An improved method for the identification of galaxy systems: Measuring the gravitational redshift by Dark Matter Haloes

Mariano Javier de León Domínguez Romero1*, Diego García Lambas1 and Hernán Muriel1

1 Instituto de Astronomía Teórica y Experimental (IATE), Consejo de Investigaciones Científicas y Técnicas de la República Argentina (CONICET), Observatorio Astronómico, Universidad Nacional de Córdoba, Laprida 854, X5000BGR, Córdoba, Argentina.

Accepted XXX Received XXX ; in original form 10 March 2013

ABSTRACT

We introduce a new method for the identification of galaxy systems in redshift surveys based on the halo model. This method is a modified version of the K-means identification algorithm developed by Yang et al. (2005). We have calibrated and tested our algorithms using mock catalogs generated using the Millennium simulations (Springel et al. 2005) and applied them to the NYU-DR7 galaxy catalog (based on the SDSS datasets). Using this local sample of groups and clusters of galaxies we have measured the effect of gravitational redshift produced by their host dark matter haloes. Our results show radial velocity decrements consistent with general relativity predictions and previous measurements by Wojtak et al. (2011) in clusters of galaxies.

Key words: galaxies: groups - galaxies: haloes - gravitation - cosmology: dark matter

1 INTRODUCTION

In a recent article (Wojtak et al. 2011) (WHH) analyze the pattern of spectroscopic redshifts for galaxies in 7800 clusters from the 7th data release of the Sloan Digital Survey (SDSS). Motion of a galaxy in a cluster generates a red or blueshift of its spectral lines; the equivalent velocities are typically on the order of 600 kms/s. WHH show that it is possible to disentangle the superimposed gravitational redshift predicted by the General relativity that corresponds to a radial velocity differences of the order 10 km/s. In their analysis the authors used a stacking data from several clusters and determined the shift of the redshift distribution’s centroid with growing radial coordinate. The positions and redshifts of cluster galaxies are derived from a Gaussian Mixture Brightest Cluster Galaxy cluster catalogue (Hao et al. 2010), one of the largest samples of galaxy clusters assembled on the basis of the SDSS-DR7. In spite of the fact that massive clusters are expected to have the largest gravitational redshift effects, these systems are also affected by large galaxy peculiar velocities and substructure. Kim and Croft (Kim & Croft 2004) suggested that it is possible to overcome partially these difficulties by averaging over many clusters and groups of relatively low mass.

The existence of volume complete samples of galaxies with redshifts measured in the nearby universe as in the SDSS main galaxy sample (Abazajian et al. 2009); lead us to explore the gravitational redshift effects using systems of galaxies in the nearby universe. Several groups have derived catalogs of systems of galaxies applying different techniques (Friends of friends FoF, Matched Filters etc.) to large data sets such as the 2dF and the SDSS.

In order to make a meaningful comparison between observation and theory, group definitions must have a reliable three dimensional counterpart. From the point of view of current theory of galaxy formation, the most direct route for defining galaxy groups is dark matter halos. More important for this work is the novel hybrid technique introduced by Yang et al. (2005), that starts with cluster candidates selected by a classical FoF algorithm and a selected sample of bright isolated galaxies. Other satellite galaxies are classified according a Maximum Likelihood criterion. Such method employs a model of the redshift space distribution of the galaxies (as function of system mass) as filter. A convolution of the model with the galaxy distribution allows to decide the membership of each galaxy to the proposed systems (provided by a suitable and calibrated FoF method) using an iterative procedure. The method relies in an assumed mass to light relation, which is a more reliable method for compute the system mass that those based on the computation of the velocity dispersion and virial radii from galaxy positions.

Yang et al. (2004) method is a basic version of the well known K-means, a clustering algorithm where points are assigned to exactly one cluster and all points assigned to a
cluster are equals in that system. Yang have carefully tested the performance of their group finder. The method was applied to the 2dFGRS and compared with those extracted from detailed mock galaxy redshift surveys. More recently Yang et al. (2007) applied this method to the fourth release of the SDSS survey in order to study the dependence of color, stellar mass, star formation rate and morphology on halo mass.

