Alex Nie

Conroy, Wellons

Astron 202a

12/11/2015

Galaxy Clustering and the Halo Model

I. Introduction

Within the study of large scale structure, galaxy clustering data encodes information about both galactic evolution as well as underlying cosmology. With the emergence of statistically large redshift surveys, as well as the convergence of observations favoring ΛCDM cosmogony as the standard model for cosmology (Spergel et al. 2003), much of the study of galactic clustering has focused on fitting and testing the hierarchical formation paradigm for galaxies (Gao et al. 2004). Within this paradigm, dark-matter “host” halos – defined as roughly spherical, virialized regions of matter overdensities about 200 times that of background density -- are seen as environments in which galaxies form and accrete, typically hosting one large central galaxy, and a number of self-bound satellite galaxies, each encased in their own dark matter “sub-halo”. In turn, much of the study of galaxy clustering focuses on relating properties of clustered galaxies to the properties of dark matter halos they inhabit. As remarked in the literature, this paper will use “distinct” halos to mean the smallest halo which encompasses a cluster and “sub-halo” to mean the smallest halo encompassing an individual galaxy (Zheng, Coil, & Zehavi 2008).

Of particular interest in the study of galaxy clustering are the correlation function (ξ(r)) and projected correlation functions, which measure the clustering present in each galaxy. While the correlation function ideally provides a measure of cluster within a given region by estimating the likelihood above expected value of finding two galaxies at a given separation, redshift distortions tend to distort clustering along the line-of-sight (Zehavi et al. 2005). As a result, the projected correlation function, similar in conception to the angular correlation function, is used. This is derived by integrating over a fixed range of redshifts deemed to fall within the range of "clustering". Mathematically, this can be defined as $P(r) = \bar{n}\big(1 + \xi(r)\big)dV,\ \xi(r) = \xi\big(r_{\perp,}\, r_{\parallel}\big),$ $w_p\big(r_p\big) = \int_0^{\pi} \xi(r_{\perp}, r_{\parallel})dr_{\parallel}$ where π is some maximum redshift difference considered the maximum separation between two galaxies in a cluster along the line-of-sight.

Several essential questions guide inquiry in examining galaxy clustering. First, how do current and potential theories of galaxy formation and cosmology explain observed galactic clustering? As demonstrated by Yang, Mo, & van den Bosch (2003), galactic clustering constrains models of general galaxy formation as well as individual galaxy histories through statistical means on two scales: that of the cluster, and that of the sub-halo population (2003). Second, given the role of dark matter halos in catalyzing galaxy formation in the ΛCDM model, how do properties of galaxies relate to properties of the host-halo? This paper reviews work by several different groups relating galaxy distribution, luminosity, and color within halos to properties of the parent halo such as circular velocity and host mass (Zheng, Coil, & Zehavi 2008; Conroy, Wechsler, & Kravtsov 2006; Zehavi et al. 2011). As dark-matter halos are not directly observable, this paper will also present common frameworks and assumptions used in the field to infer properties of the halos from the observable galaxies.

Several common challenges also appear in studying galactic clustering. The first involves the problem of halo detection. For many simulations, as dark matter does not form definitive, discrete spherical regions, a cutoff is required to establish whether two nearby galaxies inhabit the same halo or two different halos, which yields implications for their evolutionary history as well as explanations for observed clustering, as later explained (Zheng, Coil, & Zehavi 2008). As standard in any study of clustering, another hurdle in studying galactic clustering involves the detection of clustering in observational data. Again, informed by theory, distance cutoffs are often used to limit how far apart galaxy clusters can be (Zehavi et al. 2005). Moreover, this is corrected by the projection out to a finite redshift so that galaxies separated by significant distances along line-of-sight are excluded from cluster-counting. Next, in comparing halo models to observational data, currently only luminous galaxies are observationally available. This has two implications. First, data on galaxies at very small scales, and thus low luminosities, is expected to undercount the actual population of galaxies at that scale and thus skew statistics (Zehavi et al. 2005). Second, as dark matter halos are not directly observable, simulations must make assumptions based on current cosmological models about coupling of dark matter to matter. This leads to the final essential question: is the population of galaxies in general one-to-one with the halo population? (Yang, Mo, & van den Bosch 2003; Gao et al. 2004; Conroy, Wechsler, & Kravtsov 2006) As assumed in several papers, this relationship is at least surjective above a certain scale, with halos above a certain mass being automatically populated with galaxies in simulations. However, as discussed in the paper, several alternatives apply, including halos' lacking galaxies, known as "dark galaxies", and galaxies lacking halos, known as "orphan galaxies" (Gao et al. 2004)/ Due to effects such as tidal stripping and dynamical friction, these orphan galaxies can comprise a non-trivial portion of a galaxy population and must be accounted for (Gao et al. 2004).

