Galaxy Groups in the Updated 2MRS Using a Graph-Theory Based Friends-of-Friends Algorithm

By

Trystan Lambert

Supervisor: Prof. Renée C. Kraan-Korteweg

Co-Supervisor: Prof. Thomas H. Jarrett

Department Astronomy
University of Cape Town

Submitted in partial fulfillment of the requirements for the degree of M.Sc. in Astrophysics degree at the University of Cape Town

February 2020

Abstract

A galaxy group catalogue for the recently-completed 2MASS Redshift Survey (2MRS, Macri et al., 2019) is presented which consists of 44572 redshifts, including 1041 new measurements for galaxies mostly located within the Zone of Avoidance. The galaxy group catalogue is generated using a novel, graph-theory based, modified version of the Friends-of-Friends algorithm. Several graph-theory examples are presented throughout this paper, and include a new method to identify substructures within groups. The results and graph-theory methods have been thoroughly interrogated against previous 2MRS group catalogues and a Theoretical Astrophysical Observatory (TAO) mock by making use of cutting-edge visualization techniques including immersive facilities, a digital planetarium, and virtual reality. This has resulted in a stable and robust catalogue with on-sky positions and line-of-sight distances within 0.5 Mpc and 2 Mpc of the group respectively. It has recovered all major groups and clusters. The final catalogue consists of 3022 groups, resulting in the most complete "whole-sky" galaxy group catalogue to date. The 3D-positions of the groups are presented, as well as their luminosity and comoving distances, observed and corrected number of members, richness metric, velocity dispersion, and estimates of R_{200} and M_{200}. Three additional catalogues are provided: 2MRS galaxies found in groups, a catalogue of subgroups, and the catalogue of 687 new group candidates which had no counterparts in previous 2MRS-based analyses.

Dedication and acknowledgements

There are several people I would like to acknowledge. Without their help this thesis would have been impossible.

I would like to thank Dr. Lisa Crause who helped first spark my initial passion for observational astronomy. Thanks to Dr. Margaretha Pretorius for providing a constant stream of advice and support. Thanks to both believing in me throughout this process and providing with an environment which allowed me to finish this work.

I would also like to thank Janet and Warren Beech who very thoughtfully and lovingly provided support at a time when I needed it the most. Thanks to my dad for providing humour when none was to be found. Thanks to my mom who has on always stood by my side and has relentlessly supported me through every possible challenge and whom I owe the world. Thanks to the remainder of my family (which is too large to name individually but who all deserve mentions). The people who support me unquestionably despite the fact that they don't know what I really do.

I need to thank Prof. Lucas Macri. Not only for hosting me during turbulent times but also for lending his expertise throughout my research career, making me the best possible observer I could be, and being a massive inspiration. Special thanks to my co-supervisor Prof. Thomas Jarrett who has played no small part in my journey and has spent countless hours nursing me throughout my masters, both in the lab and in the pub. He has truly been a dear friend and confident throughout every difficult step. Finally, special thanks to my supervisor Prof. Renée Kraan-Korteweg for having stuck with me since even before my graduate career began. She has had an unfaltering belief in my abilities and has provided invaluable guidance throughout my career and I am truly thankful for everything that she has done through the years.

This thesis is dedicated to Anne Isaacs and Grant Engel.

Author's declaration

I hereby declare that the work in this dissertation is my own, except where indicated by specific reference in the text and that it has not been submitted for any other academic award. Work done in collaboration with, or with the assistance of others, is indicated as such. Any views expressed in the dissertation are those of the author.

Signed by candidate

SIGNED: ..... DATE: .....

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

It has long been known that galaxies are not isolated in space. They form part of larger structures such as groups, clusters, filaments and voids that together form part of the so-called cosmic web (Davis et al., 1982; Zeldovich et al., 1982; Klypin & Shandarin, 1993; Bond et al., 1996; Jarrett, 2004). The existence of these large-scale structures is predicted and expected in the current leading model of cosmology, the well known ΛCDM (Peebles, 1993). It is therefore reassuring that these structures are evident when visually examining distributions of galaxy positions in redshift space (Fairall, 1998).

Visual inspection of galaxies in redshift space immediately reveals the presence of large-scale structure and over densities of galaxies (in the form of groups and clusters; Peebles, 1980; Davis et al., 1982; Huchra & Geller, 1982; Tully & Fisher, 1987; Peebles, 1993; Fairall, 1998; Huchra et al., 2005; Crook et al., 2007; Robotham et al., 2011; Huchra et al., 2012; Tempel et al., 2016b,a, 2018). However, while visual inspection of redshift data has been used to show the existence of groups, (Hubble, 1934; Peebles, 1975; Seldner et al., 1977; de Lapparent et al., 1986a), this method of visually identifying and classifying structures is subjective and is neither robust nor quantitative (Jarrett, 2004; Crook et al., 2007; Robotham et al., 2011; Huchra et al., 2012; Alpaslan et al., 2014; Tempel et al., 2016b; Saulder et al., 2016; Tempel et al., 2016a; Jarrett et al., 2017; Kourkchi & Tully, 2017; Tempel et al., 2018). As a result, several algorithmic and computational methods for objectively identifying groups have been put forward. One of the most established and applied methods over the past decades has been the Friends-of-Friends (FoF) algorithm (Huchra & Geller, 1982). It has been extensively used to find galaxy groups in numerous magnitude-limited redshift surveys (Huchra & Geller, 1982; Ramella et al., 1997; Trasarti-Battistoni et al., 1997; Crook et al., 2007; Robotham et al., 2011).

1.1 Redshift Surveys

The Hubble-Lemaître law is the famous relationship between recessional velocities of galaxies and their distances (Hubble, 1929), due to the expansion of the universe. Measuring a galaxy's recessional velocity is a relatively easy process which can be done by measuring the Doppler shift in the galaxy's spectrum (hence redshift). This can be done for as many galaxies as the technology and telescope time allow. From these velocity measurements a distance can be calculated using the Hubble-Lemaître. Although this distance measurement is not entirely accurate (because the measured velocity of a galaxy is a combination of its peculiar velocity and cosmological velocity) the method is far easier and more applicable than any other method of calculating a galaxy's distance. This makes redshift surveys one of the most convenient and efficient tools to study the 3-dimensional distribution of galaxies.

Redshift surveys are very useful tools which lend themselves to achieving a variety of science goals. Being able to objectively identify large-scale structures within these surveys is relevant to addressing a number of unanswered questions in observational cosmology. These include accounting for the Cosmic Microwave Background (CMB) dipole anisotropy (Rauzy & Gurzadyan, 1998; Bilicki et al., 2011), establishing a relationship between galaxies and dark matter halos, probing dark matter interactions (Coutinho, 2016; Lu et al., 2016), constraining cosmological models (Peebles, 1980), and exploring the relationship of environment on galaxy evolution (Dressler, 1980; Tempel et al., 2016a, 2018).

Redshift surveys come in a variety of flavours. The two most common types are those constrained to a small angular area on the sky versus those covering a large angular area of the sky. Firstly there are the so-called pencil-beam surveys which measure galaxies through to high redshifts, examples of which are the Galaxy and Mass Assembly Survey (Baldry et al., 2010) and the Cosmic Evolution Survey (Scoville et al., 2007). Secondly, the wide-surveys that come at expense of depth such as the 6df and 2df redshift surveys (Colless et al., 2001; Jones et al., 2004, 2009). The exact nature of the redshift survey depends on the science goals at hand, instrumentation, and availability of telescope time (Fairall, 1998). Narrow and deep pencil-beam surveys provide excellent tools to probe the evolution of galaxies and their environments as a function of cosmic time. Shallow and wide surveys allow for large-scale structure of the local universe to be mapped and local galaxy environments to be examined.

Several limitations have to be taken into account when designing and creating a redshift survey, of which the largest consideration is our position within the disk of Milky Way, and that the Milky Way itself blocks a lot of light from extragalactic sources.

1.2 The Zone of Avoidance

Although numerous large-area redshift surveys have been made over the decades, none span the entire 4π sky. The overwhelming majority of so-called "whole-sky" surveys, exclude the regions close to the Milky Way plane where high stellar densities, increasingly brighter sky-background, and high levels of extinction make observing galaxies very difficult (Kraan-Korteweg & Lahav, 2000; Kraan-Korteweg et al., 2002; Kraan-Korteweg, 2005; Kraan-Korteweg et al., 2008; Staveley-Smith et al., 2016; Kraan-Korteweg et al., 2017, 2018). This region of exclusion is often referred to as the Zone of Avoidance (ZoA). As a result of the ZoA, a truly "whole-sky" redshift survey does not exist. The requirement for such a survey has been highlighted for several decades to fully explain the CMB dipole anisotropy, a well known signature caused by the peculiar motion of our own Local group with respect to the CMB (Lineweaver et al., 1996; Rauzy & Gurzadyan, 1998). This peculiar motion of our own Galaxy cannot adequately be accounted for due to the obscuration by the ZoA which hides large mass concentrations such as the Great Attractor (Lynden-Bell et al., 1988; Kraan-Korteweg & Lahav, 2000), the 3C129 cluster in Perseus-Pisces supercluster (Ramatsoku et al., 2014, 2016) and the newly discovered Vela Super Cluster (Kraan-Korteweg et al., 2015, 2017; Courtois et al., 2019). The Vela Super Cluster also highlights the importance of mapping this often overlooked region. Simply put, there might very well be unpredicted large-scale structures which have gone undetected, the consequences of which would be an incomplete understanding of the make-up and dynamics of the local universe. The physical gap left behind in most redshift surveys is an apt metaphor for the very real gap in our knowledge.

While a truly whole-sky, magnitude-limited redshift survey does not exist, imaging surveys have been made of the entire sky. Notably among these is the 2 Micron All Sky Survey (2MASS, Skrutskie et al., 2006), which uniformly mapped 99.998% of the entire sky in the J, H, and $K_{\rm s}$ bands. Near-infrared light is far less affected by dust in the Galactic plane, resulting in a less-prominent ZoA in 2MASS. The 2MASS extended source catalogue (2MASX, Jarrett, 2004; Skrutskie et al., 2006) was extracted consisting of around 1.6 million sources. However, the sheer number density of the stars in the Galactic plane makes 2MASX incomplete below $|b| < 5^\circ - 8^\circ$, depending on Galactic longitude (Jarrett, 2004).

1.3 The 2MASS Redshift Survey

The near whole-sky nature (and diminished ZoA) of 2MASX enabled the 2MASS Redshift Survey (2MRS; Huchra et al., 2005, 2012; Macri et al., 2019), a twenty-year concerted effort to measure the redshifts of ~ 45000 2MASX sources with $K_{\rm s}^{\rm o} < 11\overset{\rm m}{.}75$ and thus obtain the most comprehensive (sky-coverage wise) whole-sky redshift survey. A

preliminary data release with a magnitude limit of $K_s^o < 11\overset{m}{.}25$ consisted of $\sim$ 24000 galaxies (Huchra et al., 2005). This deeper data release reached the target magnitude depth. It contained over 44000 galaxies Huchra et al. (2012). Although it was $\sim$ 97% complete, it was still significantly incomplete at low Galactic latitudes. The final data release (Macri et al., 2019) added the remaining $\sim$ 1000 redshifts, which are mostly located at very low latitudes in the ZoA. Figure 1.1 shows the distribution of measurements from Huchra et al. (2012) and Macri et al. (2019). The recent completion of 2MRS, with most of the new redshifts lying close to the ZoA, allows for new large-scale structures in the thinly-mapped area to be uncovered.

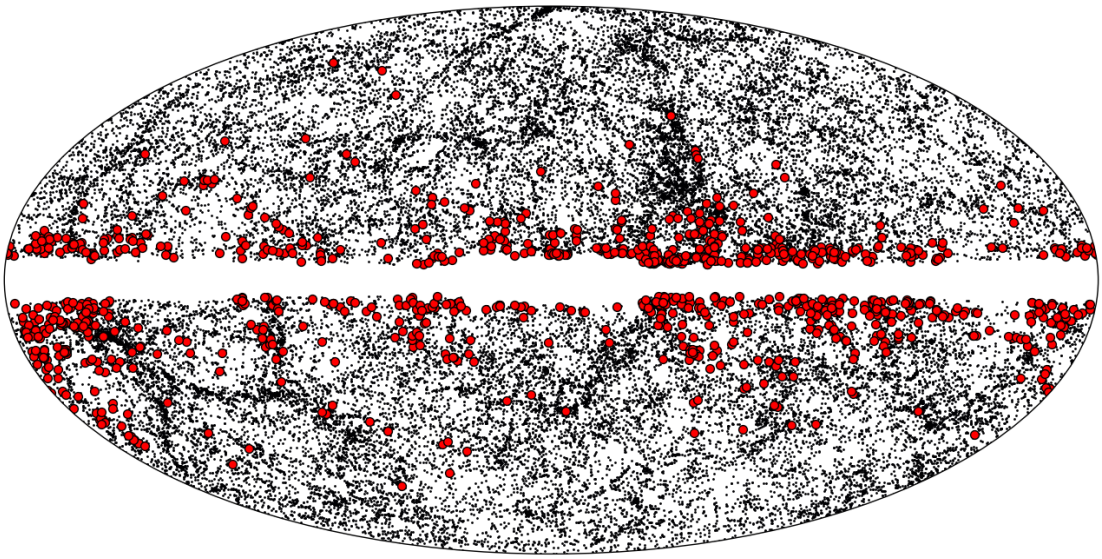

Figure 1.1: Aitoff projection of the complete 2MRS centered on Galactic co-ordinates (Macri et al., 2019). Black dots represent the galaxies with redshifts in Huchra et al. (2012), red dots the previously missing redshifts. Red dots have been scaled for visibility.

A further benefit of having a complete whole-sky galaxy catalogue covering the local Universe, is that it lends itself well to the derivation of a group catalogue, one that is more complete along the ZoA than previous work. These catalogues can be complementary to dedicated surveys within the ZoA such as dedicated HI surveys (e.g., Staveley-Smith et al., 2016; Ramatsoku et al., 2016; Kraan-Korteweg et al., 2018). A complete whole-sky local-universe group catalogue is also required to allow for quantification of other large-scale structures such as filaments and voids (Alpaslan et al., 2014), which are useful in conjunction with other studies of galaxy flows due to large-scale structures in the local Universe (Pomarède et al., 2017; Courtois et al., 2019; Tully et al., 2019). Furthermore, a complete 2MRS group-catalogue based on spectroscopic redshifts, can be compared to one based on photometric redshift estimates (2MPZ, Bilicki et al., 2014), which may improve the accuracy of the later technique.

1.4 Galaxy Groups and the Friends-of-Friends Algorithm

Identifying groups in redshift surveys in an objective manner is paramount to fully exploiting redshift surveys. One of the most successful algorithms to achieve this has over the past four decades been the Friends-of-Friends algorithm (Huchra & Geller, 1982). This algorithm has been developed to identify galaxy groups in particular. While the exact definition of a galaxy group differs in the details in which they are defined by different research teams, it is widely understood that galaxy groups are gravitationally bound systems of a number of galaxies which can be as small as 0.1 Mpc (in the case of compact groups (Zheng & Shen, 2020)) or on the scales of several Mpc (Dressler, 1980; Dressler & Shectman, 1988; Fairall, 1998; Crook et al., 2007; Tully, 2015; Tempel et al., 2016a). Galaxy clusters are similar and closely related to galaxy groups. The defining work on galaxy clusters has been Abell et al. (1989), which defines the well-known Abell radius to be 1.5 h^{-1} Mpc. The difference in the definition of a galaxy group or a galaxy cluster are not well-defined and the boundary between massive groups and low-mass irregular clusters are not clear-cut. Throughout this thesis I do not really make a distinction between groups or clusters (as is discussed later on). A lot of effort has been put into recovering both clusters and groups.

In order to characterize the large-scale structure I have modified the FoF algorithm and, after optimisation, applied it to the final (deepest and complete) version of the 2MRS to date. This modified version of the FoF algorithm aims to improve upon the already successful traditional algorithm by addressing several shortcomings such as sensitivity to initial conditions, non-generalizable parameter selection, and non-physical group membership identification. A FoF-based group finder is presented resulting in a highly accurate and robust "whole-sky" galaxy groups catalogue. The catalogue is verified by comparing the results against previous renditions of the 2MRS galaxy catalogue, a deep mock catalogue to test for various (Malmquist-like) biases in the algorithm, and applying an exhaustive visual inspection using state of the art 3-D visualization tools.

1.5 The Role of Environment on Galaxy Evolution

Galaxies are not homogeneous in their physical make-up. They come in numerous shapes, sizes, masses, morphologies, and chemical make-up (Buta, 2013; Jarrett et al., 2017). How galaxies form, evolve, build their stellar populations and form their distinct morphologies presents many challenges to our understanding of the universe. It is widely agreed that the effect effect that the environment plays on a galaxy's properties is significant (Dressler, 1980; Jarrett et al., 2017; Gordon et al., 2018; Bianconi et al., 2020; Carlesi et al., 2020; Otter et al., 2020). Studying the role that environment has on galaxy formation and evolution requires knowledge not only of the positions of the galaxies (which act

as tracers of the baryonic mass throughout the cosmic web) but also the environments in which these galaxies reside. Redshift surveys and their corresponding galaxy group catalogs make for very important tools in decoding galaxy formation and evolution (Dressler, 1980; Fairall, 1998; Alpaslan et al., 2014; Jarrett et al., 2017).

As mentioned above, there are two main types of redshift surveys, namely very narrow and deep surveys and wide and shallow surveys. Narrow surveys are excellent tools for studying galaxy evolution and formation through cosmic time while shallow surveys lend themselves very well to exploring the role of environment on galaxies' morphologies in the Local Universe ($z < 0.2$). Nearby galaxies in the local universe are most often found in galaxy groups (van de Weygaert & Bond, 2008; Robotham et al., 2011; Gordon et al., 2018). Thus, the 2MRS galaxy group catalogue (being the widest area survey to date) makes for an optimal survey to explore the effects of galaxy environments ranging from small groups (e.g., Local Group) to the largest of clusters (e.g., Coma Cluster); previous 2MRS group catalogues have been successfully applied to environmental studies of galaxies (e.g., O'Brien et al., 2018; Calderon & Berlind, 2019; Greene et al., 2019).

1.6 Thesis Overview

The structure of this thesis is as follows: In Chapter 2 I discuss the original FoF algorithm used by Huchra & Geller (1982) and explore the various methods and improvements which have implemented in order to construct the group finder presented here. Chapter 3 explores the various visualization facilities and techniques which were used during the validation stage of building the group finder. The final 2MRS Group catalogue is presented in Chapter 4 along with several supplementary catalogs. Chapter 5 contains discussions and conclusion. Throughout this thesis $H_0 = 73$ km s^{-1} Mpc^{-1}, $h = H_0/100$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

Chapter 2

Group Finder

2.1 Friends-of-Friends

The FoF algorithm has been the canonical method for identifying groups in magnitude-limited redshift-surveys because of the simplicity of the method (Huchra & Geller, 1982; Ramella et al., 1997; Crook et al., 2007; Duarte & Mamon, 2014; Tully, 2015; Tempel et al., 2016a). Although the core algorithm remains, for the most part, unchanged from the original formalism by Huchra & Geller (1982), its subsequent applications to various redshifts surveys have varied according to their wavelength coverage, completeness, depth, and area.

The FoF algorithm for the $K_{\rm s}^{\rm o}$ $<11\overset{\rm m}{.}75$ 2MRS sample was based on the one developed by Crook et al. (2007), who created a groups catalogue based on the 2MRS $K_{\rm s}^{\rm o} < 11\overset{\rm m}{.}25$ sample (Huchra et al., 2005). In its simplest form, this algorithm iteratively isolates galaxies i and j and determines whether they are in close enough proximity to each other in projected spatial and radial velocity-space to be gravitationally associated. If i and j are close in redshift space they are considered "friends". All friends of i can be found this way, followed by all the friends of friends of i, and so forth. A group is then defined according to degrees of association.

In principle galaxies should be considered "friends" if they are gravitationally bound. If the projected separation of galaxy i and galaxy j, with an angular separation θ_{ij}, defined as $D_{ij} = \sin\left(\frac{\theta_{ij}}{2}\right) \frac{v_{\rm avg}}{H_0}$, is less than some linking length D_l, then they are close enough, in terms of a plane-of-sky projection, to be associated. The line-of-sight distance is also considered, represented, in this case, by the difference in cz values of the two galaxies ($\Delta v = |v_i - v_j|$). If this distance is smaller than some linking velocity v_l, the galaxies are close enough to be associated along the line of sight.

If a pair of galaxies meets both conditions (associated both along the plane-of-sky and line-of-sight) they are considered "friends". The two parameters used to determine whether two galaxies are friends are D_l and v_l.

