Galactic Archaeology and Minimum Spanning Trees

Ben A. MacFarlane,[1] Brad K. Gibson,[2] and Chris M. L. Flynn[3]

[1]*Jeremiah Horrocks Institute, University of Central Lancashire, Preston, UK*
[2]*E. A. Milne Centre for Astrophysics, University of Hull, Hull, UK*
[3]*Centre for Astrophysics & Supercomputing, Swinburne University, Australia*

Abstract. Chemical tagging of stellar debris from disrupted open clusters and associations underpins the science cases for next-generation multi-object spectroscopic surveys. As part of the Galactic Archaeology project TraCD (Tracking Cluster Debris), a preliminary attempt at reconstructing the birth clouds of now phase-mixed thin disk debris is undertaken using a parametric minimum spanning tree (MST) approach. Empirically-motivated chemical abundance pattern uncertainties (for a 10-dimensional chemistry-space) are applied to NBODY6-realised stellar associations dissolved into a background sea of field stars, all evolving in a Milky Way potential. We demonstrate that significant population reconstruction degeneracies appear when the abundance uncertainties approach ∼0.1 dex and the parameterised MST approach is employed; more sophisticated methodologies will be required to ameliorate these degeneracies.

1. Introduction

The underlying premise of Galactic Archaeology is that surveys provide a fossil record of the evolution of the Milky Way. Mining this record entails the search for subclustering in multi-dimensional (spatial, kinematic, chemical) datasets. For systems with long dynamical times, relatively few dimensions are needed to identify clustering; e.g., energy–angular momentum phase space alone can identify the building blocks of the stellar halo (Brook et al. 2003). Unfortunately, the dominant baryonic component of the Galaxy - the thin disk - does not fall into this somewhat 'straightforward' regime.

Our thin disk is thought to have been built by many generations of now-disrupted stellar associations, the debris from which having been subsequently scattered/migrated by a convolution of processes, including systematic spiral arm- and bar-driven 'churning', and random diffusion-like kinematic heating from giant molecular clouds. Before an association has fully disrupted, identifying stellar siblings - i.e., the parent birth cloud/association - is relatively straightforward. Spatial, kinematic, and/or phase space coherency can be maintained for a few ∼100 Myrs (depending upon cluster mass, concentration, and galactocentric radius/orbit). Unfortunately, on the ∼10 Gyr timescale of the thin disk, the combined effects of diffusive scattering and radial migration quickly wash out this coherency, making sub-clustering analysis in low-order spatial and kinematic dimensions a fruitless endeavour (at least for reconstructing the birth locations of the sea of Galactic field stars and searching for our own Sun's siblings).

To combat this dimensionality 'problem', Freeman & Bland-Hawthorn (2002) proposed the use of 10-20 dimensions of chemistry-space (or 'C-space'). Dubbed 'chemical tagging', the principle hinges on the presumption that if the gas clouds from which the now-dissolved stellar associations formed were chemically homogeneous,

even after dissolution and full configuration- and phase-space mixing of the debris, the parent clouds' chemical 'fingerprints' would remain invariant and identifiable. This presumption has been shown to hold empirically, with chemical homogeneity confirmed on an element-by-element basis (at the ∼0.1 dex level, spanning a range of nucleosynthetic processes) for >20 associations (De Silva et al. 2007b, 2009).

With homogeneity confirmed, the first 'blind' chemical tagging experiments were conducted (Mitschang et al. 2013, 2014). High-resolution spectroscopic data for field stars were analysed, to attempt a probabilistic approach to identifying cluster populations lurking in the field. Though the study presented a means to analyse the datasets from next-generation surveys, no means for explicitly determining cluster/association recovery percentages was presented. Our goal within TraCD (Tracking Cluster Debris: Moyano Loyola et al. 2015) is to build on this pioneering work and characterise parametric and non-parametric approaches to multi-dimensional group finding within C-space, with the goal being the development of tools which can inform upcoming surveys such as GALAH, WEAVE, and 4MOST.

2. Method

As detailed by Moyano Loyola et al. (2015), our framework is a static 3-component (logarithmic halo, Plummer sphere bulge, Miyamoto-Nagai disk) potential; $\sim 10^5$ disk stars, equally spaced in ages up to 10 Gyrs old, are evolved with an N-body integrator with treatments of both random molecular cloud scattering and systematic spiral arm churning (Sellwood & Binney 2002) applied at each timestep. This background sea of stars possess kinematics consistent with those of the Milky Way. We employ four NBODY6 realisations of 250 $M_\odot$ stellar associations, each evolved in the same 3-component potential as the background stars; these are injected into the potential at various galactocentric radii r_0 ranging from 4 to 10 kpc; e.g., the dissolution time for a cluster injected at the solar circle (8 kpc) is ∼0.5 Gyrs. As stars escape the dissolving association, their trajectories are tracked with the same integrator advancing the positions of the background stars, and the same diffusion and churning treatments applied.