II. Methods

This section provides a brief overview of the methodology common in studying galactic clustering. In general, research involves adopting a framework for how halos are populated by galaxies, running N-body and semi-analytic simulations under these and various cosmological parameters, and comparing statistical descriptions to those from observational data (Yang, Mo, & van den Bosch 2003; Nagai & Kravtsov 2005; Zehavi et al. 2005; Conroy, Wechsler, & Kravtsov 2006).

In measuring galactic clustering, several groups have focused on modeling the two-point correlation function and projected correlation functions defined earlier for galaxies and halos, defined as the relative excess of clustering above mean value between galaxies in the sample. After performing analysis on this projected correlation function, this transformation is then inverted to yield correlation functions, typically of the form $\xi(r) = \left(\frac{r}{5.91} h^{-1} Mpc\right)^{-\alpha}$ where α is the free parameter in a power-law fit (Zehavi et al. 2005; Conroy, Wechsler, & Kravtsov 2006; Zehavi et al. 2011).

With the advent of large redshift surveys, the study of galactic clustering has evolved so that populations rather than individual histories of galaxies are studied. As such, there currently exist three major frameworks under which the relatively difficult problem of tracking individual galaxy formation and evolution is transformed into to a statistical problem of galaxy distributions. The first framework, the Halo Occupation Distribution framework (HOD), characterizes the problem through a conditional probability that a halo of mass M contains N

galaxies, $P(N|M)$, which along with descriptions of galaxy distributions within halos, relate observed clustering to theories of formation (Peebles 1974; Zehavi et al. 2005). In simulations of clustering, this is employed by using a power law to fit the mean occupation of galaxies for an observed population. In general,

$$\langle N \rangle_M = \left(\frac{M}{M_1}\right)^{\alpha}$$

where $M_1$is a power law cutoff mass and α is the logarithmic slope. For values below the cutoff mass, the average occupation is assumed to be 1 above a minimum threshold M_{min}, considered the minimum threshold at which a dark matter halo can form a galaxy. These simulations are then compared with various empirical results (Zehavi et al. 2005; Zehavi et al. 2011). As discussed later, this minimum threshold is insufficient at small scales, as disruptions of dark matter halos which cause them to dip below the threshold mass do not necessarily disrupt the galaxies they host.

The conditional luminosity function (CLF), provides a refinement on the HOD model by inspecting the number of galaxies with luminosities within the range $L \pm \frac{dL}{2}$ given host halo mass M, which yields the CLF, $\Phi(L|M)$. Using empirical results for the galaxy luminosity function $\Phi(L)$, the number of galaxies with luminosities in the range $L \pm \frac{dL}{2}$, the conditional luminosity function can then be constrained and normalized through the relation

$$\Phi(L) = \int_0^{\infty} \Phi(L|M) n(M) dM$$

along with a Schechter fit

$$\Phi(L) = \frac{\Phi^*}{L^*}\left(\frac{L}{L^*}\right)^{\alpha} e^{-\frac{L}{L^*}}$$

where $n(M)$ is the distribution of galaxies by mass (Yang et al. 2003). In addition to correlation functions, this approach is also used to calculate mass-to-light ratios for various halos using the fact that

$$\langle L \rangle (M) = \int_0^\infty \Phi(L|M) L \, dL$$

thereby allowing tests of galaxy formation and cosmological parameters (Yang et. al 2003). Yang et. al argues that this approach both avoids overgeneralizing galaxy populations given different evolutions under various physical processes like star formation and feedback. In turn, Zehavi et. al argue that the initial HOD framework is generally more flexible to global and sub-populations, yielding more five free parameters, while the CLF framework assumes a luminosity functional that may or may not be present in a given sample (Zehavi et al., 2011).

