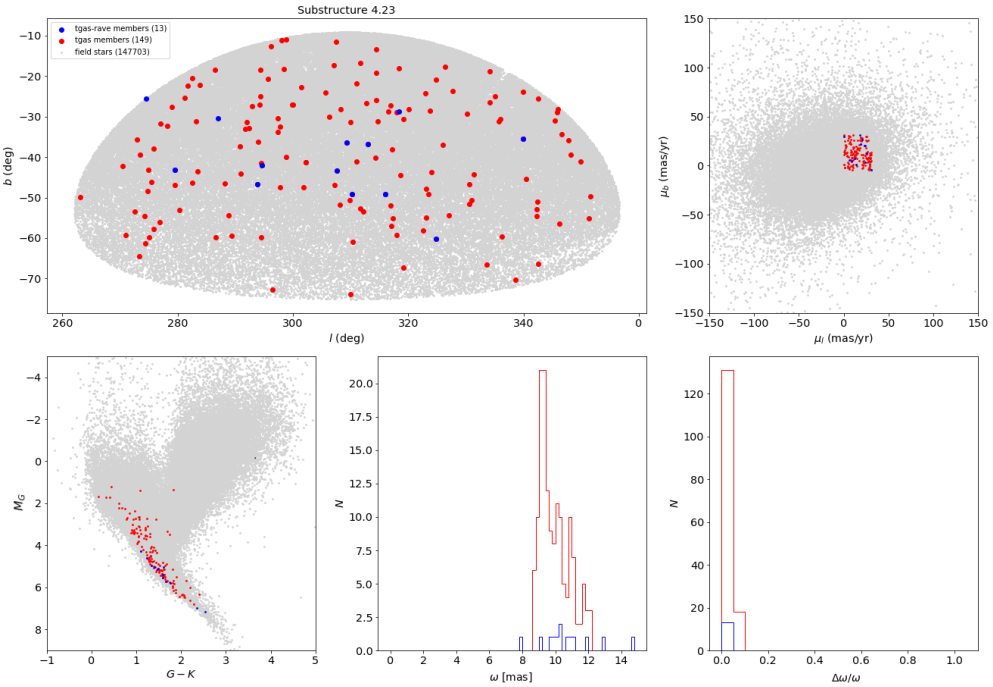


Figure 4.17: Substructure 4.23 in Experiment 4, with the same notations as Figure 4.16. More than a hundred stars are assigned as additional members from TGAS based on the above selections.

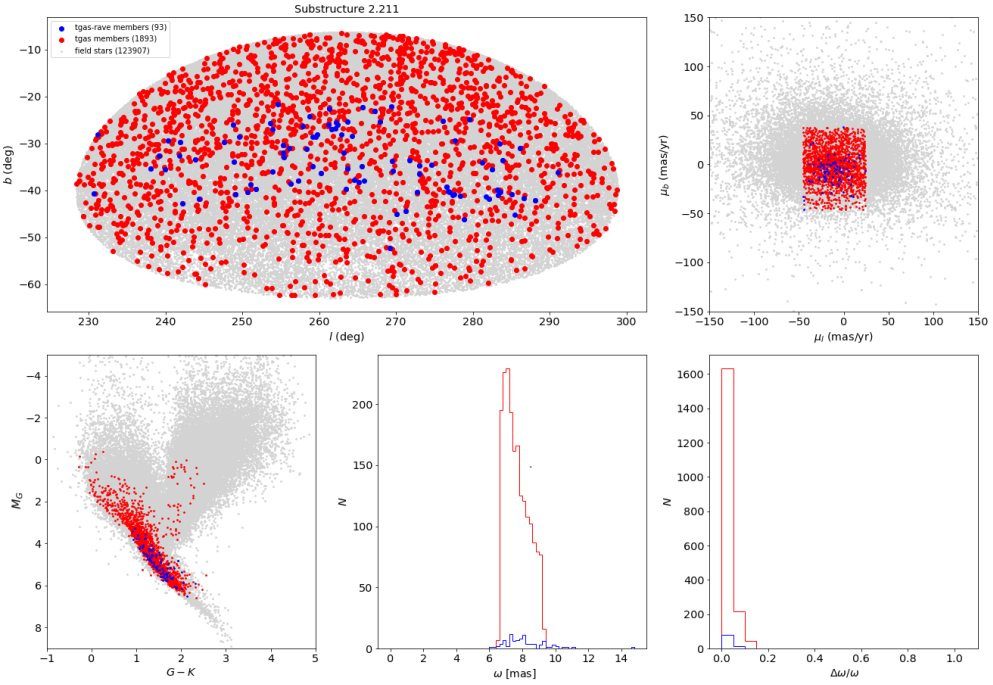


Figure 4.18: Substructure 2.211 in Experiment 2, with the same notations as Figure 4.16. The method for finding more members did not work very well for this substructure, as can be seen from the spread in the color magnitude diagram.

