Searching for galaxy groups

Searching for galaxy groups in photometric data from the CFHT Legacy Survey

By

Rachel Elizabeth Anderson, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2009) McMaster University

(Department of Physics and Astronomy) Hamilton, Ontario

TITLE: Searching for galaxy groups in photometric data from the CFHT Legacy Survey

AUTHOR: Rachel Anderson, B.Sc.

SUPERVISORS: William E. Harris & Laura C. Parker

NUMBER OF PAGES: x, 78

Abstract

Most galaxies in the universe live in galaxy groups, and we can use them to learn a lot about the influence of environment on galaxy evolution. Historically, galaxy groups have been identified by painstakingly obtaining many spectra and finding associated galaxies, however, much larger samples could be found if we had a reliable method of finding groups using photometry. We follow a variation of the Probability Friends of Friends algorithm suggested by Li & Yee (2008) to find galaxy groups in photometric redshift data. Our approach employs a continuous friends-of-friends search in the transverse direction and uses the redshift probability density in redshift space. We use this method to find galaxy groups in the 4 Canada-France-Hawaii Telescope Legacy Survey Deep Fields, and compare it to a 5th Nearest Neighbor galaxy density search algorithm applied to the same four fields. To further assess our Probability Friends of Friends method, we test it on light cones from the Millennium Simulation. We found that less than 20% of the Probability Friends of Friends groups with 12 or more members were interlopers. The average density of groups of 12 or more members is $\sim$ 54 per square degree. Our final sample of groups in the Canada-France-Hawaii Telescope Legacy Survey Deep Fields contains 33,279 groups with three or more members, 218 of which have 12 or more members.

Acknowledgements

Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/DAPNIA, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institute National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at TERAPIX and the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

The Millennium Simulation databases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory.

I would like to thank my supervisors, William E. Harris and Laura C. Parker, for their support and guidance. They have not only been of great assistance in regards to my research, but with respect to the coursework and other aspects of graduate life at McMaster.

Dedicated to Eric & Rebecca Knudson

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Galaxy groups

On large scales we know the universe to be both homogeneous and isotropic, but on small scales galaxies are distributed unevenly and many are bound in systems called groups and clusters – the largest gravitationally bound structures. A galaxy group contains less than ~ 50 members, has an average mass of about $2 \times 10^{13} h^{-1}$ $M_\odot$, and their spread in projected space is about $1.4 h^{-1}$ Mpc (Carroll & Ostlie, 2007). Here, h, the Hubble parameter, is defined such that the Hubble constant, H_0, is $100h$ km s^{-1} Mpc^{-1}; the current best value being $h = 0.719^{+0.026}_{-0.027}$ as determined by the WMAP 5 year data (Dunkley et al., 2009). A cluster, on the other hand, may have from 50 to thousands of galaxies, a spread in projected space of $6h^{-1}$ Mpc, and a mass on order of $1 \times 10^{15} h^{-1} M_\odot$ (Carroll & Ostlie, 2007). Though galaxy groups are harder to find than galaxy clusters, they are important for understanding galaxy evolution as most galaxies ($\sim$ 60%) at $z \sim 0$ live in groups of $N \geq 2$ (Eke

et al., 2004; Berlind et al., 2006; Tago et al., 2006). For these reasons we are interested in developing a reliable galaxy group finding method.

Another motivation for our study is that identifying galaxies in groups allows us to determine the role environment plays on galaxy evolution. There has been much support for the idea that galaxies which live in richer groups have a higher red galaxy fraction at earlier redshifts (Yee et al., 2007). Furthermore, a recent study by Li et al. (2009) tested for the dependence of cluster galaxy populations and their evolution on both local and global environment. They used the red galaxy fraction, (f_{red}), as the measure for evolutionary status. For the local environment they used the surface galaxy density, Σ_5, calculated from the distance to the 5th nearest neighbor. The global cluster environment was measured using the the Cluster-Centric Radius, r_{CL}, which is the radius to the nearest cluster normalized by the virial radius. It was found that f_{red} increased with Σ_5 in three different redshift bins from $0.15 < z < 0.30$, $0.30 \leq z < 0.40$, and $0.40 \leq z < 0.55$. Furthermore, f_{red} decreased when r_{CL} increased for the same redshift bins, but with a stronger dependence. Therefore they concluded that galaxy population and evolutionary history depend on the properties of the dark matter halo surrounding the galaxy.

In another study Deng et al. (2009) studied the effect that environment has on the properties of red early-type and red late-type galaxies. They used the distance to the 5th nearest neighbor to define a galaxy density in a comoving sphere using spectroscopic velocities. The results indicated that red early-type galaxies show a correlation between luminosity and density in the luminosity range $M_r \leq -21.95$, while the red late-type galaxies do not. Also,

the dependence between $g - r$ color and density of environment (objects in higher densities are redder) is stronger for red late-type galaxies (Deng et al., 2009). Clearly environments impact galaxy growth, and it is important to define these environments in a reliable manner.

1.2 Finding galaxy groups

In order to study the impact of environment on galaxy properties, it is necessary to develop an objective and physically meaningful method for finding galaxy groups and clusters. There are several methods for determining high and low density areas, including the friends-of-friends algorithm (e.g. Huchra & Geller (1982); Davis & Djorgovski (1985); Li & Yee (2008)) and the Nth Nearest Neighbor approach as described by Cooper et al. (2005) and Clark & Evans (1954).

For our purposes, a 'group' will be defined as a collection of galaxies that have a high probability of being spatially near each other, by having a close projection on the sky and similar redshifts. While the friends-of-friends method has been applied to data with spectroscopic redshifts, Li & Yee (2008) have developed a friends-of-friends method to find galaxy groups in photometric data using probability in redshift space. Thus, while large redshift errors would not allow us to be certain if a galaxy was at the same redshift as another, we can at least calculate the probability of them being at the same redshift and make a cut based on this probability in order to define a group. There will be some interlopers certainly, but with this method we can at least know the probability that there is a group at a certain position and redshift. Furthermore,

we are particularly interested in identifying high density regions, not knowing whether a specific galaxy is a member of a given group. This method will allow for those goals to be achieved.

1.3 Thesis objectives

Knowing that environment plays a strong role in galaxy evolution, and that the definition of the environment itself is crucial, we want to take a look at where clear advances can be made in this field. One is to create a reliable method to define these galaxy groups. The second is to find a method that can work with large datasets. We set out to accomplish these goals by modifying the Probability Friends of Friends (hereafter pFoF following Li & Yee (2008)) algorithm and applying it to the Canada-France Hawaii Telescope Legacy Survey (CFHTLS) Deep fields, and furthermore we calibrate it on light cones from the Millennium Simulation (MS) (Springel et al., 2005). Large current and future surveys rely primarily on photometry, especially at higher redshifts. By using probability in redshift space, we expand the ability of the friends of friends searches to utilize these surveys, which makes this work that much more important. Overall, photometric data is much easier to obtain for large areas than spectroscopic data, making this a valuable tool.

Chapter 1 offers a brief introduction to galaxy groups as well as the effect being in a galaxy group might have on a galaxy, different methods to look for galaxy groups, and finally what this thesis strives to demonstrate. The pFoF method is described in Chapter 2 as well as our method of finding dark halo groups in the MS light cones, and our use of the Nth Nearest Neighbor

technique. Chapter 3 describes the data used in this study and how it was manipulated for our purposes. Chapter 4 presents a summary of the results. Chapter 5 contains concluding remarks and discusses the potential for future study.

Chapter 2

Methods for finding high density areas

2.1 Probability Friends of Friends

Friends of Friends is an iterative method of finding galaxy groups with spectroscopic redshifts (Huchra & Geller, 1982). We use a variation of the pFoF algorithm suggested by Li & Yee (2008). What is unique about their method is that it can be used with photometric redshift data. The pFoF algorithm employs a friends-of-friends search in the transverse direction and uses the redshift probability density in redshift space. Our approach differs in that we continue the friends-of-friends iterative cycle until no more galaxies are added to the group, while Li & Yee (2008) employ a group combining step instead.

By using the method described by Li & Yee (2008), we make the assumption that group members will be at the same redshift as well as have a small separation in the transverse direction. Therefore, for a galaxy, A, to be considered a member of any group, it must meet two requirements:

1. The transverse distance between galaxy A and at least one member of the group must be less than or equal to that group's linking length, $D0$.

2. The probability that galaxy A is at the same redshift as the group, measured by P_{ratio}, must be greater than or equal to the probability cutoff, P_{ratio_crit}.

The remainder of this section is dedicated to explaining the parameters used in the requirements listed above. For the first, $D0$ is defined such that:

$$D0 = \sqrt{\frac{R_w \sum_i^N w_i}{N}} \frac{D0_{xy}}{1+z} \tag{2.1}$$

Here, $D0_{xy}$ is the linking length at a redshift of zero. With a larger $D0_{xy}$ more groups will be found, however the number of interlopers will also increase. If $D0_{xy}$ is too small we will miss many groups. Equation **??** has two completeness weights, one to correct for the varying magnitude limit throughout the survey, R_w, and the other to correct for unreliable redshifts, w_i. R_w is a scaling factor to make up for the fact that the apparent magnitude cutoff will cause fewer galaxies to be detected at higher redshift. It is defined as:

$$R_w = \left(\frac{\int_{-\infty}^{M_{cut}} \Phi(M_i) \mathrm{d}M_i}{\int_{-\infty}^{M_{lim}} \Phi(M_i) \mathrm{d}M_i} \right) \tag{2.2}$$

where $\Phi(M_i)$ is the luminosity function where M_{cut} is the absolute magnitude limit we aim for, while the apparent magnitude limit of the survey leads to the absolute magnitude limit, M_{lim}. If M_{cut} is brighter than M_{lim}, $R_w = 1$. The CFHTLS data we are working with is plenty deep out to our upper redshift limit of 1.0, thus we set $R_w = 1$. The completeness correction weight, w_i, is

calculated for each galaxy as the ratio of the total number of galaxies within $\delta m_i = 0.1$ magnitude of that galaxy's magnitude, to the total number of galaxies selected in the same magnitude range. The selection criterion is that the probability of a galaxy's redshift to be within $3\sigma_{z_{cut}}$ of its photometric redshift must be greater than 99.7%, where $\sigma_{z_{cut}} = 0.2(1+z)$. Therefore, if there are 100 galaxies within 0.1 magnitude of galaxy i, and 30 of them do not pass this selection criteria, then $w_i = 100/70$. Li & Yee (2008) also suggest a photometric redshift range as another selection criteria, however since we will be comparing results of two different datasets which cover different redshift ranges (MS data has $0.01 \leq z \leq 3.01$ and CFHTLS data has the range of $0.04 \leq z \leq 6.00$ but claims a validity domain of $0.20 \leq z \leq 1.50$), we found this addition made the two datasets greatly differ. Furthermore, Li & Yee (2008) apply an apparent magnitude cutoff at which $w_i = 2$. This correction weight would mean that half of the galaxies in galaxy i's magnitude bin are failing the selection criteria (i.e. their redshifts are not very reliable). We do not make the same cut, as for the CFHTLS data this would translate to a cutoff at $m_i = 23.635$, and 45.8% of the CFHTLS galaxies are fainter than this. Furthermore, only 0.28% of galaxies have $w_i \geq 2$. Since the purpose was to avoid galaxies of high weight, if the galaxy fails this selection criteria, or if $w_i \geq 2$, the galaxy is not included in this study.

For the second requirement, P_{ratio} is defined as:

$$P_{ratio} = \frac{\int_0^\infty P_i(z) P_{group}(z) \mathrm{d}z}{maxP} \tag{2.3}$$

$P_{group}(z)$, defined in equation 2.4, is the probability that galaxies A, B, ...,n, are all at the same redshift, z.

$$P_{group}(z) = P_A(z)P_B(z)...P_n(z) \tag{2.4}$$

where $P_A(z)$ is the probability density of galaxy A defined by the redshift of galaxy A, z_A, and the uncertainty in the galaxy's redshift, $\sigma_{z,A}$. $maxP$ is what the numerator would be if all the galaxies were at the same redshift. In other words, instead of calculating each galaxy's $P_i(z)$ using z_i and $\sigma_{z,i}$, use the group redshift along with $\sigma_{z,i}$, where the group redshift is that which maximizes $P_{group}(z)$. To better demonstrate this, Equation 2.3 is extended below in Equation 2.5.

$$P_{ratio} = \frac{\int_0^\infty P_i(z, z_i, \sigma_{z,i})P_{i+1}(z, z_{i+1}, \sigma_{z,i+1})...\mathrm{d}z}{\int_0^\infty P_i(z, z_{group}, \sigma_{z,i})P_{i+1}(z, z_{group}, \sigma_{z,i+1})...\mathrm{d}z} \tag{2.5}$$

By dividing by $maxP$ we normalize the P_{ratio} of that group by what the P_{ratio} would be if each member was at the same redshift, given the uncertainties in the redshifts. Finally, $P_{ratio,crit}$ is met when $\mid z_1 - z_2 \mid = \sigma_1 + \sigma_2$. P_{ratio} is ~ 0.37 for galaxies with $\sigma_1 = \sigma_2$, and ~ 0.50 for two galaxies when one σ is much smaller than the other (Li & Yee, 2008). The differences are shown in Figure 2.1. If $P_{ratio,crit}$ is too small, the number of false detections will increase, whereas if it is too large we will miss many groups. For the right balance, Li & Yee (2008) show that choosing $D0_{xy} = 0.25$ Mpc and $P_{ratio,crit} = 0.37$ gives the best results. If a galaxy meets this criterion for two or more groups, it belongs to the group for which it has the largest P_{ratio}.

We follow Li & Yee (2008) and employ these criteria and tools in the algorithm as follows:

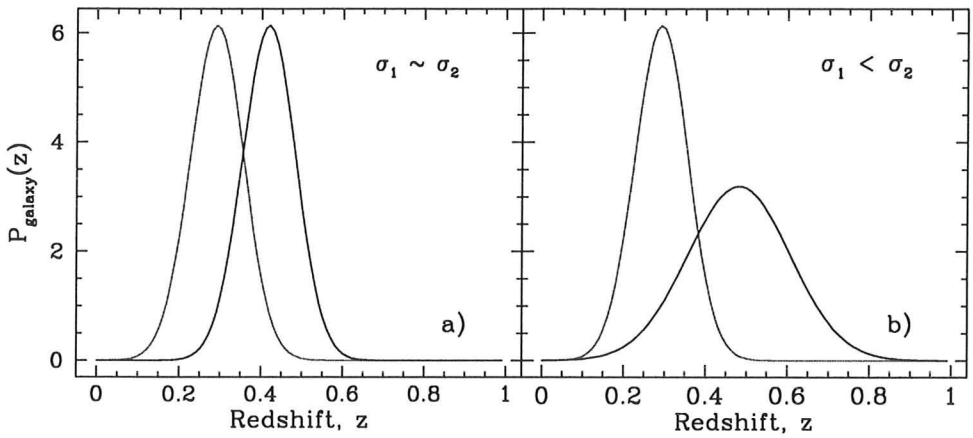

Figure 2.1 The photometric probability densities used to calculate P_{group} are plotted here for the cases where the redshift error of one galaxy is similar to that of the other as seen in (a), and (b) shows the case where the two redshift errors are different. In the case of (a), the criteria that $\mid z_1 - z_2 \mid = \sigma_1 + \sigma_2$ will be true for $P_{group} \sim 0.37$, while for (b) it will hold for $P_{group} \sim 0.50$.

1. Assign each galaxy a weight and make a magnitude cut at the apparent magnitude at which $w_i = 2$.

2. Sort the galaxy list in order of luminosity (highest luminosity first), so that more luminous galaxies are the seeds of the groups (Li & Yee (2008) instead sort by the photometric redshift probability peak values, but they find it has negligible effect).

3. Starting with the most luminous, for each galaxy, A_i, search through a list of the other galaxies, $A_j, j \neq i$. If the separation between galaxies is less than $D0$, calculated using galaxy A_i's redshift for z in Equation 2.1, then calculate P_{group} of galaxy A_j with respect to galaxy A_i.

4. Find the galaxy with the largest P_{ratio}, and add that to the group A_i. Calculate a new group redshift by taking the peak of P_{group} given in Equation 2.4. Using that, calculate a new linking length, $D0$.

5. Every time a galaxy is added or subtracted, go through the remaining group and make sure they fulfill the new linking length criterion. If not, then remove them. Put a flag on this galaxy so that it may not be added back to the group until another galaxy joins the group first.

6. With this new group, go back and search through all galaxies not in the group, and if they fulfill the new linking length criterion, then start again at Step 4 until there are no other galaxies added.

7. If galaxy A_i has no friends, move on to the the next galaxy that is not included in a group, and go to Step 3. Otherwise, go through Steps 4 to 6 with the next galaxy in the group as the new center for the linking length. The galaxies added here will be the friends of galaxy A_i's friends, and become part of group A_i.

8. Continue until each galaxy in the group has acted as the center for the linking length until no more galaxies are added or subtracted.

At this point, it is possible for a galaxy to be included in more than one group. For example, two galaxies that might be friends with respect to group A_i's $D0$ and the P_{ratio} of that galaxy with respect to the group, a group with another group redshift and set of weights and redshift errors will have another $D0$ and P_{ratio} which might not classify these galaxies as friends, creating two groups that overlap. As each galaxy is already a member of any group it could

possibly join given this criteria, the only thing we can do is subtract galaxies from groups. The steps to delete multiple group memberships are outlined below:

1. Go through each galaxy that is a member of more than 1 group, and assign it to the group with which it has the largest P_{ratio} with respect to the group. Recalculate new P_{group} for both groups, find the group redshifts, and the $D0$ for each group.

2. The subtraction of a galaxy in a group can break a group apart, thus for each remaining member, check that it fits the new $D0$ with respect to at least one other member, if not, remove that galaxy and repeat. The new P_{ratio} must also be checked. If it is less than $P_{ratio,crit}$ then remove members (based on which one lowers the P_{ratio} by the largest amount) until it meets this criterion.

3. For each galaxy removed in Step 2, if it was once a member of multiple groups but has been already assigned to the group it was just removed from based on the P_{ratio}, then check to see if it may be added back to another group it had been a member of, starting with the group with which it has the highest P_{ratio}.