A third method known as the Sub-Halo Abundance Matching (SHAM), a refinement of the HOD framework, explores the one-to-one galaxy-to-halo problem (Conroy, Wechsler, & Kravtsov 2006). In this framework it is assumed that there is a monotonic relation between luminosity of the central galaxy and the mass of its parent halo. In particular, assignment of halos is made in simulations with $n_g(> L_i) = n_h(> V_{max,i})$ where $n_g$is the number of galaxies above a threshold luminosity L_i and $n_h$is the number of dark-matter halos with a certain circular velocity at the time of accretion. This model is then used to calculate mean occupation of halos with a given mass and minimum circular velocity in simulations for comparison with empirical data, thereby reducing bias inherent in low-luminosity observations. In their paper, Conroy, Wechsler, & Kravtsov (2006) also demonstrate that this is an alternative to considerations of orphan galaxies as this framework, later adopted by Zehavi et al. (2011) reproduces observed clustering behavior as well.

For the literature reviewed, simulations concerned with generating correlation functions typically used a flat concordance ΛCDM model along currently accepted estimates for various components: $\Omega_0 \sim .3, \Omega_\Lambda = .7, \Omega_b \sim .05$ (Yang, Mo, & van den Bosch 2003; Nagai & Kravtsov 2005; Conroy, Wechsler, & Kravtsov 2006; Zehavi et al., 2011). Instead simulations differed in their inclusion of various interactions of baryonic matter to compare the effects of baryonic interaction with dark-matter interactions in dominance over clustering behavior. Example models include dissipationless models and models with gas cooling and stellar formation (Nagai & Kravtsov 2005).

In comparing with observational results, the literature reviewed primarily used recent redshift surveys, including the 2-degree Field Redshift Survey (2dFRS), the Sloan Digital Sky Survey (SDSS) and Deep Extragalactic Evolutionary Probe 2 (DEEP2) (Yang, Mo, & van den Bosch 2003; Zehavi et al., 2005; Conroy, Wechsler, & Kravtsov 2006; Zehavi et al, 2011) . Overall, observational results were pruned before fitting with simulated data to account for bias in observations. In particular, studies typically imposed a luminosity threshold for galaxies and volume threshold for galaxies and halos in order to ensure sufficient sampling of a given population at relevant scales (Zehavi et al. 2005). For example, luminosity thresholds are imposed to prevent biasing toward very bright but small galaxies on small scales, as observations will be biased towards the most luminous of small, and difficult to observe galaxies. Despite such pruning measures, given the abundance of data from recent redshift surveys, a sufficient population of galaxies (118,000 when limited by luminosity, 22,000 when limited by volume), enhanced by mock cataloguing, was enough to produce statistically meaningful results. (Zehavi et al. 2005).

IV. Major Results

Departures from Power-Law Fits of the Correlation Function

While early analyses of galaxy clustering saw deviations from power-law fits within the noise tolerance of the sampling procedure (Peebles 1974), recent improvements in observational power indicate that these deviations are systematic. (Zehavi et al 2005) As shown in Figure 1, both flux-limited and volume-limited samples from the Sloan Digital Sky Survey indicate a deviation from a best-fit power law for the projected correlation function around r_p.~ 1.5 - 4 h^{-1} Mpc.