We tag our background field stars with empirical radial abundance patterns drawn from Luck & Lambert (2011), using 10 dimensions of C-space (Al ,Mg, Si, Ca, Ti, Cr, Co, Ni, Y, Nd). Using (of order) this number of elements ($N_C \sim 10$) provides the leverage to span the breadth of nucleosynthesis sites, while minimising the search through parameter space. Because of the imposed (i) radial abundance gradients, and (ii) constant star formation history for our disk, the applied diffusion and churning means that we consequently also impose a temporal evolution pattern to the abundance gradient (Gibson et al. 2013). The gradient's dependence on time t_0 and radial position r_0 is used to tag the mean abundance of the associations injected into the background stars at any given time and location, with the user also imposing an element-by-element scatter σ_C to each pre-disrupted system. An important difference for the latter is that the chemical 'fingerprint' pattern imposed on the pre-disrupted system is not 'random', but instead unique and homogeneous, as per De Silva et al. (2007b, 2009).

Through use of a minimum spanning tree algorithm (MST - Allison et al. (2009)), we attempt to identify the debris of disrupted satellites amongst the background stellar disk. From the 10-dimensional C-space, the MST determines a level of similarity δ_C between all single stellar components:

$$\delta_C = \sum_{C}^{N_C} \frac{|A_C^i - A_C^j|}{N_C},$$

where i and j are respective stars and A_C the abundance for element C (Mitschang et al. 2013). In order to deconvolve the matrix of δ_C values into likely parent stellar clusters, MST begins building a similarity tree. To do so, stellar components of greatest δ_C value are joined as nodes, with subsequent iterations joining less similar stellar components until all stars are placed in the similarity tree. Having created the tree, a parametric exit condition is defined in which the similarity tree is pruned to a value of δ_C where the number of clusters, k, is present. Fig 1 illustrates the building of a similarity tree for a toy distribution, in which it is built (right panel) and subsequently pruned. Due to the nature of the similarity matrix, the parameterised pruning method may dissociate tree constituents. Such dissociation events thus have no cluster association.

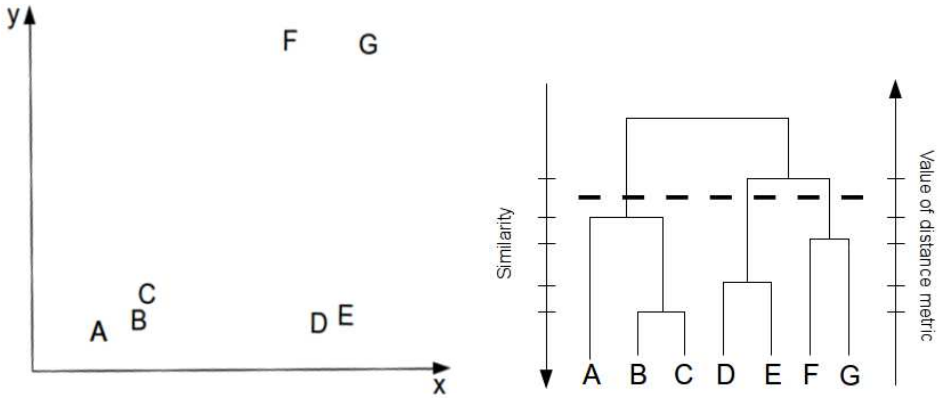

Figure 1. Toy model illustrating three clustered populations in $x-y$ space (left). MST builds a similarity tree, iteratively joining nodes of greatest δ_C until all components are included in the tree (right). A parametric cluster condition is then applied (dashed line) to define the cluster constituents.

3. Results

As a proof-of-concept, we evolved four systems with r_0 (kpc) = [4,6,8,10] for 5 Gyrs. To mimic observations, we filter stars at the end of the simulation to only include those within 3 kpc of our imposed solar neighbourhood (centred on (x, y, z)=(−8.5,0,0) kpc. To determine the success of the MST cluster 'reconstruction', four chemical abundance uncertainties σ_C (dex) were explored: [0.01,0.05,0.10,0.15]. Having tagged each star particle (see Fig 2), the MST was deployed, using a parametric exit condition of k = 4. From comparisons of the 'real' cluster constituents vs the MST associations, the successful cluster recovery rates and dissociation percentages were derived (see Tbl 1).