Figure 2.2 shows how many iterations are used for the step where we search the galaxy field for a list of possible friends of a member galaxy in order to add them to the group. On average, two iterations are necessary for each member of a group, the average number of galaxies per group is 4.5 (as shown in Chapter 4), and so for 8,320 groups per field, that is an average of 74,880 iterations

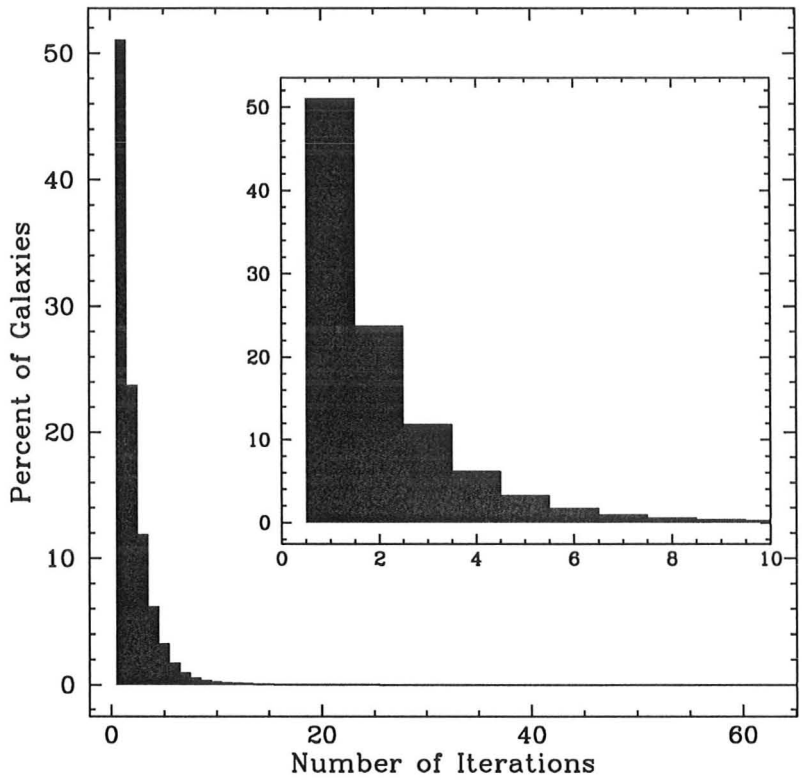

Figure 2.2 The number of iterations are shown here for Step 3 where we search for a new list of friends to add to a group in the pFoF algorithm. The inset plot is a closer look at the 0-10 iterations range.

for this step per square degree. The code was written in the programming language of C by Rachel E. Anderson, and ran on computer clusters of the Shared Hierarchical Academic Research Computing Network (SHARCNET) Consortium at McMaster University *.

Clearly this method is much more computationally intensive than the traditional spectroscopic redshift Friends-of-Friends algorithm, especially since the density of galaxies with photometric redshifts is much higher.

* *https : //www.sharcnet.ca/my/front/*

2.2 Nth Nearest Neighbor

A reliable way to find over-densities is by finding the distance to the nth nearest neighbor (Cooper et al., 2005). We chose $n = 5$, and calculated the distance from each galaxy to the nearest galaxies, and found the projected distance to the 5th nearest neighbor. By doing so we can obtain a measurement of the local galaxy density. We find $n = 5$ ideal as a lower n would be more likely to give a false 'high density' measurement to an isolated system, and a larger n will move us away from the local density regime, and into the global density. When calculating the distance to the 5th Nearest Neighbor (hereafter 5thNN), we follow the same calculations we did for the pFoF method, with an addition to account for the sparser galaxy number density at higher redshift. For the pFoF algorithm, we scaled the linking length as $D0_{xy}/(1+z)$. Here we do not have a linking length cut off, but to make the results of our 5thNN search consistent for different redshifts, we scale the distances between galaxies to follow Equation 2.1:

$$D_{AB} = \sqrt{\frac{R_w \sum_i^N w_i}{N}} \frac{D0_{AB}}{1+z} \tag{2.6}$$

where D_{AB} is the scaled distance between A and B and $D0_{AB}$ is the non-scaled distance. w_i is the same weight discussed previously. This is somewhat different than when we calculate pFoF as we do not need to sum up the weights because are only looking for the distance between galaxy A and other galaxies (no groups), so $\sum_i^N w_i$ becomes w_A and instead of using group redshift for z, we simply use $z = z_A$. N would then be unity. This gives us Equation 2.7.

$$D_{AB} = \sqrt{R_w w_A} \frac{D0_{AB}}{1+z_A} \tag{2.7}$$

We used the 5thNN method in projected space, disregarding redshift, and twice more with a cut in redshift space using the 1σ and the 3σ redshift errors. For these last two cases, to be considered as a possible 5th nearest neighbor of galaxy 1, with redshift z_1 and redshift error σ_1, a galaxy at redshift z_2 with redshift error σ_2, must satisfy:

$$\mid z_1 - z_2 \mid \leq \sqrt{{\sigma_1}^2 + {\sigma_2}^2} \tag{2.8}$$

Ideally we would like to do a 5thNN search in a sphere, but with photometric redshifts the outcome would carry too much uncertainty. However, using spectroscopic redshifts, Deng et al. (2009) found similar results for the impact of environment on galaxy properties when they used a 5thNN search in the projected plane along with a redshift slice of ± 1000 km s^{-1}, compared with when they used a comoving sphere. Thus we assume our method is likewise sufficient.

For each galaxy we apply the different redshift cuts, and from the remaining galaxies we calculate the distance to the 5th nearest neighbor using Equation 2.7.

2.3 Finding 'real' Millennium groups

The purpose of using the MS data is that the positions and redshifts of groups in the simulation are precisely known. We call these groups 'real' groups, and we use them by applying the pFoF algorithm to the fields and comparing the resulting pFoF groups to these 'real' groups. In the Millennium Simulation it is possible to identify 'real' groups as they are tagged and the

total group virial mass is calculated for each galaxy. However, each group did not have a unique central virial mass, so a 'real' group was defined to be a group of three galaxies or more that fit the following criteria:

1. The galaxies must have the same central virial mass.
2. The galaxies in the group can have a maximum redshift range of 0.001.
3. The galaxies can be a maximum distance of 0.5 Mpc from the nearest member.

More practically, galaxies in a 'real' group would appear in sequence with the other group members in the catalog. The criteria mentioned above were simply checked for each group to make certain that two groups with the same central virial mass did not happen to follow each other in the catalog listing.

Chapter 3

Data

3.1 CFHT Legacy Survey Deep fields

Our data is from the four Canada-France Hawaii Telescope Legacy Survey (CFHTLS) "Deep Survey" fields *. These fields are listed in Table 3.1. The fields collectively cover 3.6 deg^2. We consider galaxies with $i'_{AB} \leq 25$. The photometric redshifts were calculated by Ilbert et al. (2006) using 3,241 spectroscopic redshifts between 0 and 5 from the VIMOS VLT Deep Survey for calibration as well as to test the "Le Phare" redshift code they used. The bands used were $u^*g'r'i'z'$. They report the redshift accuracy for $i'_{AB} \leq 24$ to be $\sigma_{\Delta z/(1+z)} = 0.029$ with larger errors for fainter objects such that $\sigma_{\Delta z/(1+z)}$ = 0.025 and 0.034 for i'_{AB} = 17.5 to 22.5 and 22.5 to 24 respectively. Here, $\Delta z = z_{spectra} - z_{photo}$. Ilbert et al. (2006) define η to be percent of catastrophic errors where $|\Delta z|/(1 + z_{spectra}) > 0.15$ and report these to be η = 1.9% and 5.5% of catastrophic errors for the magnitude bins stated above, with starburst galaxies being responsible for 50% of the catastrophic errors. Finally, Ilbert

* *http://www.cfht.hawaii.edu/Science/CFHTLS*

Field	RA	Dec	RA Range (deg)	Dec Range (deg)
D1	02:25:59	-04:29:40	0.9780618	0.9829728
D2	10:00:28	+02:12:30	0.936543	0.984654
D3	14:19:27	+52:40:56	1.5849	0.970279
D4	22:15:31	-17:43:56	1.011673	0.989911

Table 3.1 The 4 CFHT Legacy Survey Deep fields.

et al. (2006) stated that the redshifts are valid within $0.2 \leq z \leq 1.5$. We considered galaxies where at least three bands were used for the computation of the redshift. Also, we look at galaxies where $0.2 \leq z \leq 1.5$ for the calculation of w_i only, and for the rest of the calculations we use an upper limit of 1.0 for the redshift. The 0.2 redshift lower limit was chosen so that we may avoid contamination in our results due to false redshifts (i.e. there are catastrophic errors at low and high redshift). The 1.0 redshift upper limit was chosen to avoid a sampling bias due to the fact that at higher redshifts we would see fewer galaxies (i.e. severe incompleteness). Therefore, we conclude that 0.2 $\leq z \leq$ 1.0 is a safe range.

Following Li & Yee (2008), we assign a completeness correction weight, w_i, to each galaxy in preparation for the pFoF algorithm as explained in Chapter 2. Recall that if $w_i \geq 2$, or the galaxy does not meet the criterion that the probability of a galaxy's redshift to be within $3\sigma_{z_{cut}}$ of its photometric redshift must be greater than 99.7%, the galaxy is cut from this study.

3.2 Light cones from the Millennium Simulation

The light cones (i.e. simulated photometric redshift catalogs) were created from the MS dark matter halo catalog described in Springel et al. (2005). This simulation was conducted by the Virgo Consortium which is made up of British, German, Canadian, and US astrophysicists. The Millennium Simulation was dark matter only, and was simulated from first principles while the galaxies were added later. The group used cold dark matter (Λ-CDM) as their structure formation model along with the theory of cosmic inflation to predict initial conditions for structure formation under the assumption of hierarchical growth through gravitational instability (Springel et al., 2005). The initial growth of density perturbations is linear and thus can be calculated analytically, however the collapse of fluctuations and hierarchical growth is nonlinear and requires direct numerical simulations. This catalog uses code by Kitzbichler & White (2007) for this task.

The simulation used $2,160^3$ particles with a mass of $8.6 \times 10^8 h^{-1}\mathrm{M}_\odot$, from $z = 127$ to 0 in a $(2.230 \times 10^9 \text{ ly})^3$ cube, or likewise, a $(500h^{-1} \text{ Mpc})^3$ cube in the present day ($z = 0$) epoch. Of the particles, 49.6% are in halos above a detection limit of 20 particles. The cosmological parameters used were taken from the combination of 2dFGRS and WMAP data and are as follows: $\Omega_{m,0} = \Omega_{dm,0} + \Omega_{b,0} = 0.25$, $\Omega_{b,0} = 0.045$, $h = 0.73$, $\Omega_{\Lambda,0} = 0.75$, $n = 1$, and $\sigma_{8,0} = 0.9$ (Springel et al., 2005). Here, $\Omega_{m,0}$ is the ratio of the measured density of all types of matter (baryonic and dark matter) to the the critical density ($\rho_{c,0}$), the present density required for a flat universe, where $\rho_{c,0} = 3H_0/8\pi G$, and

G is Newton's gravitational constant. Similarly, $\Omega_{b,0}$ and $\Omega_{\Lambda,0}$ represent the density ratios for baryons and dark energy. The spectral index, n, is defined by the initial power spectrum (monochromatic energy flux), $P(k) \propto k^n$, where $k = 2\pi/\lambda$. Finally, σ_8 is the root mean square fluctuation in the density field smoothed with a spherical top-hat filter of radius 8 Mpc.

The MS cones have been populated with galaxies using a semi-analytic model to watch galaxies form from gas, star, and super-massive black-hole processes using merger history trees of dark matter halos and their substructures (Springel et al., 2005). The cones were produced by using a code described in Kitzbichler & White (2007) to bring together the galaxies from a redshift range of 0 to 3 into a 2×2 degree field. The model used to generate all cones is described in Croton et al. (2006). The dust prescription used is from Kitzbichler & White (2007).

These light cones have been tested to see how well they mimic real data, with the results implying that they match the characteristics of real data quite well (Springel et al., 2005). For example, the analytic formula for the mass function of cold dark matter halos by Jenkins et al. (2001) fits the results well for $z \leq 12$ and $M \geq 1.7 \times 10^{10} h^{-1} \mathrm{M}_{\odot}$. Furthermore, the model was successful in reconstructing the relationship between clustering and magnitude, and clustering and color, in 2-degree Field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS) data. The main difference was in the amplitude of the dependence of the clustering on color, however the slope was the same. Furthermore, the galaxy 2 point correlation function, $\xi(r)$, matches that done for the 2dFGRS data by Hawkins et al. (2003).

The information given in these cones includes galaxy type (central or satellite), galaxy/halo position, halo mass (Mvir), halo velocity (Vvir), halo virial radius (Rvir), halo spin and velocity dispersion, total and bulge stellar mass, cold, hot and ejected gas mass, black hole mass, star formation rate, and hot halo cooling and AGN heating rates. The apparent and absolute (total and bulge) magnitudes are given in DEEP2 (BRIK), SDSS (ugriz), and Cousins (BVRIK) filters, for each galaxy (14 filters in all). The AB system (Oke & Gunn, 1983) is used for all magnitudes. The absolute magnitudes for the DEEP2 and the SDSS filters were calculated from the apparent magnitudes using Mike Blanton's kcorrection code (Blanton & Roweis, 2007). The selection for magnitude was that $r < 29.0$ and $M_R - 5\log h < -16.5$. There are four of these cones available, and all four were used in this study. Each cone was treated equally.

3.3 Matching the Millennium Simulation cone fields to the Legacy Survey Deep fields

In order to compare the cones to the four CFHTLS fields, we needed the two sets of data to be comparable. We use the four CFHTLS fields as a reference, and have each MS Field 1 match the size and other properties of the CFHTLS Deep Field 1, and so on. To this end the following was done:

- Each 2×2 degree cone field was split into four fields to exactly match the field of view of the CFHTLS fields, listed in Table 3.1.

- To convert the SDSS magnitudes to the MegaCam magnitudes, we used the calibration set by Pritchet & SNLS Collaboration (2006) for the *griz* bands, and for u we used that defined by the CFHT Collaboration [†]. These are shown in Equations 3.1. To each we applied a limiting magnitude cut at $i'_{AB} = 24.5$. While this is slightly lower than the magnitude cut applied to the CFHTLS fields ($i'_{AB} = 25.0$), we found it made the distribution of the redshifts, an important factor in the pFoF algorithm, more similar.

$$\begin{aligned} u_{Mega} &= u_{SDSS} - 0.241(u_{SDSS} - g_{SDSS}) \\ g_{Mega} &= g_{SDSS} - 0.153(g_{SDSS} - r_{SDSS}) \\ r_{Mega} &= r_{SDSS} - 0.024(g_{SDSS} - r_{SDSS}) \\ i_{Mega} &= i_{SDSS} - 0.085(r_{SDSS} - i_{SDSS}) \\ z_{Mega} &= z_{SDSS} + 0.074(i_{SDSS} - z_{SDSS}) \end{aligned} \tag{3.1}$$

- The CFHTLS fields contain masked regions due to saturated stars, which we duplicated in the Cone fields. In one set cuts were applied to remove circular regions in the place and of the size of their corresponding CFHTLS field. The other set was left unaltered to explore the effect the masked regions would have on the pFoF algorithm, leaving us with a total of three datasets:

  1. CFHTLS Deep fields
  2. MS cones without masked regions

[†] *http://cfht.hawaii.edu/Instruments/Imaging/MegaPrime/generalinformation.html*

3. MS cones with masked regions

Not all of the masked regions in the CFHTLS fields were duplicated to the cones, but the larger regions only. The masked regions transferred to the cones are shown in Figure 3.1.

- We included galaxies with redshifts within $0.2 \leq z \leq 1.5$ for the weight calculation, and within $0.2 \leq z \leq 1$ for all subsequent calculations, to be comparable to the CFHTLS fields. The different redshift distributions are shown in Figure 3.2. The shapes of the $N(z)$ histograms are similar beyond a redshift of 0.4 for the CFHTLS and the MS cones (i.e. the MS fields have a deficiency of galaxies below a redshift of 0.4). It is also evident that there is no visible difference between the MS cones with masked regions and without.

- Redshift errors (σ_z) were assigned to the MS cone galaxies to match the observed distribution of σ_z's per redshift bin, which is shown in Figure 3.3. To do so, each z and σ_z pair from a CFHTLS galaxy was assigned to a redshift bin of width 0.2. Then each MS cone galaxy would randomly be assigned a σ_z from the bin in which its own redshift falls. The resulting redshift error distributions are shown in Figure 3.4. Here we notice a difference in distributions between the CFHTLS galaxies and the MS galaxies. This is to be expected when referencing Figure 3.2 as we see that the MS cones have more galaxies at higher redshifts than the CFHTLS fields, and Figure 3.3 shows that this means the MS cone galaxies will have more of the larger redshift errors. Furthermore, in Figure 3.3 we see that the redshift errors assigned to the MS galaxies are

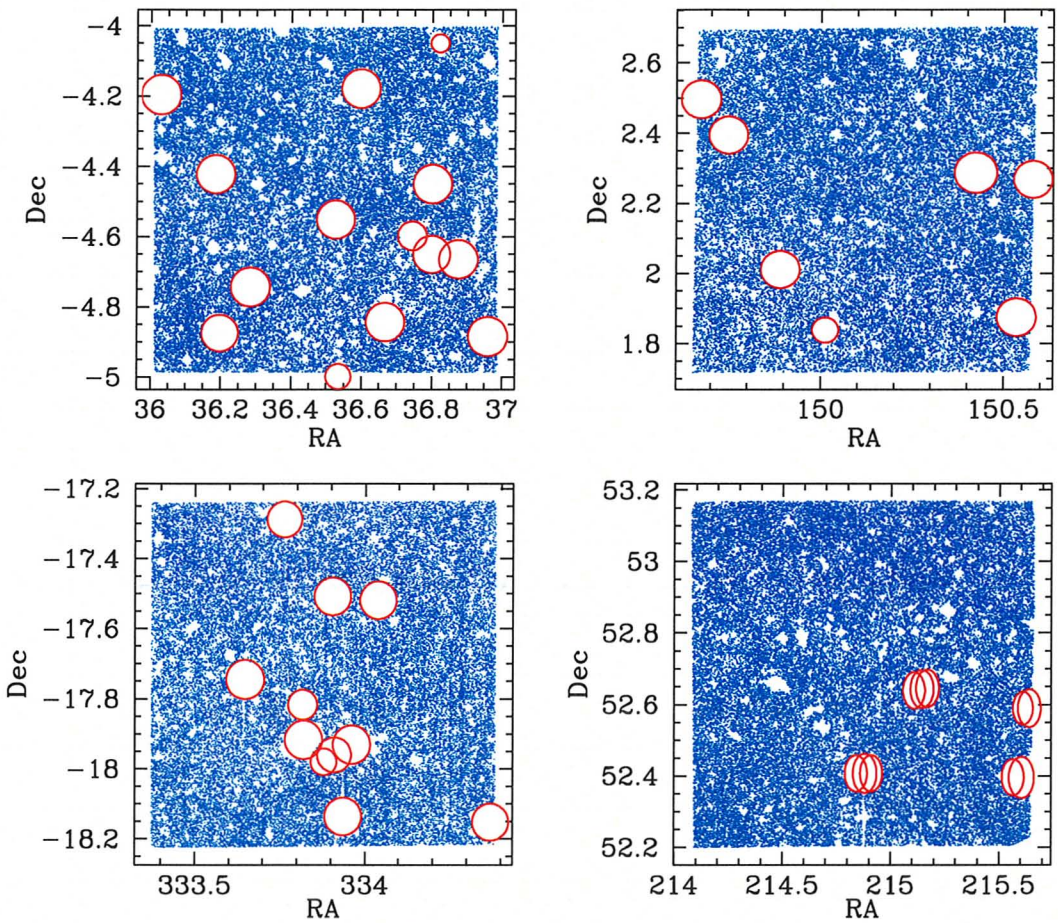

Figure 3.1 The four CFHTLS fields are shown starting with Field 1 in the upper left and counting clockwise. The red circles mark the masked regions that were removed from the MS cone fields.