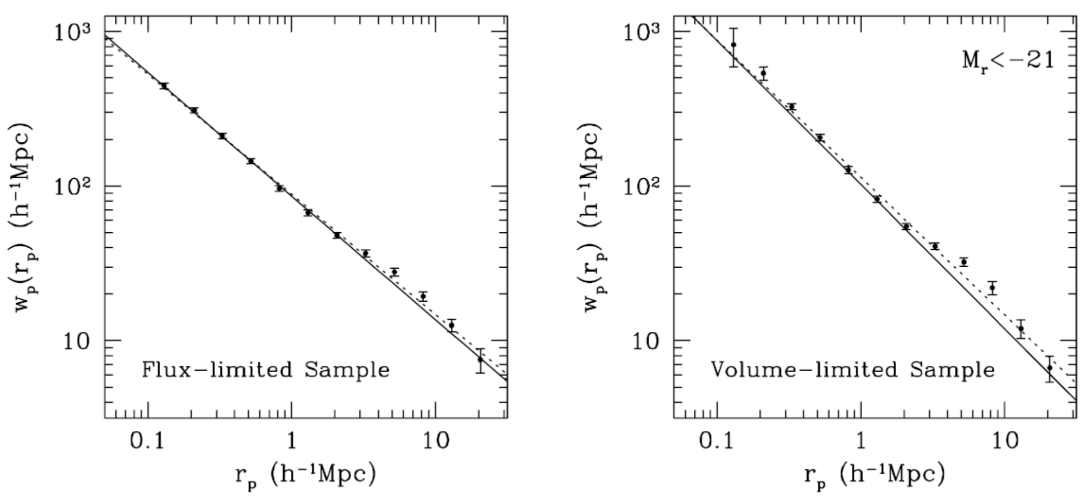

Figure 1: Projected correlation function $w_p(r_p)$ for flux-limited (left) and volume-limited (right) redshift samples from the Sloan Digital Sky Survey. Solid lines represent maximum-likelihood power-law fits while dotted lines represent least-squares fits. (Zehavi et al. 2005)

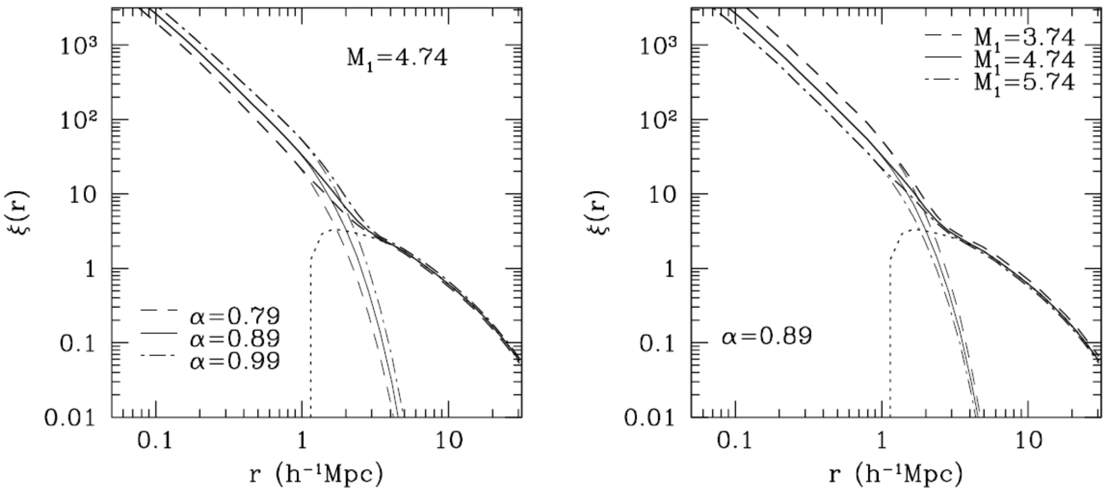

Figure 2: Correlation function $\xi(r)$for simulated HOD models varying logarithmic slope α over fixed power-law cutoff mass $M_1 = 4.74\ h^{-1}M_{\odot}$ and varying M_1 over $\alpha = .89\ h^3 Mpc^{-3}$. One-halo correlation functions populate the upper left curves while two-halo terms populate the curves showing a steep rise near r=1 and populating the lower-right.