In order to measure the gravitational redshift caused by dark matter haloes in systems of galaxies, we present in Section 2 modifications of the algorithm consisting in the introduction of a soft degree of assignment of a galaxy to each cluster. We have tested and calibrated our modified method in Section 3 with an extensive use of a galaxy mock catalog building up on the results of the Millennium simulations which mimics the SDSS-DR7 galaxy catalog. In section 4 we use a selected sample of clusters and groups of galaxies from the SDSS-NYU-DR7 datasets to compute the gravitational redshift produced by its dark matter haloes. This is followed by a brief summary. We adopt for our study a redshift produced by its dark matter haloes. This is in Section 2 modifications of the algorithm consisting in the introduction of a soft degree of assignment of a galaxy to each cluster. We have tested and calibrated our modified method in Section 3 with an extensive use of a galaxy mock catalog building up on the results of the Millennium simulations which mimics the SDSS-DR7 galaxy catalog. In section 4 we use a selected sample of clusters and groups of galaxies from the SDSS-NYU-DR7 datasets to compute the gravitational redshift produced by its dark matter haloes. This is followed by a brief summary. We adopt for our study a redshift produced by its dark matter haloes. This is

2 IDENTIFICATION OF SYSTEMS OF GALAXIES:

The aim of this section is to develop a group finder that assigns galaxies in a common halo to a single group. We will follow the methodology of Yang et al. 2003 which has advantages over other group identification algorithms, and introduce some improvements related to the K-means implementation as well as the initial seeds for the model. This method relies on the introduction of a background level B, that defines a threshold density contrast in redshift space. Ideally this value should correspond roughly to the redshift-density contrast of galaxies in redshift space. However it as a main shortcoming, is a “hard” rather than “soft” algorithm: points assigned to a cluster are equals in that cluster. Points located near the border between two or more clusters should also play a role in determining the locations of all the clusters that they could plausibly be assigned to. Nevertheless in the K-means algorithm, each borderline point is dumped in one cluster, with the same vote than all the others points in that cluster, and no vote in any other cluster. The previous criticisms of K-means motivate the “Soft K-means algorithm” used in this work (see MacKay 2003).

A fundamental hypothesis for this clustering problem is the underlying idea that the maximum likelihood method, which identifies the setting of the parameter vector $\hat{\Omega}$ that maximizes the likelihood $P(\text{Data}|\hat{\Omega}, \text{Model})$ provides a good fit to the data. Briefly, the algorithm is set in three steps: Initialization, Assignment and Update.

Step 1: Initialization In order to start the K-means, the mean values $\hat{\Omega}_{k}^{(i)}$ must be initialized in some way, we start using all the galaxies in the catalog. A representation of this assignment of galaxies to groups is given by the so called “responsibilities”, which are indicator variables $r_{k}^{(i)}$ that a given galaxy $n$ belongs partially to the $k$ group of galaxies. We firstly take all galaxies as ‘central’, and so are considered as the center of a potential dark matter halo. Using the total group luminosity $L_{\text{group}}$ and a model for the group mass-to-light ratio we can estimate the halo mass associated i.e. the halo radius $r_{200}$, the virial radius $r_{\text{vir}}$, the virial velocity $V_{\text{vir}} = (GM/r_{\text{vir}})^{1/2}$ and the line-of-sight velocity dispersion of the galaxies within the dark matter halo $\sigma = V_{\text{vir}}/\sqrt{2}$. The total luminosity of a selected potential group is estimated using (at this stage is simply the luminosity of the galaxy). $L_{\text{group}} = \sum_{i} r_{k}^{(i)} L_{i}$, where $L_{i}$ is the luminosity of each galaxy in the group, $r_{k}^{(i)}$ is the k group responsibility for each (i) galaxy (=1 initially).