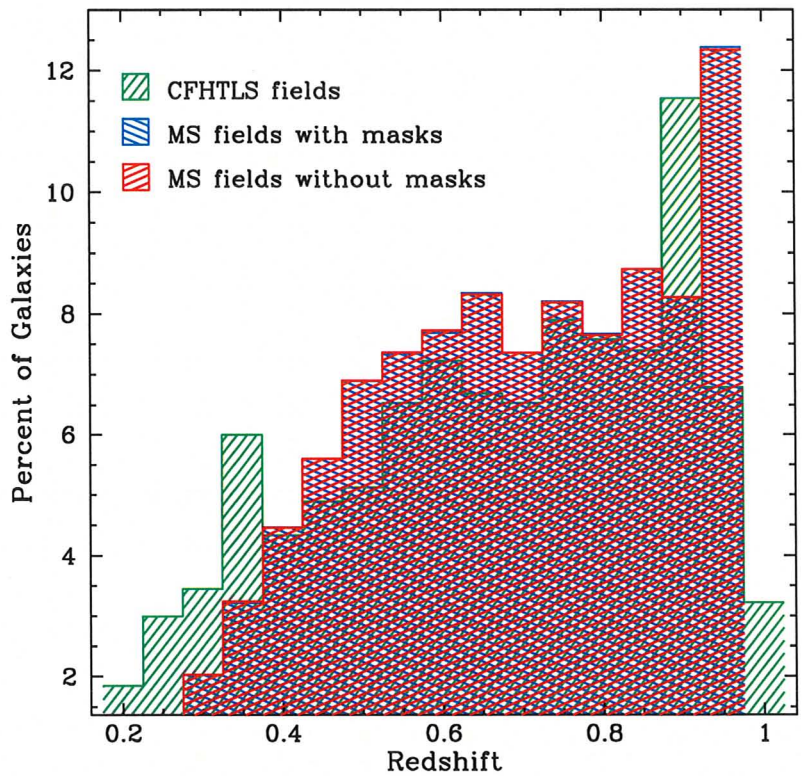

Figure 3.2 This histogram shows the redshift distribution for the three datasets: The CFHTLS galaxies, the MS cone field with masks galaxies, and the MS cone fields without masks galaxies. The shapes are similar (disregarding small fluctuations at the level expected for cosmic variance) beyond a redshift of 0.4.

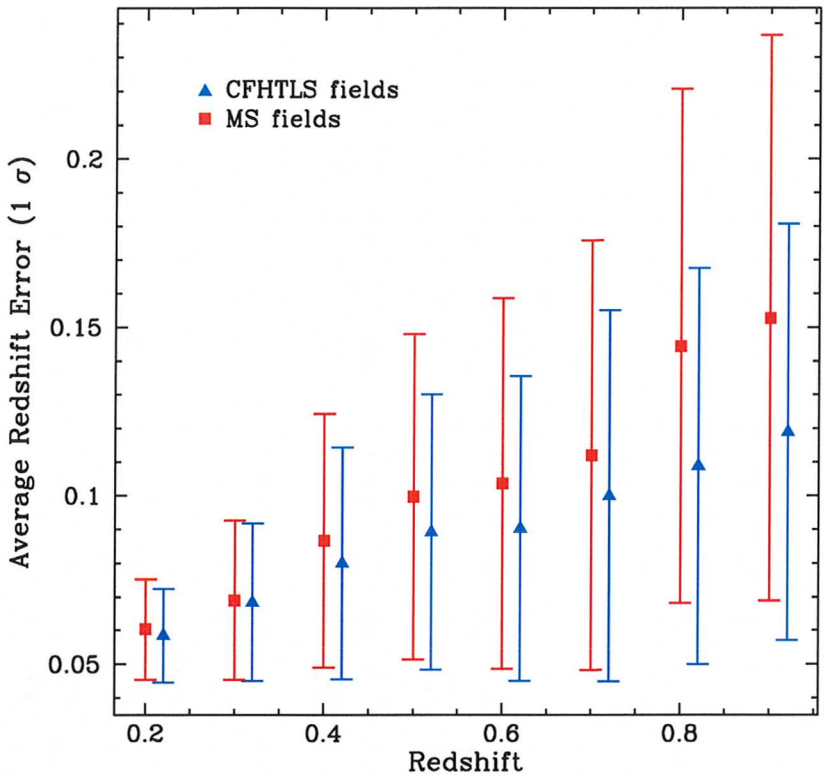

Figure 3.3 Shown here is a plot of the average redshift error for redshift bins of width 0.2, with error bars of one standard deviation. The redshift errors for the MS data were assigned based on errors from the CFHTLS data. The CFHTLS data has been offset by $z = 0.02$ for clarity.

systematically higher than the CFHTLS galaxies' redshift errors. This is due to the fact that there are more galaxies at higher redshift, and the redshift errors are greater at higher redshifts. Therefore, in each redshift bin the majority of galaxies are at the higher redshift end, and thus the majority of the redshift errors (the redshift errors which will be randomly assigned to the MS galaxies) are at the higher end.

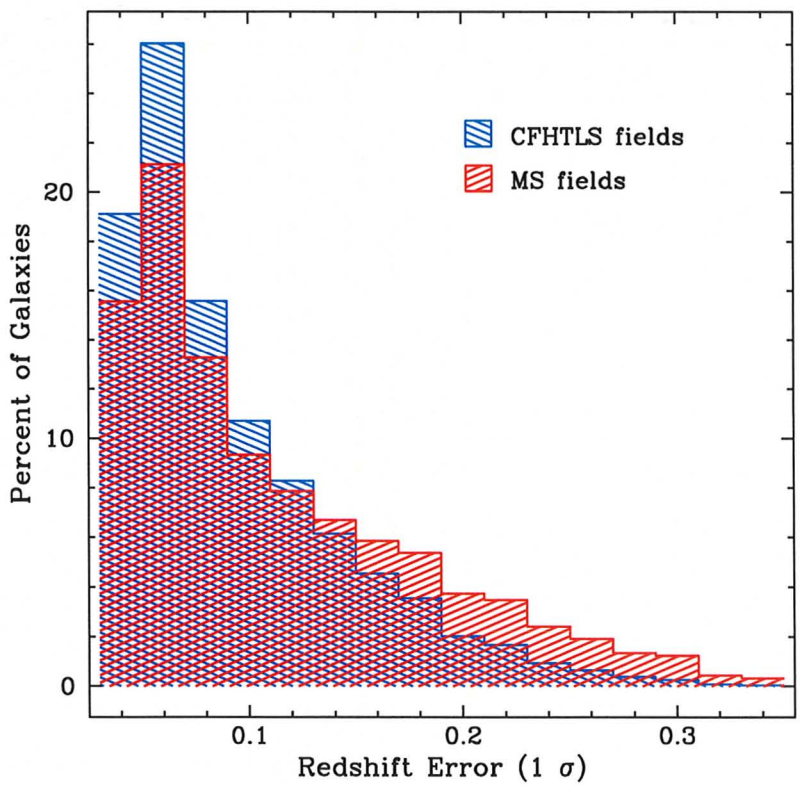

Figure 3.4 This histogram shows the redshift error distribution for the CFHTLS galaxies and the MS cone field galaxies.

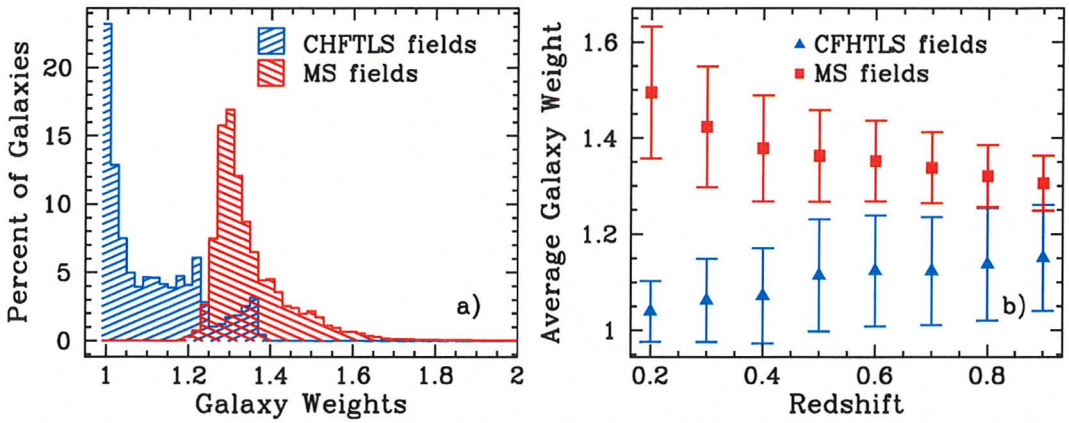

Figure 3.5 In (a) is the percent of galaxies with different galaxy weights for the CFHTLS fields and the MS fields. (b) shows the average galaxy weight as a function of redshift for the same datasets. Error bars are one standard deviation.

- Weights were calculated for each galaxy as described above for the CFHTLS field. Similarly, if any galaxy had $w_i \geq 2$, or it did not meet the selection criterion used for the calculation of w_i, it was not included in this study. Figure 3.5 shows the weight distribution and the weight-redshift relationship for the CFHTLS fields and the MS fields. The masked millennium fields and the unmasked fields show the same distribution, so they are treated as one dataset here.

From Figure 3.5 we can see that the weights assigned to the MS galaxies are higher than those assigned to the CFHTLS galaxies. This tells us that more MS galaxies are not meeting the selection criterion than the CFHTLS galaxies. We explored possible reasons for this result. The first possible explanation is that the redshift errors assigned to the MS galaxies are slightly higher on average for all redshifts than for the CFHTLS galaxies, as shown in Figure 3.3. The possible cause for this is that the width of the redshift bins we used to

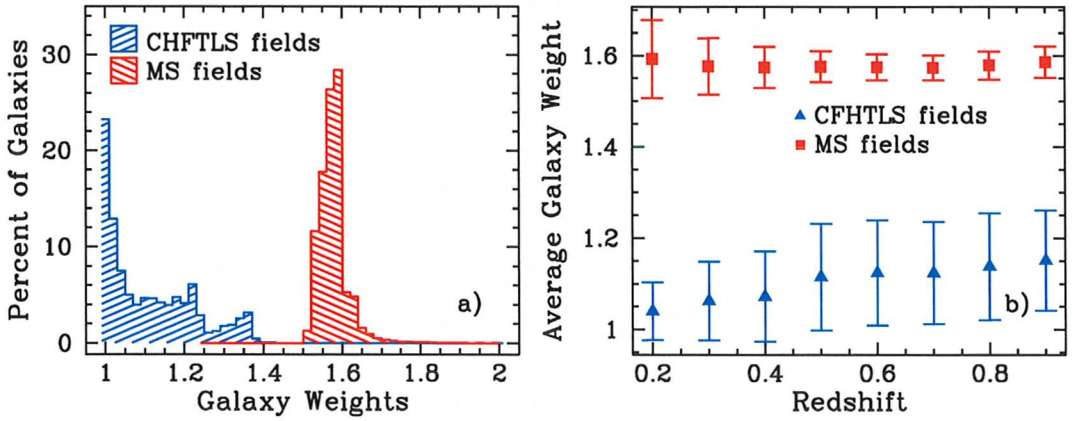

Figure 3.6 This figure shows the effect assigning redshifts to the CFHTLS fields using smaller redshift bins has on the weights. A histogram of the percent of galaxies by galaxy weight is shown in (a) for the CFHTLS fields and the MS fields, while (b) shows the average galaxy weight as a function of redshift for the same datasets. The error bars are one standard deviation.

assign the MS galaxies a σ_z (width of 0.2) was too large. Since we randomly choose a σ_z from a redshift bin where the number of galaxies slightly increases with redshift (Figure 3.2), and we also see an increase in σ_z with z (Figure 3.3), we could be assigning greater σ_z's on average. We tested this theory on two cone fields by randomly choosing σ_z's from redshift bins of width 0.05, and recalculating the weights. The result is shown in Figure 3.6. The smaller redshift bins reduces the spread in the weights, however it increases the average weight for all redshifts.

The second option we explored was to make a magnitude cutoff at $m_i = 23.64$ (where $w_i = 2$ for the CFHTLS dataset, as Li & Yee (2008) have done) for one and for both datasets. This had a negligible effect on the weights. One other difference between the CFHTLS and the MS fields is that the redshifts of the MS galaxies are known and we simply assigned a σ_z from which we calculate the probability of that galaxy being at any one redshift (but we

know what redshift it is at), while for the CFHTLS fields, each galaxy has a redshift, but it could actually be anywhere within $\pm\sigma_z$ with a probability of 68.2%. We explored this possibility by keeping the σ_z assigned to the galaxy, but randomly selecting a new redshift from within that range and calculating the weights again. This also had negligible effect on the weights. Furthermore, when we changed the width of the redshift bins *and* randomly selected a new redshift within the assigned sigma, the only noticeable effect was that caused by changing the redshift binning as shown in Figure 3.6.

Though the distribution of the weights for the CFHTLS and the MS galaxies differ, we are confident in the weights assigned to the CFHTLS galaxies and believe the difference results in a conservative estimate of the number of 'real' groups the pFoF algorithm finds. Li & Yee (2008) report a median w_i of 1.09 which is similar to our average of 1.11, while the average weight of the MS galaxies is 1.35. Finally, since the MS galaxies have higher weights, this corresponds to larger linking lengths which will increase the number of interlopers. Therefore by testing the pFoF algorithm on the MS data we are overestimating the number of interlopers found in the CFHTLS fields.

The results of the pFoF algorithm depend on the density of galaxies in the fields, so it is important to have all matching fields have similar densities. The details of the different fields are shown in Table 3.2. We choose to make the overall galaxy density equal for corresponding fields (e.g. all Field 1's have the same galaxy densities and so on). Thus, from the remaining populations in the matching fields (CFHTLS fields and MS cones of Field 1, Field 2, Field 3, and Field 4) we determined the minimum number of galaxies (which would

be a CFHTLS field or a MS cone field with masked regions) and randomly cut the difference from the remaining masked MS or CFHTLS fields. The unmasked MS fields were treated differently. For every masked field, whatever the percentage of its population that was cut, that same percentage was cut for the corresponding unmasked field.

Field	Initial Number of Galaxies	Final Number of Galaxies
deep, 1	57300	46067
deep, 2	61510	46483
deep, 3	60022	60022
deep, 4	51701	51701
cone012000masked, 1	46067	46067
cone012000masked, 2	46483	46483
cone012000masked, 3	84449	60022
cone012000masked, 4	57764	51701
cone012000, 1	51742	51742
cone012000, 2	49254	49254
cone012000, 3	87413	62128
cone012000, 4	63044	56426
cone012100masked, 1	48385	46067
cone012100masked, 2	52204	46483
cone012100masked, 3	91997	60022
cone012100masked, 4	56597	51701
cone012100, 1	54046	51456
cone012100, 2	55815	49698
cone012100, 3	95266	62154
cone012100, 4	61730	56389
cone120000masked, 1	49392	46067
cone120000masked, 2	54329	46483
cone120000masked, 3	88573	60022
cone120000masked, 4	56428	51701
cone120000, 1	55550	51810
cone120000, 2	57813	49463
cone120000, 3	91853	62244
cone120000, 4	61843	56662
cone201000masked, 1	53479	46067
cone201000masked, 2	50645	46483
cone201000masked, 3	90206	60022
cone201000masked, 4	56583	51701
cone201000, 1	60112	51780
cone201000, 2	53786	49365
cone201000, 3	93633	62302
cone201000, 4	61622	56305

Table 3.2 Listed here are the fields, the number of galaxies in each field, and the number of galaxies used in our calculations so that the average density of galaxies in each field would match.

Chapter 4

Results

Using the datasets described in Chapter 3, we calculated the 5thNN distance and pFoF for these samples, and we now look at the results.

4.1 Distance to the 5th Nearest Neighbor results

To see how well the simulated MS datasets fit the likeness of the CFHTLS fields, we take a closer look at the relative galaxy densities between the CFHTLS and the MS datasets. In Figure 4.1 we show the the distance to the 5th Nearest Neighbor for the different datasets and for different cutoffs in redshift space. From (a) we can see the difference between using a redshift cut of 3σ, a redshift cut of 1σ, and no redshift cut on the CFHTLS fields. In (b) we see the same trend for the MS datasets when using a 1σ redshift cut and no cut. (b) also shows that the masked and un-masked MS fields produce the same results. In (c) we can see the slight variation between the CFHTLS fields, though all histograms peak between 20 and 30 arcseconds. Finally, (d) shows that there is no significant difference in the density distribution between the CFHTLS

and MS datasets. This means that the distance to the 5th nearest neighbor is the same for the same percentage of galaxies in each dataset. Typically the 5thNN technique is used with spectroscopic redshifts in either projected space with velocity cuts to remove foreground and background sources, or in 3D space. Cooper et al. (2005) employs the 5thNN in this way, and defines the boundary between high and low density regions at $\log_{10}(D_5) = 0.5$, where D_5 is the 3D distance to the 5th nearest neighbor and is in units of h_{-1}Mpc (co-moving). They define this boundary condition such that the most dense 1/3 of the sample is considered high density. We apply a similar boundary condition to our 5thNN results (using the $1\sigma_z$ cut) such that the top 1/3 are in the high density region. For the CFHTLS fields, we define the galaxies to be within local high density regions if $D_{AB} \leq 24.5$ arcseconds (from Equation /refeq:dis), for our limiting magnitude of $m_i = 25.0$. This boundary for the MS data is $D_{AB} \leq 24.2$ arcseconds for the limiting magnitude of $m_i = 24.5$. Knowing the galaxy densities of the CFHTLS and MS fields are similar is an important factor as we use one field to test the results of the other.

