BAAA, Vol. 61A, 2019
R. Gamen, N. Padilla, C. Parisi, F.A. Iglesias & M.A. Sgró, eds.

Asociación Argentina de Astronomía
Boletín de artículos científicos

Siblings, friends and acquaintances: Testing the galaxy association methods

J.P. Caso[1,2] & C.A. Vega-Martínez[1,2]

[1] *Facultad de Ciencias Astronómicas y Geofísicas, UNLP, and Instituto de Astrofísica de La Plata (CONICET–UNLP), La Plata, Argentina*
[2] *Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina*

Contact / jpcaso@fcaglp.unlp.edu.ar (JPC); cnvega@fcaglp.unlp.edu.ar (CVM)

Resumen / Se presentan resultados preliminares tendientes a comprobar la exactitud de los métodos de asociación aplicados a *surveys* de galaxias, en busca de establecer sus limitaciones y determinar posibles alternativas para mejorar sus resultados. Nos enfocamos en el método de *friends-of-friends*, llevando a cabo el análisis en la simulación cosmológica de materia oscura MDPL2, del proyecto *Multidark*. Los resultados apuntan a una elevada fracción de contaminantes, en particular para halos masivos en ambientes de alta densidad. Cotas en los parámetros de asociación y la aplicación de test de subestructura pueden mitigar el número de falsos positivos.

Abstract / We present preliminary results to test the accuracy of association methods applied to galaxy surveys, in order to constraint their limitations and develop possible alternatives to improve them. We focused in the friends-of-friends (FoF) method, carrying on the analysis on the dark matter cosmological simulation MDPL2, from the Multidark project. Results point to a large fraction of contaminants from the application of the FoF method, particularly for massive haloes in high density environments. Thresholds in the association parameters and the subsequent use of tests for substructures can mitigate the occurance of fake positives.

Keywords / galaxies: statistics — distances and redshifts — clusters: general

1. Introduction

An accurate classification of the environments where galaxies belong is a key point to understand its role in galaxy evolution. Several efforts have been made to classify galaxies as probable cluster/group members, mainly by friends-of-friends (FoF) methods, but the limitations of observational astronomy might result in false positives and/or lost members. The aim of this ongoing project is to take advantage of a large sample of galaxy halos extracted from a high resolution cosmological dark matter simulation to model the available data of observational surveys and test the methods. Here we focus in the FoF method.

2. The sample

In order to achieve this goal we analyzed the MDPL2 public simulation, part of the Multidark project (Klypin et al., 2016). This simulation consists in a cubic volume of 1 h^{-1}Gpc of size, considering 3840^3 dark matter particles with mass of $1.51 \times 10^9\ h^{-1}\mathrm{M}_\odot$. The dark matter halos of the simulation were detected using the ROCKSTAR halo finder, which catalogs are also publicly available. For this project we use the list of halos found at the snapshot corresponding to the local Universe ($z = 0$). We selected halos with masses above $10^{11}\ \mathrm{M}_\odot$, and divided the simulation in cubic samples with 100 h^{-1}Mpc sides, in order to obtain samples with similar sizes than observational surveys of the nearby universe. We included an overlapped envelope of width 10 h^{-1}Mpc to avoid biases at the edges of the samples. We force each halo of the simulation to act as host of a unique galaxy. Therefore, the main halos become hosts of the central galaxies of each system, and the satellite halos are the hosts of the satellite galaxies.

We assigned to each halo a luminosity in the K band by using a simple halo occupation distribution method (HOD Vale & Ostriker, 2006), which estimates the luminosities in a non parametric way. By taking this approach, the magnitudes are obtained from the combination of the halo mass function of the simulation samples with the K luminosity function measured by Kochanek et al. (2001). In order to add uncertainties to radial velocities (V_R), we analysed V_R measurements from the 2MASS Redshift Survey (Huchra et al., 2012) to derive typical uncertainties in V_R (eV_R). Hence, we added randomly generated eV_R, assuming Gaussian distributions dependent on the K magnitude.

We applied the FoF algorithm supported on V_R and projected distances (D_p), following Crook et al. (2007). The latter value, was calculated from the angular separation and the distance in the line-of-sight estimated from the average V_R for each pair of galaxies. In order to select the thresholds for both linking parameters, we run the FoF algorithm in a subsample of the data and determine the percentage of accuracy for main haloes and satellite haloes separately. This analysis showed that setting the thresholds in $D_\mathrm{p,max} = 525$ kpc and $\Delta_\mathrm{V,max} = 1000$ km s^{-1} are commitment values to ob-

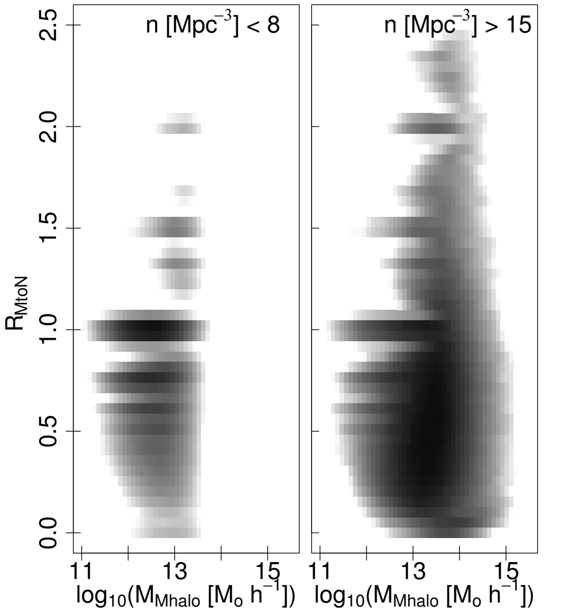

Figure 1: Distribution of ratios between real satellites and haloes associated by the method (R_{MN}) for main haloes as a function of their virial masses, for two different ranges of environmental densities.