Step 2: Assignment step Once we have a group centre, and a tentative estimate of the group size, mass, and velocity dispersion, we can assign satellite galaxies to this group according to the properties of the associated halos. If we assume that the phase-space distribution of galaxies follows that of the dark matter particles, the number density contrast of galaxies in redshift space $P(\text{Data}|\hat{\Omega}, \text{Model})$ around the k group centre at redshift $z_{\text{group}}$ is computed in a similar way than in Yang’s technique.

$$P(\text{Data}|\hat{\Omega}, \text{Model}) = P_{M}(R, \Delta z) = \frac{H_{0}}{c} \frac{\Sigma(R)}{\rho} p(\Delta z), \quad (1)$$

Here $\Delta z = z - z_{\text{group}}$ and $\Sigma(R)$ is the projected surface density of a (spherical) NFW halo. Hereafter $R$ is the projected distance at the redshift of the group. The func-
An improved method for the identification of galaxy systems: Measuring the gravitational redshift by Dark Matter Halo


tion \( p(\Delta z) d\Delta z \) describes the redshift distribution of galaxies within the halo for which standard Gaussian shape is adopted, with \( \sigma \) as the rest-frame galaxy velocity dispersion. Thus defined, \( P_R(R, \Delta z) \) is the three-dimensional density contrast in redshift space. In order to decide whether a given galaxy should be assigned as a primary or as a satellite into a particular group, for each galaxy we loop over all groups, and compute the corresponding distance \( (R, \Delta z) \) in phase space between galaxy and group centre. Firstly we classify the galaxies as primaries or satellites according to parameters \( k \) adopted, with \( \pi \) in order to obtain a "soft" weighted luminosity, which should be more representative of the real value. This procedure has been shown to be more accurate in comparing observed galaxy properties of models and observations. Since the median redshift of galaxies in SDSS is \( \sim 0.1 \), the \( r \)-band absolute magnitudes \( M_r \) of each model galaxy are corrected to \( z = 0.1 \) \( (M_{r,0}) \) using the \( K \)-correction code \( \text{(Kcorrect v3.1b) of Blanton \& Roweis (2007)} \) and the luminosity evolution model of Blanton et al. (2003). Galaxy redshifts are assigned by placing the observer at the corner of the simulation box and are determined by the comoving distance to the observer plus the galaxy peculiar velocity contribution. The corrected \( r \)-band magnitudes are given by: \( M_{r,0}^0 = -2.5 \times \log L + K_{\text{correction}} + E_{\text{correction}} - 5 \log h \). The survey geometry and the radial selection function were properly taken into account in the construction of the mock catalogue where we record galaxy redshifts, angular coordinates, 3D positions, apparent and absolute \( u,g,r,i \) and \( z \) band SDSS magnitudes, as well as an index to identify the galaxy assignment to a dark matter halo. In order to model the mass-weighted luminosity relation we select a common luminosity scale \( L_{19} \), defined as the luminosity of all group members brighter than \( M_{r,0}^0 = -19 + 5 \times \log(h) \). To calibrate this relation for the weighted luminosities we select from the mock SDSS catalog all groups with \( z < 0.068 \), corresponding to the volume limited magnitude given for a catalog with apparent magnitude limit of 17.5. Using the actual group centers and masses we compute the responsibilities of each group over all the galaxies in the catalogue. The use of responsibility weighted luminosities provides a slightly improved (with lower dispersion) mass-luminosity relation, compared to \( \text{(Yang et al. 2005)} \). Since the total group luminosity is dominated by the few brightest galaxies, the estimated mass is much less sensitive to the absence of faint group members (typical in flux-limited catalogs) and the effects of interlopers. This methodology allows to avoid one of the problem of the maximum likelihood, the well known phenomenon of overfitting. For all groups selected below the redshift limit \( z < 0.068 \) we compute \( L_{19} \) directly from the selected members with \( M_{r,0}^0 = -19 + 5 \times \log(h) \leq -19.0 \). For groups at higher redshifts, we compute \( L_{\text{group}} \) and use the average relation between \( L_{19} \) and \( L_{\text{group}} \) to estimate the former.