4.2 Comparison between CFHTLS pFoF groups and MS pFoF groups

In Chapter 3 we prepared three datasets of galaxy fields. In Chapter 2 we described the pFoF algorithm. Here we look at the results of the pFoF algorithm applied to these datasets. There are three main group characteristics calculated by the pFoF algorithm, and they are the P_{ratio}, the number of

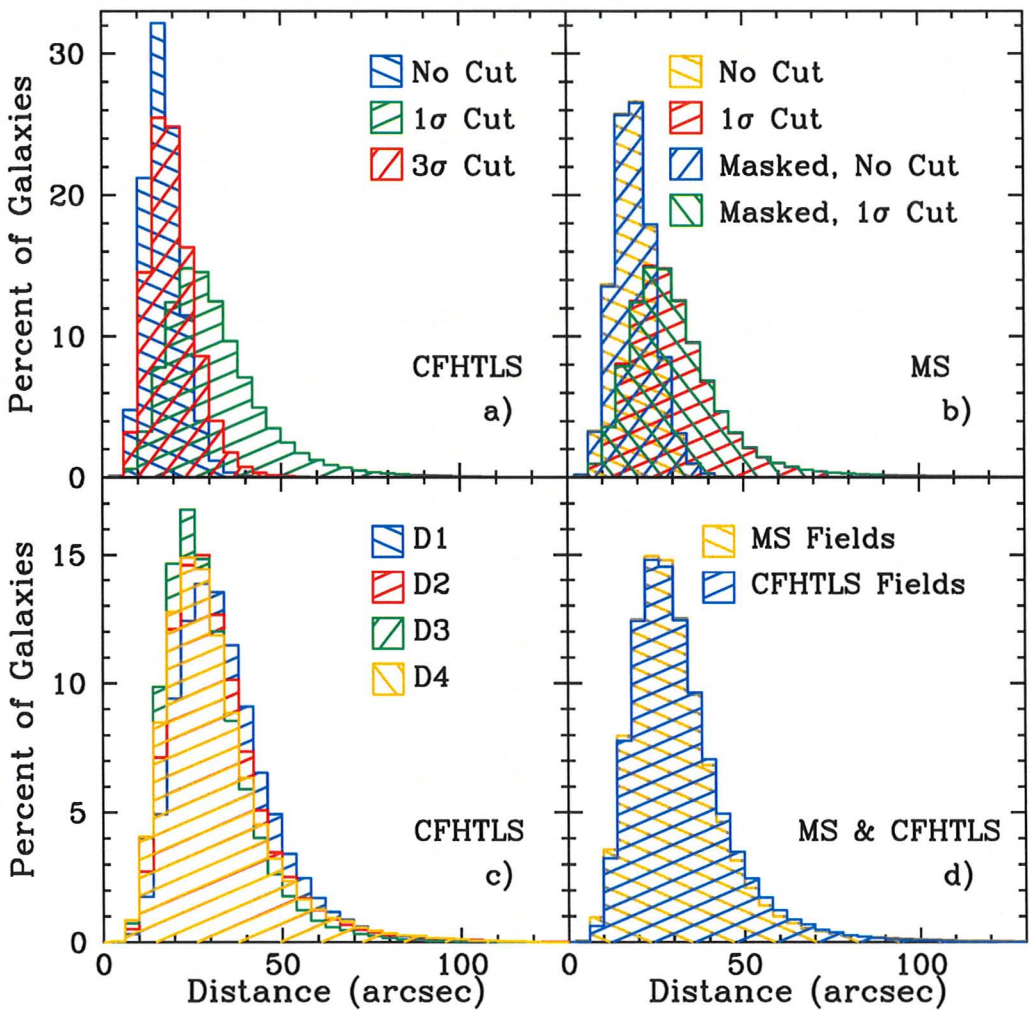

Figure 4.1 In (a) the distance to the 5th nearest neighbor is shown for different σ cuts for the CFHTLS data. In (b) we show the effect of the 1σ redshift cut and no redshift cut for the MS dataset with masked regions, and without. The results from the masked and unmasked fields are almost indistinguishable. (c) shows the difference between the CFHTLS fields for the 1σ redshift cut. Finally, (d) shows a comparison between the CFHTLS fields and the MS fields with a 1σ redshift cut.

galaxies in the group, and the group redshift. We look at each in turn and discuss the differences between the datasets.

First, we will compare the P_{ratio} of the CFHTLS and MS pFoF groups, as illustrated by Figure 4.2. The peak at 1 for the MS fields illustrates one of the main differences between the MS and the CFHTLS datasets: all MS galaxies that are in a group are at the same redshift, while the CFHTLS galaxies are possibly at the same redshift. Because of this, more MS pFoF groups will have a $P_{ratio} \sim 1$; a peak which is absent in the CFHTLS pFoF groups. The peak would move to a lower P_{ratio} if we were to select redshifts randomly within the errors we assigned to the MS galaxies, as we did to test the effect this would have on the weights in Chapter 2. Finally, we note that the two MS datasets have a similarly shaped distribution with only slight variations indicating that the masked regions make little difference.

The second item to consider is the number of galaxies in a pFoF group. This is shown in Figure 4.3. From this figure we see that 2% more of the the CFHTLS groups have three members, while the MS pFoF groups have a slightly higher percentage of larger groups ($\sim$ 0.1% more of the MS pFoF groups have 20 members). The average number of galaxies in a group is 4.5 for the CFHTLS pFoF groups, and 5.0 for the MS pFoF groups. The larger MS pFoF groups are possibly a result from the better P_{ratio} as well as the greater linking length due to larger weights. Again, the masked regions appear to have a negligible effect.

The final item we want to look at is group redshifts, shown in Figure 4.4. Here we notice that the group redshift distributions match the distributions

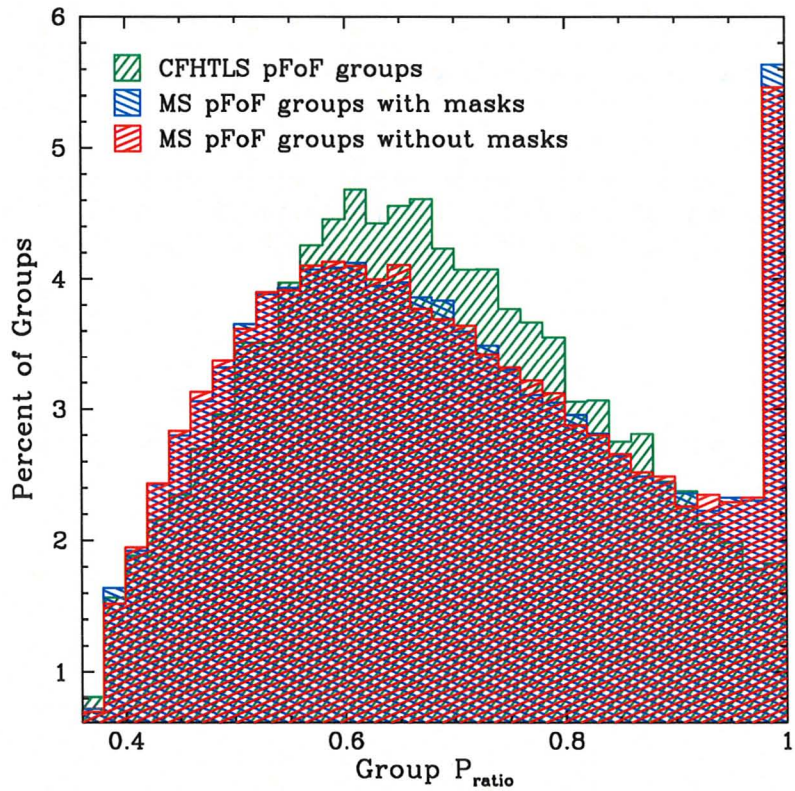

Figure 4.2 Shown here is a histogram of the P_{ratio} of the pFoF groups of the three different datasets.

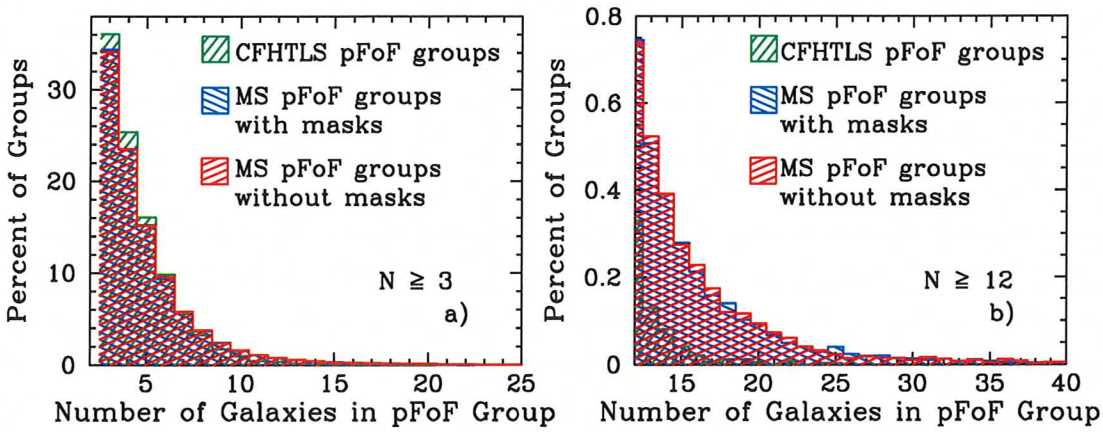

Figure 4.3 The number of members in groups are plotted. The limits are fixed for plot (a), but the cone pFoF groups have up to 65 members, the cone data with masks pFoF groups have up to 61 members, and the CFHTLS pFoF groups have up to 25 members. Plot (b) show groups with 12 or more members.

from the galaxy catalogs, as shown in Figure 3.2. The MS datasets have a slightly higher percentage of their groups at higher redshifts than the CFHTLS dataset.

From these comparisons we can say that the MS datasets are comparable to the CFHTLS datasets keeping two things in mind:

1. The redshift distributions are similar above redshift of 0.4 for the galaxies and the galaxy groups.

2. The MS groups are larger possibly due to better P_{ratio}'s and greater linking lengths from higher weights.

Furthermore, we note that the two MS datasets give similar results, and in subsequent sections we will combine their results and treat them as one.

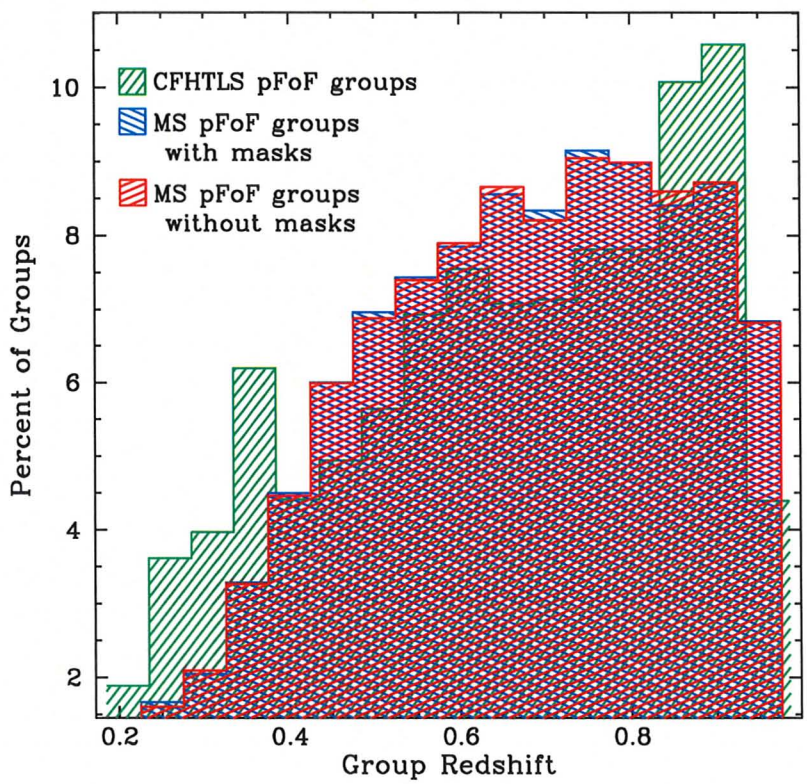

Figure 4.4 A histogram of the group redshifts.

4.3 Comparison between 'real' groups and pFoF groups in the Millennium light cone fields

We are interested in using the Millennium Simulation to test the pFoF method, and since we have shown that the MS and CFHTLS datasets are comparable, we hope to apply the results of this test to the CFHTLS data. We now take a look at those results, and see how well the pFoF algorithm does at picking out groups defined by the Millennium Simulation to be 'real'.

The 'real' MS groups typically have $P_{ratio} \sim 1$, as shown in Figure 4.5, because here there is no uncertainty in the redshifts, and there are no interloper galaxies. These galaxies truly are at the same redshift as discussed in Section 4.2. We explore the effect of a higher $P_{ratio,crit}$ for the pFoF algorithm in Section 4.5.

On the other hand, there is no noticeable difference when looking at the group redshifts of 'real' and pFoF groups, as shown in Figure 4.6. This shows that there is no obvious redshift trend that would contribute to the interloper population.

The distributions of the number of galaxies in the pFoF and the 'real' groups are shown in Figure 4.7. Here we see that $\sim 43\%$ of the 'real' groups have three members while only $\sim 34\%$ of the pFoF groups do. The 'real' groups also have a higher percentage of large groups, where $\sim 8\%$ of the 'real' groups have 12 or more members, compared to $\sim 3\%$ for the pFoF groups. However, $\sim 60\%$ of the pFoF groups have 4-9 member galaxies, while here only $\sim 46\%$ of the 'real' groups fall within this range.

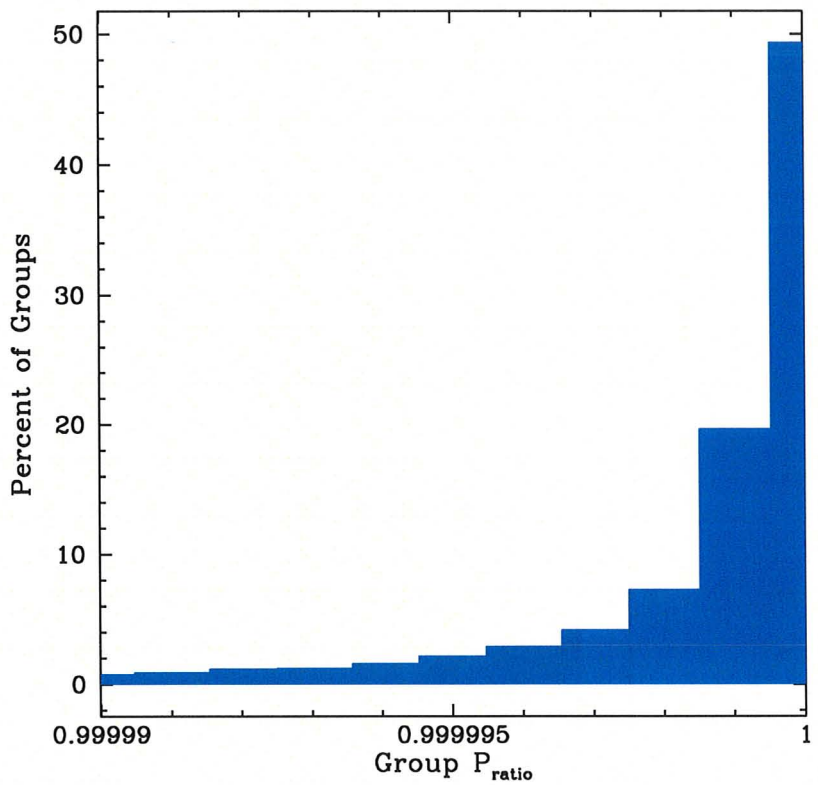

Figure 4.5 Histogram of the group P_{ratio} of the 'real' groups from the MS light cones. The values start at 0.998753, but lower P_{ratios} are not shown here so more attention may be made to the bulk of the population.

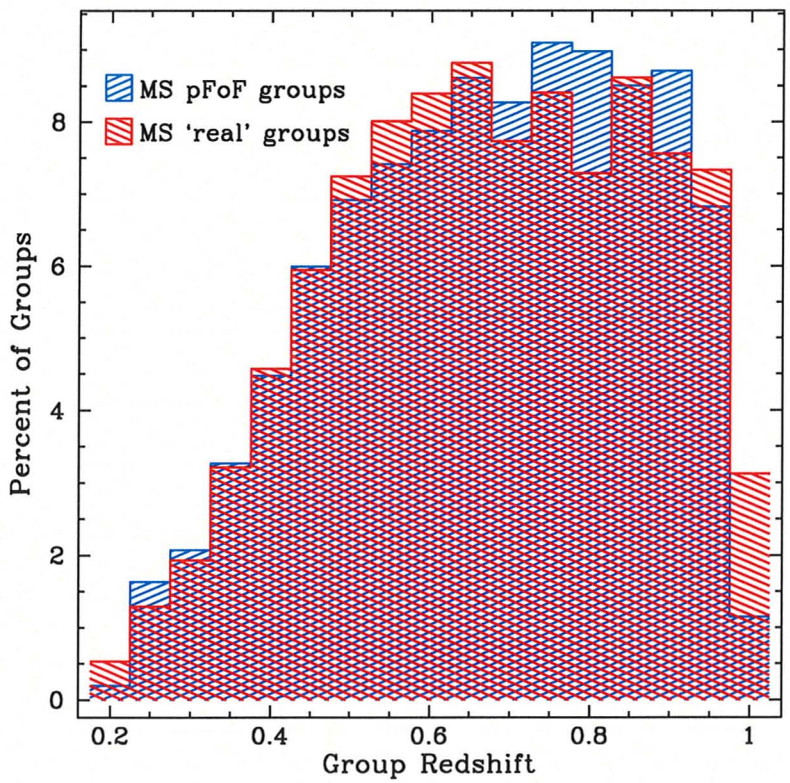

Figure 4.6 A histogram of the group redshifts for both 'real' and pFoF groups in the MS cone fields.