The simplest approach would be to keep these values constant. However, this does not take into account a variety of biases and only corrects for the change in projected separation with distance (Huchra & Geller, 1982).

Following their work, as well as Crook et al. (2007), D_l has instead been scaled in a manner which takes the sampling density as a function of redshift into account, given the survey magnitude limit. Namely,

$$D_l = D_0 \left[\frac{\int_{-\infty}^{M_{ij}} \Phi(M) dM}{\int_{-\infty}^{M_{\rm lim}} \Phi(M) dM} \right]^{-\frac{1}{3}}, \tag{2.1}$$

where

$$M_{\rm lim} = m_{\rm lim} - 25 - 5\log\left(\frac{v_f}{H_0}\right) \tag{2.2}$$

and (2.3)

$$M_{ij} = m_{\rm lim} - 25 - 5\log\left(\frac{v_{ij}}{H_0}\right), \tag{2.4}$$

and v_f, H_0, $m_{\rm lim}$, and $\Phi(M)$ are the fiducial velocity (set to 1000 km s^{-1} Huchra & Geller (1982),Crook et al. (2007)), the Hubble constant, the apparent-magnitude limit of the sample, and the Schechter luminosity function (Schechter, 1976) respectively:

$$\Phi(M) = 0.4\ln(10)\Phi^* 10^{0.4(\alpha+1)(M*-M)} \exp\left(-10^{0.4(M^*-M)}\right). \tag{2.5}$$

As proposed by Kochanek et al. (2001), $\alpha = -1.02$, $M^* = -24.2$, $\Phi^* = 1.08 \times 10^{-2} h^3 \mathrm{Mpc}^{-3}$ to represent the parameters of the K-band luminosity function.

Defining the linking-length D_l according to Eq. 2.1 scales the surveyed volume by the number density of galaxies which can maximally be observed at that distance (Huchra & Geller, 1982). In the case of v_l, a scaling, similar to that of Eq. 2.1 could be implemented; however, it has been argued that this is unnecessary because the velocity dispersion of a galaxy group is independent of redshift (Huchra & Geller, 1982). Thus, setting v_l to some constant value is sufficient and will not introduce a bias in velocity dispersion as a function of distance (Huchra & Geller, 1982; Crook et al., 2007). Therefore, this work adopts $v_l = v_0$, where v_0 is a constant.

D_0 can be represented as the radius of the sphere used in calculating the density contrast ($\delta\rho/\rho$) in Eq. 2.6 (Huchra & Geller, 1982):

$$\frac{\delta\rho}{\rho} = \frac{3}{4\pi D_0^3}\left[\int_{-\infty}^{M_{\rm lim}} \Phi\left(M\right) dM\right]^{-1} - 1 \tag{2.6}$$

Determining the appropriate D_0 and v_0 requires careful selection and optimization of the parameters. If D_0 and v_0 are too small, all the galaxies will be put into their own individual groups. Similarly, in the extreme case, if D_0 and v_0 are too large, every galaxy will be put into one giant group consisting of the whole data set. The optimal choice of these parameters lies somewhere in between. A physical case can be made for mostly all choices, but a certain level of arbitrariness cannot be avoided. This issue of a fixed parameter choice has been accepted by previous works. In this thesis however, a new method of parameter selection is proposed by spanning through the FoF parameter space and using the parameter space itself to statistically identify groups, opposed to relying on a fixed set of v_0 and D_0 values which has several short-comings as discussed in the following subsections.

Ramella et al. (1997) found that linking lengths of $D_0 = 0.56$ Mpc and $v_0 = 350$ km s^{-1} most accurately recover galaxy groups. This was shown by Ramella et al. (1989) who measured the number of groups, total number of galaxies in groups, the median velocity dispersion, and the first and third quartiles of the median velocity dispersion for the 6° and 12° CfA (Huchra & Geller, 1982) survey as well as a mock catalog of the survey by de Lapparent et al. (1986b). This was done for varying linking lengths with $D_0 = 0.56$ Mpc and $v_0 = 350$ km s^{-1} found to be the optimum choice of linking lengths at limiting the number of interlopers but not breaking apart large clusters. These same optimum linking lengths were later verified again in Ramella et al. (1997) on the Northern CfA2 survey (Huchra et al., 1995), who measured numerous group properties for several different linking lengths. These were compared to a geometric simulation of the sample. These linking lengths were once again validated and used in Ramella et al. (2002) on both the Updated Zwicky Catalog and Southern Sky Red-shift Survey (Falco et al., 1999; Ochsenbein et al., 2000). Finally, this combination of linking lengths was been applied successfully to earlier versions of the 2MRS (namely Crook et al., 2007). Given that this linking length choice has been successfully used over four different surveys (including an earlier 2MRS survey) and has been validated through several independent methods, we choose $D_0 = 0.56$ Mpc and $v_0 = 350$ km s^{-1}. This choice is later evaluated against our own mock catalog for further validation.

In order to incorporate the graph structure (discussed shortly), D_0 and v_0 are varied between 0.56 ± 0.1 Mpc and 350 ± 100 km s^{-1}, respectively, over 100 runs (using 2 kpc

and 2 km s^{-1} steps, respectively). The values are within the search grid used previously by Ramella et al. (1997, 1989) and are also validated on mocks later in this thesis.

2.1.1 Shortcomings of the Traditional FoF Algorithm

Selecting a particular set of values for v_0 and D_0 can statistically alter the results when applying the traditional FoF algorithm. Firstly, no particular static choice of parameters can be completely generalized. Galaxy groups are all different and vary in size and dynamics, thus different sets of parameter choices work better for different groups and no single set can be optimized for all groups.

In some cases it may be appropriate to use tighter parameters when groups are small. However, larger groups may then be "shredded" into smaller, non-physical groups. Likewise, with a larger set of parameters, two physically distinct groups may be found as belonging to one group. A static parameter choice is unable to deal with the range in size, density, and dispersion of galaxy groups and clusters.

The FoF algorithm can sometimes exhibit "runaway". Even with a strict set of linking lengths, large, non-physical groups are found because new friends are constantly being identified as being part of the group. This is because the FoF algorithm does not take into account what the current group already looks like before searching for new members. Runaway may also occur in systems where galaxies are incorrectly included within groups because of their chance alignments between neighboring groups and other large-scale structures. This results in groups which are too large and often too elongated to be considered physical.

Furthermore, the FoF algorithm renders all results (groups) with equal confidence. Groups with a large number of members which are very tightly constrained in redshift space are considered equally likely compared to groups consisting of a very low number of members, that are loosely constrained in redshift space.

For these reasons, a modified version of the FoF has been developed that considers a many different sets of linking lengths, thereby rendering the group finder statistically robust.

2.1.2 Hard Limits

To prevent runaway connections between groups, hard limits are introduced in both redshift space ($v_{\max}$) and projected on-sky distance ($\delta_{\max}$). This constrains how large any one particular group can grow. After the first friends are found, the center of the proto-group $(\bar{RA}, \bar{Dec}, \bar{v})$ is determined, where $\bar{RA}$, $\bar{Dec}$, and $\bar{v}$ are the mean of RA, Dec, and velocity of the current group members. Friends-of-friends are only added for galaxies within the proto-group with $|v_i - \bar{v}| < v_{\max}$ and with projected angular distance

$\delta_i = \sin\left(\frac{\Delta\theta}{2}\right) \frac{v_i}{H_0} < \delta_{\max}$, where $\Delta\theta$ is the angular separation between the galaxy i and the center of the proto-group. The center of the proto-group is recalculated with every new addition of a member until no new members are found.

Although $v_{\max}$ and $\delta_{\max}$ are not dissimilar to the linking lengths v_l and D_l, their choice is not as physically ambiguous. The values $v_{\max}$ and $\delta_{\max}$ can be thought of as the maximum velocity distribution of a group and the maximum radius of a group, respectively. Both properties have been studied extensively for well-known groups and clusters (Dressler, 1980; Huchra & Geller, 1982; Colless et al., 2001; Jones et al., 2004; Crook et al., 2007; Robotham et al., 2011; Alpaslan et al., 2014; Ramatsoku et al., 2016). $\delta_{\max}$ is set to the Abell radius, $R_A = 1.5\ h^{-1} = 2$ Mpc (Abell et al., 1989). Typical group and cluster velocity distributions (σ_v) have been shown to vary between 500 km s^{-1} and 1000 km s^{-1} (Peebles, 1980; Sandage & Tammann, 1981; Dressler & Shectman, 1988; Abell et al., 1989; Ramella et al., 1997; Loeb & Narayan, 2008; Tempel et al., 2016b; Coutinho, 2016). Therefore, $v_{\max}$ is set to represent the 3σ level for a typical cluster, i.e. $3\sigma_v$= 3000 km s^{-1}. The limit along the line-of-sight direction is purposely large in order to account for the observational bias associated with redshift measurements in that direction. The higher quality accuracy in the plane-of-sky positions of the galaxies allows for a stricter limit.

The hard limits are chosen to be the size of distinct clusters since this is the largest that a group could physically grow. This ensures that "shredding" of clusters and particularly large groups does not occur, while at the same time inhibiting runaways. Smaller groups will still be identified.

By introducing these hard limits, the modified FoF algorithm will migrate towards the highest densities of galaxies first rather than around randomly-chosen galaxies. This has the advantage that large collections of galaxies will preferentially shift the center to the cores of groups, reducing any bias that the algorithm may have with respect to the randomly-selected starting galaxy.

In summary, the introduction of hard limits improves the FoF algorithm by: inhibiting runaways, allowing large groups and clusters to be found without shredding them into smaller groups, and increasing its sensitivity to higher densities rather than randomly-selected starting points.

An unfortunate drawback from introducing hard limits is that the group finder is no longer robust to initial conditions. This means that the order which the group finder is run alters the results. In order to correct this (as well as other issues), the group finder is stabilized over many runs.

2.1.3 Stabilizing the Algorithm Over Many Runs

To stabilize the algorithm many different runs of the modified FoF algorithm (with hard limits) are executed for varying sets of linking length parameters and random starting points. The results of the numerous runs are averaged and conglomerated into one final result. Accepting groups which are found over several runs ensures that the results are robust to initial conditions (correcting the issue arising from introducing hard limits) and that statistical anomalies are removed. Secondly, by working with a range of parameters, larger groups and massive clusters that are not a chance agglomeration of smaller groups can still be identified without losing sensitivity to smaller groups. The execution of numerous runs furthermore allows the construction of confidence intervals. Associations of galaxies found under the strictest of parameters are more reliable than ones based on a more relaxed choice of parameters.

Every run produces a new group catalogue but not every group appears in every run. Groups can be split up, combined, have varying members, or disappear completely from run to run. Because of this, galaxy-galaxy pairs are tracked throughout all the runs. Since the defined galaxy groups themselves change with every run, they cannot be as easily tracked. However, the number of times galaxy i and galaxy j were found in a group together can be tracked reliably through any number of trials. In this way the strength of the connection between any i and j galaxy pair can be assessed.

Recording galaxy-galaxy associations results in a list of vectors. A single vector comprises the i^{th} and j^{th} galaxies, and the integer parameter w describes how many times galaxies i and j were found to be members of the same group. After k runs on a data set containing n galaxies, a final list of $n(n-1)/2$ elements is created where each element takes the form of (i, j, w) where $i, j \in [1, n], i \neq j$ and $w \in [0, k]$. This list can then be used to recreate the groups, taking into account the varying parameter space covered throughout the runs. This type of data set, being topological in nature, lends itself well to graph-theory, which is used to interpret pairwise interactions.

2.2 FoF Graph Theory Adaptation

To manage and assess the resulting group catalogs that are averaged over, a mathematical graph is constructed from the final list of FoF results. The nodes of the graph represent the galaxies in the redshift survey, and the connecting lines, or weighted edges between the nodes, represent the number of times a pair of galaxies is found in the same group. Due to the topological nature of the resulting data set, mathematical graphs are an optimal way to manage and quantify results from multiple runs, and provide a powerful visualization tool (Pascucci et al., 2011). To ensure stability against statistical anomalies, edges with a w value less than 0.5 (i.e., 50 % repeatability) are removed.

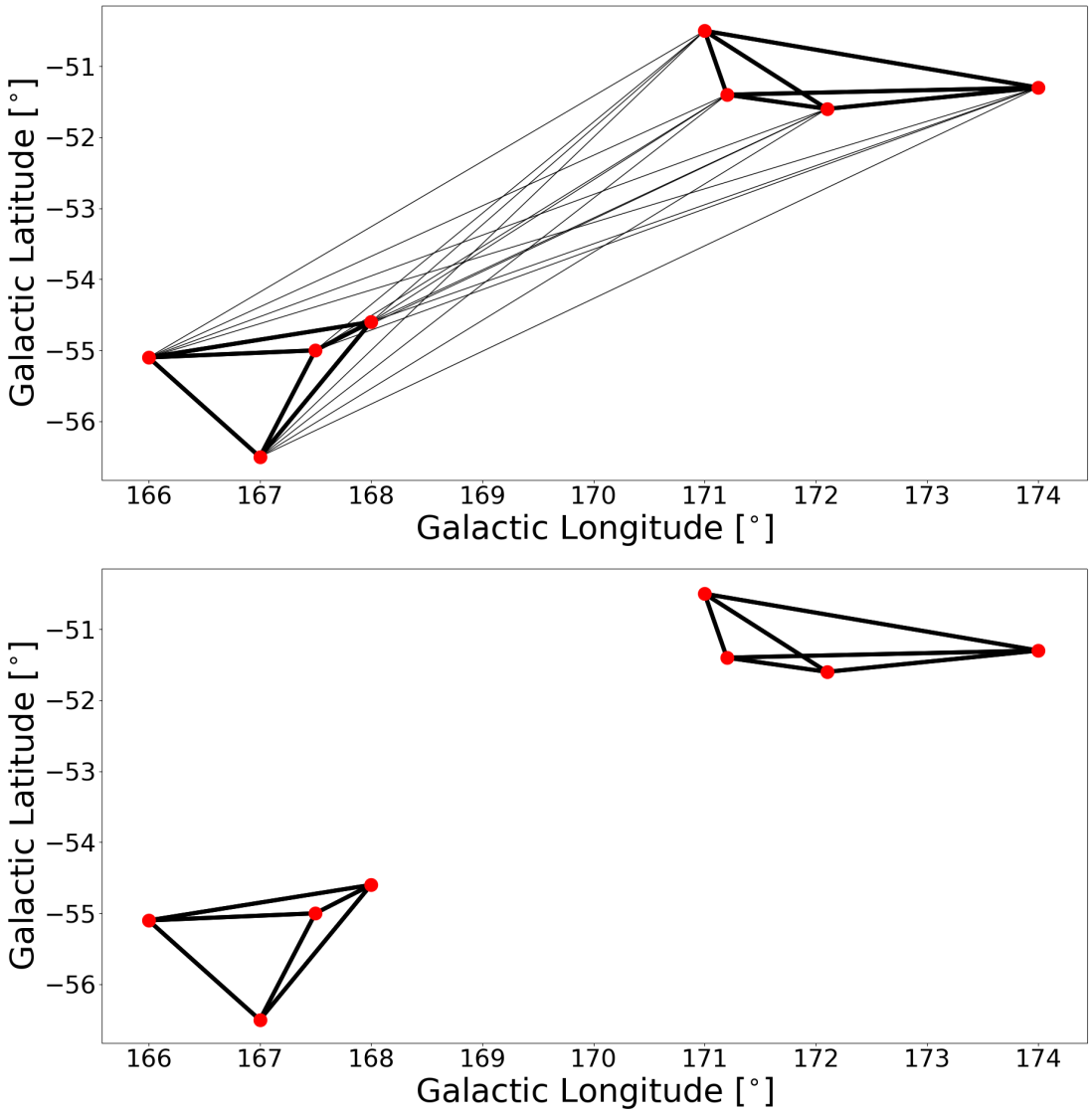

Figure 2.1: Graph representation of a single group after averaging over many runs. Points represent galaxies in the 2MRS survey. Edges are weighted according how often the connected pair were found in the same group. Thick edges represent pairs where galaxies were found $> 90\%$ of the time, thin edges pairs which were identified $< 10\%$. The top panel shows the group before weak edges are removed. The bottom panel shows how the group is split into two groups after the weak edges were removed.

This results in a single large graph which consists of many smaller disconnected subgraphs. The final group catalogue is generated by identifying all of these disconnected subgraphs by identifying all isolated graph objects within the main graph, using the `Graph.subgraph` function in the `networkx` Python package. The remaining subgraphs are the groups selected for further inspection. At this point every group in the catalogue can be represented as an independent, self-contained graph.

An example of how removing weak edges results in statistical robustness is shown in Fig. 2.1, where the graph represents an average over many runs for a particular group. The points (galaxies) are plotted in Galactic coordinates. Visual inspection of the points imply two separated groups; note that they were found as a single group in less than 10% of the runs. This can be seen from the weighting of the weakest edges (lines) connecting the two groups. Edges are also closely related to the linking length choice; in this example, the algorithm finds one group for a large linking length while a more moderate linking length results in two separate groups. A continuous range of linking lengths ($D_0 = 0.56 \pm 0.1$ Mpc and $v_0 = 350 \pm 100$ km s^{-1}) are executed over several runs. This allows the removal of weak edges, hence eliminating statistical anomalies and rejecting non-physical linking lengths.

This visualization technique provides a synergistic medium between human and computer understanding. A simple inspection of the the points in Fig. 2.1 (in this particular example) clearly demonstrates that these two groups really are separate entities when overlaid with weighted edges. Averaging over many runs using graph theory and only considering statistically significant groupings and physical linking lengths (by removing weak edges) results in a more stable and robust algorithm over the traditional FoF.

Edges are powerful tools; however, a lot of additional information can also be extracted from individual nodes in the respective group graphs which provide another angle of visual interrogation. In a similar way to edges, the nodes of all the graphs are weighted to take into account their number of connections and how strong said connections are. This is accomplished by defining the "connectedness score" of an individual node in a graph as:

$$S_i = C_i W_i, \tag{2.7}$$

where C_i is the percentage of the graphs edges connected to node i and W_i is the average weight of all the edges connected to node i, defined respectively as

$$C_i = c_i \left[\frac{n\,(n-1)}{2} - 1 \right]^{-1} \tag{2.8}$$

and

$$W_i = \frac{\sum_{j=1}^{c_i} w_{ij}}{c_i}, \tag{2.9}$$

where c_i is the total number of edges connected to node i, n is the total number of nodes in the graph, and w_{ij} is the weight of edge j connected to node i (the percentage of runs in which galaxy i was found in the same galaxy group as galaxy j.

Two metrics are defined with respect to the graph itself, namely the total and average connectedness score. These two global group statistics quantify the reliability of a group given the strength of its edges and the number of members. Indeed, the total connectedness score can be thought of as the corresponding number of nodes which make a mathematical "complete" graph (a graph where each node is connected to every other node). The average connectedness score would represent the number of members of a group found in every single run. For example: a group might have 20 group members, with a total connectedness score of only 10. This would imply that it could be a group of 10 members found in every single run, with an additional 10 members found at a lower weighting or connectedness, while still satisfying the criterion of at least 50% repeatability.

These are the metrics that are later used to identify groups which seem unstable, and require further inspection by visual examination.

2.3 Identifying Substructure

The graph description of each individual group furthermore allows for an innovative method of identifying substructure within groups, therefore extending the capabilities of the original FoF algorithm.

Cutting at a fixed significance level ensures robustness. However, cutting at higher significance levels can, in some cases, result in a group being sub-divided further. In order to combat the pseudo-randomness of the significance cut, every group is visually inspected to determine whether a stricter significance cut would result in separating groups. If this is the case then these are identified as substructures in a given group.

For every group, edges are iteratively removed from the graph from lowest to highest weighting, and with every iterative cut the number of subgraphs with three or more nodes is calculated. Once all edges have been removed, the maximum number of subgraphs with the lowest edge-cut become the subgroups.

Fig. 2.2 shows an example of such a case. The entire system is identified as a single group because all edges in Fig. 2.2 are within the cut-off limit of 50%. However, if the cut-off limit would instead be at 70 %, then the weak edges would have been removed, resulting in two independent groups (in the same way as the groups in Fig. 2.1). Removing edges sequentially from weakest to strongest results in two separate structures being identified. These substructures are classified as subgroups. In Fig. 2.2 the subgroups are represented by the two large red circles.