3 APPLICATION TO SIMULATED AND SDSS CATALOGS

Numerous mock catalogs have been produced from full-blown semi-analytic model of galaxy formation \( \text{(Wang et al. 2006)} \). The Millennium Simulation \( \text{(Springel et al. 2005)} \) used in this work, is one of the largest simulation of cosmic structure growth carried out so far. In what follows, we will use the semi-analytic galaxy catalogue at redshift zero constructed from the Millennium simulation by \( \text{Croton et al. (2006)} \) \( \text{(http://www.mpa-garching.mpg.de/galform/agnpaper/)} \), who considered a detailed model for cooling, star formation, supernova feedback, galaxy mergers and metal enrichment as well as a simple treatment of heating by a central AGN. These authors find that several observed properties of the galaxy population at \( z = 0 \) are suitably reproduced indicating that galaxy characteristics and stellar ages are in much better agreement with observation than for models without AGN feedback. We build up mock catalogs of the SDSS-DR7 (spectroscopic) survey in order to calibrate our system identification method as well as comparing observed galaxy properties of models and observations. Since the median redshift of galaxies in SDSS is \( \sim 0.1 \), the \( r \)-band absolute magnitudes \( M_r \) of each model galaxy are corrected to \( z = 0.1 \) \( (M_{r,0}) \) using the \( K \)-correction code \( \text{(Kcorrect v3.1b) of Blanton \& Roweis (2007)} \) and the luminosity evolution model of Blanton et al. (2003). Galaxy redshifts are assigned by placing the observer at the corner of the simulation box and are determined by the comoving distance to the observer plus the galaxy peculiar velocity contribution. The corrected \( r \)-band magnitudes are given by: \( M_{r,0}^0 = -2.5 \times \log L + K_{\text{correction}} + E_{\text{correction}} - 5 \log h \). The survey geometry and the radial selection function were properly taken into account in the construction of the mock catalogue where we record galaxy redshifts, angular coordinates, 3D positions, apparent and absolute \( u,g,r,i \) and \( z \) band SDSS magnitudes, as well as an index to identify the galaxy assignment to a dark matter halo. In order to model the mass-weighted luminosity relation we select a common luminosity scale \( L_{19} \), defined as the luminosity of all group members brighter than \( M_{r,0}^0 = -19 + 5 \times \log(h) \). To calibrate this relation for the weighted luminosities we select from the mock SDSS catalog all groups with \( z < 0.068 \), corresponding to the volume limited magnitude given for a catalog with apparent magnitude limit of 17.5. Using the actual group centers and masses we compute the responsibilities of each group over all the galaxies in the catalogue. The use of responsibility weighted luminosities provides a slightly improved (with lower dispersion) mass-luminosity relation, compared to \( \text{(Yang et al. 2005)} \). Since the total group luminosity is dominated by the few brightest galaxies, the estimated mass is much less sensitive to the absence of faint group members (typical in flux-limited catalogs) and the effects of interlopers. This methodology allows to avoid one of the problem of the maximum likelihood, the well known phenomenon of overfitting. For all groups selected below the redshift limit \( z < 0.068 \) we compute \( L_{19} \) directly from the selected members with \( M_{r,0}^0 = -19 + 5 \times \log(h) \leq -19.0 \). For groups at higher redshifts, we compute \( L_{\text{group}} \) and use the average relation between \( L_{19} \) and \( L_{\text{group}} \) to estimate the former.

3.1 Performance of the method

A real group is defined as the set of galaxies that reside in the same halo of the mock catalog. In order to quantify the group finder performance we introduce the completeness, defined as the ratio between the number of true members selected by the group finder and the total number of true
group members, and the contamination, defined as the ratio between the number of false members (interlopers) selected by the group finder and the total number of real members. We also measure the fragmentation, defined as the mean number of extra galaxy groups per dark matter halo having mass at least 0.1 times that of their true associated galaxy group and compute the purity of the systems defined as the ratio between the true members and the group members.

Figures 1 gives an overview of the completeness, contamination, fragmentation and purity for our group finder. Dots correspond to individual groups (halos), while large triangles indicate the average for all the systems included inside the window function without restrictions in the absolute magnitude limits. As can be observed, for high mass/population systems it is difficult to recover all the members in the outer regions. Nevertheless, these objects are preferentially infalling galaxies that are difficult to recover by any method. Fragmentation of groups should be considered if these subgroups comprise a considerable mass fraction. If this fraction is $\sim 10\%$ the effect is negligible for high mass systems but affects intermediate mass/occupation systems. With the modifications introduced, the improved algorithm is significantly more efficient than the popular FoF identification algorithms or its variants in terms of high completeness and purity, low contamination and fragmentation as is shown in Figure 1. Notice that our test are based on a mock sample from a full N-body SAM based model, which presumably results in a more realistic scenario than a CLF based one as is the 2dF mock catalog used by Yang et al. (2007).