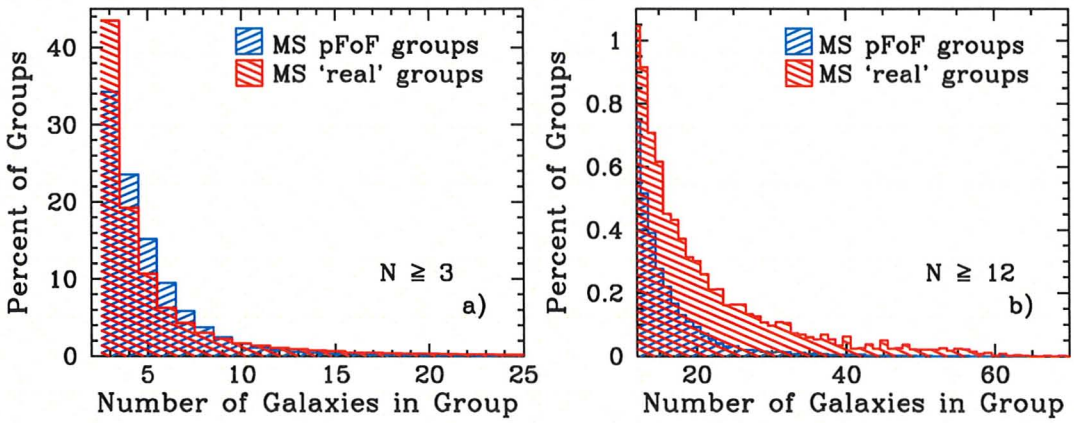

Figure 4.7 The number of members in groups are plotted. The cone pFoF groups have up to 65 members and the 'real' groups have up to 300 members. Plot (a) shows up to 25 group members to focus on the smaller groups, while plot (b) shows groups with 12 or more members.

Figure 4.8 shows the percent of MS galaxies that the pFoF algorithm considers to be in a group, but the simulation does not. For example, in $\sim 20\%$ of the pFoF groups, $\sim 34\%$ of that group is in a group halo according to the Millennium Simulation, which means that $\sim 66\%$ of that group is contamination. The largest peak was at 0%, with 61% of the pFoF groups consisting of galaxies that are not in a 'real' group. Note that this dominant peak is not shown in the figure. Therefore, the pFoF algorithm considers many galaxies which are not actually in a group to be part of a group.

This difference in number of galaxies in the group is further demonstrated by Figure 4.9 which shows the percent of galaxies that are in a group as a function of redshift for the 'real' and pFoF groups. Of the number of galaxies in MS pFoF groups, 26.8% are in 'real' groups, while the remaining 73.2% are interloper galaxies. From this we see that pFoF considers a massive number of galaxies to be in a group which the MS simulation does not consider to be

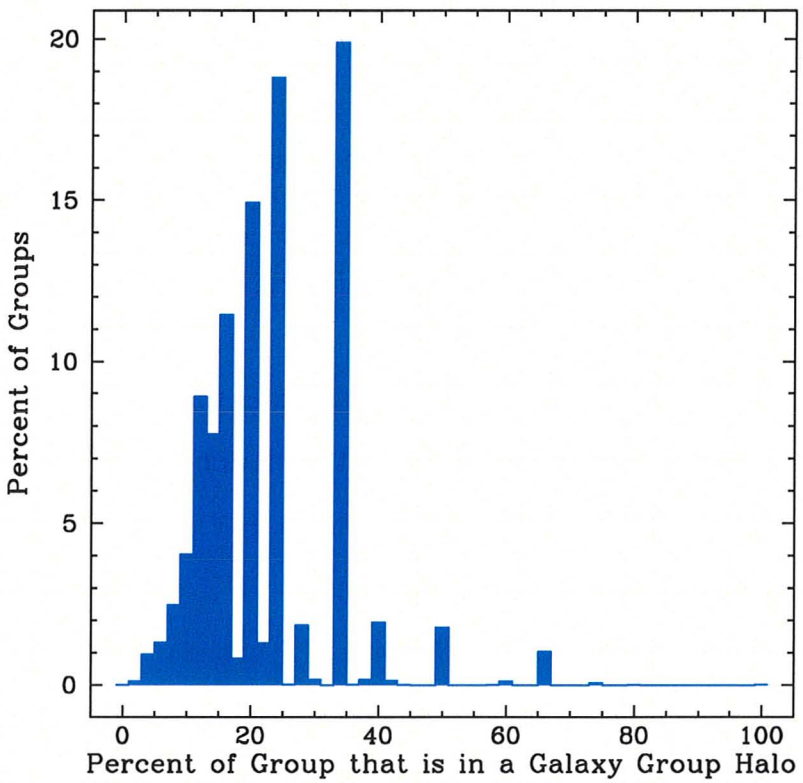

Figure 4.8 The percentage of pFoF groups that MS considers to be part of a 'real' group. For example, in $\sim 15\%$ of the pFoF groups, $\sim 20\%$ of that group is in a group halo according to the Millennium Simulation, which means that $\sim 80\%$ of that group is contamination. The dominant peak at 0% is not shown here.

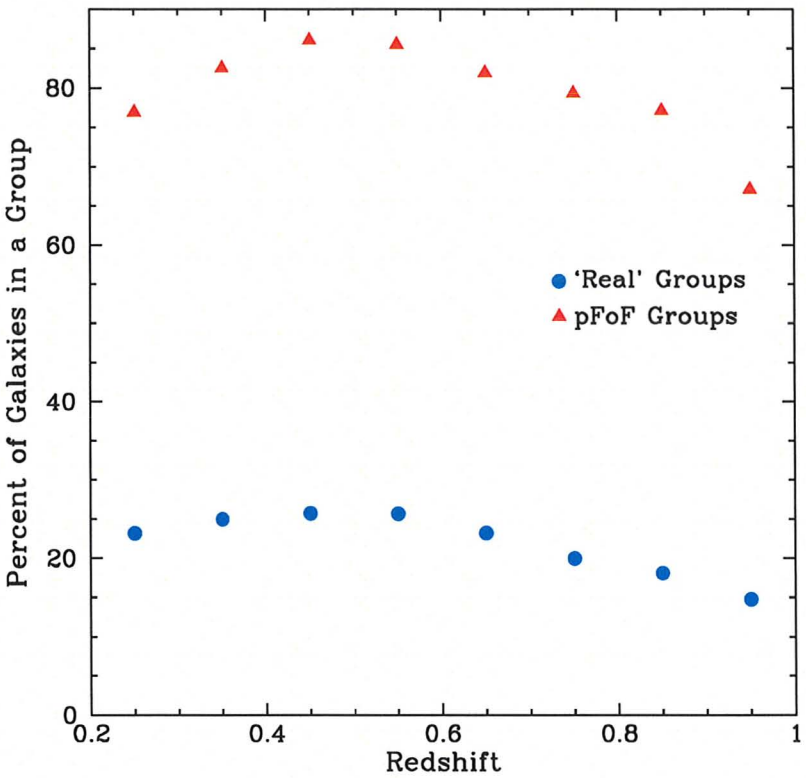

Figure 4.9 A histogram of the percent of galaxies that are members of a group for different redshift bins for both 'real' and pFoF MS group is shown.

in a common dark matter halo. As stated in Chapter 1, $\sim 60\%$ of galaxies should be in a group of $N \geq 2$ at $z \sim 0$ (Eke et al., 2004; Berlind et al., 2006; Tago et al., 2006). While the pFoF algorithm surpasses this, the number of galaxies in 'real' MS groups is well below. However, since we are considering groups of $N \geq 3$ only, and we have a range of redshift from 0.2 to 1.0, we can still conclude that the pFoF algorithm selects more galaxies to be members of groups than there should be.

Figure 4.10 shows the percent of 'real' groups found as a function of the number of galaxies in that 'real' group. As expected the larger 'real' groups are more likely to be found, with groups of 60 members or more always found. Though success at finding 'real' groups is high, they still have interloper galaxies.

Also note that in Figure 4.10, there were 97 systems with more than 100 members, and we found that each of these systems had a central virial mass of at least $1\times10^{14}h^{-1}\mathrm{M}_{\odot}$. However, 11 systems have more than 200 members with the central virial mass greater than $3 \times 10^{14}h^{-1}\mathrm{M}_{\odot}$, indicating they are more likely to be clusters than group-sized systems. This is expected as Bartelmann (2001) has shown that it is typical to have 1.5 galaxy clusters every square degree, and we have a total of 16 square degrees.

Figure 4.11 is helpful in identifying when the pFoF is a reliable indicator of a high density region. We define a pFoF group as a found 'real' group when three or more galaxies in the pFoF group are in a common 'real' group. On the other hand, if a pFoF group does not contain at least one set of three or more galaxies that are known to be in the same 'real' group, than that pFoF group is an interloper group. We have 207,757 groups total and 172,713 interloper groups, which means 83.13% of the pFoF groups are interlopers. However, according to Figure 4.11, if we want the interloper population to be below 20%, for example, then we need to focus on groups of 12 or more. Therefore, unlike spectroscopic studies where 3-4 members are typically all that are used to locate groups, we need to link more members together in order to be confident we have a real group. If we combine this with Figure 4.3,

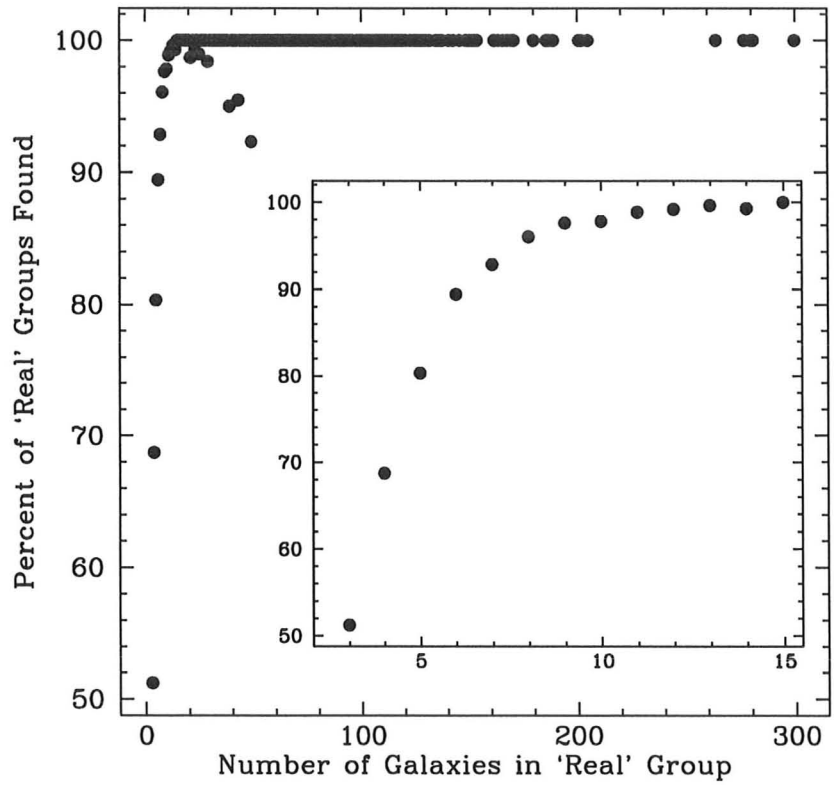

Figure 4.10 Plotted here is the percent of 'real' groups found as a function of the number of galaxies in that 'real' group. The inset is a closer look of the groups with 3 to 15 members. Success at finding groups is high, but these groups have interloper members.

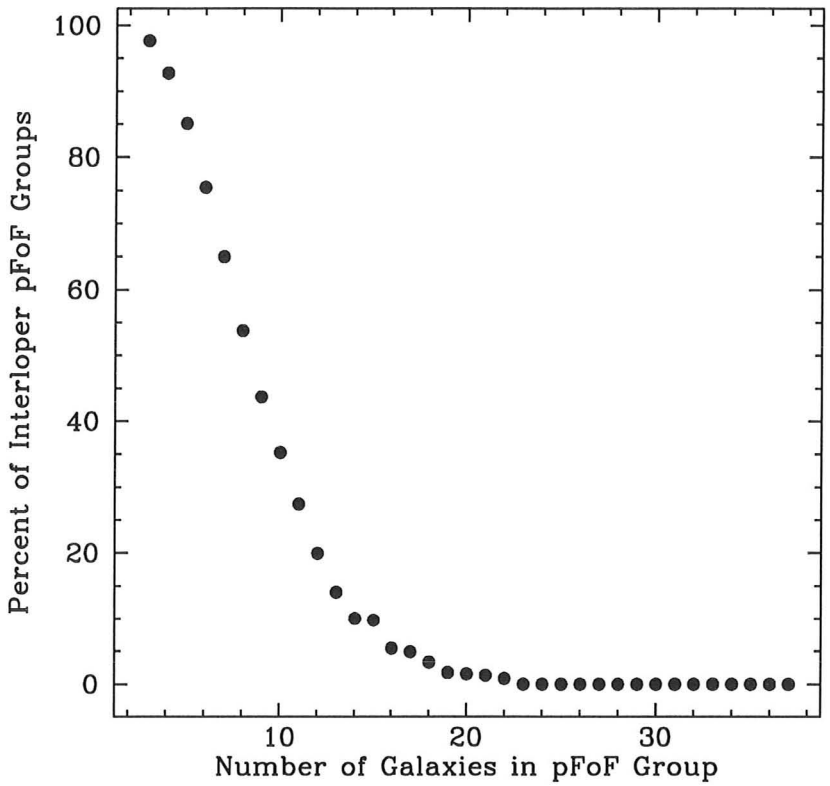

Figure 4.11 The percent of interloper groups is shown here as a function of the number of galaxies in the groups.

we see that only 0.655% of our pFoF CFHTLS groups and 3.06% of our pFoF MS groups (including masked and unmasked) have $N \geq 12$. Finally, recall from Section 3.3 that the MS galaxies are assigned higher weights than the CFHTLS galaxies on average, and that this leads to an overestimation of the number of interlopers in these figures and tables.

Figure 4.11, gives the percent of interloper groups as a function of the number of galaxies in a group. Of the 71,007 groups with three members, 69,364 are interlopers, and of these, 55,731 are interloper groups where none

$N_{galaxies}$ in Group	N_{groups}/deg^2	% Interloper Groups
12	27.5	19.9
13	10.8	14.0
14	6.8	10.0
15	3.5	9.8
16	1.8	5.5
17	0.5	4.9
18	1.0	3.3
19	1.0	1.8
20	0.5	1.6
21	0.0	1.4
22	0.8	0.9

Table 4.1 Shown here is the reliability and frequency of pFoF groups based on the number of galaxies in that group. For groups of 23 members or more, none of the groups have been found to be interlopers, however they are rare.

of the 3 galaxies have any real friends. The pFoF algorithm has more galaxies in groups than there "should be" as shown by Figure 4.9, and these galaxies in small groups (low density areas) are not considered clustered by the MS cones. Groups of three members make up 34.24%, 97.69% of these groups are interlopers, and 80.35% of these groups consist of galaxies that should not be in any group. Therefore, pFoF groups of three are not to be trusted. Table 4.1 details the reliability of the pFoF method. In this table, the number of interloper groups is based on those found in the MS data. The number of groups found per square degree is based on the results of the CFHTLS data as this data is better at representing the true number of groups we will find with photometric redshifts.

With this information in hand, we made a preliminary color-magnitude diagram (CMD) with the CFHTLS galaxies that are members of groups with

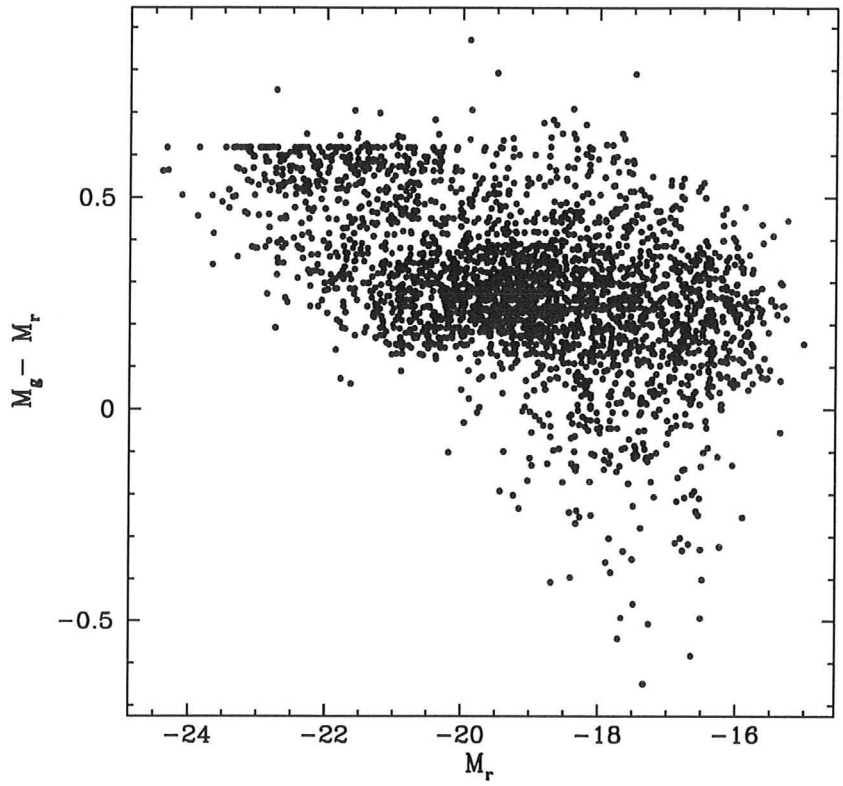

Figure 4.12 CMD of galaxies in pFoF groups of 12 or more members.

12 or more members, shown in Figure 4.12. The sharp feature in the red sequence is not physical, and further investigation is necessary to determine its cause. However, we notice a clear red sequence between M_r of -20 to -24, at a $M_g - M_r$ of ~ 0.62. This is an excellent sign, as galaxies in groups should have a defined red sequence.

Finally, in Figure 4.13 we investigate the average number of galaxies in a group as a function of redshift. As the redshift increases the average number of galaxies in a group drops slightly for the 'real' MS groups, whereas for the pFoF groups it remains about the same. Even though the galaxies at higher

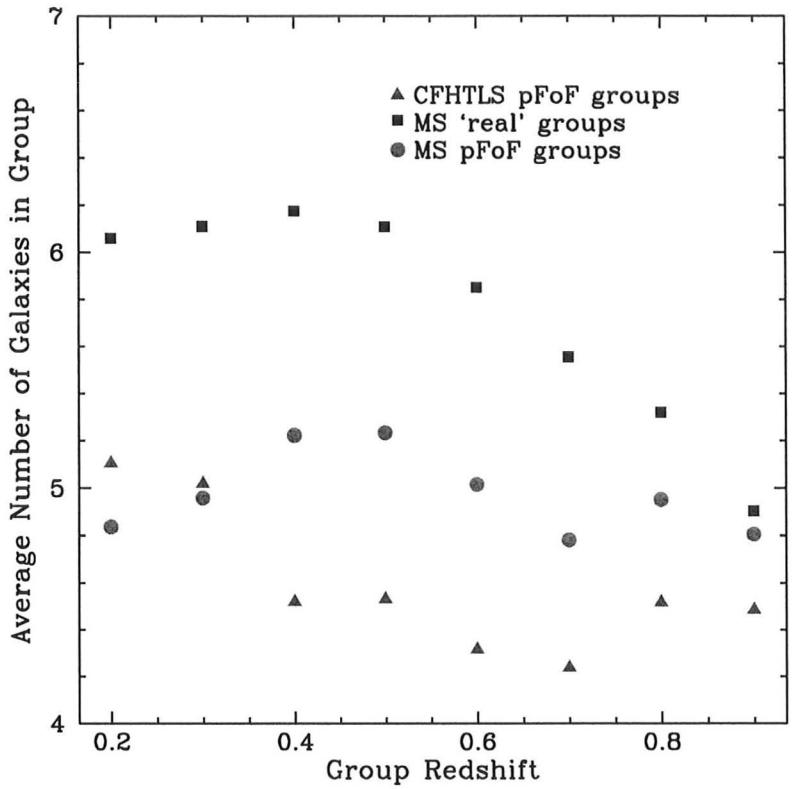

Figure 4.13 The average number of galaxies in a group is plotted here as a function of redshift for the different populations.

redshifts have larger redshift errors, the pFoF method is reliable in that it does not result in falsely large groups.