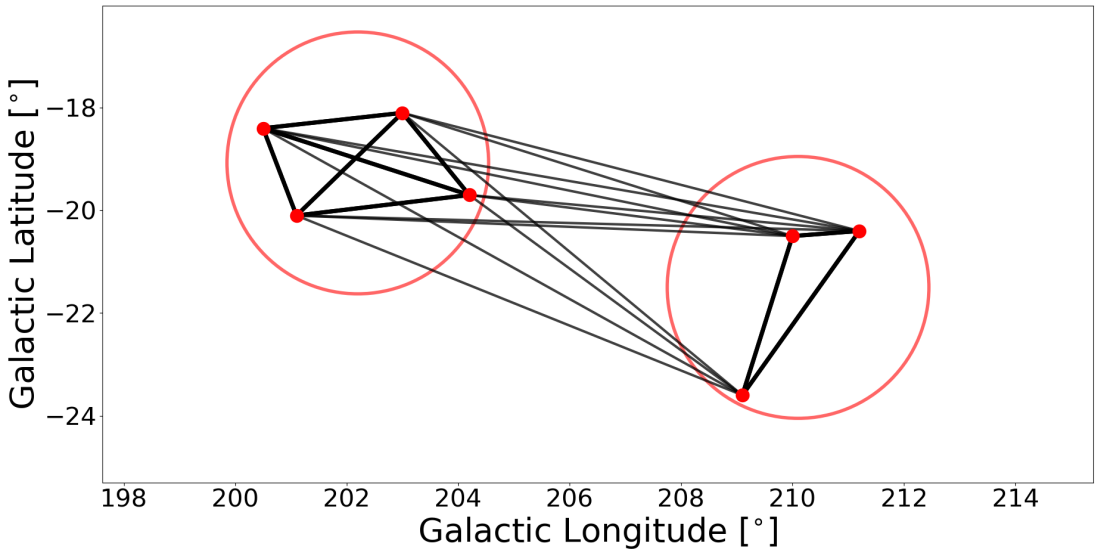

Figure 2.2: Graph representation of a group with sub-groups. Nodes represent galaxies in the 2MRS. Edges are weighted according to the percentage of the runs in which a galaxy pair is found for the same group. Strong edges represent $> 90\%$, weak edges between $60\% - 70\%$. The two circles show the two subgroups identified when using this method.

2.4 Correcting for Unreliable Connections

A lot of effort has been put into address shortcomings of the FoF algorithms, since it is so easy to connect nonphysical groups because of chance alignments. The graph method provides a tool to correct for this. Nevertheless, a remnant of this issue remains for more complicated systems (such as the Virgo Cluster); it is still possible for two or more obviously disconnected groups to be assigned to a single group and in some cases a low number of edges/connections can remain after a significance cut.

In some cases, visual evidence seems to suggest that two groups may be distinct and are arbitrarily connected through a single connection. However, computationally the entire system is seen as a single group and it is non-trivial to separate it into two distinct entities. This problem is compounded by the fact that there might be more than one unreliable connection. Determining what is an acceptable level of connections between these two groups before they are thought of as a single group is a challenging exercise. Furthermore, removing single edges and looking for a split in the group quickly becomes computationally exhaustive and is near impossible for cases where there is more than one connection.

In order to solve this issue the average connectedness score of every group was looked at and whether or not those groups had any subgroups. The average connectedness of groups becomes a useful metric to flag these cases: when two obviously-separated group

are connected by very few connections, the number of connections in the group are less than the total number of possible connections. Thus, the average connectedness score is very sensitive to such situations and looking for groups with low values of this statistic that have subgroups immediately allows us to flag this issue. Groups with average connectedness scores of $< 50\%$ and contained subgroups, were corrected.

While this artifact exists for less than 0.5% of the total groups, it is still important to correct for it. The best solution is to simply allow the subgroups to become their own groups. Any galaxies not identified as belonging to a subgroup were assigned to the nearest on-sky subgroup, as long as the R_{200} (defined later on) of that subgroup was greater than the distance to the left-over galaxy.

Figure 2.3 is an example of a situation which is automatically flagged and corrected by the group finder and shows the 3D representation of the Virgo Super Cluster as found in the 2MRS. There are two visually evident structures however, both of these structures were identified as a single group (before applying the correction described in Sect. 2.4). The edges shown in Fig. 2.3 are all $> 70\%$ and the structures are connected to one another because of a tiny connection (highlighted with a red arrow in Fig. 2.3) connecting at the $> 70\%$ level. Thus cutting at stricter and stricter cuts does not split the independent structures. This is the problem that the correction described in Sect. 2.4 aims to address. After this correction all three structures are identified as their own, independent groups.

The linking length parameter choice was validated against a 2MRS-like mock catalogue, generated by the Theoretical Astrophysical Observatory (TAO) (see next chapter). The mock is complete up to $K_{\rm s}^{\rm o} = 14\overset{\rm m}{.}75$, i.e. three magnitudes deeper than the 2MRS (Bernyk et al., 2016). The TOA mock contains both cosmological and peculiar redshifts.

The validity of the selected linking lenghts was confirmed by using state of the art visualization techniques including Virtual Reality. This is discussed in detail in Sect. 3.

2.5 Validation of Algorithm on a Mock catalogue

The choice of linking length parameters and all other additions to the algorithm were validated against a 2MRS-like mock catalogue, generated by the Theoretical Astrophysical Observatory (TAO). This is an online virtual laboratory which easily allows users to quickly and simply generate their own simulated catalogs (Bernyk et al., 2016). The mock is complete up to $K_{\rm s}^{\rm o} = 14\overset{\rm m}{.}75$, i.e. three magnitudes deeper than the 2MRS and contains both cosmological and peculiar redshifts (Bernyk et al., 2016).

The TAO mock catalog was used to 1) evaluate the choice of linking lengths set by Ramella et al. (1997) and 2) substantiate how accurately groups are recovered in the 2MRS given the relatively modest magnitude limit of $K_{\rm s}^{\rm o} < 11\overset{\rm m}{.}75$ and the effect of incompleteness as a function of magnitude limit and observed recession velocity.

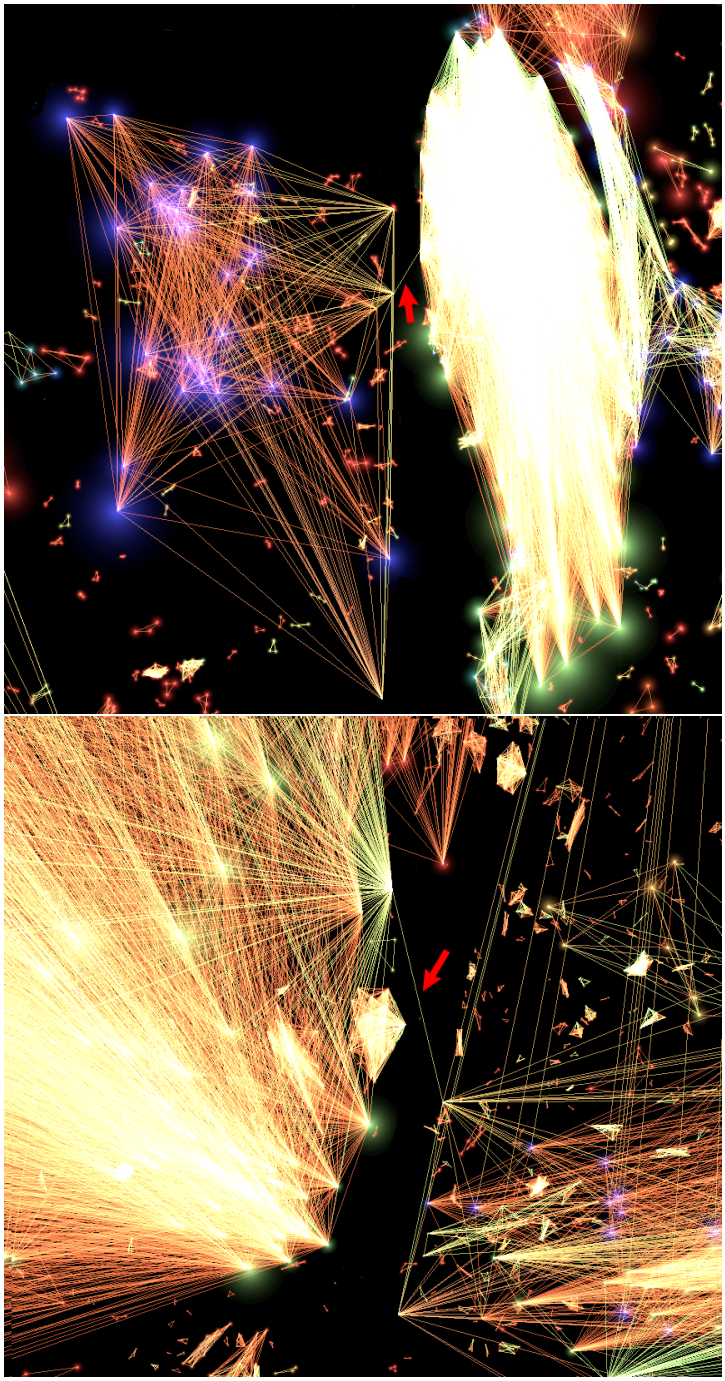

Figure 2.3: 3D visualization of the Virgo cluster. The two large (independent structures) are both identified as a single group before corrections (as described in Sect. 2.4 are made. Green, orange, and red points represent galaxies with a scores of $> 70\%$, $> 80\%$, and $> 90\%$ respectively. The top and bottom panels shows a normal and zoomed in view of the false connection respectively.

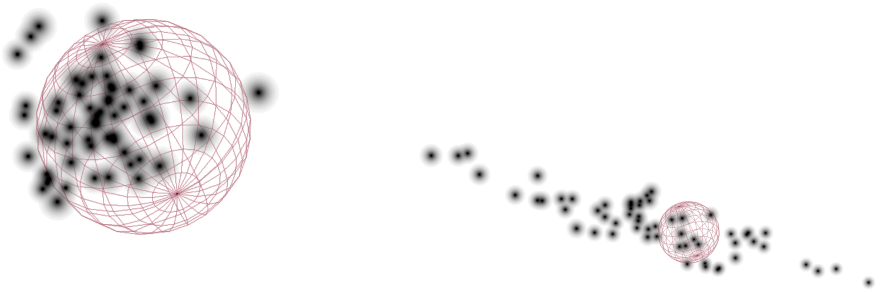

Figure 2.4: Example of a group identified in the TOA mock. The pink sphere shows the location of the group as found in the distorted mock (right), and is overlaid in the undistorted mock (left). This shows that the FoF algorithm finds physical groups whilst being run on a redshift distorted data set.

2.5.1 Validating Parameter Choice

The group finder was run with numerous different linking length pairs and compared to groups which were identified in the cosmological TAO mock catalog (free from any redshift distortion). There are many ways to compare galaxy group catalogs. Three quantitative measures are utilized namely Reliability, Completeness, and Variation of Information (VI) all of which have been used successfully by Stothert et al. (2019). Completeness quantifies which galaxies are found in a group based on distances according to the cosmological mock, compared to groups based on the redshifts of the same mock catalog. Conversely, reliability quantifies to what extent galaxies found in a group in the redshift distorted mock catalog are also found in a group in the purely cosmological mock catalog. As per Stothert et al. (2019), Completeness and Reliability can be calculated as:

$$C = \frac{1}{\sum_{i=1}^{N_G} \sum_{j=1}^{N_H} n_{ij}} \sum_{j=1}^{N_H} \max_i (n_{ij}) \tag{2.10}$$

and

$$R = \frac{1}{\sum_{i=1}^{N_G} \sum_{i=1}^{N_G} n_{ij}} \sum_{j=1}^{N_H} \max_j (n_{ij}) \tag{2.11}$$

respectively, where n_{ij} is the number of members found in both group i (from the redshift distorted mock catalog) and group j (from the purely cosmological mock catalog) and N_G and N_H represent the total number of groups in the redshift distorted mock and the purely cosmological mock respectively.

The ideal group catalog comparison (i.e. recovering all groups perfectly) would render both completeness and reliability to 100%. However, in reality usually only a compromising balance between the two metrics is possible. Completeness and Reliability are combined into a single metric called Variation of Information (Stothert et al., 2019). This could be thought of as a quantification of the amount of information which can be inferred from one clustering set to another. In an ideal matching case this would be 0 (see Meilă (2005) & Meilă (2007) for further discussion). VI can be calculated as:

$$
\begin{aligned}
VI = &- \sum_{j=1}^{N_H} \left(\frac{\sum_{i=1}^{N_G} n_{ij}}{\sum_{i=1}^{N_G} \sum_{j=1}^{N_H} n_{ij}} \ln \left(\frac{\sum_{i=1}^{N_G} n_{ij}}{\sum_{i=1}^{N_G} \sum_{j=1}^{N_H} n_{ij}} \right) \right) \\
&- \sum_{i=1}^{N_G} \left(\frac{\sum_{j=1}^{N_H} n_{ij}}{\sum_{i=1}^{N_G} \sum_{j=1}^{N_H} n_{ij}} \ln \left(\frac{\sum_{j=1}^{N_H} n_{ij}}{\sum_{i=1}^{N_G} \sum_{j=1}^{N_H} n_{ij}} \right) \right) \\
&- 2 \sum_{i=1}^{N_G} \sum_{j=1}^{N_H} \left(\frac{n_{ij}}{\sum_{i=1}^{N_G} \sum_{j=1}^{N_H} n_{ij}} \ln \left(\frac{(n_{ij}) \left(\sum_{i=1}^{N_G} \sum_{j=1}^{N_H} n_{ij} \right)}{\left(\sum_{j=1}^{N_H} n_{ij} \right) \left(\sum_{i=1}^{N_G} n_{ij} \right)} \right) \right)
\end{aligned}
\tag{2.12}
$$

All three metrics were used in order to justify our choice of central linking lengths (used by Ramella et al. (1997) & Crook et al. (2007)). The results of running numerous different parameters and comparing them to the groups found in the purely cosmological mock are shown in Fig. 2.5. The top left and top right panels in Fig. 2.5 show reliability and completeness respectively. When linking lengths are strict, the reliability score is large, whereas having large linking lengths has the opposite effect. The inverse is true for the completeness metric where small linking lengths result in a low score and large linking lengths tend towards higher scores. This is makes intuitive sense. As linking lengths are small, fewer groups make the cut (reducing completeness), however, the groups that do make the cut are very likely to be real and present in the purely cosmological mock. Alternatively, when linking lengths are large many groups pass the cut, however, the probability of false detections increase and a drop in reliability is seen. The completeness and the reliability results both provide a useful sanity check.

Reliability tends to be lower than that of completeness in general. When visually inspecting the mock catalog, the effect of redshift distortion is evident. Chance alignment of galaxies, in combination with a large spread in velocity due to peculiar motions in groups and large-scale cosmic flows, cause several false detections. This effect cannot be removed and is instead a consequence of identifying groups in redshift space. It is important to keep this in mind when using this (or any other) redshift groups catalog.

The VI results show where we could optimize both completeness and reliability. The gradient present in the results is another sanity check, validating the metric. We overlay the parameter choices used by several different authors also running FoF algorithms. Of particular interest is the black star which represents the parameter choice used by

Ramella et al. (1997) and later Crook et al. (2007) (which we have opted to use). However, unlike previous surveys, our group finder runs over several parameter ranges using the graph theory adaptation described in section 2.2. The range of parameters spanned by this group finder is demarcated in Fig. 2.5 as a black rectangle.

The parameters used by the group finder are not unjustified. Using the Ramella et al. (1997) parameters results in a reliability of more than 0.7 and a completeness of more than 0.8. These parameters also lie low on the VI plot implying a reasonably good recovery of the groups in the purely cosmological mock when running the group finder with these parameters. The Ramella et al. (1997) linking lengths are also not far off other choices of linking lengths used by different authors. All linking lengths chosen by previous authors seem to hover around the VI "well" in Fig. 2.5. It is very difficult to argue for the superiority of any given choice.

It is tempting to use the VI statistic alone in order to justify an optimum choice of linking lengths, i.e. selecting the pair of linking lengths which minimizes VI. However, it is important to remember that 1) there are many metrics, statistics, and parameters which could be used to quantify how well a catalog recovers real groups and this is often dependent on the survey or science being done and 2) these metrics are dependent on mock catalogs which means that these metrics will very likely vary across different mocks and thus will not be entirely robust, and will be dependent on the physics, and group finding methods used on the mock itself.

With this in mind, VI, completeness and reliability are not used to determine the parameter choice but are instead used to justify the parameter choice. Figure 2.5 seems to suggest that adopting the Ramella et al. (1997) parameters (as done in this thesis) is well justified and reasonable. This is cemented further when put into the context of other works. Furthermore, Fig. 2.5 shows that being able to probe an area of parameter space (instead of a singular point) is beneficial and is a further validation of the graph theory adaptation introduced in §2.2.

2.5.2 Evaluating Incompleteness

In order to evaluate the effect of incompleteness as a function of magnitude limit and observed recession velocity, the group finder was run on the TAO mock with a magnitude limit of $K_{\rm s}^{\rm o} < 11\overset{\rm m}{.}75$. Groups with three members (representing the extreme case of low number statistics) were identified providing the base for all future comparisons. The group finder was then run on increasingly deeper subsets of the mock in steps of $\Delta m = 0\overset{\rm m}{.}25$ up to $K_{\rm s}^{\rm o} < 14\overset{\rm m}{.}75$ (see x-axis of Fig. 2.6).

The groups with three members found at $K_{\rm s}^{\rm o} < 11\overset{\rm m}{.}75$ were re-identified in subsequent, deeper runs of the group finder on the mock. Their positions (both on-sky and line-of-sight), the total number of members at each magnitude cut, and the difference of

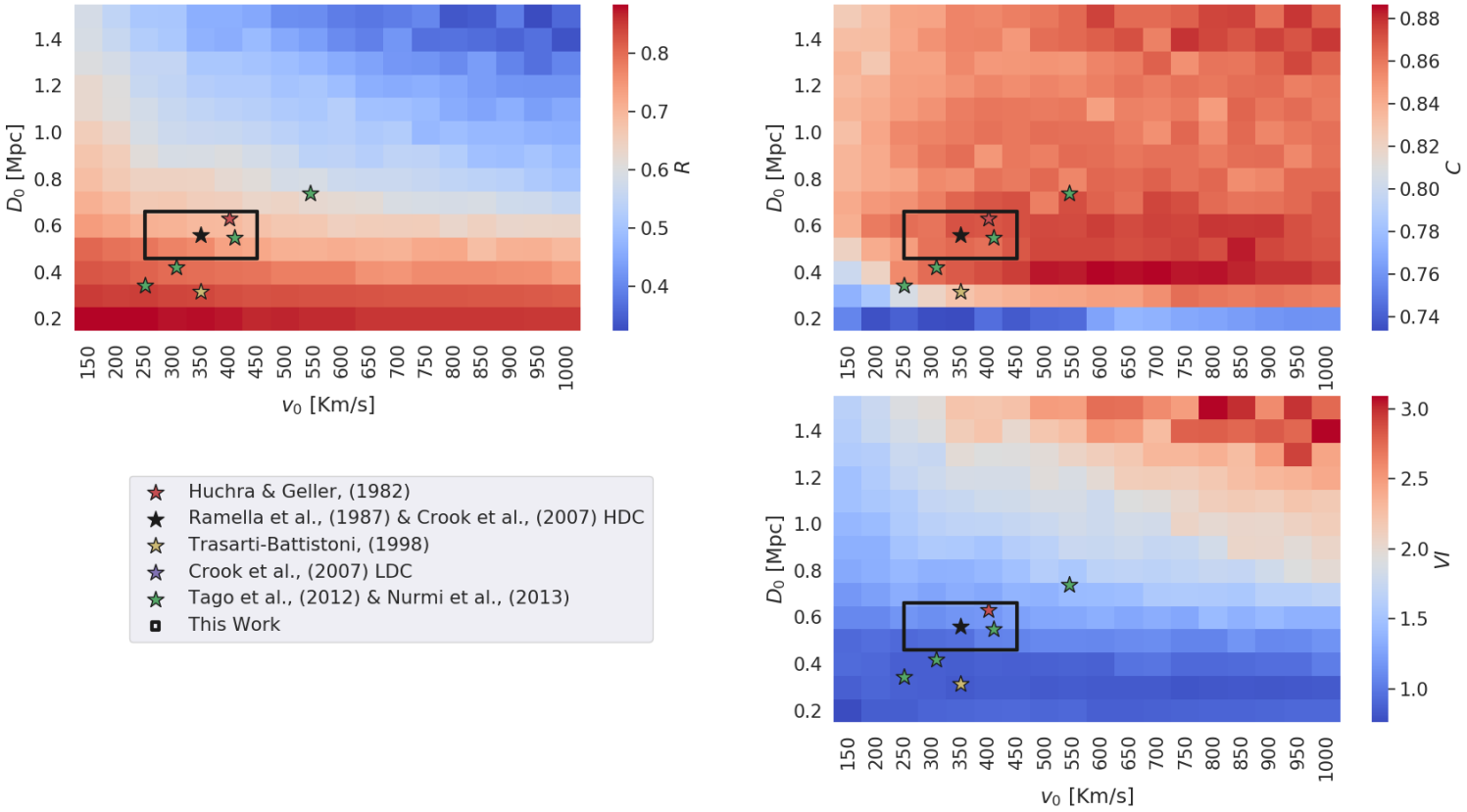

Figure 2.5: Completeness, Reliability and Variation of Information results of comparing groups found in the purely cosmological TAO mock catalog to groups found by running our group finder on the redshift distorted TAO mock over several different parameter choices. Each pixel represents a comparison. The v_0 and D_0 parameter choices for each individual run form the x and y axes respectively. Stars represent the parameter choices of previous works. The black square shows the parameter range which is probed by this work when implementing the graph theory method discussed in §2.2. The top left, top right and bottom right panels show the Completeness metric, Reliability metric, and Variation of Information metric respectively.

various group parameters compared to the $K_s^o < 11\overset{m}{.}75$ limit were noted. These groups were then grouped into three distance bins of 50, 100 and 150 Mpc. Finally, the difference in the group properties of the actual 2MRS were compared to the stepwise increased magnitude limit in the mock for each bin and are presented in the three panels of Fig 2.6.