Table 4.5: Master Table of the Substructures

Sub ID	N star	l [deg]	b [deg]	d [kpc]	Size [pc]	σ_v [km/s]	v_r [km/s]	$\mu_l \cos b$ [mas/yr]	μ_b [mas/yr]	[M/H]
1.100	14	166.71	-22.99	0.14	3.46	2.30	2.72	45.21	-20.26	-0.22
1.145	10	-154.21	33.49	0.17	3.45	3.17	32.61	0.81	-38.02	0.13
1.204	22	-29.43	-45.36	0.13	10.25	7.20	-7.92	-39.26	-32.30	-0.30
1.206	20	-43.19	-26.70	0.19	12.27	6.46	-0.82	-32.55	-10.13	-0.21
1.207	54	-29.23	-35.47	0.16	18.86	7.91	9.69	3.90	-5.92	-0.26
1.210	15	-25.16	-54.50	0.11	9.72	6.46	6.27	-22.07	-0.54	-0.27
1.213	66	-43.16	-34.47	0.10	18.35	7.21	7.42	15.80	-4.69	-0.21
2.205	17	-72.15	-48.88	0.19	10.40	5.97	5.17	-14.58	-5.23	-0.16
2.206	11	-44.64	-38.59	0.22	11.35	4.74	-4.30	-20.57	-1.15	-0.21
2.208	16	-58.50	-38.93	0.16	14.09	5.49	29.45	15.13	6.52	-0.28
2.209	58	-42.25	-35.50	0.20	16.29	8.36	12.12	1.22	-6.97	-0.37
2.211	93	-96.22	-33.78	0.13	19.28	7.63	6.31	-13.25	-7.02	-0.25
2.212	24	-63.34	-41.02	0.09	19.19	5.17	0.77	-84.61	6.87	-0.03
2.213	104	-69.22	-40.51	0.14	24.10	6.56	3.88	-54.75	3.14	-0.18
2.215	66	-77.70	-36.62	0.13	20.61	5.71	7.14	26.11	-9.94	-0.31
2.218	13	9.17	-32.83	0.14	10.52	7.46	22.39	-3.38	1.94	-0.23
3.1	21	-115.38	-36.19	0.12	21.74	4.97	57.15	-21.97	28.59	-0.06
3.2	13	-40.18	-38.35	0.28	28.76	5.12	-2.73	-45.38	-21.55	-0.23
3.3	15	-124.02	-40.78	0.33	22.74	6.11	32.73	-12.94	16.17	-0.21
3.4	16	-32.59	-40.39	0.42	27.31	4.26	-23.37	-11.47	-9.01	-0.15
3.5	28	-5.99	-35.09	0.10	27.51	6.92	36.31	-15.23	34.11	-0.26
3.7	48	-71.26	-48.56	0.16	23.54	3.67	7.45	-40.52	12.47	-0.10
3.10	79	-86.28	-40.39	0.26	22.66	6.36	7.50	-14.59	-6.65	-0.23
3.13	32	81.76	-71.32	0.13	20.96	4.50	17.87	10.43	-14.08	-0.22
3.17	41	-34.76	-31.28	0.19	25.08	4.21	-7.87	-10.42	-9.38	-0.20
3.18	81	-73.14	-47.59	0.19	21.67	5.61	4.87	-22.30	-11.07	-0.34
3.22	43	-79.24	39.97	0.15	19.02	4.31	0.65	-19.31	-4.21	-0.18
3.25	20	-4.48	32.80	0.10	16.96	3.67	-14.77	-12.84	5.17	-0.22
3.28	38	-115.20	-37.89	0.15	20.54	5.41	0.91	31.22	-15.19	-0.28
3.35	63	-23.72	35.08	0.18	21.40	6.42	-2.86	-3.02	-4.93	-0.33

Master Table of the Substructures (cont.)