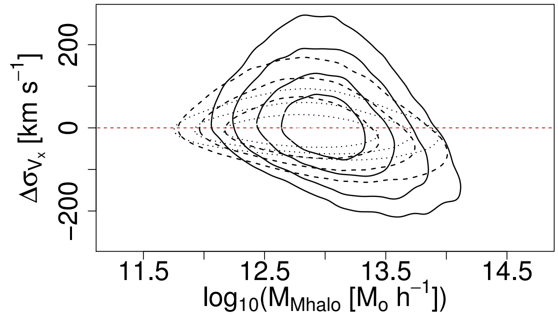

Figure 2: Distribution of the change in velocity dispersion ($\Delta\sigma_{V_R}$) for main haloes when fake positives are considered. The line types differentiate between main haloes largely contaminated ($R_{MtoN} < 0.33$, solid curves), moderately contaminated ($0.33 < R_{MtoN} < 0.66$, dashed curves) and less contaminated ($0.66 < R_{MtoN} < 1$), dotted curves).

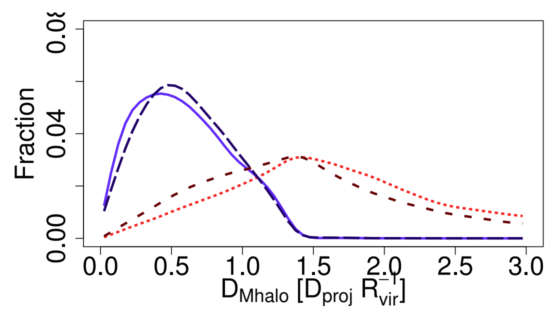

Figure 3: Distribution of projected distances to the main halo (D_{Mhalo}) in terms of the virial radius, for real satellites and fake positives in two different mass ranges. Main haloes with virial mass of $1-5\times10^{12}$ $h^{-1}M_\odot$ are represented with solid and dotted curves, respectively. Main haloes ranging $3-20\times10^{13}h^{-1}M_\odot$ are shown with long-dashed and dashed curves, respectively.

tain a high accuracy for main and satellite haloes.

3. Results

Fig. 1 presents the distribution of ratios between the number of real satellites and haloes associated by the method (R_{MtoN}) for main haloes as a function of their virial masses. We focused on main haloes with at least two satellites, assuming this as a simplistic definition of galaxy groups. The sample was split in equally populated ranges of environmental density, calculated as the number of haloes lying in a surrounding sphere of 1.75 Mpc h^{-1}. The left panel corresponds to haloes in sparse environments, with numerical densities $n < 8$. The right panel is an analogue for denser environments, typical of clusters of galaxies ($n > 15$). The grey scale ranges from the maximum of the distribution and a hundredth of it . The fraction of main haloes around $R_{MtoN} = 1$ decreases with environmental density, representing 40% and 7%, respectively. This is due to high environmental densities favor the occurrence of "fake positives", with 62% of the main haloes presenting $R_{MtoN} < 1$ in the right panel.

The occurrence of fake positives in the association method can produce changes in the velocity dispersions for a main halo (i.e., "group/cluster of galaxies‘"), with respect to that obtained from their real members. To study this effect, Fig. 2 shows the distribution of the change in velocity dispersion ($\Delta\sigma_{V_R}$) when fake positives are considered. We split the sample in three bins, depending on the value of R_{MtoN}, main haloes largely contaminated ($R_{MtoN} < 0.33$, solid curves), moderately contaminated ($0.33 < R_{MtoN} < 0.66$, dashed curves) and less contaminated ($0.66 < R_{MtoN} < 1$), dotted curves). As expected, in these latter cases $\Delta\sigma_{V_R}$ comprise a more restricted range of values. For largely contaminated main haloes, the distribution of $\Delta\sigma_{V_R}$ is significantly asymmetric, with main haloes more massive than few times 10^{13} $h^{-1}M_\odot$ preferring a decrease in σ_{V_R}.

We also analyzed the distribution of projected distances to the main halo (D_{Mhalo}) in terms of the virial radius (Fig 3), particularly for main haloes in the virial mass ranges of $1-5\times10^{12}$ $h^{-1}M_\odot$ and $3-20\times10^{13}$ $h^{-1}M_\odot$. These ranges correspond to the estimated masses for the Local Group (Carlesi et al., 2017) and nearby clusters of galaxies like Fornax (Drinkwater et al., 2001), respectively. The projected distributions for real satellites (solid and long-dashed curves, respectively) and fake positives (dotted and dashed curves, respectively) were plotted separately. In both cases the distribution of real satellites vanished at ≈ 1.5 R_{vir}, coincidentally with the typical maximum for the distribution of fake positives. These latter distributions are more extended, which might be partially solved by applying an upper limit in projected distance to the association method.