Finally the difference in the three group properties of the 2MRS $K_s^o < 11\overset{m}{.}75$ limit were compared to the stepwise increased magnitude limit in the mock for each bin and are presented in the three panels of Fig 2.6.

The top panel in Fig. 2.6 shows how the line-of-sight position of the group changes with deeper magnitude cuts. Note that line-of-sight positional offsets, i.e., the radial velocity axis, tends to increase with depth and slowly level off around 2 Mpc. This relatively large offset in line-of-sight position can be attributed to the well known "finger-of-god" effect. It never exceeds an Abell radius though.

The middle panel shows the projected on-sky offset with depth. The average offset is less than 0.5 Mpc. This is much less than the error in the radial direction because there is no velocity distortion along the plane of the sky, other than the diminutive Kaiser flattening effect. No "finger-of-god" effect is present. Both the line-of-sight and on-sky offsets appear to be independent of the distance of the groups.

The bottom panel shows the percentage of the group present in a survey limited at $K_s^o \leq 11\overset{m}{.}75$. Unsurprisingly, there is a very clear trend with distance. Nearby groups are less affected by incompleteness than those at large distances. Thus, the fraction of recovered members is greater for nearby groups.

It is important to keep these offsets and incompleteness in mind when making use of this group catalogue.

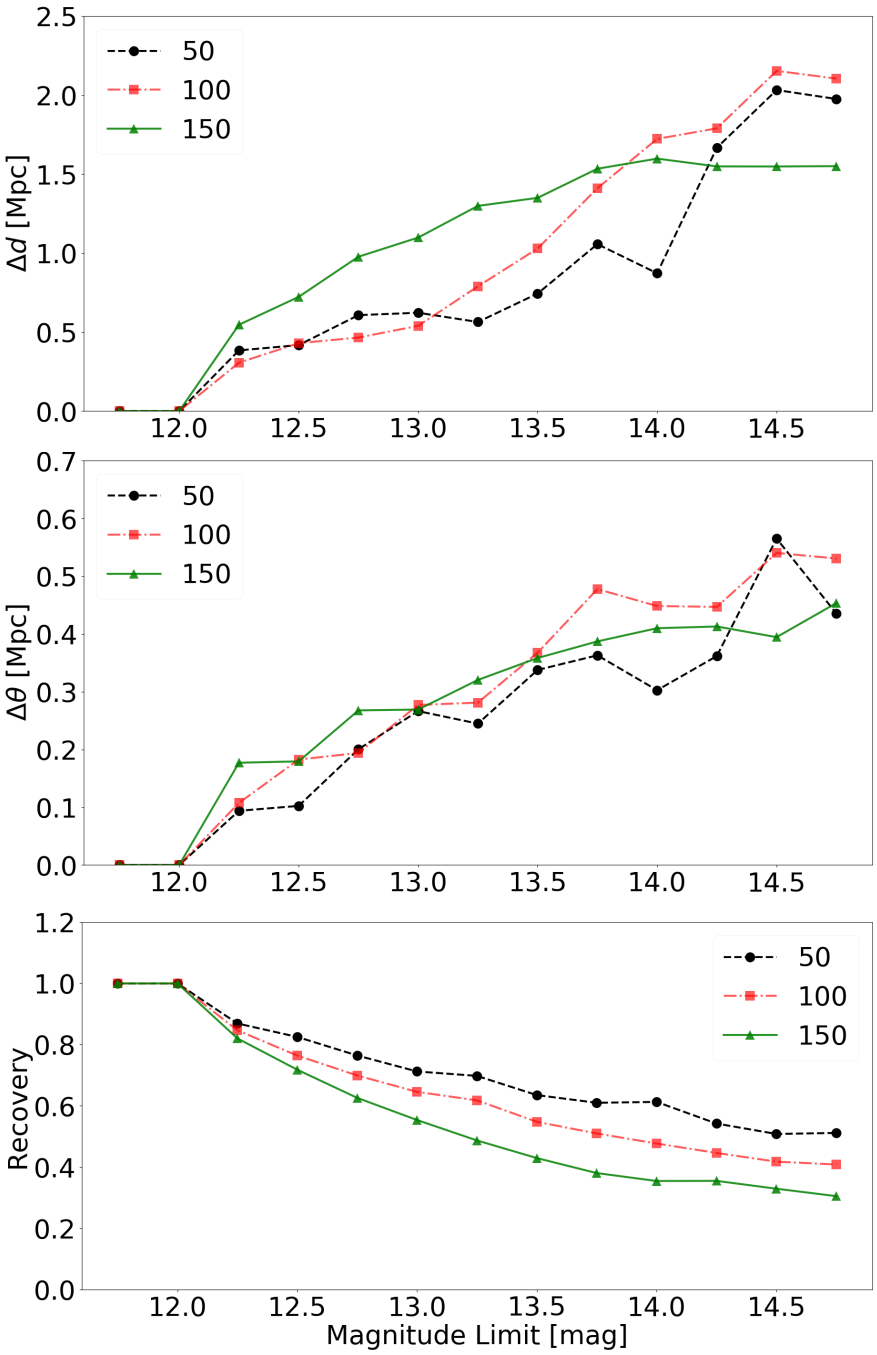

Figure 2.6: Trends of galaxy groups with three members found in the TAO mock catalogue at the 2MRS magnitude limit of $K_s^o < 11\overset{m}{.}75$, as a function of magnitude completeness limit (x-axis) and distance (different curves). The solid, dashed and dotted lines represent groups with a co-moving distance of $0 < d \leq 50$ Mpc, 50 Mpc $< d \leq 100$ Mpc, and 100 Mpc $< d <$ 150 Mpc respectively. Top: offset in line-of-sight position with changing depth. Middle: projected on-sky offset with changing depth. Bottom: percentage of the total group recovered at a magnitude limit of $K < 11\overset{m}{.}75$. This plot quantifies how accurately positions of groups are recovered and what percentage of the group members are found.

Chapter 3

Visualization Techniques

The state-of-the-art visualization laboratory hosted by the Institute of Data Intensive Astronomy (IDIA) was used, which includes immersive displays and notably a Virtual Reality (VR) system that is ideal for 3D datasets. The group results were efficiently compared (based on running the algorithm with the linking length choice) against the intrinsic large-scale structure distribution provided by the mock. Fig. 2.4 shows how well the algorithm recovers real groups using the linking length parameter choice, by overlaying the groups that were found over the positions of the galaxies in the TAO mock.

These visualization techniques allow for cutting edge quality insurance during the construction of the FoF group finder as well as the final data product. These techniques allow the entire data set to be interrogated thoroughly and for problematic case studies to be quickly identified and debugged. In the age of incredibly large redshift surveys (e.g. DESI, which will provide (Collaboration et al., 2016)) normal methods of assurance control will undoubtedly be unable to cope with the large data volumes. In this thesis we make use of visualization to guide parameter choice and graph theory adaptation which is later objectively applied to the entire catalog. The final data product is ruthlessly interrogated and refined in this manner.

3.1 Visualization Facilities

As mentioned, the catalogue was tested and verified using the new state-of-the-art visualization facilities hosted by IDIA. This included the use of Virtual Reality (VR), a panorama immersive facility, and the 8k digital planetarium dome housed at the Cape Town Iziko Museum. Using these facilities provides an additional supplement to the

standard methods of bug-finding. This is achieved by overlaying several data sets, plus a variety of visual markers. This can consist of several other markers that can be used as diagnostic tools such as spheres (designating radius), lines connecting associated galaxies of the same group, colour palettes characterizing distance.

Every run of the algorithm and technical change made to the group finder was interrogated against previous renditions using the visualization lab, to inspect the results and ensure that: 1) no logical errors were present in the updated version, and 2) that changes rendered better results than previous iterations.

3.2 Rebuilding the Traditional FoF algorithm on Crook et al., (2007)

A group-finder using the same method described in Crook et al. (2007) was first built and then applied to the early (shallower and incomplete) version of 2MRS. The two group catalogues were then compared using virtual reality in addition to running the new algorithm on this same data set. The galaxies from Crook et al. (2007) were portrayed as small points in three dimensions using Cartesian Galactic coordinates. Both group catalogues were displayed as large blue and red spheres (for Crook et al., 2007, and these results, respectively). The spheres mix colors in the virtual environment creating an overlap of purple (in this particular case). In this way it is possible to see at a glance (a) an agreement of the two catalogues (purple spheres) and (b) numerous types of discrepancies including line-of-sight offsets, on-sky offsets, groups found by the catalogue but not by Crook et al. (2007), and vice versa.

3.3 Visualizing Graphs

Using graph theory to average over the runs allowed us to think of each group as a topological data set, and as such they optimally lend themselves to visualization (Pascucci et al., 2011). This was accomplished by looking at pairs of galaxies, and deriving the frequency with which pairs were put in the same group. This results in a completely novel way of constructing a group catalogue, and solving the degeneracy issue arising from averaging over a three-dimensional dataset.

Furthermore, a unique approach in visualizing the multiple results was developed by translating the mathematical graph objects from graph space into redshift space. Visual examination of the dataset then provides instantaneous feedback of both the underlying group finder and simultaneously the underlying mechanisms that led to those results.

Fig. 3.1 shows an example of how the results of the group finder were visualized for any given group. Having the graphs displayed in this fashion allowed for exploration on how the graph structures themselves could be better used to improve the group results. This optimization was done by making use of: (1) the mock catalogue, which has both

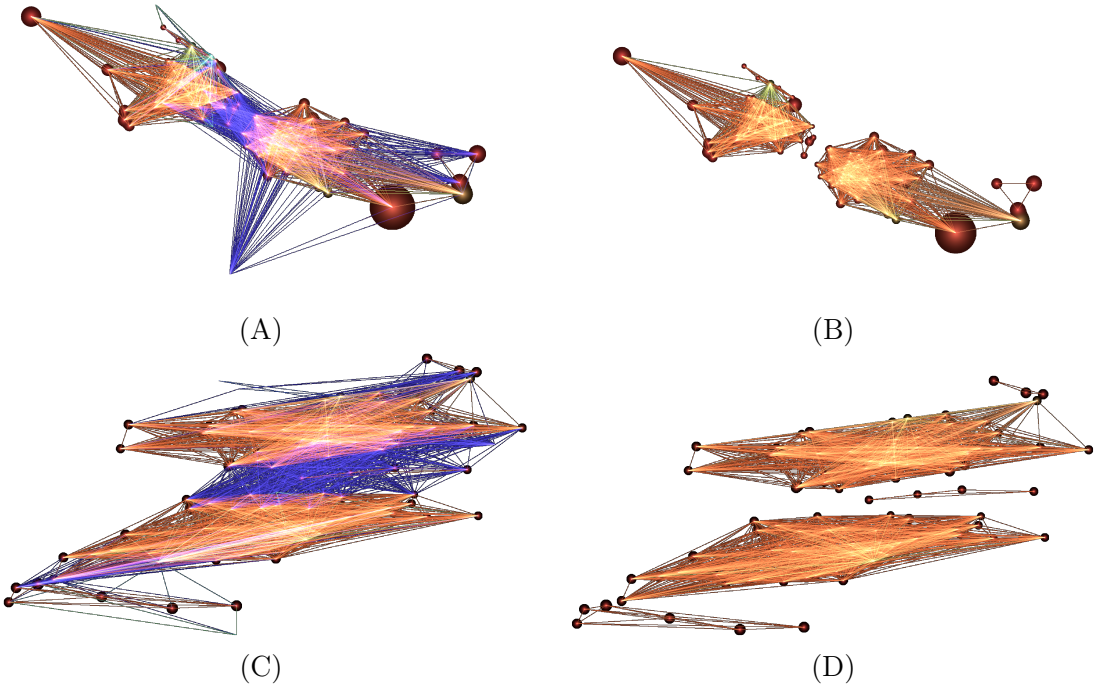

Figure 3.1: Three-dimensional example of a group with the resulting graph overlaid. Points represent the positions of 2MRS galaxies. Edges represent the number of connections of galaxies in the same group. Red edges represent 90% to 100% and blue edges represent 0% to 10%. Frames (A) and (B) display the on-sky plane the galaxy group, (C) and (D) the lateral view, i.e. along the line-of-sight. Frames (A) and (C) show the group before the cut was implemented, frames (B) and (D) after the cut. A video displaying these connections in 3D is available here.

redshift and distance information; (2) well-known clusters and groups in the full 2MRS catalogue; and (3) any examples of groups which were found to be suspicious. The latter option was of particular importance, and possible only thanks to the powerful 3D visualization techniques. Finding examples of failures of the algorithm (e.g., where two obviously separate groups were connected as one) is straightforward when using visual inspection in VR and allows for exploration of intrinsic weaknesses of the group finder algorithm.

3.4 Removing Statistical Outliers

Visual inspection reveals that averaging over all results is not good enough. While the robustness of the method is assured, groups can become too large because of a single connection/edge between two obviously distinct groups. It also is unrealistic to accept every single connection. A galaxy which is associated with a group for 10% of the runs

but with another one for the remaining 90% is much more likely to form part of the latter. In other words, while the robustness of the method is assured, robustness to statistical outliers is not and a significance cut must be made. A careful analysis based on both the mock catalogue and well-known groups and clusters indicated that accepting connections at the $> 50\%$ level occurrence gives the best results. Using the graph tool was highly effective here; a removal of all edges with a weighting $< 50\%$ was sufficient to disconnect subgraphs. At this point, the group finder is completely robust. An example of this process is shown in Fig. 3.1. Edges are coloured according to the percentage of runs in which a galaxy pair was assigned to the same group. Frames (A) and (B) display the on-sky plane the galaxy group, (C) and (D) the lateral view, i.e. along the line-of-sight. Frames (A) and (C) show the group before the cut was implemented, frames (B) and (D) after the cut. Note that if the cut had not been implemented (i.e. removal of low-weighted, blue edges) then the entire system would have been classified as a single group.

3.5 Subgroups

On the other hand, several groups that seemed to have two or more concentrations of galaxies. If the statistical cut would have been stricter than 50% (say 60% or 70%) then they would have split into separate groups. To account for these cases, a higher statistical cut was applied to each group to check the likelihood of them having real substructure, and/or whether some of these subgroups are real. Differentiation of these cases was easy in VR; however, translating this visual information into an algorithmic equivalent is quite a challenge. The connectedness score described in Eq. 2.7 achieved this goal – coloring the galaxies by this property made the groups that required attention stand out clearly. Based on this information, groups with low average connectedness score and which also contained subgroups could be identified. In these cases, these subgroups became their own groups.

Figure. 3.2 gives an illustration of a group with subgroups. From top to bottom, lines of consecutive strengths are being cut away until no more lines can be cut. The number of separate structures changes from one in the top frame to two in the others. Because the maximum number of isolated systems is two, subgroups are chosen to be the first instance at which the maximum number of isolated systems occurs (frame (B) in this particular example).

The use of numerous visualization techniques enabled us to identify issues affecting the traditional FoF algorithm and make numerous updates to address them. These included: (1) running the algorithm many times; (2) averaging over all runs using graph-theory; (3) removing statistical outliers by implementing a significance cut; (4) applying

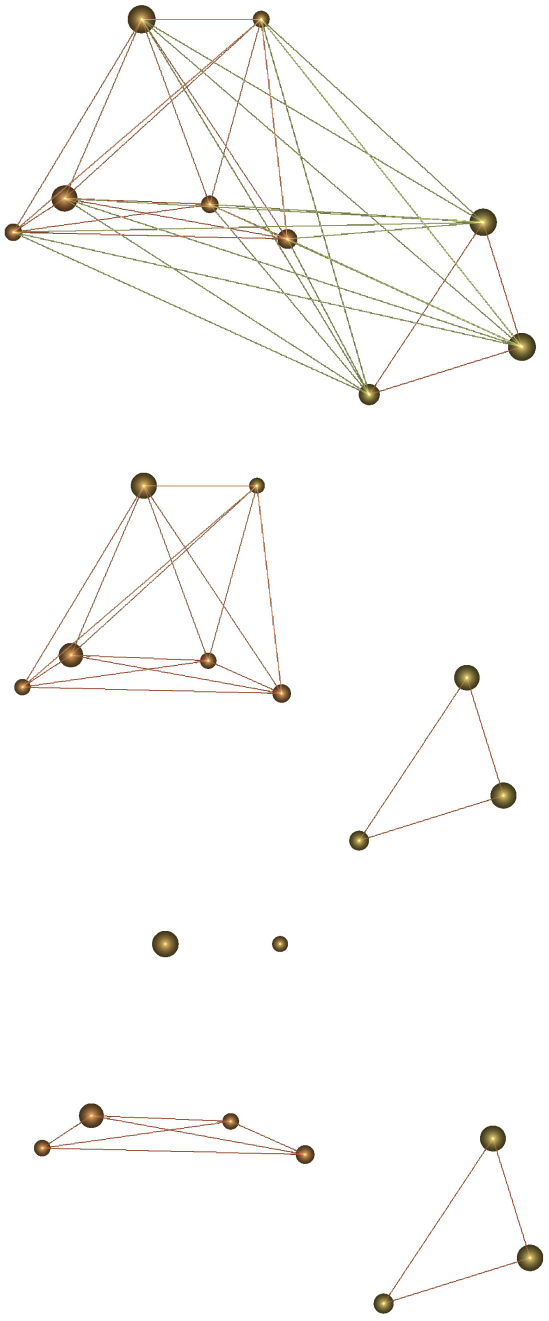

Figure 3.2: Example of a galaxy group containing two subgroups and highlighting the method of identifying subgroups by incrementally removing the weakest edges. The top panel shows the group with all edges $\geq 50\%$ still present, the middle shows the group after edges with a weighting of 60% and less are removed, and the bottom shows the group when edges with a weighting of 80% and less are removed, resulting in only the strongest edges remaining (in this case edges with a weighting of 1). A 3D-animation of these examples is available here.

a recursive cut to graphs to find any subgroups; and (5) identifying weak and nonphysical connections and correcting these groups.

Chapter 4

The 2MRS Group catalogue

The final 2MRS group catalogue contains 3022 entries and is the most complete "whole-sky" galaxy group catalogue to date. Group properties were measured and calculated, such as: 3D positions, luminosity and comoving distances, observed and corrected number of members, richness metric, velocity dispersion, and estimates of R_{200} and M_{200}.

4.1 Measured Group Properties

The measured group properties include the velocity dispersion and the projected central position, which in turn affect many of the other derived quantities (discussed further below).

4.1.1 Velocity Dispersion

Two methods for calculating the velocity dispersion were used, namely the simple standard deviation of the velocities of the group members and the root mean square (RMS) of the peculiar velocity distribution along the line of sight, described as

$$v_i = c\frac{z_i - \bar{z}}{1 + \bar{z}} \tag{4.1}$$

where v_i, $\bar{z}$ are the velocity of galaxy i and the mean redshift of the galaxy group respectively (Navarro et al., 1995; Jarrett et al., 2017).

4.1.2 Projected Group Center

There have been several proposed methods of determining the projected group center amongst different surveys and catalogs (Crook et al., 2007; Poggianti et al., 2010;

Robotham et al., 2011; Tempel et al., 2018). Several techniques were explored using the TAO mock catalogue including calculating the center of light position, the flux weighted geometric center, and the simple geometric center. It was found that taking the average RA and Dec of the group members rendered the most accurate central on-sky position, both when inspecting the mock catalogue and identifying well known, existing groups and clusters.