4.4 pFoF groups in the CFHTLS data

The groups are plotted for $N \geq 12$ in Figure 4.14, and for $8 \leq N \leq 12$ in Figure 4.15. Furthermore, Tables 4.2, 4.3, 4.4, and 4.5, give lists of the pFoF groups with $N \geq 12$ found in the CFHTLS Deep fields using $P_{ratio,crit} = 0.37$. Every group of $N \geq 3$ is assigned a group number in the order that they are found. The

group RA and Dec are luminosity-weighted. The P_{ratio}'s of the groups range from the limits of ~ 0.37 to ~ 1. The redshifts of these groups and the number of galaxies in the groups have previously been presented in Figure 4.13, and Figure 4.3 respectively. The group luminosities are on order of 1×10^9 $L_\odot$ to 1×10^{11} $L_\odot$. If we assume a mass-to-light ratio of $200h$ $M_\odot$ ${L_\odot}^{-1}$ (Parker et al., 2005), this corresponds to a range of mass from $1 \times 10^{13} h^{-1}$ $M_\odot$ to $1 \times 10^{14} h^{-1}$ $M_\odot$. This fits with the suggested mass for galaxy groups of $2 \times 10^{13} h^{-1}$ $M_\odot$, and lies below the high end of a galaxy cluster's mass of $1 \times 10^{15} h^{-1}$ $M_\odot$ described in Chapter 1 (Carroll & Ostlie, 2007).

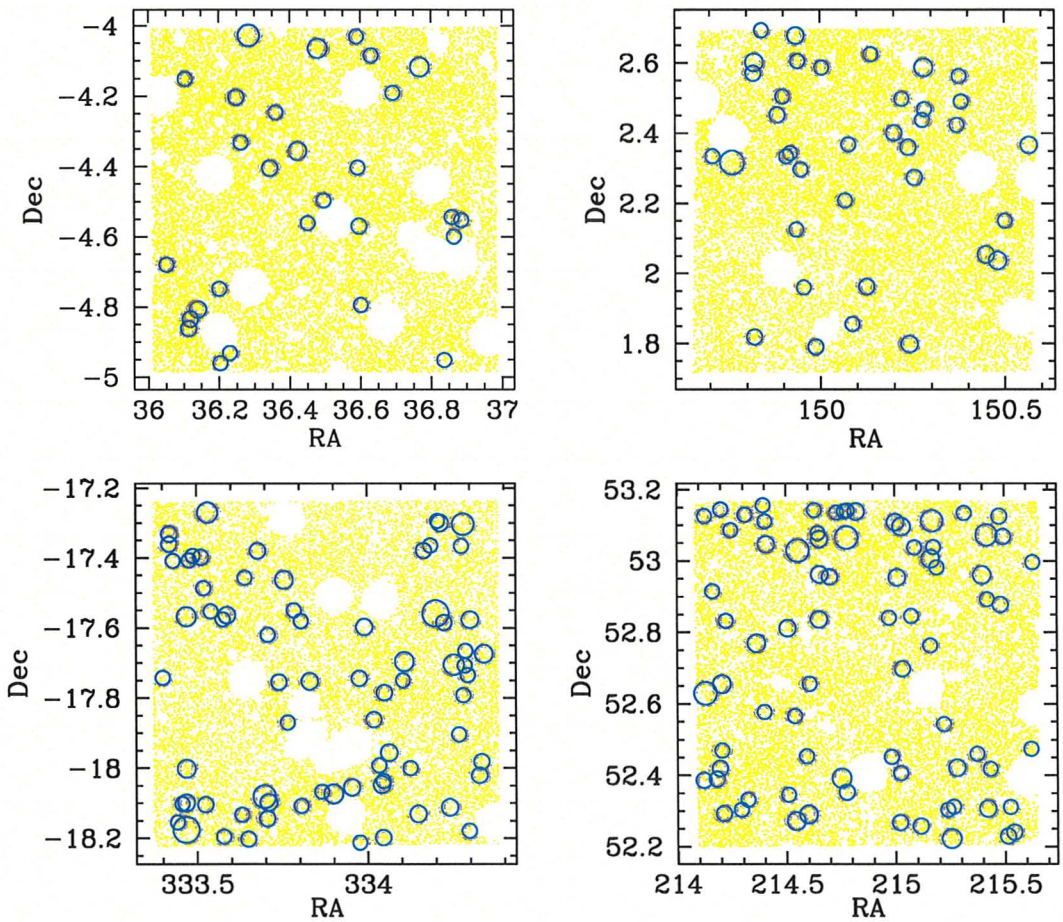

Figure 4.14 The four CFHTLS fields are shown starting with Field 1 in the upper left and counting clockwise. The blue circles mark the pFoF groups of $N \geq 12$ in the field, and the sizes are scaled by $N/8$. The yellow marks show the galaxies in the field.

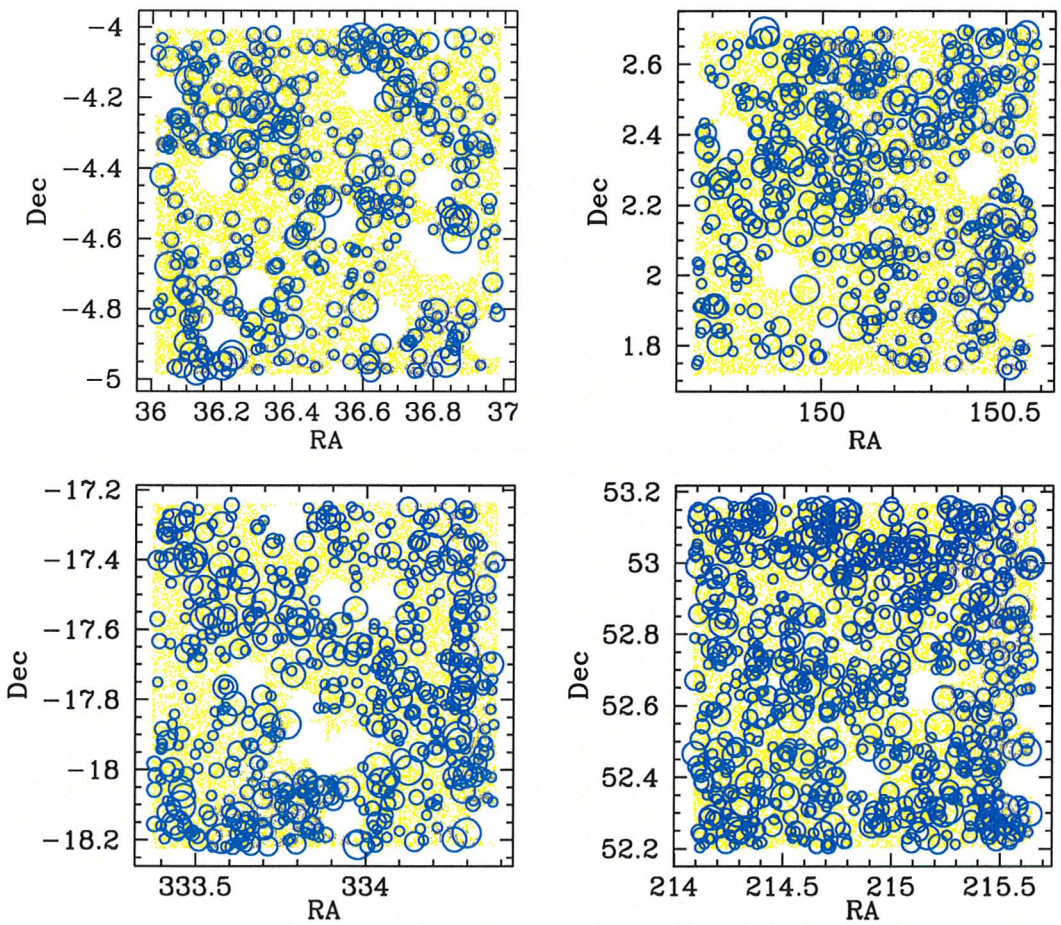

Figure 4.15 pFoF groups of $8 \leq N \leq 12$ are shown here in the four CFHTLS fields starting with Field 1 in the upper left and counting clockwise. The blue circles mark the groups, and the sizes are scaled by $N/2 - 3$. The location of the galaxies (and masked regions) in the field can be seen by the the yellow marks.

Table 4.2 pFoF groups from the CFHTLS Deep Field 1.

Group #	RA_{center}	Dec_{center}	N_{gal}	z_{group}	P_{group}	$L_{group}(10^{10}L_{\odot})$
D1-00389	+02:24:012.29	-04:40:045.37	12	0.450	0.544	29.122
D1-00259	+02:24:024.99	-04:09:004.41	12	0.870	0.655	32.540
D1-00181	+02:24:027.20	-04:51:041.24	13	0.530	0.558	63.567
D1-00169	+02:24:028.41	-04:50:004.87	13	0.520	0.545	48.126
D1-05103	+02:24:033.53	-04:48:025.47	14	0.550	0.533	12.573
D1-04775	+02:24:047.95	-04:44:054.11	12	0.530	0.530	11.922
D1-02083	+02:24:048.51	-04:57:037.21	12	0.620	0.391	35.838
D1-07059	+02:24:055.00	-04:55:053.32	12	0.310	0.484	5.339
D1-02049	+02:24:059.47	-04:12:011.31	13	0.630	0.650	19.857
D1-03889	+02:25:002.59	-04:19:055.81	12	0.950	0.588	16.469
D1-00807	+02:25:007.97	-04:01:041.19	18	0.240	0.371	27.147
D1-01737	+02:25:022.30	-04:24:018.41	13	0.360	0.628	18.602
D1-45440	+02:25:026.25	-04:14:047.87	12	0.230	0.507	4.169
D1-02675	+02:25:041.09	-04:21:021.92	15	0.230	0.474	9.014
D1-01501	+02:25:047.99	-04:33:041.61	12	0.910	0.392	27.621
D1-00064	+02:25:054.68	-04:03:053.22	16	0.730	0.434	65.940
D1-04272	+02:25:058.79	-04:29:049.62	12	0.950	0.514	14.877
D1-23881	+02:26:021.11	-04:01:051.38	12	0.360	0.712	3.749
D1-15777	+02:26:022.08	-04:24:011.00	12	0.340	0.648	2.825
D1-01634	+02:26:022.96	-04:34:010.96	13	0.320	0.655	15.905
D1-00930	+02:26:024.09	-04:47:040.55	12	0.370	0.493	16.267
D1-12237	+02:26:031.09	-04:05:003.97	12	0.320	0.776	5.501
D1-02533	+02:26:045.80	-04:11:027.22	12	0.950	0.528	14.164
D1-02543	+02:27:004.34	-04:07:004.19	16	0.240	0.578	15.374
D1-03012	+02:27:020.67	-04:57:006.33	12	0.230	0.592	9.238
D1-42713	+02:27:026.08	-04:32:038.51	12	0.330	0.719	10.840
D1-00260	+02:27:027.42	-04:35:057.92	12	0.340	0.620	51.384
D1-24740	+02:27:032.52	-04:33:009.76	12	0.280	0.759	6.098

Table 4.3 pFoF groups from the CFHTLS Deep Field 2.

Group #	RA_{center}	Dec_{center}	N_{gal}	z_{group}	P_{group}	$L_{group}(10^{10}L_{\odot})$
D2-45746	+09:58:049.67	+02:20:003.97	12	0.220	0.689	13.091
D2-01992	+09:59:002.34	+02:18:058.82	20	0.360	0.591	38.283
D2-05077	+09:59:016.23	+02:34:014.21	13	0.730	0.535	19.595
D2-20417	+09:59:016.72	+01:49:008.68	13	0.520	0.410	12.216
D2-02034	+09:59:016.88	+02:36:005.52	15	0.220	0.455	11.821
D2-01712	+09:59:021.62	+02:41:031.89	12	0.320	0.717	14.289
D2-00961	+09:59:031.96	+02:27:007.25	13	0.370	0.468	30.520
D2-02007	+09:59:035.40	+02:30:015.45	12	0.730	0.622	35.219
D2-01428	+09:59:037.79	+02:20:003.04	12	0.890	0.579	28.290
D2-07050	+09:59:040.45	+02:20:037.81	12	0.350	0.601	7.626
D2-01194	+09:59:044.21	+02:40:039.53	14	0.800	0.533	17.892
D2-03170	+09:59:044.33	+02:07:031.61	12	0.640	0.605	13.917
D2-00292	+09:59:045.25	+02:36:018.21	12	0.400	0.619	65.948
D2-02238	+09:59:047.22	+02:17:049.04	12	0.650	0.551	19.233
D2-01259	+09:59:049.08	+01:57:036.29	12	0.930	0.622	25.923
D2-11666	+09:59:056.77	+01:47:026.28	13	0.360	0.725	16.603
D2-01916	+10:00:000.94	+02:35:011.62	12	0.600	0.587	23.960
D2-02742	+10:00:016.17	+02:12:029.57	12	0.220	0.481	9.581
D2-00639	+10:00:018.46	+02:22:004.91	12	0.300	0.455	25.329
D2-00208	+10:00:020.79	+01:51:021.34	12	0.550	0.400	30.532
D2-00467	+10:00:030.16	+01:57:046.36	13	0.380	0.452	29.898
D2-40413	+10:00:033.05	+02:37:027.33	12	0.500	0.761	2.940
D2-01623	+10:00:048.01	+02:24:002.93	13	0.860	0.396	24.387
D2-11814	+10:00:052.99	+02:29:051.68	12	0.340	0.499	4.739
D2-05094	+10:00:057.18	+02:21:036.34	13	0.350	0.465	12.828
D2-10016	+10:00:057.63	+01:47:054.40	14	0.870	0.484	15.675
D2-02515	+10:01:001.16	+02:16:026.20	13	0.210	0.858	10.269
D2-03745	+10:01:006.34	+02:26:016.19	12	0.350	0.584	7.694
D2-11053	+10:01:007.30	+02:35:013.41	15	0.340	0.642	4.463
D2-12578	+10:01:007.68	+02:28:005.50	12	0.220	0.727	2.478
D2-42033	+10:01:028.83	+02:25:025.42	12	0.370	0.694	4.460
D2-04163	+10:01:030.18	+02:33:045.05	12	0.350	0.525	8.698
D2-01269	+10:01:031.57	+02:29:025.42	12	0.460	0.563	16.528
D2-00018	+10:01:047.61	+02:03:015.11	14	0.380	0.582	78.803
D2-00625	+10:01:055.26	+02:02:013.50	15	0.550	0.600	28.034
D2-02402	+10:02:000.09	+02:09:001.76	12	0.220	0.575	16.323
D2-06403	+10:02:015.93	+02:22:000.75	14	0.360	0.728	6.168

Table 4.4: pFoF groups from the CFHTLS Deep Field 3.

