

# Group finding algorithms for galaxy cluster substructure detection from spectroscopy

Tia Bäcker

April 22nd, 2026

Galaxy clusters are the largest found gravitationally bound objects, making their observations irreplaceable in research concerning the evolution and large-scale structure of the universe. Because of their abundance in cosmological space and time scales, extensive surveys of galaxy clusters allow us to place constraints on cosmological model parameters, properties of dark energy and large-scale structure evolution. Within clusters, measuring the distribution of their members reveals information about how galaxies populate dark matter halos and their formation history.

However, the number of methods available for identification of merging events and estimating clustering statistics are limited, and their performance tests are often still in their early stages. Additionally, there are generally two options for galaxy cluster studies from galaxy observations: spectroscopic data and photometric data. Photometric redshift ( $z_{\text{photo}}$ ) measurements are significantly less precise, as they are obtained from empirical models, so substructure identification from photometric data is often limited to 2-dimensional analysis, losing information in the velocity space. Hence, spectroscopic redshift ( $z_{\text{spec}}$ ), which is calculated from the complete spectrum of an object, should be used for reliable 3-dimensional analysis. Spectroscopic data is, however, significantly limited in abundance compared to photometry, consequently making  $z_{\text{spec}}$  completeness in member catalogs a relevant factor of consideration within this study. Understanding at which completeness levels spectroscopic data compares to photometry would improve the efficacy with which  $z_{\text{spec}}$  measurements can be utilized.

My thesis aims to address these issues by investigating four group-finding algorithms for the purpose of substructure analysis using spectroscopic redshift measurements of galaxies. These group finding algorithms are built on different principles, namely shifting gapper (Lopes et al., 2009) halo-based methods (Tinker, 2021), Friends of Friends (Tempel et al., 2017), and Gaussian mixture modelling with `mclust` (Scrucca et al., 2023). A preliminary study reveals that `mclust` is the most suitable within the cluster environment. Hence this algorithm is focused on in the thesis as we proceed to more rigorous statistical analysis of its performance.

This is done using simulated large-scale cluster data from the dark matter halo simulation Uchuu (Ishiyama et al., 2021), populated with galaxies using techniques developed by Comparat et al. (2020). The subset of data is then sampled from at varying completeness thresholds in the range 10% – 100%, and `mclust` is applied to the cluster. Then, focusing on the most massive subcomponents, the recovery of substructure is investigated using five different Cluster Validity Indexes as functions of completeness. These are then compared to the results of a similar analysis done on photometric data by Sepúlveda et al. (in prep.). Preliminary results suggest that `mclust` on spectroscopic data outperforms substructure detection on photometric data already around 20%  $z_{\text{spec}}$  completeness. However, further study is needed into the effects of the underlying properties of the different simulations to make a thorough comparison. Future prospects include study of substructure dynamical properties and applying the algorithm on observational catalogs.

## References

- Comparat, J., Eckert, D., Finoguenov, A., Schmidt, R., Sanders, J. S., Nagai, D., Lau, E. T., Käfer, F., Pacaud, F., Clerc, N., Reiprich, T. H., Bulbul, E., Chitham, J. I., Chiang, C.-H., Ghirardini, V., Gonzalez-Perez, V., Gozaliasl, G., Fitzpatrick, C. C., . . . Yepes, G. (2020). Full-sky photon simulation of clusters and active galactic nuclei in the soft X-rays for eROSITA. *The Open Journal of Astrophysics*, *3*(1), Article 13, 13. <https://doi.org/10.21105/astro.2008.08404>
- Ishiyama, T., Prada, F., Klypin, A. A., Sinha, M., Metcalf, R. B., Jullo, E., Altieri, B., Cora, S. A., Croton, D., de la Torre, S., Millán-Calero, D. E., Oogi, T., Ruedas, J., & Vega-Martínez, C. A. (2021). The Uchuu simulations: Data Release 1 and dark matter halo concentrations., *506*(3), 4210–4231. <https://doi.org/10.1093/mnras/stab1755>
- Lopes, P. A. A., de Carvalho, R. R., Kohl-Moreira, J. L., & Jones, C. (2009). NoSOCS in SDSS - I. Sample definition and comparison of mass estimates., *392*(1), 135–152. <https://doi.org/10.1111/j.1365-2966.2008.13962.x>
- Scrucca, L., Fraley, C., Murphy, T. B., & Raftery, A. E. (2023). *Model-based clustering, classification, and density estimation using mclust in R*. Chapman; Hall/CRC. <https://doi.org/10.1201/9781003277965>
- Sepúlveda, G. K., Barriga, N., & Rojas, D. (in prep.). *Machine learning tools for the characterization of galaxy evolution using photometric data*.
- Tempel, E., Tuvikene, T., Kipper, R., & Libeskind, N. I. (2017). Merging groups and clusters of galaxies from the sdss data - the catalogue of groups and potentially merging systems. *AA*, *602*, A100. <https://doi.org/10.1051/0004-6361/201730499>
- Tinker, J. L. (2021). A Self-Calibrating Halo-Based Group Finder: Application to SDSS., *923*(2), Article 154, 154. <https://doi.org/10.3847/1538-4357/ac2aaa>