Sub ID	N star	l [deg]	b [deg]	d [kpc]	Size [pc]	σ_v [km/s]	v_r [km/s]	$\mu_l \cos b$ [mas/yr]	μ_b [mas/yr]	[M/H]
4.2	12	-13.59	29.65	0.16	22.79	5.05	13.26	-33.87	-48.41	-0.59
4.3	11	-50.29	-27.06	0.41	19.97	4.01	7.11	-26.30	-5.69	-0.37
4.4	11	44.07	-34.48	0.26	27.16	4.89	12.86	1.70	-8.34	-0.29
4.5	20	-134.10	-32.08	0.31	28.21	4.66	-9.55	10.03	-1.54	-0.28
4.6	27	71.43	-62.31	0.17	31.05	4.63	4.84	-16.72	8.86	-0.27
4.7	30	-35.59	-30.21	0.33	29.39	5.72	6.15	8.69	0.61	-0.21
4.8	24	-6.31	-31.09	0.34	24.03	5.86	2.20	-5.99	-4.08	-0.16
4.9	60	-154.43	-32.59	0.27	25.26	5.07	15.54	2.63	1.23	-0.18
4.10	13	-56.94	-65.04	0.15	22.03	2.78	9.21	-9.52	-20.28	-0.29
4.11	16	-29.82	32.24	0.09	20.37	2.71	-27.00	-68.94	7.74	-0.11
4.12	35	-73.69	40.40	0.14	39.85	3.73	-9.76	-21.44	-20.49	-0.20
4.13	11	4.98	31.66	0.14	13.43	2.60	-45.07	-13.66	19.50	-0.18
4.14	16	-50.68	37.55	0.21	18.12	4.05	-5.12	-25.86	-6.51	-0.30
4.15	35	-119.13	-36.44	0.21	19.85	5.34	4.93	14.03	0.17	-0.23
4.17	33	-113.55	-40.31	0.19	32.75	3.01	8.22	-4.81	-16.65	-0.21
4.18	57	-139.23	-29.77	0.26	24.58	6.87	5.59	3.15	-7.75	-0.31
4.19	14	-123.34	-59.48	0.20	20.34	3.32	9.79	-2.61	-3.53	-0.30
4.20	14	-135.72	-39.46	0.21	19.19	3.39	1.74	0.22	4.26	-0.23
4.21	44	-93.48	30.93	0.10	21.67	4.77	-5.15	1.06	-2.04	-0.21
4.22	20	16.31	-30.36	0.16	16.04	4.51	18.23	-8.94	5.92	-0.26
4.23	13	-52.49	-43.83	0.10	18.12	3.79	-6.68	17.45	13.93	-0.28
4.26	96	13.70	-70.12	0.22	39.03	4.24	4.80	-2.05	-14.28	-0.22
4.29	20	-27.16	-46.66	0.21	23.59	2.47	-0.12	-10.14	-17.56	-0.23
4.31	88	-71.77	41.03	0.16	27.65	4.87	-4.30	-49.57	-11.92	-0.05
4.34	54	-118.77	-27.95	0.10	26.99	3.00	28.89	17.03	-3.72	-0.23
4.42	88	10.58	-30.66	0.20	26.76	5.81	-10.53	-18.82	-12.76	-0.20
4.43	34	-23.29	32.30	0.11	20.84	5.38	-1.22	12.40	-21.92	-0.25

Chapter 5

Conclusion

5.1 Conclusion

We used a friends-of-friends algorithm to find substructures in the cross-matched data set between TGAS and RAVE based on the 6-dimensional phase space. A quality cut has been applied to the data set to select stars with reliable physical properties. We have shown that friends-of-friends algorithm can recover the known substructures in the data set whose properties are comparable to those from previous studies. We have conducted experiments to test the appropriate choice of linking parameters in order to optimize the outputs. We found more than half a thousand substructures from different experiments, and we have selected substructures that are most compact in space and most coherent in velocity, resulting in 57 substructures as our final sample. A significance test has been done to see the probability of finding substructures with such properties by chance. We concluded that all of the substructures in our final sample are significant. We performed a visual analysis of the color magnitude diagrams to further check if the substructures may have the same origin. Based on the physical sizes, velocity dispersions, and color magnitude diagrams, we concluded that these substructures are likely to be candidates of open cluster remnants. We applied a selection based on 5-dimensional quantities to find more members in the TGAS data set. We found this method can only perform well to the substructures that have tight distributions in the positions, proper motions, and parallaxes. We summarized the properties of the 57 substructures in our final sample in a master table.

5.2 Future prospects

In principle, one can use the ROCKSTAR algorithm to find substructures in 5-dimensional phase space, using positions, parallaxes, and proper motions, without the information of the radial velocities. This will be possible by modifying the weighting parameters for phase space density metric.

In the meantime, Gaia collects spectra for all sources brighter than $G \approx 17$ using its Radial Velocity Spectrometer (RVS). Although the results from this instrument are not published in Gaia DR1, the median radial velocities are expected to be available in the second data release for sources brighter than $G = 12$. We can expect to find more substructures using this information of radial velocities to complement the highly accurate positions, parallaxes, and proper motions from Gaia.