4.1.3 Projected Radius

The projected radius was calculated as the maximum projected on-sky distance of all the galaxies from the projected group center. While this method is not robust against outliers, it was found that those could be robustly identified and corrected using the new visualization techniques as they caused spheres representing groups to be hugely exaggerated, immediately highlighting problematic groups and methods which were then ultimately corrected, resulting in the complete catalogue having no such issues.

4.2 Calculated Group Properties

4.2.1 R_{200} and M_{200}

Staple properties in many group catalogues are R_{200} and M_{200}, respectively defined as the radius of the sphere with 200 times the critical density and the total mass enclosed within said sphere (Poggianti et al., 2010).

By assuming the virial theorem R_{200} and M_{200} can be calculated as

$$R_{200} = 1.73 \frac{\sigma}{1000 \text{ km s}^{-1}} \frac{1}{\sqrt{\Omega_\Lambda + \Omega_0 (1+z)^3}} h_{100}^{-1} \text{ Mpc} \tag{4.2}$$

$$M_{200} = 1.2 \times 10^{15} \left(\frac{\sigma}{1000 \text{ km s}^{-1}} \right)^3 \frac{1}{\sqrt{\Omega_\Lambda + \Omega_0 (1+z)^3}} h_{100}^{-1} M_\odot \tag{4.3}$$

where σ is the line-of-sight velocity dispersion (Poggianti et al., 2010).

4.2.2 Comoving and Luminosity Distances

Comoving and Luminosity distances, as described by Peebles (1993), are calculated as

$$d_c = \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_M (1+z')^3 + \Omega_k (1+z')^2 + \Omega_\Lambda}} \tag{4.4}$$

and

$$d_L = d_c(1+z) \tag{4.5}$$

respectively.

4.2.3 Richness

A simple definition of richness similar to Andreon (2016) is included in the catalogue. Richness is defined, in this catalogue, as the number of galaxies within a group with absolute magnitude of $M_K^o > -23\overset{m}{.}5$. This definition is maintained up to 100 Mpc where the survey is still reasonably complete. The richness is still calculated for groups greater than 100 Mpc but this value would represent the minimum richness possible and is reported as such in the final catalogue.

4.2.4 Corrected Group Members

I provide a simple correction to the number of members in a group in order to gain a relative estimate of the true number of members were there no incompleteness. This is done by first assuming that the Virgo Supercluster is complete. The number of corrected group members for a given group ($N'_{\rm mem}$) can then be calculated as

$$N'_{\rm mem} = N_{\rm mem}\left(\frac{N_{\rm v}}{N_{\rm v}(d)}\right) \tag{4.6}$$

where $N_{\rm mem}$ and $N_{\rm v}$ are the number of members found in the given group and the number of members found in Virgo respectively. $N_{\rm v}(d)$ is the number of group members that Virgo would have if it were at the same distance (d) of the given group. This is strictly dependent on the luminosity function of Virgo.

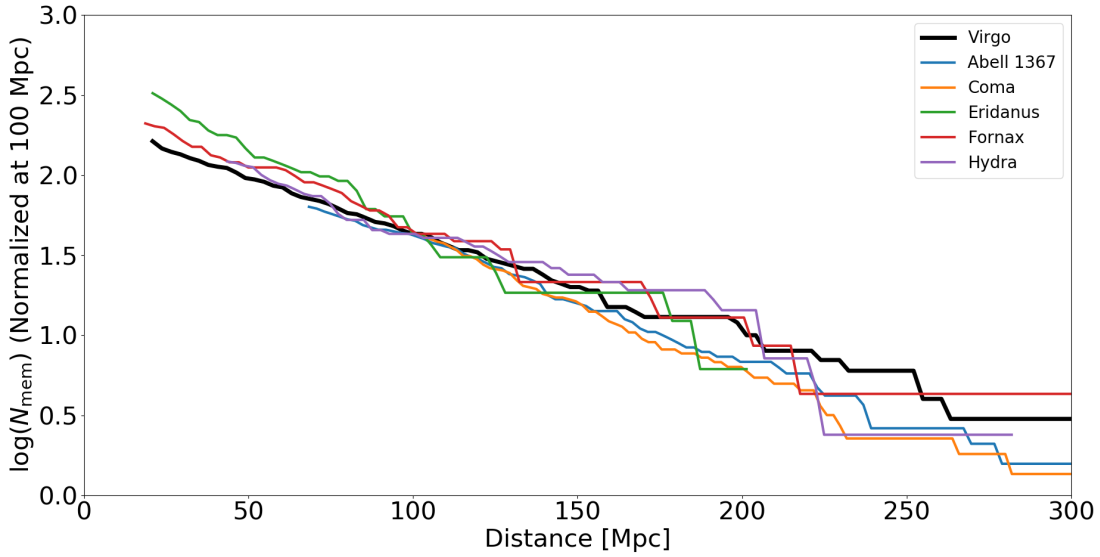

Figure 4.1: Plot showing the number of members of well-known clusters which would be found by the 2MRS at certain distances. The tracks have been corrected to the Virgo Cluster (thick black line) and normalized at 100 Mpc.

Figure 4.1 gives an intuitive understanding of how the correction works. The assumption is made that Virgo (sitting at 21 Mpc away and thus relatively nearby) is complete in the 2MRS. Therefore, by comparing other clusters to Virgo a measure of said clusters incompleteness can be gained. "Tracks" (shown in Fig.4.1) are created by first working out the absolute magnitudes of the galaxy members in each cluster (taking the distance of the cluster to represent the distance of all the members and using the distance modulus) and then seeing how many cluster members fall below the apparent magnitude limit of 2MRS at further and further distances.

These tracks are compared to the track of Virgo (marked in Fig. 4.1 as the thick black line) and then normalize the tracks to a distance of 100 Mpc. The essence of Fig. 4.1 is seeing how the slopes of the various clusters compare to Virgo.

The stagnation at the ends of the tracks (appearing to be more discrete and "step-like") is tantamount to the large and intrinsically bright galaxies which reside within the clusters. As the cluster is moved further and further away these giants cling on, still being visible in the 2MRS, creating the step-like appearance of the tracks at these further distances.

4.3 Catalogues

Along with the main 2MRS group catalogue, two additional supplementary catalogues are included. Each one is described in detail below. Due to their size I include subsamples in this section of the thesis with more complete examples in the Appendix. The

full csv versions of these tables are available at `https://www.dropbox.com/home/2MRS_Catalogues` or upon request.

4.3.1 The 2MRS Catalogue

The final 2MRS group catalogue is presented in Table 4.1. The first 20 largest groups by number of members are shown for explanatory purposes. The full table is available online.

(1) - 2MRS group ID as defined in this work.
(2) - Names (where available) of well known groups and clusters in other catalogs.
(3) - Group RA in degrees, calculated as the average RA of the group members.
(4) - Group Dec in degrees, calculated as the average Dec of the group members.
(5) - Galactic longitude of the group in degrees, calculated as the average of the Galactic longitude of the members.
(6) - Galactic latitude of the group in degrees, calculated as the average of the Galactic latitude of the members.
(7) - Number of 2MRS galaxies in the group.
(8) - Corrected number of members in the group from correcting the group to the position of Virgo as described in Eq. 4.6.
(9) - Richness of the group. Defined as the number of members in the group with an absolute magnitude of $M_K^o > -23\overset{\text{m}}{.}5$.
(10) - Recession velocity corrected to the CMB reference frame.
(11) - Co-moving distance, calculated using Eq. 4.4.
(12) - Luminosity distance, calculated using Eq. 4.5.
(13) - Velocity dispersion calculated using Eq. 4.1
(14) - Velocity dispersion determined by taking the standard deviation of the velocity of the members of the group.
(15) - R_{200} measurement of the group calculated using Eq. 4.2.
(16) - M_{200} measurement of the group calculated using Eq. 4.3.
(17) - Number of subgroups within the main group as defined in chapter 2.3.

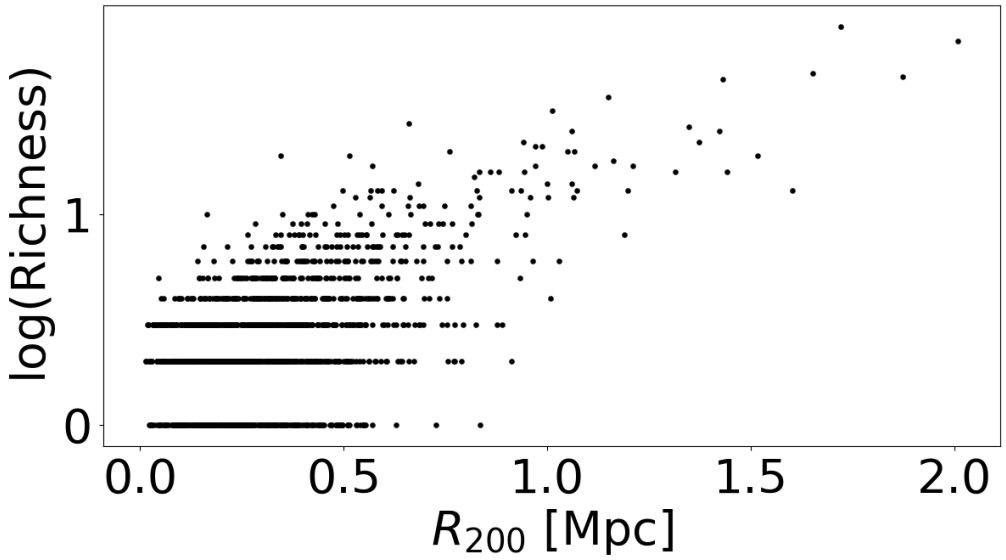

Figure 4.2: Relationship between the richness of a group (defined as the number of galaxies with $M_{\rm k}^{\rm o} > -23\overset{\rm m}{.}5$) and R_{200}

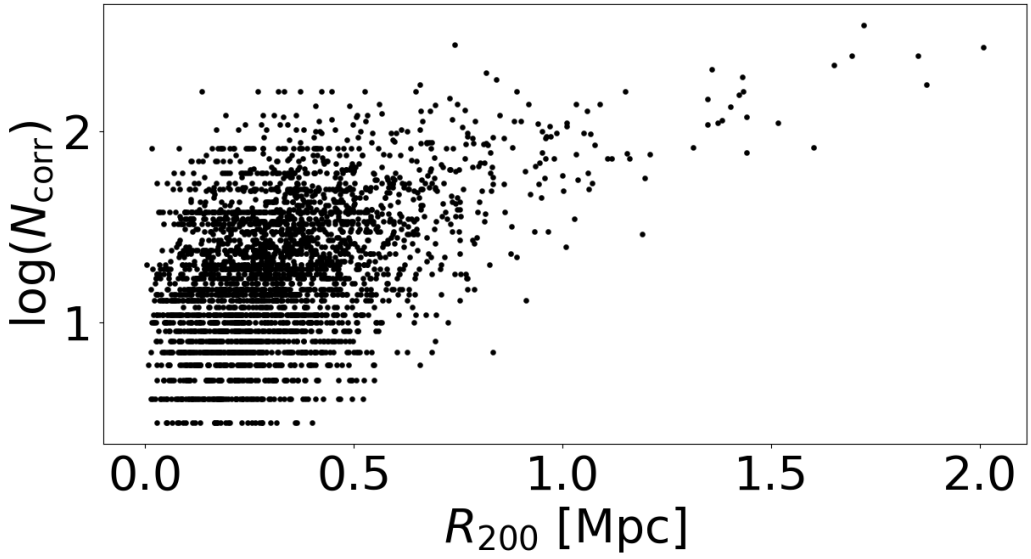

Figure 4.3: Relationship between the number of corrected members in a group and the R_{200} metric.

The main catalogue consists of 3022 groups.

Both richness (column 9 in Table 4.1) and the number of corrected members,$N'_{\rm mem}$ (column 8 in Table 4.1) have been plotted against R_{200} in Fig. 4.2 and Fig. 4.3 respectively. R_{200} is determined solely by the dispersion of the group unlike Richness and

$N'_{\rm mem}$ which are entirely independent of dispersion. It therefore bodes well to have both the number of members and richness increase with R_{200}. This definition of richness also renders a relatively tight relation with R_{200} which is a good validation of the richness definition. The extreme outlier present in Fig. 4.3 is the Virgo Super Cluster. Figure 4.2 and Fig. 4.3 help in showing the integrity of the main catalogue.

The mass distribution of the 2MRS according to our catalog is also included in Fig. 4.4. The blue distribution in Fig. 4.4 shows the mass distribution of M_{200} as reported in the 2MRS catalog (this work). The peak of the M_{200} mass distribution is between $10^{12}M_{\odot}$ and $10^{14}M_{\odot}$. This is in a reasonable range but is slightly lower than expected. In general the entire distribution ranges from $10^{11}M_{\odot}$ to $10^{14}M_{\odot}$ which is again slightly lower than expected. M_{200} is dependent entirely on velocity distribution σ (see Eq. 4.3). This mass underestimation is due to a velocity dispersion underestimation. This is most likely due to a selection bias within the 2MRS which arises due to its relatively bright magnitude limit. Larger, brighter galaxies are preferentially found and these types of galaxies tend to concentrate towards the center of groups and clusters, thus biasing the velocity distribution and therefore M_{200}.

The virial mass can also be calculated from the 2MRS catalog as done by Crook et al. (2007). This virial measurement of mass peaks at around 13.5 $M_{\odot}$ and ranges from 12 $M_{\odot}$ to $15M\odot$. This is fairly reasonable.

The Mass distribution from Lim et al. (2017) for an earlier version of the 2MRS is also shown in Fig. 4.4 for comparison purposes. The Lim et al. (2017) results roughly agree with the two different methods within this catalog. This bodes well as to the reliability of the results.

4.3.2 The 2MRS Sub Group Catalogue

A subgroup catalogue is also included, with a random sub-sample shown in Table 4.2. The subgroup catalogue includes Group ID, RA, Dec, Galactic coordinates, $\rm v_{cmb}$, and $N_{\rm mem}$ as described in Table 4.1. The sub ID, (1) in Table 4.2, demarcates the unique identifier for each individual subgroup. The ID is made up of two parts, namely the host galaxy group ID (the group to which said subgroup belongs) and the numbered subgroup within said host group.

4.3.3 The 2MRS Galaxies in Groups Catalogue

A catalogue consisting of all the galaxies that are member of a group in Table 4.3 is also included. This catalogue includes the 2MASS ID, RA, Dec, Galactic coordinates and $\rm v_{cmb}$, as given in Macri et al. (2019), as well as the Group ID from Table 4.1 and

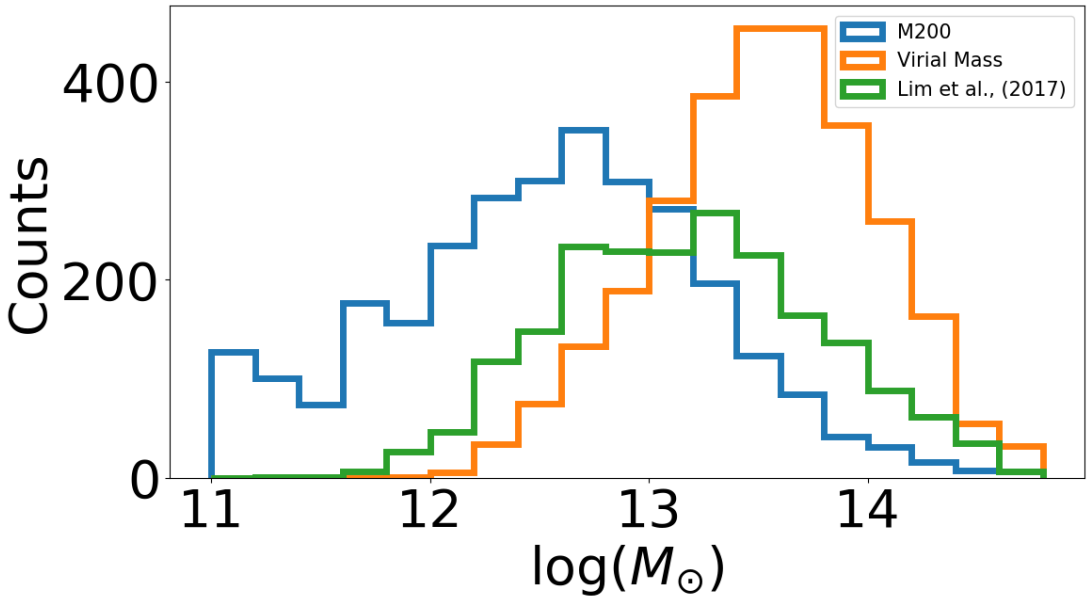

Figure 4.4: Distribution of various mass metrics within the 2MRS catalog. Blue and Orange histograms show the M_{200} and Virial mass metrics calculated from the 2MRS catalog directly. The green histogram shows the mass distribution as per Lim et al. (2017) for comparison.

the Sub ID from Table 4.2. A random sub-sample of galaxies are shown in Table 4.3 for reference.

The on-sky distribution of the final 2MRS group catalogue is shown in Fig. 4.5. All the well-known, large-scale structures have been recovered and, in particular, the intricate substructures which comprise them. This bodes well in validating the group finding algorithm as the structures which were recovered are similar to previous 2MRS-based catalogues (Crook et al., 2007; Saulder et al., 2016; Tempel et al., 2016a; Kourkchi & Tully, 2017; Tempel et al., 2018).

4.4 Comparisons to Literature

In validating the results it is important to compare the 2MRS group catalogue with previous results based on earlier versions of 2MRS, and to compare this method with other techniques. For the former, the catalogue was compared against Crook et al. (2007), Tempel et al. (2016a), and Tempel et al. (2018). The cross-matching search was done using tolerances of 3 and 6 Mpc for plane-of-sky and line-of-sight distances, respectively. The resulting matches and discrepancies were investigated using the IDIA visualization lab.

Table 4.1: 2MRS Galaxy Group catalogue

Group ID	Other Names	α	δ	l	b	$N_{\rm mem}$	$N'_{\rm mem}$	R	$v_{\rm cmb}$	d_c	d_L	RMS	σ	R_{200}	M_{200}	$N_{\rm subs}$
			[deg, J2000]						[km s^{-1}]	[Mpc]		[km s^{-1}]		[Mpc]	[$\times 10^{11} M_\odot$]	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
2987	VIRGO CLUSTER	187.80	11.41	285.13	73.58	163	163	36	1523	20.8	20.9	486	484	1.15	188.5	0
153	ABELL 3627	243.67	−60.89	325.27	−7.17	121	278	67	4931	67.3	68.4	854	861	2.01	1015.6	3
50	ABELL 3526B	192.03	−41.21	302.26	21.66	102	177	45	3784	51.7	52.3	795	792	1.87	820.8	3
3031	ABELL 0426	49.76	41.38	150.53	−13.45	95	224	47	5158	70.4	71.6	702	714	1.65	564.4	0
228	COMA CLUSTER	194.90	28.00	59.10	87.99	95	360	78	7260	98.9	101.3	734	751	1.72	643.8	3
179	-	159.12	−27.66	269.64	26.35	85	156	25	4043	55.2	56.0	604	611	1.42	360.2	0
271	ABELL 0262	28.44	36.23	136.75	−24.96	61	129	25	4630	63.2	64.2	450	457	1.06	148.7	0
3011	-	181.50	46.19	145.21	68.92	61	61	4	993	13.6	13.6	193	193	0.46	11.9	0
500	ABELL 1367	176.15	19.97	234.45	73.12	56	194	44	6787	92.5	94.6	610	623	1.43	368.4	0
151	ABELL S0805	281.49	−63.03	332.53	−23.38	56	111	31	4392	60.0	60.8	429	424	1.01	129.2	0
25	ERIDANUS CLUSTER	54.21	−20.12	211.53	−51.64	53	53	5	1517	20.8	20.9	251	250	0.59	26.1	0
84	HYDRA CLUSTER	157.79	−35.36	273.17	19.27	51	76	17	3155	43.1	43.6	513	517	1.21	221.0	2
4	FORNAX CLUSTER	53.93	−35.02	236.00	−54.21	49	49	9	1357	18.6	18.7	296	295	0.70	42.3	0
202	-	17.31	32.65	127.26	−30.08	48	107	21	4838	66.0	67.1	412	417	0.97	114.3	2
150	-	193.67	−13.77	304.14	49.10	48	106	20	4718	64.4	65.4	323	328	0.76	54.8	3
369	OPH CLUSTER	258.06	−23.39	0.53	9.30	46	250	≥ 46	8757	119.2	122.6	792	812	1.85	804.5	2
101	-	20.83	33.54	130.51	−28.86	46	97	20	4614	63.0	64.0	453	460	1.07	151.8	0
261	-	193.35	−9.26	303.75	53.61	45	92	16	4473	61.1	62.0	365	369	0.86	79.4	2
355	-	107.01	49.74	167.35	22.93	44	128	27	6032	82.2	83.9	281	286	0.66	36.1	3
142	-	206.98	−30.39	317.13	30.92	44	98	21	4853	66.2	67.3	419	424	0.99	120.2	0

Table 4.2: 2MRS SubGroup catalogue

Sub ID	Group ID	α	δ	l	b	v_{cmb}	N_{mem}
		[deg, J2000]				[km s^{-1}]	
1000-1	1000	140.72	24.44	203.96	43.31	10337	4
1000-2	1000	141.60	23.89	204.98	43.94	10145	3
1015-1	1015	107.14	−49.10	259.72	−17.53	12790	7
1015-2	1015	106.12	−48.98	259.35	−18.11	12997	3
1031-1	1031	241.04	69.77	103.49	39.28	7540	3
1031-2	1031	239.39	70.74	104.93	39.22	7486	6
1091-1	1091	52.65	41.59	152.24	−12.05	5367	7
1091-2	1091	52.56	40.53	152.81	−12.95	4772	5
1093-1	1093	241.24	17.57	31.31	44.52	10395	17
1093-2	1093	241.48	18.00	31.99	44.46	11589	11
1100-1	1100	251.94	58.81	88.16	38.82	5256	3
1100-2	1100	249.56	57.86	87.30	40.25	5349	6
1154-1	1154	52.59	41.79	152.08	−11.92	4325	3
1154-2	1154	52.13	40.29	152.68	−13.34	4107	3
1186-1	1186	195.68	−56.21	304.51	6.63	6225	4
1186-2	1186	196.91	−57.33	305.13	5.47	6030	13
119-1	119	142.76	−61.89	281.56	−7.66	2862	17
119-2	119	137.59	−64.03	281.46	−10.82	2076	4
1201-1	1201	173.22	−9.62	272.85	48.61	6630	21
1201-2	1201	174.68	−9.32	274.62	49.52	6155	4

4.4.1 Crook et al., (2007)

Crook et al. (2007) based their group finder on the classic FoF algorithm as used by Huchra & Geller (1982) without any modification. This group finder was applied to an earlier and shallower 2MRS catalogue which had a significantly shallower limit of $K_s^o < 11\overset{m}{.}25$. Nevertheless, since this group finder was built upon their techniques for the analysis, a comparison is appropriate.