Group #	RA_{center}	Dec_{center}	N_{gal}	z_{group}	P_{group}	$L_{group}(10^{10}L_{\odot})$
D3-23825	+14:16:029.05	+52:23:009.84	13	0.470	0.784	11.951
D3-00956	+14:16:029.95	+53:07:033.89	12	0.820	0.474	29.849
D3-50928	+14:16:031.00	+52:37:044.83	19	0.230	0.781	2.484
D3-00367	+14:16:038.71	+52:54:054.17	12	0.880	0.567	51.880
D3-00553	+14:16:043.39	+52:23:022.27	13	0.550	0.648	22.478
D3-11397	+14:16:046.85	+52:25:009.31	13	0.830	0.542	9.352
D3-25497	+14:16:048.00	+53:08:040.94	12	0.220	0.551	1.151
D3-00968	+14:16:049.00	+52:39:018.33	15	0.590	0.390	32.426
D3-00816	+14:16:049.07	+52:28:010.15	12	0.880	0.499	37.677
D3-00317	+14:16:051.03	+52:17:035.30	13	0.440	0.691	24.297
D3-08739	+14:16:053.38	+52:49:055.18	12	0.350	0.556	6.154
D3-00514	+14:16:058.85	+53:05:011.11	12	0.910	0.628	62.500
D3-02616	+14:17:010.14	+52:18:013.25	12	0.460	0.375	14.507
D3-00888	+14:17:014.58	+53:07:049.12	12	0.350	0.496	66.433
D3-01487	+14:17:017.32	+52:19:054.68	12	0.460	0.547	31.224
D3-16292	+14:17:027.06	+52:46:009.65	15	0.230	0.679	2.367
D3-01400	+14:17:034.53	+53:09:023.63	12	0.650	0.627	25.266
D3-00443	+14:17:035.73	+52:34:037.72	12	0.560	0.791	23.169
D3-09613	+14:17:036.40	+53:06:040.87	12	0.610	0.571	9.789
D3-36202	+14:17:037.79	+53:02:048.93	14	0.620	0.562	3.024
D3-05785	+14:18:001.67	+52:48:041.91	14	0.350	0.434	9.334
D3-02394	+14:18:002.27	+52:20:044.94	12	0.350	0.706	9.435
D3-00852	+14:18:009.43	+52:33:057.75	12	0.710	0.383	23.702
D3-01363	+14:18:010.65	+52:16:018.77	15	0.490	0.621	33.147
D3-00832	+14:18:012.98	+53:01:045.09	19	0.440	0.389	34.912
D3-00030	+14:18:022.37	+52:27:012.05	12	0.340	0.429	55.806
D3-26457	+14:18:024.20	+52:17:026.71	15	0.220	0.582	2.072
D3-30722	+14:18:025.60	+52:39:020.79	12	0.240	0.762	0.997
D3-00061	+14:18:031.32	+53:08:034.92	12	0.400	0.591	74.344
D3-01144	+14:18:035.02	+53:04:042.19	12	0.390	0.652	40.811
D3-04426	+14:18:036.05	+52:50:012.35	14	0.890	0.617	22.121
D3-03155	+14:18:036.57	+53:03:041.19	14	0.440	0.708	13.295
D3-00856	+14:18:036.86	+52:57:044.17	14	0.600	0.578	31.891
D3-39370	+14:18:048.08	+52:57:020.99	13	0.340	0.702	1.580
D3-01876	+14:18:055.87	+53:08:010.33	12	0.610	0.731	17.104
D3-00341	+14:19:000.93	+52:23:031.78	16	0.290	0.589	50.627

Continued on next page

Table 4.4 – continued from previous page

Group #	RA_{center}	Dec_{center}	N_{gal}	z_{group}	P_{group}	$L_{group}(10^{10}L_{\odot})$
D3-10217	+14:19:004.14	+53:08:023.41	12	0.430	0.730	9.725
D3-00463	+14:19:006.41	+52:21:011.55	13	0.300	0.433	39.981
D3-00476	+14:19:006.74	+53:03:057.83	19	0.340	0.561	12.108
D3-53840	+14:19:007.44	+53:08:031.37	12	0.330	0.603	6.142
D3-03069	+14:19:017.22	+53:08:020.46	14	0.240	0.601	9.954
D3-00305	+14:19:053.17	+52:50:028.67	12	0.930	0.506	28.113
D3-01358	+14:19:055.88	+52:27:005.14	12	0.510	0.515	22.522
D3-00484	+14:20:000.29	+53:06:032.28	14	0.230	0.569	30.742
D3-02809	+14:20:002.22	+52:57:013.61	14	0.340	0.523	13.878
D3-00531	+14:20:004.85	+52:16:003.23	13	0.850	0.673	52.252
D3-53864	+14:20:005.88	+52:24:021.73	12	0.220	0.869	0.484
D3-39243	+14:20:006.97	+53:05:048.11	15	0.220	0.793	2.521
D3-01562	+14:20:007.81	+52:41:052.31	13	0.240	0.778	8.853
D3-03003	+14:20:017.58	+52:50:045.40	12	0.620	0.654	19.027
D3-15308	+14:20:021.33	+53:02:017.15	12	0.820	0.629	16.025
D3-29438	+14:20:027.47	+52:15:029.68	13	0.230	0.706	2.510
D3-02200	+14:20:038.33	+52:45:047.66	12	0.320	0.648	17.233
D3-10869	+14:20:038.93	+53:00:022.54	15	0.220	0.688	6.137
D3-01408	+14:20:040.60	+53:06:047.62	18	0.220	0.506	24.580
D3-01000	+14:20:042.02	+53:02:020.43	12	0.460	0.469	32.051
D3-00172	+14:20:045.55	+52:58:056.23	12	0.390	0.665	66.181
D3-28284	+14:20:053.17	+52:32:033.65	12	0.240	0.810	1.494
D3-00921	+14:20:057.28	+52:18:013.66	12	0.770	0.467	21.350
D3-08794	+14:21:001.22	+52:13:022.32	16	0.910	0.672	13.365
D3-08345	+14:21:003.64	+52:18:045.09	12	0.880	0.523	9.318
D3-07512	+14:21:007.49	+52:25:017.13	14	0.750	0.520	9.313
D3-02433	+14:21:015.78	+53:08:004.93	12	0.530	0.674	17.970
D3-23263	+14:21:029.65	+52:27:036.95	12	0.220	0.687	0.795
D3-01175	+14:21:035.28	+52:57:039.21	15	0.220	0.652	23.433
D3-30177	+14:21:040.40	+53:04:022.87	18	0.230	0.689	2.314
D3-04233	+14:21:040.97	+52:53:033.21	12	0.890	0.554	18.123
D3-09477	+14:21:041.36	+52:18:025.56	14	0.230	0.681	4.196
D3-02824	+14:21:044.13	+52:25:005.23	12	0.740	0.607	24.108
D3-03266	+14:21:054.50	+53:07:033.07	13	0.370	0.432	26.233
D3-24154	+14:21:055.69	+52:52:041.12	13	0.240	0.656	2.305
D3-03388	+14:21:058.80	+53:04:007.04	13	0.930	0.668	15.338
D3-00116	+14:22:003.43	+52:13:047.74	13	0.410	0.384	55.227
D3-00309	+14:22:006.25	+52:18:041.13	12	0.350	0.492	9.966

Continued on next page

Table 4.4 – continued from previous page

Group #	RA_{center}	Dec_{center}	N_{gal}	z_{group}	P_{group}	$L_{group}(10^{10}L_{\odot})$
D3-04462	+14:22:010.40	+52:14:027.82	13	0.720	0.372	22.806
D3-01247	+14:22:029.16	+52:28:026.85	12	0.550	0.680	31.186
D3-10508	+14:22:031.30	+52:59:048.64	12	0.230	0.580	3.486

Table 4.5: pFoF groups from the CFHTLS Deep Field 4.

Group #	RA_{center}	Dec_{center}	N_{gal}	z_{group}	P_{group}	$L_{group}(10^{10}L_{\odot})$
D4-49518	+22:13:036.19	-17:44:034.23	12	0.340	0.917	0.718
D4-05378	+22:13:040.83	-17:21:038.26	13	0.530	0.775	9.931
D4-01252	+22:13:040.94	-17:19:053.40	13	0.650	0.621	16.729
D4-15158	+22:13:043.51	-17:24:030.24	12	0.570	0.809	3.770
D4-01728	+22:13:046.33	-18:09:017.70	12	0.910	0.529	35.096
D4-28345	+22:13:049.65	-18:06:007.57	12	0.220	0.718	0.567
D4-06051	+22:13:052.41	-18:10:034.43	22	0.230	0.553	8.311
D4-27119	+22:13:052.62	-18:05:056.54	14	0.880	0.744	9.190
D4-01011	+22:13:052.75	-18:00:006.33	15	0.930	0.515	25.088
D4-00898	+22:13:052.86	-17:34:005.09	16	0.950	0.560	32.204
D4-09930	+22:13:054.78	-17:24:025.26	12	0.550	0.541	18.351
D4-05277	+22:13:057.49	-17:23:038.50	12	0.890	0.451	25.892
D4-06762	+22:14:003.07	-17:23:054.09	13	0.330	0.703	25.204
D4-00567	+22:14:004.87	-17:29:011.71	12	0.950	0.390	35.495
D4-03141	+22:14:005.96	-18:06:015.80	13	0.900	0.714	32.396
D4-04863	+22:14:007.79	-17:16:012.13	17	0.600	0.617	16.118
D4-01285	+22:14:009.84	-17:33:010.36	12	0.880	0.659	50.072
D4-02467	+22:14:018.34	-17:34:033.09	12	0.370	0.472	13.044
D4-03587	+22:14:018.73	-18:11:044.50	12	0.930	0.565	19.905
D4-04952	+22:14:021.26	-17:33:048.16	14	0.310	0.661	11.501
D4-04823	+22:14:031.48	-18:07:058.97	12	0.890	0.639	17.383
D4-06093	+22:14:033.71	-17:27:024.35	12	0.900	0.717	24.024
D4-01819	+22:14:035.95	-18:12:010.79	13	0.640	0.687	23.862
D4-08393	+22:14:042.88	-17:22:047.31	13	0.360	0.694	18.673
D4-00869	+22:14:047.37	-18:04:055.17	19	0.220	0.483	4.790
D4-09966	+22:14:049.34	-18:08:039.99	13	0.890	0.546	9.594
D4-27318	+22:14:049.82	-17:37:012.37	12	0.220	0.490	1.686
D4-00826	+22:14:050.29	-18:05:053.07	14	0.790	0.622	36.440
D4-51479	+22:14:057.57	-17:45:018.02	13	0.230	0.624	24.849
D4-03923	+22:15:001.39	-17:27:046.46	15	0.910	0.586	17.712
D4-00033	+22:15:003.81	-17:52:011.87	12	0.470	0.601	77.778
D4-00515	+22:15:008.18	-17:32:059.36	12	0.890	0.761	11.201
D4-39246	+22:15:013.17	-17:34:050.85	12	0.890	0.704	9.610
D4-06429	+22:15:013.31	-18:06:027.95	12	0.600	0.660	8.281
D4-00611	+22:15:019.27	-17:45:011.59	14	0.880	0.620	44.912
D4-02423	+22:15:027.78	-18:04:004.15	12	0.510	0.413	20.560

Continued on next page

Table 4.5 – continued from previous page

Group #	RA_{center}	Dec_{center}	N_{gal}	z_{group}	P_{group}	$L_{group}(10^{10}L_{\odot})$
D4-01972	+22:15:035.82	-18:04:027.00	16	0.480	0.674	25.769
D4-05996	+22:15:048.95	-18:03:019.34	14	0.240	0.437	6.422
D4-09454	+22:15:054.27	-17:44:040.02	13	0.340	0.577	3.560
D4-01618	+22:15:054.52	-18:12:047.05	12	0.900	0.618	26.709
D4-14839	+22:15:057.95	-17:35:052.59	14	0.220	0.633	3.809
D4-00031	+22:16:004.50	-17:51:043.78	13	0.920	0.403	73.976
D4-03711	+22:16:008.18	-17:59:034.91	12	0.930	0.679	19.698
D4-04700	+22:16:009.86	-18:02:049.40	14	0.890	0.635	23.979
D4-02935	+22:16:010.82	-18:11:054.95	13	0.340	0.490	28.338
D4-01622	+22:16:011.04	-18:02:013.08	12	0.600	0.634	15.611
D4-00024	+22:16:011.79	-17:47:007.59	13	0.880	0.475	10.191
D4-06223	+22:16:015.05	-17:57:023.01	14	0.900	0.507	18.147
D4-00871	+22:16:024.95	-17:45:000.98	12	0.900	0.565	38.538
D4-08033	+22:16:025.94	-17:41:046.86	16	0.880	0.438	26.406
D4-06818	+22:16:030.16	-18:00:001.11	12	0.920	0.657	24.519
D4-01141	+22:16:035.58	-18:07:050.07	14	0.230	0.456	23.315
D4-03238	+22:16:039.79	-17:22:044.34	13	0.350	0.374	12.514
D4-07504	+22:16:044.54	-17:21:052.64	12	0.220	0.629	4.657
D4-00558	+22:16:048.17	-17:33:032.03	22	0.900	0.422	41.639
D4-05405	+22:16:049.53	-17:17:038.47	12	0.560	0.451	15.303
D4-14576	+22:16:051.23	-17:18:000.31	14	0.220	0.440	3.515
D4-00594	+22:16:053.97	-17:35:004.72	13	0.920	0.523	44.041
D4-01404	+22:16:057.83	-18:06:043.05	14	0.230	0.675	17.959
D4-00786	+22:17:000.78	-17:42:019.21	17	0.920	0.645	25.845
D4-04960	+22:17:004.41	-17:54:014.74	12	0.900	0.724	13.600
D4-05269	+22:17:006.16	-17:21:057.32	12	0.230	0.710	6.046
D4-05064	+22:17:007.45	-17:47:032.78	12	0.560	0.678	7.909
D4-18808	+22:17:007.45	-17:18:011.62	18	0.900	0.730	15.393
D4-04266	+22:17:008.46	-17:42:029.22	12	0.910	0.420	16.496
D4-00704	+22:17:008.95	-17:39:056.64	12	0.920	0.706	21.991
D4-01395	+22:17:010.50	-17:44:006.76	12	0.870	0.550	40.967
D4-08207	+22:17:011.48	-18:10:044.61	12	0.900	0.484	12.683
D4-00659	+22:17:012.11	-17:34:036.01	14	0.910	0.380	31.119
D4-00765	+22:17:018.65	-18:01:015.33	13	0.910	0.617	52.873
D4-02446	+22:17:020.17	-17:58:054.17	13	0.510	0.523	21.403
D4-02184	+22:17:021.85	-17:40:027.42	15	0.940	0.407	27.341

For a full list of groups and their member galaxies, please contact lparker@mcmaster.ca.

4.5 The effects of a tighter constraint on P_{ratio}

From the above results, it is obvious that the pFoF method is adding more galaxies into groups than the Millennium Simulation says there are, as well as considering more than 60% of galaxies to be in groups. Not only do the groups contain galaxies who should not be in a group, but most of the galaxies in smaller groups do not really belong to a group. There are two main parameters that can be varied in order to produce the best results. One is $D0_{xy}$ and the other $P_{ratio,crit}$. Decreasing $D0_{xy}$ would help to remove galaxies that truly do not belong in a group, however, we have already set it to 0.25 Mpc, which is rather low, and upon a careful look at the MS 'real' groups many of these galaxies would be removed as well under such a restriction. The other option would be to increase $P_{ratio,crit}$ so that galaxies must have a higher probability ratio of being at the same redshift as a group to be considered part of that group. We choose $P_{ratio,crit} = 0.50$ and explore the effect this will have on our results. As discussed in Section 2.1, this is the $P_{ratio,crit}$ at which $\mid z_1 - z_2 \mid = \sigma_1 + \sigma_2$ when one σ is much smaller than the other (Li & Yee, 2008). Since this is the largest we can make $P_{ratio,crit}$ and still have $\mid z_1 - z_2 \mid = \sigma_1 + \sigma_2$, we consider 0.50 to be the upper limit on $P_{ratio,crit}$ and likewise 0.37 to be the lower. Thus by exploring the results of $P_{ratio,crit} = 0.37$ and $P_{ratio,crit} = 0.50$, we have studied the boundaries of $P_{ratio,crit}$'s range.

In Figure 4.16 we can see that the tighter $P_{ratio,crit}$ does result in a higher percentage of groups of three members (thus elsewhere the percentage is slightly

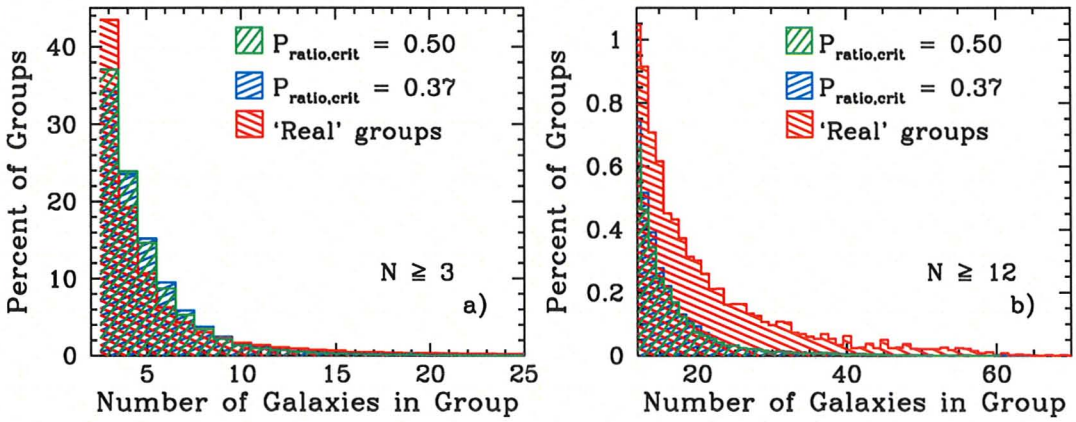

Figure 4.16 The number of members in groups are plotted for MS 'real' groups, and MS pFoF groups found using different $P_{ratio,crit}$'s. Plot (a) shows up to 25 group members to focus on the smaller groups, while plot (b) shows groups with 12 or more members.

decreased) but the overall effect is slight compared to the difference between the 'real' groups and the pFoF groups. Furthermore, since the groups of three are more likely to be interlopers, the increase in their number is not useful.

In Figure 4.17, while the peaks in the histograms may be shifted slightly, once again 61% of the pFoF groups are made up with galaxies that are not considered to be in a group by the Millennium Simulation.

The overall percentage of galaxies in a pFoF group is shown to decrease with a tighter cut on the P_{ratio} in Figure 4.18. This is expected, as fewer groups will be formed under this cut. However, the result is slight, and from Figure 4.17 we already know that the percentage of galaxies that have been wrongly assigned to a group has not changed with the $P_{ratio,crit}$.

The effect that a higher $P_{ratio,crit}$ will have on the number of 'real' groups found is shown in Figure 4.19. No difference is found between the two for 'real' groups of six or more members, however there is an interesting effect for 'real'

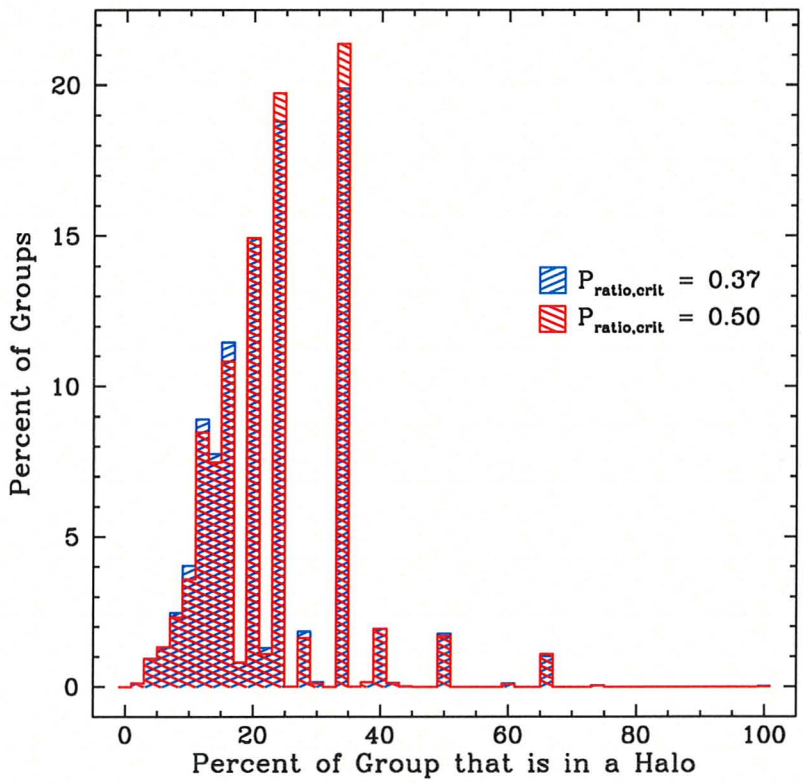

Figure 4.17 The percentage of pFoF groups that MS considers to be part of a 'real' group is shown here for different $P_{ratio,crit}$'s. For both cases, the dominant peaks at 0% are not shown here.