### 3.2 System Identification in the SDSS-NYU-DR7

Using a dedicated 2.5 meter telescope, SDSS provide the largest and most complete photometric and spectroscopic galaxy survey available. We have used the large-scale structure sample sample14 from the NYU Value Added Galaxy Catalog (NYU-VAGC; Blanton et al. 2005) DR7 version as our primary galaxy sample, a complete sample of galaxies with reddening corrected r magnitudes brighter than 17.5. We take into account fiber collisions by giving each collided galaxy their photometric redshift estimate. In order to limit the effects of incompleteness on our group identification procedure, we restrict our sample to regions of the sky where the completeness (ratio of obtained redshifts to spectroscopic targets) is greater than 90%. Galaxy magnitudes are corrected for Galactic extinction using the dust maps of Schlegel et al. (1998), absolute magnitudes are k-corrected Blanton & Roweis (2007) and corrected for passive evolution Blanton et al. (2003) to rest-frame magnitudes at redshift $z = 0.1$. We use the fit $Q(z) = 1.6(z - 0.1)$, which takes into account the effects of the luminosity evolution. We select all galaxies with extinction corrected apparent magnitude brighter than $r = 17.5$, redshifts in the range 0.015 $\leq z \leq 0.15$, and with redshift measurement completeness $\geq 0.9$. We use a lower redshift limit $z > 0.015$ to alleviate some of the problems associated with the accuracy of the photometry of some nearby extended galaxies. For a uniform magnitude limit, our selection function $n(r)$, can be separated into the product of an angular and a radial part: $n(r) = n(\hat{r})n(\hat{r})$, where $R \equiv \hat{r} \hat{r}$ and $\hat{r}$ is a unit vector. The angular part may take any value between 0 and 1, and provides the completeness as a function of position, i.e., the fraction of all survey selected galaxies for which reliable redshifts are given. In order to compute the selection function, we use Mangle, http://casa.colorado.edu/ ajsh/mangle. Such codes (Hamilton & Teske 2004), are designed to deal accurately and efficiently with complex angular mask. We apply the procedure suggested by Berlind et al. (2006) to the groups identified by our algorithm on the NYU-SDSS-DR7 galaxy sample using the mask in order to avoid completeness issues due to the angular footprint.

### 4 A MEASURE OF GRAVITATIONAL REDSHIFTS

Using the systems of galaxies identified in the SDSS-NYU-DR7 galaxy redshift survey we measured the gravitational redshift due to the dark matter gravitational profile. A standard stacking technique is used in order to disentangle the kinematic Doppler effect from the gravitational redshift, given that the latter shifts the centroid of the observed velocity distribution. Coordinates and redshifts of cluster centers need no corrections since our identification methodology provides reliable values. We search for all galaxies within a window in phase space of 3 virial radii and 3\sigma in velocity around the cluster centres. As the final step we combine redshift data of all clusters into one. Since we are interested in the detection of the GR effect in intermediate mass systems, we have considered two samples of systems of galaxies: low mass ($10^{13} M_\odot$ to $10^{14} M_\odot$; groups of galaxies) and high mass ($10^{14} M_\odot$), comparable to the cluster sample analyzed.
An improved method for the identification of galaxy systems: Measuring the gravitational redshift by Dark Matter Halo

by WHH). Our final sample comprises systems of galaxies in the range $0.015 \leq z \leq 0.15$ with a mean redshift of 0.1 hosting on average 28 galaxies with spectroscopic redshift measurements for clusters and 14 for groups. Those groups with less than 4 redshift determinations were not included in the sample. Since the presence of interlopers in our sample is negligibly small, it is possible to measure directly the mean value of the velocity distribution of cluster galaxies, $\Delta$. Assuming spherical symmetry and no strong inhomogeneities, the gravitational redshift profile of a system can be calculated using the following formula (Cappi 1995)

$$\Delta(R) = \frac{2}{c\Sigma(R)} \int_{R}^{\infty} \![\Phi(r) - \Phi(0)] \frac{\rho(r)rdr}{\sqrt{r^2 - R^2}},$$ (3)