Crook et al. (2007) used two different choices of linking length, resulting in two catalogues (respectively referred to as high- and low-density). The results were only compared against the high-density catalogue since it used the same linking lengths proposed by Ramella et al. (1997) and adopted by us. Groups which were found in this previous work that were based on artificial galaxies, which they used to fill the ZoA, were not considered.

All groups detected by Crook et al. (2007) were recovered. There are several groups

Table 4.3: 2MRS Group Member catalogue

2MASS ID	α	δ	l	b	v_{cmb}	Group ID	Sub ID
		[deg, J2000]			[km s^{-1}]		
02284905 + 3810005	37.20436	38.16689	143.22255	−20.83781	11189	1541	1541-2
10032864 − 1530044	150.86940	−15.50126	254.13606	31.03521	9998	1763	-
04375557 − 0931092	69.48154	−9.51926	205.99341	−33.93850	5110	724	-
02362379 + 3142410	39.09909	31.71140	147.68501	−26.06654	4705	152	152-2
16175726 − 6055229	244.48875	−60.92303	325.53091	−7.47124	3597	153	153-3
08333766 + 5535322	128.40689	55.59229	162.22017	36.37661	11300	1971	-
01293373 + 1717532	22.39054	17.29812	135.76518	−44.62148	12761	2535	-
14165292 + 1048264	214.22060	10.80737	357.96130	64.11185	7650	334	-
10081231 + 0958374	152.05138	9.97701	229.02748	47.94034	8532	2188	-
02442183 + 3121169	41.09096	31.35475	149.54651	−25.61624	5203	152	152-1
14034273 − 3243006	210.92809	−32.71685	320.07907	27.73697	4028	85	-
09202342 + 5456313	140.09766	54.94210	161.67014	43.04770	13827	2358	-
17281491 − 6640152	262.06204	−66.67094	325.71674	−17.07140	13589	193	-
23000358 + 1558493	345.01495	15.98034	87.56570	−39.12365	1828	354	-
13295512 − 3119561	202.47968	−31.33226	312.50024	30.82352	15142	1782	-
21073236 − 2538350	316.88477	−25.64310	21.20206	−40.26195	11998	818	-
12492664 − 4127463	192.36107	−41.46286	302.53073	21.40733	5086	50	50-1
13331326 + 3306350	203.30525	33.10971	68.98088	79.17372	7636	149	-
17030344 + 6102381	255.76439	61.04391	90.49467	36.52942	3391	839	-
19324667 − 6431251	293.19467	−64.52367	331.77798	−28.70660	4233	806	-
02094273 − 1011016	32.42808	−10.18383	174.08235	−64.95735	3623	413	-
05393634 + 1259047	84.90136	12.98459	192.95486	−9.49321	7229	2956	-
13170188 − 1046121	199.25781	−10.76999	313.08246	51.59659	3153	1447	-
18492430 − 4846088	282.35126	−48.76904	347.37582	−19.73818	5244	2405	-
22120645 + 3720024	333.02695	37.33405	91.02950	−15.47828	5632	189	189-2
03053667 + 0459396	46.40281	4.99440	173.28879	−44.36787	8407	452	-
10512393 + 2806429	162.84979	28.11196	203.77824	63.45928	1562	124	-
06254878 − 5359361	96.45339	−53.99345	262.68002	−25.29278	14006	2636	-
09045016 + 1333424	136.20901	13.56185	215.39915	35.63015	8778	1879	-
02303366 + 3210342	37.64031	32.17624	146.21729	−26.17465	4552	593	-

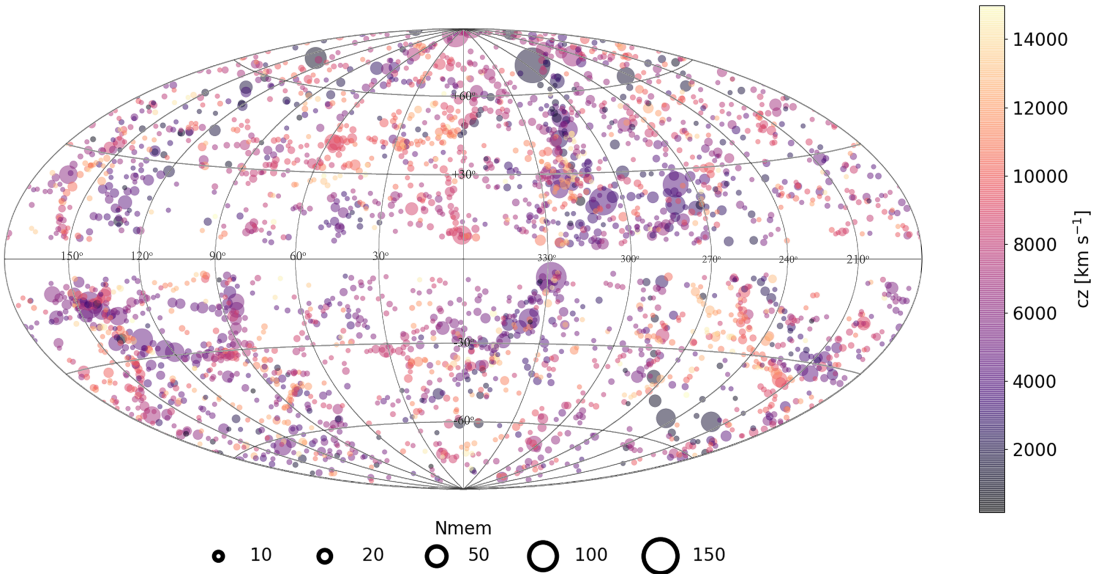

Figure 4.5: Equal-area Aitoff plot in Galactic coordinates of the completed 2MRS group catalogue. Points represent groups in the catalogue, their sizes are determined by the number of members. They are colored by their recession velocities. Only groups up to $cz = 15000$ km s^{-1} are included in this plot (for color scale).

consisting only of galaxies with $K_s^o > 11\overset{m}{.}25$ which are present in this catalogue but naturally are not in Crook et al. (2007). Interestingly, several groups in Crook et al. (2007) were identified at the lower and upper velocity ranges of large clusters. In other cases, large clusters (such as Virgo and Coma) that were made up of numerous groups in Crook et al. (2007) only contained one single large cluster in this catalogue.

4.4.2 Tempel et al., 2016 and 2018

Tempel et al. (2016a) created a modified version of the FoF algorithm in which membership refinement was included, making use of: (1) multimodality analysis (in an attempt to identify substructure within groups) and (2) determining the virial radius of groups and identifying any outliers from the escape velocity of the members (in order to not unjustifiably add members to groups). These techniques have direct analogies within the presented method to achieve many of the same goals.

Their modified version of the FoF was applied to the Huchra et al. (2012) version of 2MRS, as well as to other redshift surveys. They compared the resulting 2MRS group catalogue with the Tully (2015) 2MRS group catalogue between 3000 and 10000 km s^{-1}.

Tempel et al. (2016a) and Tempel et al. (2018) allowed groups with just two members, instead of the stronger requirement of 3 or more. These criterion (of 3 or more members) were applied to their results to enable a consistent comparison with this work.

The most significant difference between the Tempel et al. (2016a) and Tempel et al. (2018) catalogues are the methods used to identify groups. Whilst Tempel et al. (2016a) relied on a modified FoF algorithm, Tempel et al. (2018) made use of a Bayesian group finder similar to those used on other datasets (Robotham et al., 2011; Alpaslan et al., 2014). Once again, their algorithm was applied to the Huchra et al. (2012) catalogue.

As was the case in Crook et al. (2007), both Tempel et al. (2016a) and Tempel et al. (2018) had incidents in which large and well-known clusters were separated into numerous groups, whereas these are identified as a single cluster by these procedures. This further validates the success of the graph theory implementation. Groups found by Tempel et al. (2016a) or Tempel et al. (2018) and not found by us generally contained very few members and were often marginally connected. These cases are attributed (in Tempel et al. (2016a)) to a larger set of linking lengths which allowed for more marginal detections. Despite this, more groups are recovered in total than either Tempel et al. (2016a) or Tempel et al. (2018) which can be seen in the distribution in Fig. 4.6.

4.4.3 New Groups

New groups were identified as those with no match (within the tolerances) in the aforementioned catalogues. Fig. 4.6 shows the distribution of these new candidate groups as a function of redshift. Notably, their distribution is identical to the other catalogues, implying that new groups are being found around existing large-scale structure. This can be further verified from their spatial distribution, as seen in Fig. 4.7. For the most part it seems that the new galaxy group candidates are aligned with already well-known structures; however, there are several new ones that are found in under-dense regions.

Of particular interest is the ZoA, where the majority of the new redshifts in the final release of 2MRS were located. Fig. 4.8 shows a zoomed-in plot of the ZoA ($|b| < 25°$). Once again there are several new galaxy group candidates in regions of higher density, as well as several along the "lip" of the exclusion Zone. These particular groups will be interesting to follow up, as many align themselves with known structures deep within the ZoA ($|b| < 10°$, Staveley-Smith et al., 2016; Ramatsoku et al., 2016; Courtois et al., 2019).

As an additional exercise in examining the validity of the new galaxy group candidates, an extreme set of cross-matching criteria were chosen, namely 6 Mpc around plane-of-sky and 12 Mpc along the line-of-sight, in order to identify the most likely new galaxy group candidates. Of these candidate groups, only those with 5 or more members were selected, resulting in 20 groups. The positions of these groups were searched in the NASA Extragalactic Database (NED). Of the 20 groups, only 5 were identified in NED, of which 2 happened to be substructures of the well-known Shapley cluster (Zabludoff et al., 1993; Ramella et al., 1997; Bardelli et al., 1998; Mahdavi et al., 2000; Ramella

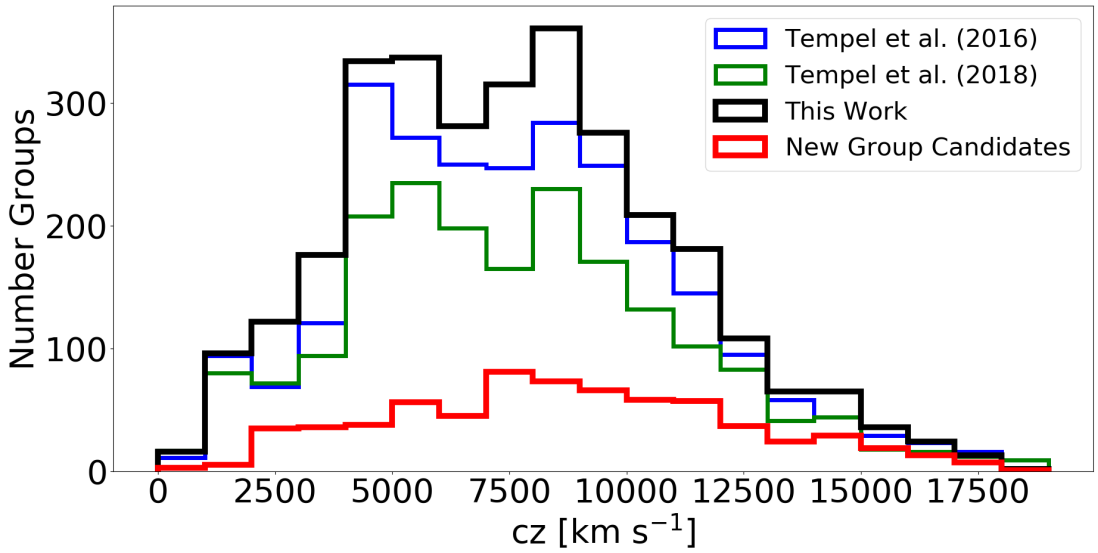

Figure 4.6: The distribution along redshift space of several previous renditions of the 2MRS group catalogue.

et al., 2002; Ragone et al., 2006; Díaz-Giménez & Zandivarez, 2015). The other 3 groups were all identified within Ramella et al. (2002) in the Northern CfA redshift survey.

75% of these new group candidates had no counterpart in NED and thus make for reasonable follow-up candidates. This also goes to show that several new groups exist within the catalogue but more work would be needed to identify and verify them as such. The new 2MRS groups that were previously identified in the literature bode well for the methods used in this work, as substructure within the Shapley Supercluster could already be identified from this rather shallow survey, unlike previous analyses of 2MRS.

A catalogue of new galaxy group candidates based on the comparisons with the aforementioned previous work is included. A small sub-sample is shown in Table 4.4 and include Group ID, RA, Dec, l, b, v_{cmb}, N_{mem}, and N'_{mem} as described in Table 4.1. Two separate cross matchings were done under strict (S) conditions, where no groups were identified in any catalog within a 3 Mpc projected on-sky radius as well $\pm$ 6 Mpc along line-of-sight, or lenient (L) where no groups were identified in any catalog within a 3 Mpc projected on-sky radius as well $\pm$ 6 Mpc along line-of-sight. A Strict/Lenient flag is included in column 9. Also included are candidate groups with 3 or more members but would recommend only considering those with 4 or more members.

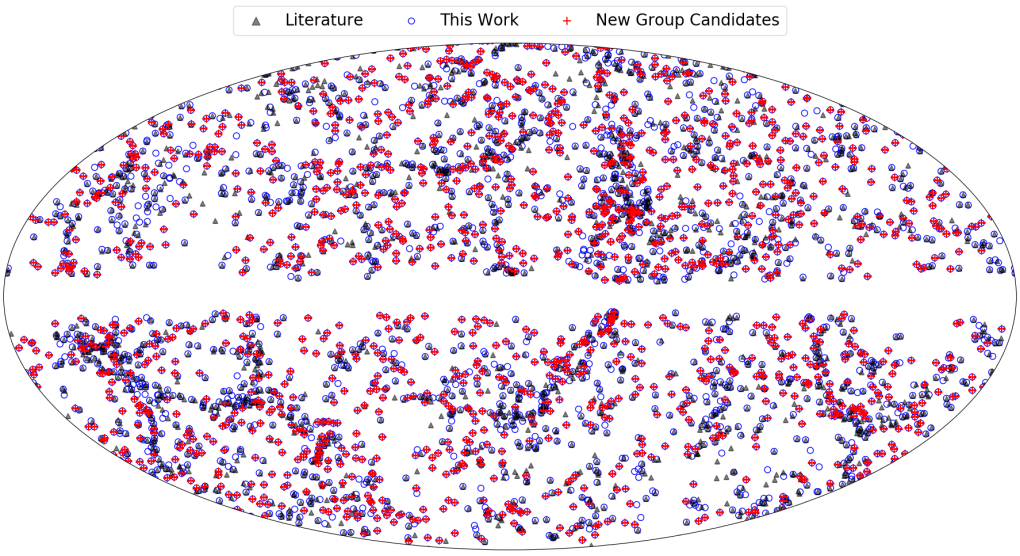

Figure 4.7: Galactic Aitoff projection showing the on-sky distribution of galaxy groups from Crook et al. (2007),Tempel et al. (2016a), and Tempel et al. (2018) as well as the groups found in this work. Black triangles show literature groups, blue circles show groups from this work, and red crosses demarcate possible new groups not found in literature.

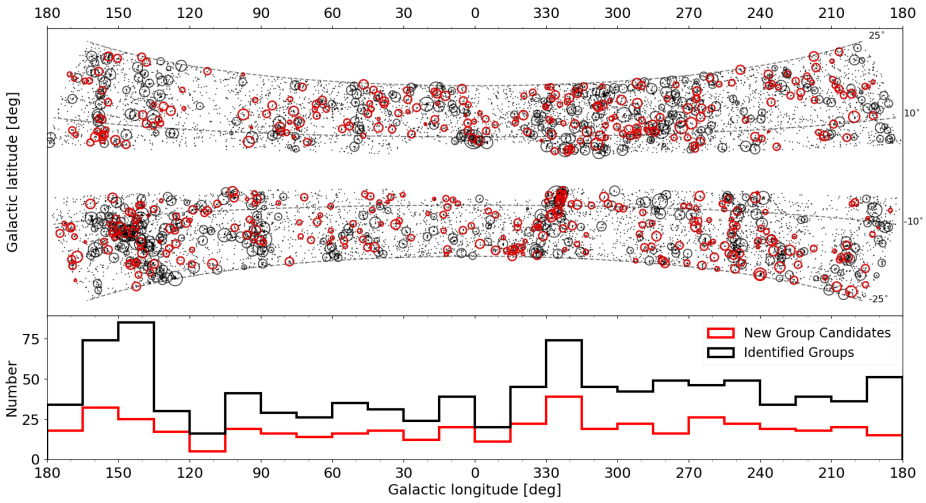

Figure 4.8: Zoomed-in view of the band around the ZoA ($|b| < 25°$). Top panel shows the on-sky distribution of galaxies and galaxy groups found in the ZoA. Circles represent groups found in this work. Black circles have a matching group in the literature. Red circles are groups without a match and are new group candidates. Bottom panel shows the distribution of the groups along the Galactic plane. Black line shows the distribution of all the groups in the 2MRS while the red shows the distribution of the new groups only.