Acknowledgement

Many people have played a role in the realization of my Master's thesis. First I'd like to thank my supervisors, Jovan Veljanoski and Lorenzo Posti. You were literally around the corner when I had a question, and your day-to-day help was of great importance. The weekly discussions we had were particularly useful for the progress of the project. Secondly, I would like to thank my professor, Amina Helmi. Your scientific insight was the foundation of this project, and your support throughout has been invaluable. Thank you for your patience and guidance, especially when the deadline was closing in. I hope we can work together again, sometime in the future.

Just writing and reading papers is not all that comes with a Master's thesis. Amina's group has shown me what it feels like to be in a research environment, which is important to me, since I want to continue in academia. Thank you for the interesting discussions and adding perspective to my time here at Kapteyn. Similarly I want to thank the students in the student office, for the nice time we had studying and having fun together. I am also very grateful for the emotional support from Sebastiaan, and his contribution of the beautiful cover image.

I'd also like express my gratitude to the Indonesia Endowment Fund for Education for their financial support for 24 months during my Master's degree. Without their help I would never have been able to study in the Netherlands, getting a Master's degree, and meeting all these wonderful people.

Bibliography

P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. The Rockstar phase-space temporal halo finder and the velocity offset of cluster cores. *The Astrophysical Journal*, 762:109, 2013.

J. Binney, B. Burnett, G. Kordopatis, P. J. McMillan, S. Sharma, T. Zwitter, O. Bienayme, J. Bland-Hawthorn, M. Steinmetz, G. Gilmore, M. E. K. Williams, J. Navarro, E. K. Grebel, A. Helmi, Q. Parker, W. A. Reid, G. Seabroke, F. Watson, and R.F.G.Wyse. New distances to RAVE stars. *Monthly Notices of the Royal Astronomical Society*, 437:351, 2014. doi: 10.1093/mnras/stt1896.

J. Bovy and D. W. Hogg. The Velocity Distribution of Nearby Stars from Hipparcos Data. II. The Nature of the Low-velocity Moving Groups. *The Astrophysical Journal*, 717:617–639, 2010. doi: 10.1088/0004-637X/717/2/617.

A. G. A. Brown, A. Blaauw, R. Hoogerwerf, J. H. J. de Bruijne, and P. T. de Zeeuw. OB Associations. 540:411, 1999.

G. Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, F. Mignard, R. Drimmel, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, and et al. Gaia Data Release 1 - Summary of the astrometric, photometric, and survey properties. *Astronomy and Astrophysics*, 595(A2), 2016.

C. Conrad, R.-D. Scholz, N. V. Kharchenko, A. E. Piskunov, S. Röser, E. Schilbach, R. S. de Jong, O. Schnurr, M. Steinmetz, E. K. Grebel, T. Zwitter, O. Bienaymé, J. Bland-Hawthorn, B. K. Gibson, G. Gilmore, G. Kordopatis, A. Kunder, J. F. Navarro, Q. Parker, W. Reid, G. Seabroke, A. Siviero, F. Watson, and R. Wyse. A RAVE investigation on Galactic open clusters II. Open cluster pairs, groups and complexes. *Astronomy and Astrophysics*, 600(A106), 2017. doi: 10.1051/0004-6361/201630012.

W. S. Dias, B. S. Alessi, A. Moitinho, and J. R. D. Lépine. New catalogue of optically visible open clusters and candidates. *Astronomy and Astrophysics*, 389:871–873, 2002. doi: 10.1051/0004-6361:20020668.

ESA, editor. *The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission*, volume 1200 of *ESA Special Publication*, 1997.

A. Helmi, J. Veljanoski, M. A. Breddels, H. Tian, and L. V. Sales. A box full of chocolates: The rich structure of the nearby stellar halo revealed by Gaia and RAVE. *Astronomy and Astrophysics*, 598 (A58), 2017. doi: 10.1051/0004-6361/201629990.

E. Høg, C. Fabricius, V. V. Makarov, S. Urban, T. Corbin, G. Wycoff, U. Bastian, P. Schwekendiek, and A. Wicenec. The Tycho-2 catalogue of the 2.5 million brightest stars. *Astronomy and Astrophysics*, 355:L27–L30, 2000.

R. Hoogerwerf and L. A. Aguilar. Identification of moving groups and member selection using Hipparcos data. *Monthly Notices of the Royal Astronomical Society*, 306:394, 1999.