4. Conclusions

We employ a parametric form of the MST algorithm to search for now-dissolved stellar associations, phase-mixed with a background exponential disk of field stars. Both the background sea of stars and the associations themselves were chemically tagged with empirically-motivated abundance patterns and a variety of chemical uncertainties and association injection radii were explored. Not surprisingly, in the absence of chemical uncertainties, the association reconstruction accuracy is high; equally unsurprising, the accuracy drops dramatically with increasing abundance uncertainties. While not meant to be exhaustive, the work presented here is a successful proof-of-concept. Various weighting schemes, and non-parametric approaches, urgently need to be explored,

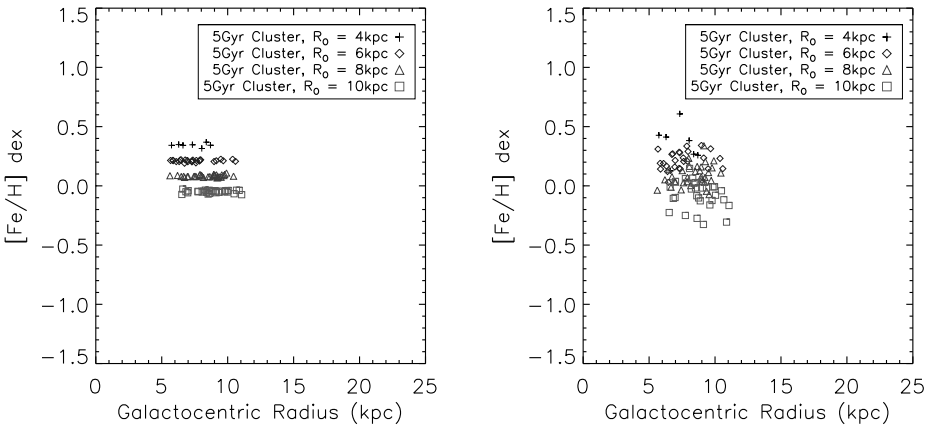

Figure 2. Present-day chemistry-radius distributions of four dissolved 250 $M_\odot$ stellar associations injected into the disk 5 Gyrs ago at 4 different galactocentric radii, with chemical uncertainties of $\sigma_C = 0.01$ dex (left) and $\sigma_C = 0.1$ dex (right).

Table 1. MST recovery accuracy and dissociation population percentages for different chemical abundance uncertainties.

σ_C (dex)	Recovery Accuracy (%)	Dissociated Population (%)
0.01	100	0
0.05	84	13
0.10	50	17
0.15	14	10

alongside more sophisticated multi-dimensional group finding algorithms (Sharma & Johnston 2012; Mitschang et al. 2013, 2014).

Acknowledgments. The support of our colleagues Jarrod Hurley, Guido Moyano Loyola, Simon Goodwin, and Rory Smith, is gratefully acknowledged.

References

Allison, R.J., Goodwin, S.P., Parker, R.J., et al. 2009, MNRAS, 395, 1449
Brook, C.B., Kawata, D., Gibson, B.K. & Flynn, C. 2003, ApJ, 585, L125
De Silva, G.M., Freeman, K.C., Bland-Hawthorn, J., et al. 2007, AJ, 113, 694
De Silva, G.M., Freeman, K.C. & Bland-Hawthorn, J. 2009, PASA, 26, 11
Freeman, K. & Bland-Hawthorn, J. 2002, ARA&A, 40, 487
Gibson, B.K., Pilkington, K., Brook, C.B., et al. 2013, A&A, 554, A47
Luck, R.E. & Lambert, D.L. 2011, AJ, 142, 136
Mitschang, A.W., De Silva, G., Sharma, S. & Zucker, D. 2013, MNRAS, 428, 2321
Mitschang, A.W., De Silva, G., Zucker, D.B., et al. 2014, MNRAS, 438, 2753
Moyano Loyola, G.R.I., Flynn, C., Hurley, J.R. & Gibson, B.K. 2015, MNRAS, 449, 444
Sellwood J.A. & Binney J.J. 2002, MNRAS, 336, 785
Sharma, S. & Johnston, K.V., 2009, ApJ, 703, 1061
Ting, Y.-S., Freeman, K., Kobayashi, C., De Silva, G., et al. 2012, MNRAS, 421, 1231