Table 4.4: 2MRS New Galaxy Group catalogue

Group ID	α	δ	l	b	v_{cmb}	N_{mem}	N'_{mem}	S/L
		[deg, J2000]			[km s^{-1}]			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1782	202.17	−31.52	285.13	73.58	15560	14	285	L
2051	87.07	−25.58	325.27	−7.17	11479	12	130	L
2186	54.99	42.49	302.26	21.66	9894	8	54	L
2150	86.74	−25.60	150.53	−13.45	12844	8	100	L
1579	267.92	7.70	59.10	87.99	6349	7	22	L
576	89.65	−52.36	269.64	26.35	9901	7	48	L
1435	254.18	−6.31	285.13	73.58	8699	6	31	S
1692	193.89	−11.85	304.14	49.10	6591	6	20	L
2339	54.96	−2.56	317.13	30.92	10392	6	47	L
2307	240.56	−62.21	136.75	−24.96	13528	6	75	L
225	325.56	−70.96	332.53	−23.38	3686	6	10	L
2645	245.84	39.89	234.45	73.12	9679	6	38	L
439	220.02	−37.06	236.00	−54.21	4559	6	13	L
1844	338.07	51.93	273.17	19.27	11315	6	65	L
1704	207.24	−50.68	59.10	87.99	8586	6	31	S
2374	106.14	53.95	142.70	−63.07	11190	6	61	L
1613	170.55	−1.11	130.51	−28.86	7878	6	26	L
1984	240.53	36.70	325.27	−7.17	9364	6	36	S
97	116.21	−51.12	145.21	68.92	1217	6	6	L
1628	155.34	−4.57	150.53	−13.45	12209	6	75	S
1157	39.55	35.39	167.35	22.93	9166	6	36	L
1508	91.52	−35.81	140.83	−17.36	9824	6	41	L
2522	256.37	25.09	302.26	21.66	11465	6	65	S
338	12.60	−2.13	319.22	26.81	3660	6	10	L
2597	318.00	−48.49	8.70	−27.09	9333	5	30	L

Chapter 5

Discussion and Conclusion

The main scientific goal of this thesis was to create a group catalog for the newly completed 2MRS. Several interesting results have been discovered after the generation of the new 2MRS group catalog. Figure 4.5 shows the distribution Galactic coordinates of the galaxy groups from Table 4.1, where the size of the symbols are scaled according to the number of members ($N_{\rm mem}$). All the major well-known large-scale structures have been recovered, including the Virgo Supercluster, the Coma cluster, the Perseus-Pisces complex, the Fornax cluster, the Norma cluster (Great Attractor), and the Ophiuchus cluster, to name a few. This bodes well for validating the presented method for structures on all scales, i.e. the smallest groups to the largest clusters were found. The major clusters were all identified in previous analyses of 2MRS as well (Jarrett, 2004; Skrutskie et al., 2006; Crook et al., 2007; Huchra et al., 2012; Macri et al., 2019).

The traditional FoF algorithm developed by Huchra & Geller (1982) has several shortcomings, mostly arising from a static set of linking lengths. This leads to large clusters being identified as several smaller groups or several small groups falsely being identified as a single large cluster. Several modifications of the algorithm over the years attempted to address some of these issues: from using two different sets of linking lengths and reporting both low and high density catalogues (Crook et al., 2007), to adopting membership refinement based on escape velocities (Tempel et al., 2016a).

Even more interesting is examining the new groups within the ZoA, shown in Fig. 4.8. Several structures within the ZoA were mentioned in Macri et al. (2019) and are aligned with structures found in the inner ZoA (Kraan-Korteweg et al., 2018),such as the Perseus-Pisces complex and its two extensions through the ZoA at $l\sim$ 90° and 165°. Fig. 4.8 confirms that the galaxies not only map out an extension of the Perseus-Pisces chain from $l\sim$ 150 ° to $l\sim$ 160° (as discussed in Macri et al., 2019) but so do the galaxy groups

which later connect to Lynx (Ramatsoku et al., 2016; Kraan-Korteweg et al., 2018; Macri et al., 2019). There are also several new group candidates at this crossing, substantiating the importance of filling in the ZoA. The second crossing of the Perseus-Pisces chain, at $l\sim$ 90 °, also includes several new groups and is well defined in the group catalogue. Macri et al. (2019) also highlights a surprising new density at $(l, b)\sim (100°, -5°)$. In Fig. 4.8 groups are seen in this over-density, several of which have been identified as new, and may therefore may be worthy of additional spectroscopy follow-up observations.

5.1 Conclusion

A graph-theory based modification to the traditional FoF algorithm has been developed in order to address various shortcomings in earlier methods. This group finder is:

1. Robust to artificial changes such as order in which the algorithm is run.
2. Reliably identifies large clusters as a single body without nonphysically identifying smaller groups as a single cluster.
3. Reliably identifies small groups without nonphysically shredding large clusters into smaller groups.
4. Robust to outliers and chance alignments of galaxies.
5. Provides a unique method of identifying substructures within larger groups and clusters.

The group finder was run on the recently-completed 2MRS (Macri et al., 2019) and as such differs from previous analyses being the deepest to date and first 2MRS galaxy group catalogue based on the 100% complete 2MRS up to a magnitude limit of $K_s^o \leq 11\overset{\text{m}}{.}75$. Comparisons with previous work show that this method is able to recover most, if not all, the groups of previous catalogues and is able to identify numerous more, including substructures of the Shapley Supercluster for the first time. A catalogue of new, previously-unidentified group candidates is included. These provide an interesting follow-up opportunity, since additional observations are required to validate them.

This novel adaptation lent itself to several visual techniques: both the new method and the final galaxy group catalogue have been extensively and exhaustively interrogated using the new IDIA visualization laboratory – including immersive screen technology, virtual reality, and the newly upgraded Iziko Digital Dome Planetarium.

5.2 Future Work

I am currently in the process of generalizing the group finder so that it can be run on any magnitude-limited redshift survey. This will be released as a Python package, allowing users to generate their own group catalogues quickly and reliably.

While the current 2MRS galaxy group catalogue is the most complete "whole-sky" catalogue to date, there still remains a large gap within the ZoA which Macri et al. (2019) excluded. It did not form part of the original 2MRS, and follow-up of these obscured galaxies were deemed too difficult. Recently, Schröder et al. (2019) has created an updated catalogue of bright 2MRS galaxies (complete to extinction-corrected magnitudes of $K_s^o < 11\overset{m}{.}25$) within the ZoA ($|b| < 10°$). These redshifts are currently in the process of being obtained and being added to a final 2MRS galaxy group catalogue which will reduce the ZoA even further.

Furthermore, in light of the Square Kilometer Array (SKA) path finder surveys being conducted now and in the nearby future, such as the Widefield ASKAP L-band Legacy All-sky Blind surveY (WALLABY; Lee-Waddell et al., 2019; Koribalski et al., 2020), The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey (Jarvis et al., 2017, Maddox et al., (in prep)) and the Apertief surveys (Verheijen et al., 2009), as well as several other large HI projects (Kraan-Korteweg et al., 2016; Zhu et al., 2016), fine tuning the group finder for HI detections is a reasonable and beneficial adaptation. I hope to make this adaptation to the new group finder, in particular relying on the WALLABY survey which covers 3/4 of the sky and thus will provide the best data set which will allow for the comparison between a NIR-selected group catalog directly with a HI-selected survey which covers the same volume. This has the possibility to be a very informative and useful tool to be used on new MeerKAT, and eventually, SKA data.

Bibliography

Abell, G. O., Corwin, Jr., H. G., & Olowin, R. P. 1989, ApJS, 70, 1

Alpaslan, M., Robotham, A. S. G., Driver, S., et al. 2014, MNRAS, 438, 177

Andreon, S. 2016, A&A, 587, A158

Baldry, I. K., Robotham, A. S. G., Hill, D. T., et al. 2010, MNRAS, 404, 86

Bardelli, S., Pisani, A., Ramella, M., Zucca, E., & Zamorani, G. 1998, MNRAS, 300, 589

Bernyk, M., Croton, D. J., Tonini, C., et al. 2016, ApJS, 223, 9

Bianconi, M., Smith, G. P., Haines, C. P., et al. 2020, MNRAS, 492, 4599

Bilicki, M., Chodorowski, M., Jarrett, T., & Mamon, G. A. 2011, ApJ, 741, 31

Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., & Steward, L. 2014, ApJS, 210, 9

Bond, J. R., Kofman, L., & Pogosyan, D. 1996, Nature, 380, 603

Buta, R. J. 2013, Galaxy Morphology, ed. T. D. Oswalt & W. C. Keel, Vol. 6, 1

Calderon, V. F., & Berlind, A. A. 2019, MNRAS, 490, 2367

Carlesi, E., Hoffman, Y., Gottlöber, S., et al. 2020, MNRAS, 491, 1531

Collaboration, D., Aghamousa, A., Aguilar, J., et al. 2016, The DESI Experiment Part I: Science,Targeting, and Survey Design, arXiv:1611.00036

Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039

Courtois, H. M., Kraan-Korteweg, R. C., Dupuy, A., Graziani, R., & Libeskind, N. I. 2019, MNRAS Letters, 490, 57

Coutinho, B. 2016, PhD thesis, Northeastern University

Crook, A. C., Huchra, J. P., Martimbeau, N., et al. 2007, ApJ, 655, 790

Davis, M., Huchra, J., Latham, D. W., & Tonry, J. 1982, ApJ, 253, 423

de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986a, , 302, L1

—. 1986b, , 302, L1

Díaz-Giménez, E., & Zandivarez, A. 2015, A&A, 578, A61

Dressler, A. 1980, ApJ, 236, 351

Dressler, A., & Shectman, S. A. 1988, AJ, 95, 284

Duarte, M., & Mamon, G. A. 2014, MNRAS, 440, 1763

Fairall, A. P., ed. 1998, Large-scale structures in the universe, ed. A. P. Fairall

Falco, E. E., Kurtz, M. J., Geller, M. J., et al. 1999, VizieR Online Data Catalog, J/PASP/111/438

Gordon, Y. A., Pimbblet, K. A., Owers, M. S., et al. 2018, MNRAS, 475, 4223

Greene, J. E., Veale, M., Ma, C.-P., et al. 2019, ApJ, 874, 66

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

—. 1934, ApJ, 79, 8

Huchra, J., Martimbeau, N., Jarrett, T., et al. 2005, in IAU Symposium, Vol. 216, Maps of the Cosmos, ed. M. Colless, L. Staveley-Smith, & R. A. Stathakis, 170

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

Huchra, J. P., Geller, M. J., & Corwin, Harold G., J. 1995, ApJS, 99, 391

Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, ApJS, 199, 26

Jarrett, T. 2004, PASA, 21, 396

Jarrett, T. H., Cluver, M. E., Magoulas, C., et al. 2017, ApJ, 836, 182

Jarvis, M. J., Taylor, A. R., Agudo, I., et al. 2017, The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey, arXiv:1709.01901

Jones, D. H., Saunders, W., Colless, M., et al. 2004, MNRAS, 355, 747

Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683

Klypin, A., & Shandarin, S. F. 1993, ApJ, 413, 48

Kochanek, C. S., Pahre, M. A., Falco, E. E., et al. 2001, ApJ, 560, 566

Koribalski, B. S., Staveley-Smith, L., Westmeier, T., et al. 2020, arXiv e-prints, arXiv:2002.07311

Kourkchi, E., & Tully, R. B. 2017, ApJ, 843, 16

Kraan-Korteweg, R. C. 2005, in Reviews in Modern Astronomy, ed. S. Röser, Vol. 18, 48

Kraan-Korteweg, R. C., Cluver, M. E., Bilicki, M., et al. 2017, MNRAS, 466, L29

Kraan-Korteweg, R. C., Henning, P. A., & Schröder, A. C. 2002, A&A, 391, 887

Kraan-Korteweg, R. C., Jarrett, T. H., Elagali, A., et al. 2015, in SALT Science Conference 2015 (SSC2015), 40

Kraan-Korteweg, R. C., & Lahav, O. 2000, A&A, 10, 211

Kraan-Korteweg, R. C., van der Heyden, K. J., Cluver, M. E., & Woudt, P. A. 2008, in First Middle East-Africa Regional IAU Meeting, 8

Kraan-Korteweg, R. C., van Driel, W., Schröder, A. C., Ramatsoku, M., & Henning, P. A. 2018, MNRAS, 481, 1262

Kraan-Korteweg, R. C., Elson, E., Blyth, S., et al. 2016, in MeerKAT Science: On the Pathway to the SKA, 21

Lee-Waddell, K., Koribalski, B. S., Westmeier, T., et al. 2019, MNRAS, 487, 5248

Lim, S. H., Mo, H. J., Lu, Y., Wang, H., & Yang, X. 2017, MNRAS, 470, 2982

Lineweaver, C. H., Tenorio, L., Smoot, G. F., et al. 1996, ApJ, 470, 38

Loeb, A., & Narayan, R. 2008, MNRAS, 386, 2221

Lu, Y., Yang, X., Shi, F., et al. 2016, ApJ, 832, 39

Lynden-Bell, D., Faber, S. M., Burstein, D., et al. 1988, ApJ, 326, 19

Macri, L. M., Kraan-Korteweg, R. C., Lambert, T., et al. 2019, ApJS, 245, 6

Mahdavi, A., Böhringer, H., Geller, M. J., & Ramella, M. 2000, ApJ, 534, 114

Meilă, M. 2005, in In ICML ‚Äô05: Proceedings of the 22nd international conference on Machine learning (ACM Press), 577–584

Meilă, M. 2007, Journal of Multivariate Analysis, 98, 873

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1995, MNRAS, 275, 56

O'Brien, A. N., Norris, R. P., Tothill, N. F. H., & Filipović, M. D. 2018, MNRAS, 481, 5247

Ochsenbein, F., Bauer, P., & Marcout, J. 2000, , 143, 23

Otter, J. A., Masters, K. L., Simmons, B., & Lintott, C. J. 2020, MNRAS, 492, 2722

Pascucci, V., Tricoche, X., Hagen, H., & Tierny, J. 2011, Topological Methods in Data Analysis and Visualization: Theory, Algorithms, and Applications, 1st edn. (Springer Publishing Company, Incorporated)

Peebles, P. J. E. 1975, ApJ, 196, 647

—. 1980, The large-scale structure of the universe

—. 1993, Principles of Physical Cosmology

Poggianti, B. M., De Lucia, G., Varela, J., et al. 2010, MNRAS, 405, 995

Pomarède, D., Hoffman, Y., Courtois, H. M., & Tully, R. B. 2017, ApJ, 845, 55

Ragone, C. J., Muriel, H., Proust, D., Reisenegger, A., & Quintana, H. 2006, A&A, 445, 819

Ramatsoku, M., Kraan-Korteweg, R., Schröder, A., & van Driel, W. 2014, Proc. of SAIP 2012, arXiv:1412.5324

Ramatsoku, M., Verheijen, M. A. W., Kraan-Korteweg, R. C., et al. 2016, MNRAS, 460, 923

Ramella, M., Geller, M. J., & Huchra, J. P. 1989, ApJ, 344, 57

Ramella, M., Geller, M. J., Pisani, A., & da Costa, L. N. 2002, AJ, 123, 2976

Ramella, M., Pisani, A., & Geller, M. J. 1997, AJ, 113, 483

Rauzy, S., & Gurzadyan, V. G. 1998, MNRAS, 298, 114

Robotham, A. S. G., Norberg, P., Driver, S. P., et al. 2011, MNRAS, 416, 2640

Sandage, A., & Tammann, G. A. 1981, A revised Shapley-Ames Catalog of bright galaxies

Saulder, C., van Kampen, E., Chilingarian, I. V., Mieske, S., & Zeilinger, W. W. 2016, A&A, 596, A14

Schechter, P. 1976, ApJ, 203, 297

Schröder, A. C., van Driel, W., & Kraan-Korteweg, R. C. 2019, MNRAS, 482, 5167

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJJS, 172, 1

Seldner, M., Siebers, B., Groth, E. J., & Peebles, P. J. E. 1977, AJ, 82, 249

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Staveley-Smith, L., Kraan-Korteweg, R. C., Schröder, A. C., et al. 2016, AJ, 151, 52

Stothert, L., Norberg, P., & Baugh, C. M. 2019, MNRAS, 485, L126

Tempel, E., Kipper, R., Tamm, A., et al. 2016a, A&A, 588, A14

—. 2016b, A&A, 588, A14

Tempel, E., Kruuse, M., Kipper, R., et al. 2018, A&A, 618, A81

Trasarti-Battistoni, R., Invernizzi, G., & Bonometto, S. A. 1997, ApJ, 475, 1

Tully, R. B. 2015, AJ, 149, 171

Tully, R. B., & Fisher, J. R. 1987, Atlas of Nearby Galaxies

Tully, R. B., Pomarède, D., Graziani, R., et al. 2019, ApJ, 880, 24

van de Weygaert, R., & Bond, J. R. 2008, Observations and Morphology of the Cosmic Web, ed. M. Plionis, O. López-Cruz, & D. Hughes, Vol. 740, 24

Verheijen, M., Oosterloo, T., Heald, G., & van Cappellen, W. 2009, in Panoramic Radio Astronomy: Wide-field 1-2 GHz Research on Galaxy Evolution, 10

Zabludoff, A. I., Franx, M., & Geller, M. J. 1993, ApJ, 419, 47

Zeldovich, I. B., Einasto, J., & Shandarin, S. F. 1982, Nature, 300, 407

Zheng, Y.-L., & Shen, S.-Y. 2020, ApJS, 246, 12

Zhu, M., Pisano, D. J., Ai, M., & Jiao, Q. 2016, in Astronomical Society of the Pacific Conference Series, Vol. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015, ed. L. Qain & D. Li, 49

Appendix

Table 1: Full Group Catalogue

Group ID	Other Names	α	δ	l	b	$N_{\rm mem}$	$N'_{\rm mem}$	R	$v_{\rm cmb}$	d_c	d_L	RMS	σ	R_{200}	M_{200}	$N_{\rm subs}$
			[deg, J2000]						[km s^{-1}]	[Mpc]		[km s^{-1}]		[Mpc]	[$\times 10^{11} M_\odot$]	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
2313	-	0.02	32.75	110.61	−28.89	4	163	>4	9821	133.5	137.9	228	227	0.53	19.2	0
2534	-	0.06	−30.26	14.32	−78.35	9	278	>9	8640	117.6	121	156	160	0.36	6.2	2
690	-	0.1	47.12	113.98	−14.86	13	177	6	4888	66.7	67.8	133	135	0.31	3.8	2
1831	-	0.14	−2.11	94.81	−62.17	4	224	3	7115	96.9	99.2	164	167	0.39	7.2	0
561	-	0.19	−27.46	28.31	−78.74	4	360	4	7840	106.8	109.5	237	241	0.55	21.5	0
2813	-	0.36	−43.79	330.87	−70.52	3	156	>3	11565	157	163.1	144	131	0.34	4.9	0
1426	-	0.47	51.79	115.18	−10.34	3	129	>3	12722	172.6	179.9	117	118	0.27	2.6	0
1944	-	0.47	−36.08	349.49	−76.19	3	61	>3	14592	197.7	207.3	155	150	0.36	5.9	0
1686	-	0.66	31.35	110.85	−30.38	4	194	2	4503	61.5	62.4	79	74	0.18	0.8	0
2421	-	0.95	8.46	103.3	−52.59	5	111	>5	11522	156.5	162.5	86	87	0.2	1	0
2230	-	0.96	−45.53	327.05	−69.3	4	53	>4	11293	153.4	159.2	180	185	0.42	9.4	0
2895	-	0.97	−53.08	318.52	−62.58	3	76	>3	9643	131.1	135.4	83	85	0.19	0.9	0
2672	-	0.98	10.76	104.43	−50.4	3	49	2	7526	102.5	105.1	18	18	0.04	0	0
2110	-	1	22.06	108.64	−39.49	4	107	1	6250	85.2	87	85	86	0.2	1	0
879	-	1.19	7.29	103.09	−53.77	4	106	2	5888	80.3	81.9	92	94	0.22	1.3	0
2711	-	1.2	27.35	110.37	−34.38	5	250	3	7236	98.6	101	61	59	0.14	0.4	0
1549	-	1.22	53.73	116	−8.51	3	97	>3	11215	152.3	158	237	245	0.55	21.5	0
2567	-	1.42	−35.98	347.8	−76.85	3	92	>3	8237	112.1	115.2	90	82	0.21	1.2	0
1441	-	1.42	5.15	102.32	−55.88	3	128	1	4979	67.9	69.1	36	33	0.08	0.1	0
1578	DC_0003-50	1.45	−50.47	320.32	−65.09	8	98	>8	10210	138.8	143.5	180	186	0.42	9.4	0
2713	-	1.57	−13.61	82.74	−72.86	3	85	0	5395	73.6	74.9	15	15	0.04	0	0
2240	-	1.66	8.55	104.45	−52.7	3	111	>3	11302	153.5	159.3	138	131	0.32	4.2	0
1556	-	1.69	−52.05	318.53	−63.7	3	41	>3	10001	136	140.5	201	188	0.47	13.1	0
2380	-	2.14	10.28	105.96	−51.17	3	94	2	6152	83.9	85.6	135	115	0.32	4	0
2459	-	2.17	−37.45	341.37	−76.3	3	57	>3	8367	113.9	117.1	72	68	0.17	0.6	0
782	-	2.3	33.13	112.83	−28.92	6	148	5	4563	62.3	63.2	112	109	0.26	2.3	0
2541	-	2.5	−46.63	322.94	−68.85	3	98	>3	9509	129.3	133.4	67	67	0.16	0.5	0
2665	-	2.61	28.54	112.14	−33.47	8	83	8	7834	106.7	109.5	167	170	0.39	7.5	0
1279	-	2.89	−57.02	313.56	−59.28	6	211	>6	9440	128.4	132.4	171	176	0.4	8.1	0
45	-	3.01	16.15	109.32	−45.68	3	72	0	568	7.8	7.8	111	95	0.26	2.3	0
811	-	3.27	30.95	113.32	−31.2	3	93	1	4480	61.2	62.1	41	35	0.1	0.1	0
503	-	3.29	−23.2	55.24	−80.49	3	111	0	155	2.1	2.1	73	73	0.17	0.6	0
1021	-	3.29	28.28	112.78	−33.83	6	78	4	6928	94.4	96.6	245	250	0.57	24	0
1619	-	3.33	22.3	111.44	−39.72	3	66	2	5689	77.6	79.1	111	100	0.26	2.2	0
294	-	3.88	−7.19	98.16	−68.32	5	204	3	5113	69.8	71	139	127	0.33	4.4	0
2101	-	4.1	−59.1	311.29	−57.46	3	110	>3	9027	122.8	126.5	186	187	0.43	10.4	0
2004	-	4.18	1.18	105.11	−60.49	3	249	>3	13004	176.4	184	131	114	0.3	3.6	0
798	-	4.2	48.08	116.97	−14.39	12	69	4	4909	67	68.1	148	149	0.35	5.3	0
305	-	4.2	17.76	111.37	−44.33	7	139	2	5256	71.7	73	203	189	0.48	13.7	2
2270	-	4.26	−32.83	353.44	−80.63	3	163	3	7353	100.2	102.6	124	125	0.29	3.1	0
816	-	4.26	−6.42	99.85	−67.74	4	28	3	7582	103.3	105.9	55	54	0.13	0.3	0
1221	-	4.4	29.89	114.24	−32.41	19	59	13	6519	88.9	90.8	352	357	0.83	71.1	0
2057	-	4.57	53.82	117.99	−8.74	4	39	>4	11147	151.4	157	83	85	0.19	0.9	0
1636	-	4.77	−3.85	103.16	−65.47	7	45	5	5813	79.3	80.8	102	96	0.24	1.7	2
762	-	4.78	10.47	110.08	−51.59	3	35	3	4974	67.9	69	73	64	0.17	0.6	0
1888	-	4.9	−23.03	60.78	−81.71	3	83	3	7214	98.3	100.7	181	163	0.42	9.6	0
421	-	5.32	22.4	113.83	−39.94	18	61	6	5448	74.3	75.7	244	248	0.57	23.7	0
1453	-	5.51	38.12	116.59	−24.39	3	120	>3	10447	142	146.9	196	181	0.46	12.3	0
603	-	5.69	29.98	115.57	−32.48	4	56	2	4429	60.5	61.4	57	58	0.14	0.3	0
2324	-	5.69	26.9	115.08	−35.53	3	25	>3	8924	121.4	125	69	61	0.16	0.5	0
2092	-	6.07	13.61	112.9	−48.73	3	50	1	4882	66.6	67.7	76	75	0.18	0.7	0
2422	-	6.19	−14.25	95.82	−75.7	3	135	3	6974	95	97.2	33	31	0.08	0.1	0
2363	-	6.21	21.25	114.69	−41.19	3	45	1	5031	68.7	69.8	67	67	0.16	0.5	0
657	-	6.32	−48.65	314.47	−67.89	5	100	1	3137	42.9	43.3	66	64	0.16	0.5	0
2214	-	6.54	39.7	117.66	−22.91	5	115	>5	10721	145.7	150.9	215	222	0.5	16	0
2670	-	6.54	−32.91	344.8	−82.05	3	62	>3	14658	198.6	208.3	50	45	0.12	0.2	0