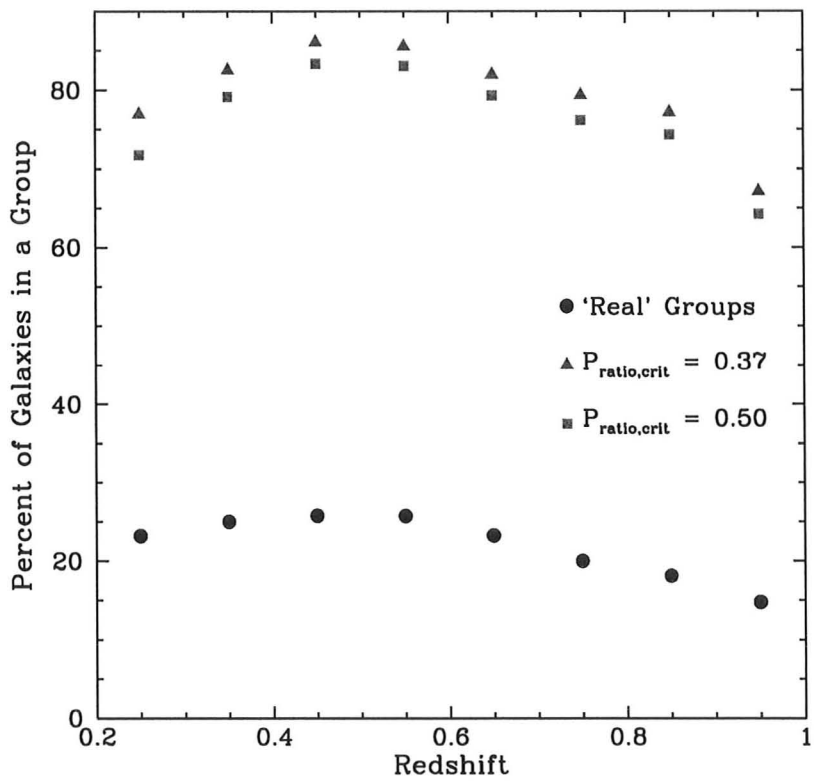

Figure 4.18 A histogram of the percent of galaxies that are members of a group is shown for different redshift bins for 'real' MS groups and pFoF MS groups found using different $P_{ratio,crit}$'s.

groups of up to five members. On first thought we might expect that the lower $P_{ratio,crit}$ would find more 'real' groups as it will allow for more links to be made between galaxies. However, we see here that in fact the higher number of links can be harmful. The reason is that while the number of overall links will be increased with the lower $P_{ratio,crit}$, so will the number of false links. Therefore, once a galaxy has been added to one or more groups, and it is time to decide which group this galaxy belongs best with, these false links can be harmful. This is especially true for small 'real' groups, where every link becomes that much more important for the discovery of the group.

Figure 4.20 demonstrates that while a $P_{ratio,crit}$ of 0.50 decreases the number of interloper pFoF groups, the overall effect is slight and restricted to groups of fewer than 18 galaxies. However, for those interested in group of fewer than 18 members, it can be useful to use the higher $P_{ratio,crit}$.

Table 4.6 displays the percent of interloper pFoF groups and the expected abundance of pFoF groups based on the number of galaxies in a group. Similar to Table 4.1, the percent of interlopers was taken from the MS data, while the expected number of groups per degree squared was taken from the CFHTLS data. The percent of interlopers decreases only slightly from what we saw in Table 4.1, and as expected from Figure 4.20, however it comes at a cost. The number of groups expected for a given number of member galaxies per square degree drops by 29.7% on average.

Figure 4.21 displays the average number of galaxies in a group as a function of redshift, and shows that there is no difference between the pFoF groups found using $P_{ratio,crit} = 0.37$ and $P_{ratio,crit} = 0.50$. This is good, as it shows

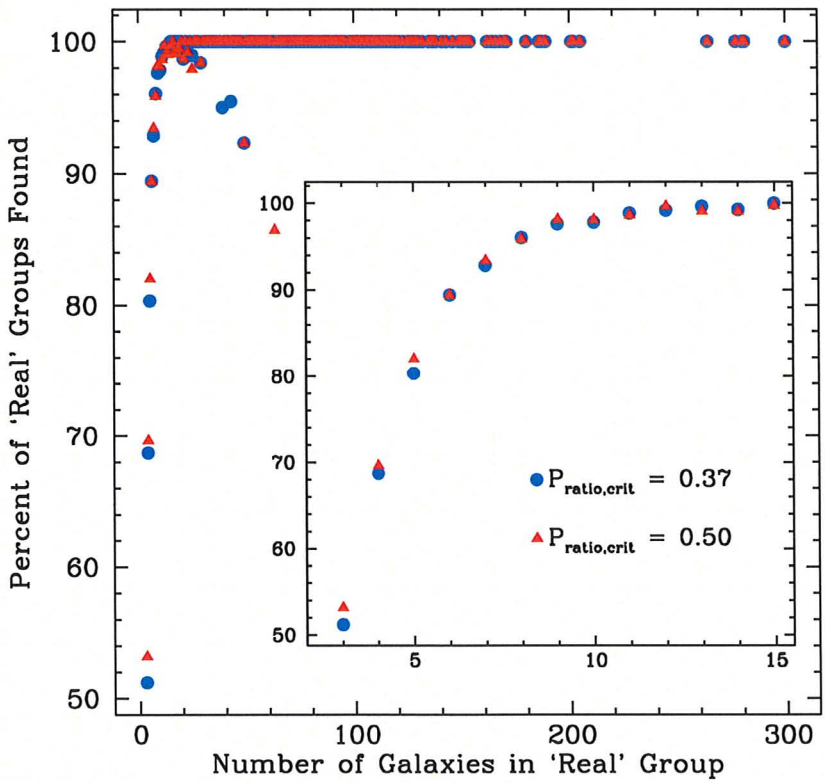

Figure 4.19 Plotted here is the percent of 'real' groups found as a function of the number of galaxies in that 'real' group. Results are shown for a $P_{ratio,crit}$ of 0.37 and 0.50. The inset is a closer look of the groups with 3 to 15 members.

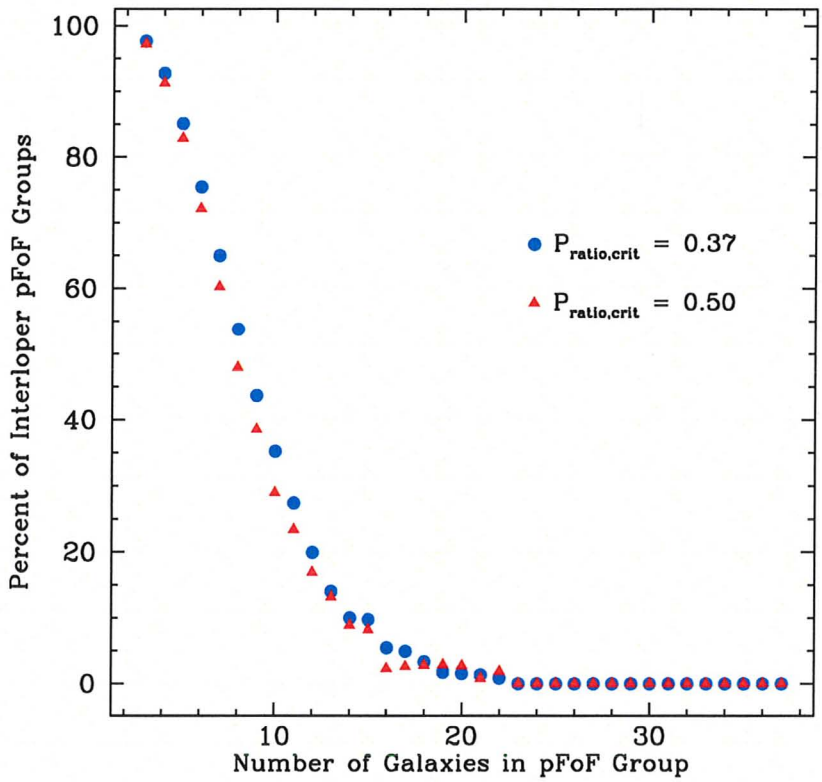

Figure 4.20 The percent of interloper pFoF groups is shown here as a function of the number of galaxies in a group for pFoF groups found using $P_{ratio,crit} = 0.37$ and 0.50.

$N_{galaxies}$ in Group	N_{groups}/deg^2	% Interloper Groups
12	14.3	17.7
13	8.0	11.3
14	2.5	10.6
15	2.0	7.4
16	1.0	5.5
17	1.0	5.0
18	0.8	3.2
19	0.3	1.9
20	0.0	4.0
21	0.5	1.3
22	0.3	0.9

Table 4.6 Shown here is the reliability and frequency of pFoF groups based on the number of galaxies in that group for a $P_{ratio,crit} = 0.50$. For the rare case of a group of 23 members or more, none of the groups have been found to be interlopers.

that even though we are making a tighter cut on P_{ratio}, it does not break apart groups at high redshift (where the galaxies tend to have larger σ_z's) to a noticeable extent.

The results presented thus far show that a very slight improvement may be made by raising the $P_{ratio,crit}$. However, this change has little effect and simply decreases the number of interloper groups but not the percent of interloper galaxies in a group. Furthermore, the higher $P_{ratio,crit}$ mostly affects groups with fewer than 16 members. These results are based on running the pFoF algorithm on the MS dataset, but as stated in Section 4.3, the galaxies in real CFHTLS groups have a range of redshifts because we really do not know where the true redshift lies within the range given by the redshift error, while we know the redshifts of the galaxies in the MS groups (and they are all nearly the same). The result is that we may make a tight cut in redshift space, by increasing $P_{ratio,crit}$, and it will have little effect on the 'real' MS

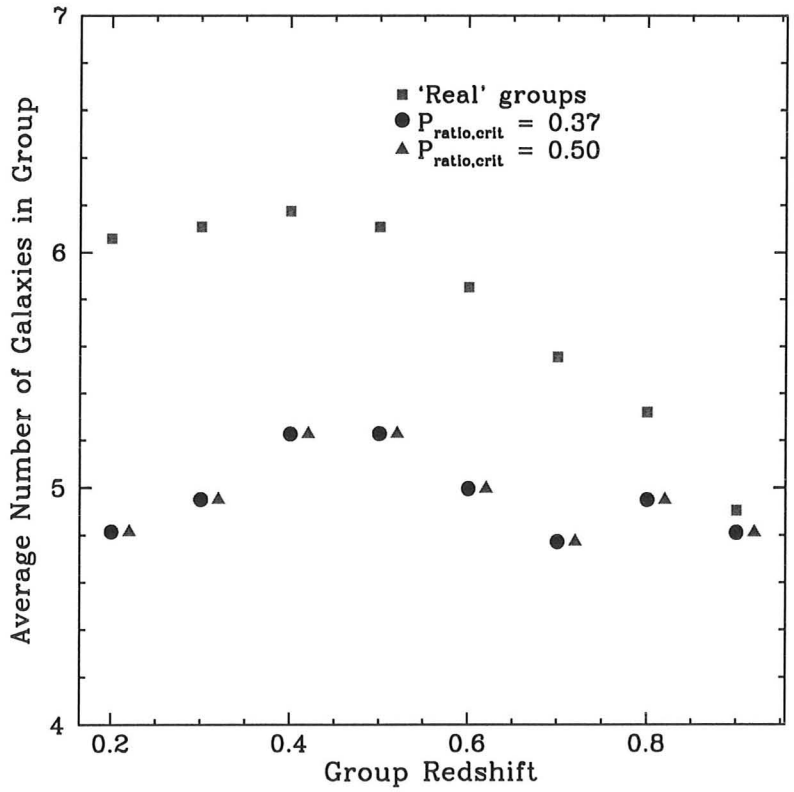

Figure 4.21 The average number of galaxies in a group is plotted here as a function of redshift for the MS 'real' groups, and the pFoF groups found with $P_{ratio,crit} = 0.37$ and 0.50. Since the difference between these last two cannot be distinguished, we have offset the $P_{ratio,crit} = 0.50$ data by $z = 0.02$.

groups. However the same cut on the CFHTLS galaxies could break apart a real system, especially if that system is elongated in the line of sight.

Chapter 5

Conclusion

5.1 Summary of results

We have presented our extended version of the pFoF algorithm by Li & Yee (2008) as well as our variation. The pFoF algorithm is a useful group identifying tool, unique in that it allows for the use of photometric redshifts. We applied this method to the four CFHTLS Deep Fields and used data from the Millennium Simulation to test the pFoF method. To make the datasets comparable, we manipulated the MS light cones to match the characteristics of the CFHTLS Deep fields. Furthermore, we explored our success and limits at making these datasets similar. Once we applied the pFoF algorithm to the MS datasets, we analyzed the results and found the range where the pFoF algorithm works best, and applied that knowledge to the CFHTLS fields. Finally we studied the effect the masked regions in the CFHTLS fields have on the pFoF results.

In doing so, we have shown:

- The masked and un-masked MS fields gave similar results, proving that the masked regions on the CFHTLS data have negligible effect on the pFoF method.

- The 5th Nearest Neighbor technique shows that the galaxy densities of the CFHTLS and MS fields are similar, which is necessary when comparing results from the pFoF method.

- The redshift distributions of the galaxies in the CFHTLS and MS fields are similar past a redshift of 0.4.

- The slight difference in the redshift distribution of the galaxies below a redshift of 0.4 causes similar differences in the pFoF group redshift distributions.

- The MS had more groups with $P_{ratio} \sim 1$ as a result of their accurate redshifts. An improvement could be made by scattering the redshift within the errors.

- The pFoF method results in an excess of galaxies being considered part of a group.

- Increasing $P_{ratio,crit}$ to an upper limit does little to improve the results, and this may only improve the MS pFoF groups.

- Due to uncertainties with the photometric redshifts, the pFoF algorithm is unreliable for smaller groups, so we must only look at the larger groups.

- Our most significant result is that the pFoF algorithm works best for groups of 12 galaxies or more, with less than 20% interlopers among $\geq$ 12 member groups.

- We have found 218 groups with 12 members or more within the $\sim$ 4 square degrees of the CFHTLS Deep fields. More than 80% of these groups should be real based on our analysis of the MS cones.

- The MS galaxies have larger weights than the CFHTLS galaxies on average, resulting in a conservative estimation on the number of real groups found by the pFoF method.

5.2 Future work

More work may be done in analyzing the galaxy densities. Another analysis tool useful in comparing the distribution of galaxies in the datasets is to measure the two point angular correlation function (hereafter 2PACF). In the method described by Landy & Szalay (1993) and Sinnott & Carlberg (2007), the 2PACF, $\omega(\theta)$, is defined by the function dP, which is the probability of finding an object at an angular separation on the sky in a solid angle:

$$dP = N^2[1 + \omega(\theta)]d\Omega \tag{5.1}$$

where $d\theta$ is the separation, and $d\Omega$ is the solid angle. This has been calculated for the CFHTLS fields before by McCracken et al. (2001), showing that in the range of $0.02 < \theta < 3.00$, the results follow the expected power law of

$\omega(\theta) = A_\omega \theta^{-\delta}$, with $\delta = 0.8$. However, how this fits to the MS data is yet to be seen.

The group catalog produced here is also a valuable tool, on which much work can also be done. These results can be used to study the photometric properties of these groups, for example, the dependence of the red galaxy fraction on redshift or on galaxy local or global density. We can also determine if the largest groups can be detected by lensing, and if so, use this to calculate the masses. Furthermore, observing how these groups appear in the x-ray or with spectroscopy would make an interesting study. One could also look at the dependence of galaxy properties (i.e. the Spectral Energy Distribution (SED), and star formation history) on the environment. The pFoF algorithm should also be applied to fields with complete spectroscopy to further calibrate the method. For example, we can determine the photometric group size which would correspond to a spectroscopic group of five members, using spectroscopic data from surveys like the Canadian Network for Observational Cosmology Cluster Redshift Survey Catalogs II (CNOC2 Cluster Redshift Survey) *. Finally, another interesting study may be done by comparing galaxy and group properties between samples selected with photometry, spectroscopy, or X-rays. With our own Milky Way galaxy traveling in the Local Group, it is easy to see the importance of a dependable group finding technique that will help expand our samples of galaxy groups on which we may do these studies.

* http://www.astro.utoronto.ca/ cnoc/index.html

Bibliography

Bartelmann, M. 2001, A&A, 370, 754

Berlind, A. A., et al. 2006, ApJS, 167, 1

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734

Carroll, B. W., & Ostlie, D. A. 2007, An Introduction to Modern Astrophysics (2 ed.) (Addison-Wesley)

Clark, P. J., & Evans, F. C. 1954, Ecology, vol. 35, p. 445-453 (1954)., 35, 445

Cooper, M. C., Newman, J. A., Madgwick, D. S., Gerke, B. F., Yan, R., & Davis, M. 2005, ApJ, 634, 833

Croton, D. J., et al. 2006, MNRAS, 365, 11

Davis, M., & Djorgovski, S. 1985, ApJ, 299, 15

Deng, X.-F., He, J.-Z., Chen, Y.-Q., & Jiang, P. 2009, Astroparticle Physics, 31, 255

Dunkley, J., et al. 2009, ApJS, 180, 306

Eke, V. R., et al. 2004, MNRAS, 348, 866

Hawkins, E., et al. 2003, MNRAS, 346, 78

Huchra, J. P., & Geller, M. J. 1982, ApJ, 257, 423

Ilbert, O., et al. 2006, A&A, 457, 841

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., & Yoshida, N. 2001, MNRAS, 321, 372

Kitzbichler, M. G., & White, S. D. M. 2007, MNRAS, 376, 2

Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64

Li, I. H., & Yee, H. K. C. 2008, AJ, 135, 809

Li, I. H., Yee, H. K. C., & Ellingson, E. 2009, ArXiv e-prints

McCracken, H. J., Le Fèvre, O., Brodwin, M., Foucaud, S., Lilly, S. J., Crampton, D., & Mellier, Y. 2001, A&A, 376, 756

Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713

Parker, L. C., Hudson, M. J., Carlberg, R. G., & Hoekstra, H. 2005, ApJ, 634, 806

Pritchet, C., & SNLS Collaboration. 2006, Colour Term Coefficients and Zeropoints, Retrieved July 27, 2009, from: http://www.astro.uvic.ca/~pritchet/SN/Calib/ColourTerms-2006Jun19/index.html#Sec04

Sinnott, B., & Carlberg, R. G. 2007, Undergraduate Thesis at University of Toronto

Springel, V., et al. 2005, Nature, 435, 629

Tago, E., et al. 2006, Astronomische Nachrichten, 327, 365

Yee, H. K. C., Gladders, M. D., Gilbank, D. G., Majumdar, S., Hoekstra, H., Ellingson, E., & the RCS-2 Collaboration. 2007, ArXiv Astrophysics e-prints