From Fig. 3 we can infer that constraints on the projected distances might be useful to reduce the number of fake positives. In order to do this, we selected as up-

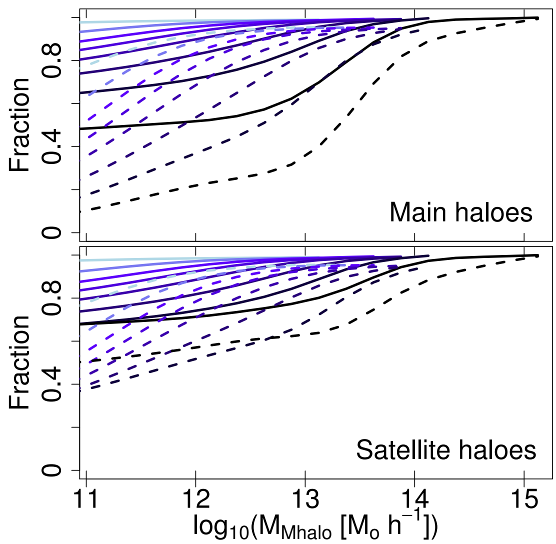

Figure 4: Fraction of accurately classified haloes, as a function of their virial masses. The gradient in the colour curves gets darker for environments with increasing numerical density. The original fractions are plotted with dashed curves, while results from the run with constraints correspond to solid curves.

per limit the 95th percentile of the projected distance from the main haloes centre to their farthest satellite $D_{\rm Mh,max}$. An equivalent upper limit can be set up for the radial velocity differences between the main haloes and their satellites. We repeated the halo association algorithm in our data sample but considering these new constraints. Although a considerable fraction of fake positives survive, results show an improvement with respect to the previous implementation. To understand the impact of considering theses constraints, the Fig. 4 shows the fraction of accurately classified main haloes (upper panel) and satellites only (lower panel), as a function of their virial masses. The gradient in the colour curves gets darker for environments with increasing numerical density. The original fractions are plotted with dashed curves, while results from the run with constraints correspond to solid curves. In general, the accuracy in the association tends to increases with virial mass, but it decreases with environmental density in both main and satellite halos. The inclusion of the new constraints increases the fraction in all the cases. The worst reported values increases from ≈ 0.1 to ≈ 0.5 for main haloes, and from ≈ 0.4 to ≈ 0.7 for satellites.

In order to analyze options to improve the accuracy of haloes association, we selected from the simulation main haloes in dense environments with more than ten satellites and virial masses from $\approx 10^{13}$ to $\approx 10^{14}$ $M_\odot$, which correspond to the mass range of nearby clusters of galaxies like Fornax and Virgo. From this sample, we selected those presenting at least five fake positives corresponding to the same main halo, representing low-mass groups wrongly classified by the method as satellites of a more massive system. When running the substructure test from Colless & Dunn (1996), in approximately 30 per cent of them, the presence of substructure cannot be rule out at the 0.1 confidence level, and the proportion increases to nearly 50 per cent for the 0.2 confidence level. A control sample was chosen from the main haloes in similar environments and mass range, but presenting a fraction of real members to fake positives from 0.9 to 1.1. From these, in 1 % of the cases the presence of substructure cannot be rule out at the 0.1 confidence level. Hence, despite the percolation algorithm results in an overpopulation of satellites for massive haloes, subsequent analysis might improve the results.

4. Summary

We use the MDPL2 dark matter simulation to reproduce observational results from FoF analysis of observational surveys of galaxies. We summarize the preliminary results in the following.

- The fraction of contaminants in groups/clusters of galaxies can be large, particularly for massive haloes in high density environments.
- These contaminants can lead to uncertain observational parameters, like radial velocity dispersions and radial projected distributions.
- The use of constraints in the algorithm mitigates the occurrence of the fake positives, but it remains in a large percentage in high density environments.
- Subsequent analysis for substructure could help to detect some cases.

Acknowledgements: This research was funded with grants from Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina (PIP 112-201101-00393), Agencia Nacional de Promoción Científica y Tecnológica (PICT-2013-0317), and Universidad Nacional de La Plata (UNLP 11-G150), Argentina.

References

Carlesi E., et al., 2017, MNRAS, 465, 4886
Colless M., Dunn A.M., 1996, ApJ, 458, 435
Crook A.C., et al., 2007, ApJ, 655, 790
Drinkwater M.J., Gregg M.D., Colless M., 2001, ApJ, 548, L139
Huchra J.P., et al., 2012, ApJS, 199, 26
Klypin A., et al., 2016, MNRAS, 457, 4340–4359
Kochanek C.S., et al., 2001, ApJ, 560, 566–579
Vale A., Ostriker J.P., 2006, MNRAS, 371, 1173