(1) – 2MRS group ID as defined in this work. (2) – Names (where available) of well known groups and clusters in other catalogs. (3) – Group RA in degrees, calculated as the average RA of the group members. (4) – Group Dec in degrees, calculated as the average Dec of

the group members. (5) – Galactic longitude of the group in degrees, calculated as the average of the Galactic longitude of the members. (6) – Galactic latitude of the group in degrees, calculated as the average of the Galactic latitude of the members. (7) – Number of 2MRS galaxies in the group. (8) – Corrected number of members in the group from correcting the group to the position of Virgo as described in Eq. 4.6. (9) – Richness of the group. Defined as the number of members in the group with an absolute magnitude of $M_K^o > -23\overset{m}{.}5$. (10) – Recession velocity corrected to the CMB reference frame. (11) – Comoving distance, calculated using Eq. 4.4. (12) – Luminosity distance, calculated using Eq. 4.5. (13) – Velocity dispersion calculated using Eq. 4.1. (14) – Velocity dispersion determined by taking the standard deviation of the velocity of the members of the group. (15) – R_{200} measurement of the group calculated using Eq. 4.2. (16) – M_{200} measurement of the group calculated using Eq. 4.3. (17) – Number of subgroups within the main group as defined in chapter 2.3. The full csv version of this table is available at `https://www.dropbox.com/sh/13jvusgloyh9k2r/AAAgd2Cxaut_dD8lNL0egBhXa?dl=0` or upon request.

Table 2: Full Subgroup Catalogue

Sub ID	Group ID	α	δ	l	b	v_{cmb}	N_{mem}
		[deg, J2000]				[km s^{-1}]	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1000-1	1000	140.72	24.44	203.96	43.31	10337	4
1000-2	1000	141.6	23.89	204.98	43.94	10145	3
1015-1	1015	107.14	−49.1	259.72	−17.53	12790	7
1015-2	1015	106.12	−48.98	259.35	−18.11	12997	3
1031-1	1031	241.04	69.77	103.49	39.28	7540	3
1031-2	1031	239.39	70.74	104.93	39.22	7486	6
1091-1	1091	52.65	41.59	152.24	−12.05	5367	7
1091-2	1091	52.56	40.53	152.81	−12.95	4772	5
1093-1	1093	241.24	17.57	31.31	44.52	10395	17
1093-2	1093	241.48	18.0	31.99	44.46	11589	11
1100-1	1100	251.94	58.81	88.16	38.82	5256	3
1100-2	1100	249.56	57.86	87.3	40.25	5349	6
1154-1	1154	52.59	41.79	152.08	−11.92	4325	3
1154-2	1154	52.13	40.29	152.68	−13.34	4107	3
1186-1	1186	195.68	−56.21	304.51	6.63	6225	4
1186-2	1186	196.91	−57.33	305.13	5.47	6030	13
119-1	119	142.76	−61.89	281.56	−7.66	2862	17
119-2	119	137.59	−64.03	281.46	−10.82	2076	4
1201-1	1201	173.22	−9.62	272.85	48.61	6630	21
1201-2	1201	174.68	−9.32	274.62	49.52	6155	4
121-1	121	200.07	−47.49	307.96	15.1	3307	9
121-2	121	198.85	−46.1	307.27	16.56	3390	4
1215-1	1215	258.88	43.74	69.03	35.21	8464	7
1215-2	1215	258.48	42.93	68.01	35.42	8076	4
125-1	125	330.52	−32.03	14.86	−53.08	2286	13
125-2	125	328.52	−34.76	10.29	−51.53	2365	3
127-1	127	46.93	41.86	148.42	−14.17	2654	4
127-2	127	45.11	43.2	146.53	−13.68	2393	3
1302-1	1302	202.39	−28.0	313.1	34.11	10707	3

1302-2	1302	201.88	−27.64	312.64	34.55	10615	13
135-1	135	73.99	−4.66	203.36	−27.73	3913	7
135-2	135	72.36	−6.13	203.98	−29.85	4589	5
1355-1	1355	44.65	45.42	145.14	−11.9	8980	4
1355-2	1355	43.78	45.76	144.43	−11.89	9481	3
136-1	136	149.37	−31.76	264.97	17.99	2975	21
136-2	136	150.52	−34.11	267.31	16.79	3114	6
1364-1	1364	67.51	−12.37	208.12	−36.92	9604	8
1364-2	1364	67.03	−12.28	207.78	−37.3	9122	3
139-1	139	56.73	−3.87	191.76	−42.13	3843	10
139-2	139	55.56	−4.58	191.68	−43.5	4112	3
14-1	14	68.41	−59.11	269.17	−40.53	1258	7
14-2	14	64.9	−55.37	264.93	−43.32	1234	14
1414-1	1414	61.5	70.17	138.09	13.26	4214	3
1414-2	1414	57.43	69.22	137.66	11.62	4939	3
143-1	143	222.41	−18.49	338.32	36.12	2977	5
143-2	143	225.22	−17.11	342.01	35.76	3254	3
1463-1	1463	166.14	45.18	165.58	62.1	6560	6
1463-2	1463	165.11	46.0	164.88	61.06	6719	4
1471-1	1471	85.23	−26.58	230.8	−26.43	11953	3
1471-2	1471	85.63	−26.11	230.43	−25.94	11669	8
1488-1	1488	189.27	26.76	218.34	86.77	7512	4
1488-2	1488	189.25	27.87	199.01	86.7	7775	3
149-1	149	203.6	34.35	73.6	78.28	7573	20
149-2	149	203.86	32.94	67.03	78.85	7518	3
150-1	150	193.52	−13.04	303.93	49.82	4775	23
150-2	150	193.56	−14.36	303.97	48.5	4547	18
150-3	150	195.06	−14.51	306.13	48.31	4974	3
152-1	152	40.99	32.24	149.01	−24.87	4716	20
152-2	152	38.84	32.63	147.03	−25.34	4548	10

(1) – individual subgroup ID with the first half representing the main group and the second half representing the unique identifier within that group. (2) – The Group ID where the subgroup is found. (3) – RA J2000. (4) – Dec J2000. (5) – Galactic longitude. (6) – Galactic latitude. (7) – Recession velocity corrected to the CMB reference frame. (8) - The number of members in the subgroup. The full csv version of this table is available at `https://www.dropbox.com/home/2MRS_Catalogues` or upon request.

Table 3: 2MRS Group Member Catalogue

2MASS ID	α	δ	l	b	v_{cmb}	Group ID	Sub ID
		[deg, J2000]			[km s^{-1}]		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
00000168 + 4716282	0.00695	47.27456	113.95532	−14.69903	4739	690	690-1
00000701 + 0816448	0.02926	8.27913	101.78542	−52.47268	11239	2556	-
00000914 + 3244182	0.03814	32.73844	110.62066	−28.90421	10045	2313	-
00001295 + 4657543	0.05404	46.96514	113.92233	−15.00847	5087	690	690-1
00001627 + 3247332	0.06772	32.7926	110.66226	−28.85704	10005	2313	-
00002880 + 3246563	0.12003	32.78236	110.70856	−28.8769	9476	2313	-
00003564 − 0145472	0.14849	−1.76318	95.13649	−61.85964	6964	1831	-
00003794 + 2823041	0.15808	28.38454	109.57114	−33.16557	8367	1974	-
00004498 + 0809292	0.18737	8.15806	101.96467	−52.63781	11403	2556	-
00004696 + 2824071	0.19569	28.40202	109.61443	−33.15629	8426	1974	-
00005298 − 4323316	0.22083	−43.39207	331.82632	−70.7887	11749	2813	-
00005317 − 3559104	0.22161	−35.9863	350.30841	−76.07829	14737	1944	-
00005891 + 2854421	0.2455	28.91172	109.80592	−32.67063	6562	2262	-
00011378 − 4400426	0.3074	−44.01183	330.58795	−70.32169	11454	2813	-
00011748 − 5300348	0.32276	−53.00967	319.35657	−62.49016	9544	2895	-
00011907 − 2725116	0.32942	−27.41995	28.5576	−78.86381	7541	561	-
00012334 + 4733537	0.34737	47.56496	114.24901	−14.46112	4961	690	690-1
00012677 + 3126016	0.36155	31.43374	110.59126	−30.23711	4481	1686	-
00013005 + 3126306	0.37521	31.44188	110.60651	−30.23178	4438	1686	-
00013415 − 3619001	0.39226	−36.31675	348.86942	−75.99117	14386	1944	-
00013584 − 2734355	0.39929	−27.57649	27.75035	−78.92843	7723	561	-
00014401 − 3025082	0.43328	−30.4189	13.18496	−78.62929	8317	2534	2534-2
00015583 − 2737382	0.48258	−27.62723	27.48804	−79.00261	8195	561	-
00015787 + 2137162	0.49125	21.62112	107.90981	−39.80305	6353	2110	-
00020314 − 4521288	0.51302	−45.35804	328.06674	−69.28413	11418	2230	-
00020544 − 3037082	0.52273	−30.61901	12.10372	−78.65454	8777	2534	2534-2
00021051 − 4358398	0.54387	−43.97774	330.22983	−70.44679	11493	2813	-
00023231 − 3037181	0.63468	−30.62166	11.96378	−78.74693	8708	2534	2534-2
00024370 + 2725454	0.68219	27.42926	109.8483	−34.20185	7207	2711	-

00030565 − 0154495	0.77358	−1.9138	96.22283	−62.25193	6945	1831	-
00030887 + 3102108	0.78701	31.0364	110.90174	−30.70468	4466	1686	-
00030962 + 2157368	0.79005	21.96022	108.36713	−39.54215	6229	2110	-
00031298 − 3556132	0.80408	−35.93708	349.28857	−76.48599	14654	1944	-
00031331 + 5352149	0.80545	53.87077	115.7833	−8.33027	11524	1549	-
00031494 + 1608428	0.81228	16.14523	106.40895	−45.17482	702	45	-
00032047 + 0837075	0.83527	8.61877	103.20203	−52.40073	11540	2421	-
00032138 − 5004494	0.83896	−50.08049	321.56412	−65.27246	10137	1578	-
00032922 + 2721063	0.87179	27.35172	110.02559	−34.31534	7350	2711	-
00033586 + 1036138	0.89954	10.60383	104.24539	−50.52399	7548	2672	-
00033775 + 0838069	0.90729	8.63525	103.32143	−52.4066	11677	2421	-
00033830 + 5146404	0.90966	51.77794	115.4481	−10.39776	12689	1426	-
00033881 + 5340004	0.91169	53.66673	115.80707	−8.54245	11196	1549	-
00035421 + 0818099	0.97591	8.3028	103.26245	−52.744	11422	2421	-
00035635 + 1052347	0.98486	10.87635	104.49463	−50.28721	7503	2672	-
00035902 + 0821567	0.99599	8.36578	103.32515	−52.68998	11500	2421	-
00040066 − 4540009	1.00279	−45.66691	326.78452	−69.20225	11026	2230	-
00040883 + 0722458	1.03686	7.37938	102.88808	−53.6418	5839	879	-
00041061 + 0821568	1.04418	8.36587	103.4002	−52.70426	11468	2421	-
00041299 + 1047258	1.0541	10.79052	104.55825	−50.38892	7526	2672	-
00042247 + 5140299	1.09369	51.67496	115.54251	−10.52015	12597	1426	-
00042449 + 3128193	1.1021	31.47199	111.31509	−30.33657	4629	1686	-
00042463 − 5257316	1.10272	−52.95879	318.4693	−62.72777	9634	2895	-
00042665 + 4729250	1.11115	47.49031	114.75687	−14.634	4993	690	690-2
00043232 + 0806175	1.13466	8.105	103.41179	−52.97998	6017	879	-
00043594 − 4528463	1.14977	−45.47957	326.80753	−69.41582	11504	2230	-
00044038 − 4537443	1.16819	−45.62894	326.55923	−69.29411	11225	2230	-
00044453 + 2649569	1.18562	26.83246	110.21674	−34.88361	7216	2711	-
00045778 + 0507245	1.24079	5.12344	101.99665	−55.84768	4996	1441	-

(1) – Unique 2MASS ID of the galaxy. (2) – RA, J2000. (3) – Dec, J2000. (4) – Galactic longitude. (5) – Galactic latitude. (6) Recession velocity corrected to the CMB reference frame. (7) – The unique identifier of the Group in the 2MRS Group Catalogue to which the galaxy belongs. (8) – The unique identifier of the subgroup to which the galaxy belongs (if available). The full csv version of this table is available at `https://www.dropbox.com/home/2MRS_Catalogues` or upon request.

Table 4: 2MRS New Galaxy Group Catalogue

Group ID	α	δ	l	b	v_{cmb}	N_{mem}	N'_{mem}	S/L
		[deg, J2000]			[km s^{-1}]			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2813	0.36	−43.79	353.66	56.02	11565	3	33	L
1944	0.47	−36.08	133.21	−13.54	14592	3	61	S
879	1.19	7.29	120.37	−32.89	5888	4	11	L
2567	1.42	−35.98	155.68	−12.39	8237	3	14	L
2713	1.57	−13.61	253.77	31.01	5395	3	8	L
2380	2.14	10.28	315.31	38.36	6152	3	9	L
2541	2.5	−46.63	203.0	89.7	9509	3	19	S
45	3.01	16.15	92.29	38.64	568	3	3	L
2004	4.18	1.18	52.29	19.46	13004	3	38	S
2324	5.69	26.9	353.8	53.4	8924	3	17	S
2775	6.74	24.06	13.99	−52.8	6986	3	11	S
2149	7.67	−5.23	197.57	11.13	7321	3	11	L
1972	7.85	53.33	174.08	−44.46	8647	3	15	L
2379	8.86	−32.72	193.81	50.19	4173	3	6	L
1896	9.26	−23.2	34.79	−35.25	3543	4	7	S
2603	9.53	−8.14	294.08	−70.55	11030	4	34	S
614	9.65	−19.15	304.99	5.75	3599	3	5	S
1139	11.22	−6.7	213.96	−34.97	5462	4	10	L
1149	11.79	−86.29	199.11	−20.49	4910	3	7	L
338	12.6	−2.13	319.22	26.81	3660	6	10	L
2278	12.76	24.46	252.28	−26.51	9781	3	20	S
2071	14.06	−1.24	321.75	6.14	12121	4	50	L
2470	14.37	7.87	255.62	−83.57	11169	3	27	S
1468	14.76	43.89	305.95	40.44	5089	3	7	L
858	14.92	−8.91	311.34	45.77	4186	4	8	S
2127	15.87	14.21	330.56	−25.62	11974	3	35	L
1591	16.4	−6.33	171.29	−22.49	12165	3	38	S
2907	16.53	−19.22	90.64	28.43	16000	3	70	L
1251	18.06	−3.88	4.3	−38.14	5273	3	7	L

2608	18.09	8.79	251.56	52.84	10440	3	24	S
1673	18.56	−7.73	210.16	86.8	5126	3	7	L
2481	18.63	−45.0	305.35	30.27	9286	4	24	L
2824	20.29	−7.01	120.37	−32.89	10485	3	24	S
1144	20.91	−32.71	161.04	31.8	8972	4	22	L
2595	21.34	−35.73	148.06	12.49	10248	3	23	L
2890	21.46	−13.07	306.66	9.97	15439	3	61	L
2589	21.58	−4.74	143.04	−11.08	14022	4	59	S
2117	21.69	−18.62	25.36	−30.24	9445	3	19	S
1630	22.06	−4.11	4.15	27.17	5410	3	8	L
2535	22.44	16.96	85.74	71.54	12745	3	38	S
2628	22.82	19.78	355.13	11.75	12500	3	38	L
2893	22.86	41.15	273.3	−37.14	2378	3	4	L
650	23.08	45.16	3.21	62.33	4822	3	7	L
2648	23.13	21.31	161.72	−37.62	9675	3	19	L
1297	23.67	−68.11	76.2	−55.88	19609	3	163	S
1589	24.55	5.5	269.46	41.46	2922	3	4	L
2432	25.23	−14.38	295.57	15.64	11658	3	33	L
2296	25.38	−33.6	351.37	−32.67	8682	4	20	L
1724	25.73	77.53	285.87	42.23	13684	3	38	S
1353	25.96	−65.88	113.98	−14.86	7935	3	14	S
2789	26.89	32.18	133.63	24.66	14417	3	54	S
328	26.95	27.91	203.0	89.7	3356	5	8	L
932	27.31	22.68	162.38	20.04	9607	3	19	S
2200	27.54	43.58	159.15	−12.78	6068	4	12	L
82	27.72	5.42	162.38	20.04	1316	3	3	L
398	27.89	−49.09	195.68	−17.68	6145	5	15	L
2251	29.25	−2.72	5.68	26.24	8494	3	15	S
1334	29.64	42.11	316.48	49.7	14963	3	61	S
1936	29.82	0.43	317.93	5.84	6643	3	10	L
2045	30.36	−31.56	83.32	31.08	5307	3	7	L

(1) – The unique identifying ID of the group in the 2MRS Group Catalogue. (2) – RA, J2000. (3) – Dec, J2000. (4) – Galactic longitude. (5) – Galactic latitude. (6) – Recession velocity corrected to the CMB reference frame. (7) – The number of members in the group. (8) – The corrected number of members in the group according to Eq. 4.6. (9) – Tag highlighting whether the new group was identified under strict (S) or lenient (L) conditions. The full csv version of this table is available at `https://www.dropbox.com/home/2MRS_Catalogues` or upon request.