G. Kordopatis, A. Recio-Blanco, P. de Laverny, A. Bijaoui, V. Hill, G. Gilmore, R. F. G. Wyse, and C. Ordenovic. Automatic stellar spectra parameterisation in the IR Ca ii triplet region. *Astronomy and Astrophysics*, 535:A106, 2011. doi: 10.1051/0004-6361/201117372.

A. Kunder, G. Kordopatis, M. Steinmetz, et al. The Radial Velocity Experiment (RAVE): Fifth data release. *The Astrophysical Journal*, 153:75, 2017.

H. B. Liu, B. C. Hsieh, P. T. P. Ho, L. Lin, and R. Yan. A new galaxy group finding algorithm: Probability friends-of-friends. *The Astrophysical Journal*, 681(1046), 2008.

A. M. Mel'nik and A. K. Dambis. Kinematics of OB-associations in Gaia epoch. *Monthly Notices of the Royal Astronomical Society*, 472:3887–3904, 2017. doi: 10.1093/mnras/stx2225.

D. Michalik, L. Lindegren, and D. Hobbs. The tycho-gaia astrometric solution: How to get 2.5 million parallaxes with less than one year of Gaia data. *Astronomy and Astrophysics*, 574:A115, 2015.

S. More, A. V. Kravtsov, N. Dalal, and S. Gottlober. The overdensity and masses of the friends-of-friends halos and universality of halo mass function. *The Astrophysical Journal Supplement Series*, 195:4, 2011.

V. K. Narayanan and A. Gould. Correlated Errors in HIPPARCOS Parallaxes toward the Pleiades and the Hyades. *The Astrophysical Journal*, 523:328–339, 1999. doi: 10.1086/307716.

S. Oh, A. M. Price-Whelan, D. W. Hogg, T. D. Morton, and D. N. Spergel. Comoving Stars in Gaia DR1: An Abundance of Very Wide Separation Comoving Pairs. *The Astrophysical Journal*, 153:257, 2017. doi: 10.3847/1538-3881/aa6ffd.

K. M. Perrett, D. A. Stiff, and D. A. Hanes. Substructure in the Andromeda galaxy globular cluster system. *The Astrophysical Journal*, 589:790, 2003.

R. A. Proctor. Preliminary Paper on Certain Drifting Motions of the Stars. *Proceedings of the Royal Society of London Series I*, 18:169–171, 1869.

M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock. The Two Micron All Sky Survey (2MASS). *The Astrophysical Journal*, 131:1163–1183, 2006. doi: 10.1086/498708.

M. B. Taylor. TOPCAT and STIL: Starlink Table/VOTable Processing Software. 347:29, 2005.

K. P. Tian, F. van Leeuwen, J. L. Zhao, and C. G. Su. Proper motions of stars in the region of the Orion Nebula cluster (C 0532-054). *Astronomy and Astrophysics Supplement Series*, 118:503–515, 1996.

F. van Leeuwen. Parallaxes and proper motions for 20 open clusters as based on the new Hipparcos catalogue. *Astronomy and Astrophysics*, 497:209, 2009.

M. Zhai, H. Abt, G. Zhao, and C. Li. The Shape Evolution of Galactic Open Clusters from Observations Under Galactic External Forces. *The Astrophysical Journal*, 153:57, 2017. doi: 10.3847/1538-3881/153/2/57.

Appendix A

Plots of positions and velocities

This appendix contains plots of distribution and kinematics of the substructures found by ROCKSTAR, after the selection of the most compact in space and the most coherent in velocity (in total of 57 pages). The first panel shows the distribution in the sky, in Galactic coordinates, along with the information of the number of member from TGAS×RAVE stars, mean distance and its sigma, mean position, and the parameter used in ROCKSTAR. The second and the third panel show the distribution of stars in the Cartesian coordinate, with arrows indicating the velocity vectors of the stars, scaling with a constant number such that $v = v_* \times 10^{-3}$. The red and the blue circles define the area of radius 50 pc and 100 pc, respectively. The last rows show the histogram of the velocity in three Cartesian components, with the information of the mean velocity and its sigma.

Appendix B

Plots of proper motions, color magnitude diagrams, and metallicities

This appendix contains plots of distribution, kinematics, and photometry properties of the substructures found by ROCKSTAR. The first panel show the distribution of stars in the Galactic coordinates. The middle panel show the point vector diagram for proper motions (left) and the color magnitude diagram (right). In these plots we compare the stars in the substructures (denoted by blue dots) with the stars in the field (denoted by grey dots), as described in section 4.7. The two panels in the bottom row are the distribution of parallaxes and their relative errors.