where $R$ is the projected cluster-centric distance, $\Phi(r)$ is the gravitational potential, $\rho(r)$ and $\Sigma(R)$ are the 3D and surface (2D) density profiles of galaxies. In order to estimate this effect for the data combined from a cluster sample, it is needed to convolve this expression with the distribution of cluster masses in the sample. In order to test the possibility of determining this effect in our sample we introduce the corresponding blueshift in the mock catalog of galaxies in their parent haloes. Given that the dark matter haloes follow an NFW density profile (Navarro et al. 1996), the gravitational potential results:

$$\Phi(r) = -(GM_v/r_v)^{1/2} g(c_v)^{1/2} \ln(1 + r/v_c)$$ (4)

and $g(c_v) = 1/\ln(1 + c_v) - c_v/(1 + c_v)$ with $c_v$ the concentration depending on the halo mass. Using formulas 3 and 4 we introduce the corresponding gravitational redshift effect by adding $\Delta$ to the line of sight velocities of the member galaxies (Hubble flow plus peculiar velocities) that populated each dark matter halo in the Millennium simulation. In the left panel of figure 2 we show $\Delta$ as a function of the projected radial distance and its comparison to the imposed effect in the systems identified in our mock catalogs. This should be compared with the measurement showed in the right panel for the SDSS-NYU-DR7 systems. Upper/Lower panels correspond to low/high mass systems. Although the predicted signal is significantly smaller for groups than for clusters of galaxies (showed in dashed lines), the quality and size of the sample of low mass systems made possible to clearly detect the gravitational redshift effect predicted by the General Relativity. We notice that the high mass results are in agreement with Woitak et al. 2011.

5 CONCLUSIONS

- Our measurements of the gravitational redshift effect by dark matter haloes show a good agreement with the predictions of General Relativity.

REFERENCES

Abazajian K. N., Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Allende Prieto C., An D., Anderson K. S. J., Anderson S. F., Annis J., Baehm N. A., et al. 2009, Astrophys. J. Suppl., 182, 543

Berlind A. A., Frieman J., Weinberg J., SDSS Collaboration 2006, Astrophys. J. Suppl., 167, 1

Blanton M. R., Hogg D. W., Bahcall N. A., Brinkmann J., 2003, Astrophys. J., 592, 819

Blanton M. R., Roweis S., 2007, Astron. J., 133, 734

Blanton M. R., Schlegel D. J., Strauss M. A., Brinkmann J., Finkbeiner D., Fukugita M., Gunn J. E., Hogg D. W., Ivezic Z., Knapp G. R., Lupton R. H., Munn J. A., Schneider D. P., Tegmark M., Zehavi I., 2005, Astron. J., 129, 2562

Cappi A., 1995, Astroparticle Physics, 301, 6

Croton D. J., Springel V., White S. D. M., De Lucia G., Frenk C. S., Gao L., Jenkins A., Kauffmann G., Navarro J. F., Yoshida N., 2006, Mon. Not. R. Astron. Soc., 365, 11

Hamilton A. J. S., Tegmark M., 2004, Mon. Not. R. Astron. Soc., 349, 115

Hao J., McKay T. A., Koester B. P., Rykoff E. S., Rozo E., Annis J., Wechsler R. H., Evrard A., Siegel S. R., Becker M., Busha M., Gerdes D., Johnston D. E., Sheldon E., 2010, Astrophys. J. Suppl., 191, 254

Kim Y.-R., Croft R. A. C., 2004, Astrophys. J., 607, 164
ACKNOWLEDGMENTS

We thank the referee for their comments and helpful suggestions for manuscript changes. These have improved both the content and clarity of the manuscript. This work has been partially supported by Consejo de Investigaciones Científicas y Técnicas de la República Argentina (CONICET) and the SeCyT-UNC. This research has made use of NASA’s Astrophysics Data System. The Millennium Run simulation used in this paper was carried out by the Virgo Supercomputing Consortium at the Computing Centre of the Max-Planck Society in Garching. The semi-analytic galaxy catalogue is publicly available at [http://www.mpa-garching.mpg.de/galform/agnpaper].

Funding for the creation and distribution of the SDSS Archive has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the NASA, the NSF, the U.S. DoE, the JM, and the MPS. The SDSS Web site is [http://www.sdss.org/].