One proposed explanation for this deviation is the transition between dominant populations at varying separations between clusters. That is, at small distances, clustering tends to occur between galaxies within the same distinct halo, either through central galaxy-satellite or satellite-satellite counts while at larger distances, clustering occurs between galaxies located in different halos (i.e. no halo contains the two galaxies counted). In Figure 2, several simulated correlation functions are shown, revealing excellent agreement in shape for the correlation function. Figure 3 compares projected correlation functions between simulations and observations for the

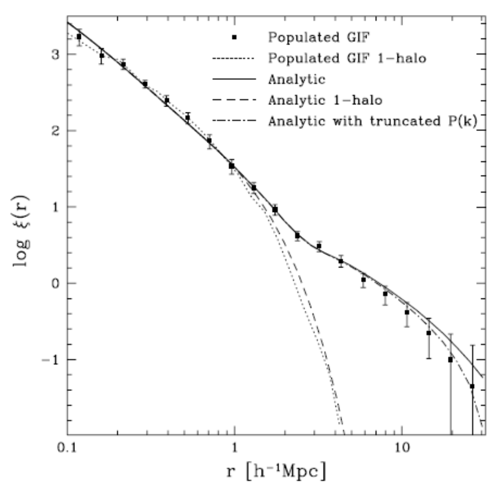

Figure 3: Tests of observed correlation function for flux-limited SDSS data $M_r < -21$ against GIF N-body simulations of Jenkins et al. (1998), (Zehavi et al. 2005)

SDSS. The 2005 study found a best-fit of $\xi(r) = \left(\frac{r}{6.40\ h^{-1}Mpc}\right)^{-1.89}$ for a maximum likelihood fit, while a least-squares regression found $\xi(r) = \left(\frac{r}{5.91\ h^{-1}Mpc}\right)^{-1.93}$. A similar analysis performed by Zheng, Coil, & Zehavi (2008) on later released SDSS and DEEP2 data corroborates the one-halo and two-halo term model:

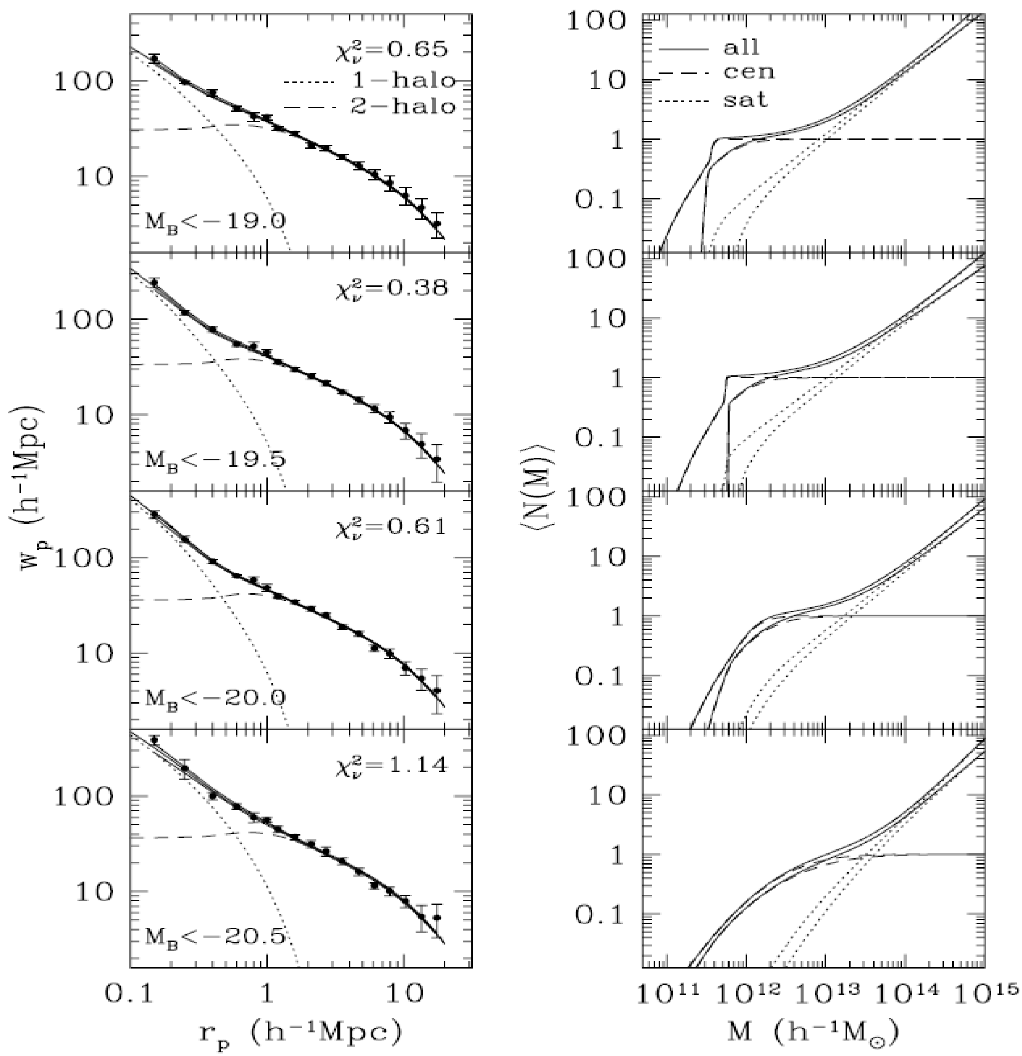

Figure 4: Projected correlation functions over varying flux-limited samples from DEEP2 (Zheng, Coil, & Zehavi 2008).

Luminosity, Color, and the Halo Model

By subsequently increasing the flux threshold of the analyzed sample, Zehavi et al. (2011) demonstrate that the projected correlation function increases in amplitude with increasing luminosity. For galaxies with a brighter magnitude than $M_r = -20.44$, projected correlation function increases rapidly with increasing luminosity, while it does so slowly for less luminous galaxies (see Figure 5). This is consistent with current cosmological and galactic formation models as the projected correlation function exhibits an increase in both the threshold occupancy mass and the power-law cutoff mass with increasing luminosity.

With respect to color, Zehavi et al. (2011) find that at fixed luminosity, bluer galaxies tend to have a low and shallow-amplitude correlation function while the reddest galaxies have a steeper correlation function, with very strong clustering at small scales (see Figure 6). While this difference is explained by differences in satellite fraction (nearly 15% for the bluest galaxies vs. 75% for the reddest galaxies), poor fitting for clustering in the high luminosity regime may

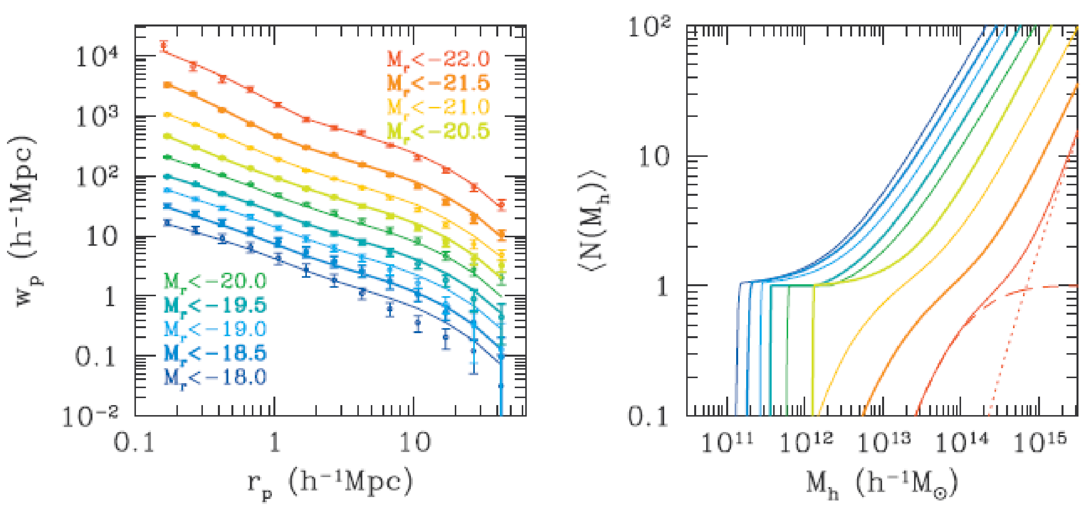

Figure 5: Projected correlation function (left) and average occupancy (right) for galaxy populations using various flux thresholds. On the right, solid lines indicate the occupancy statistic for the satellite population while the dashed line indicates occupancy for the central galaxy population.

indicate a need for a refinement of the HOD framework. Overall, the luminosity dependence of red galaxies is minimal while blue galaxies exhibit an increasing amplitude for projected correlation with luminosity (Zehavi et al. 2011)

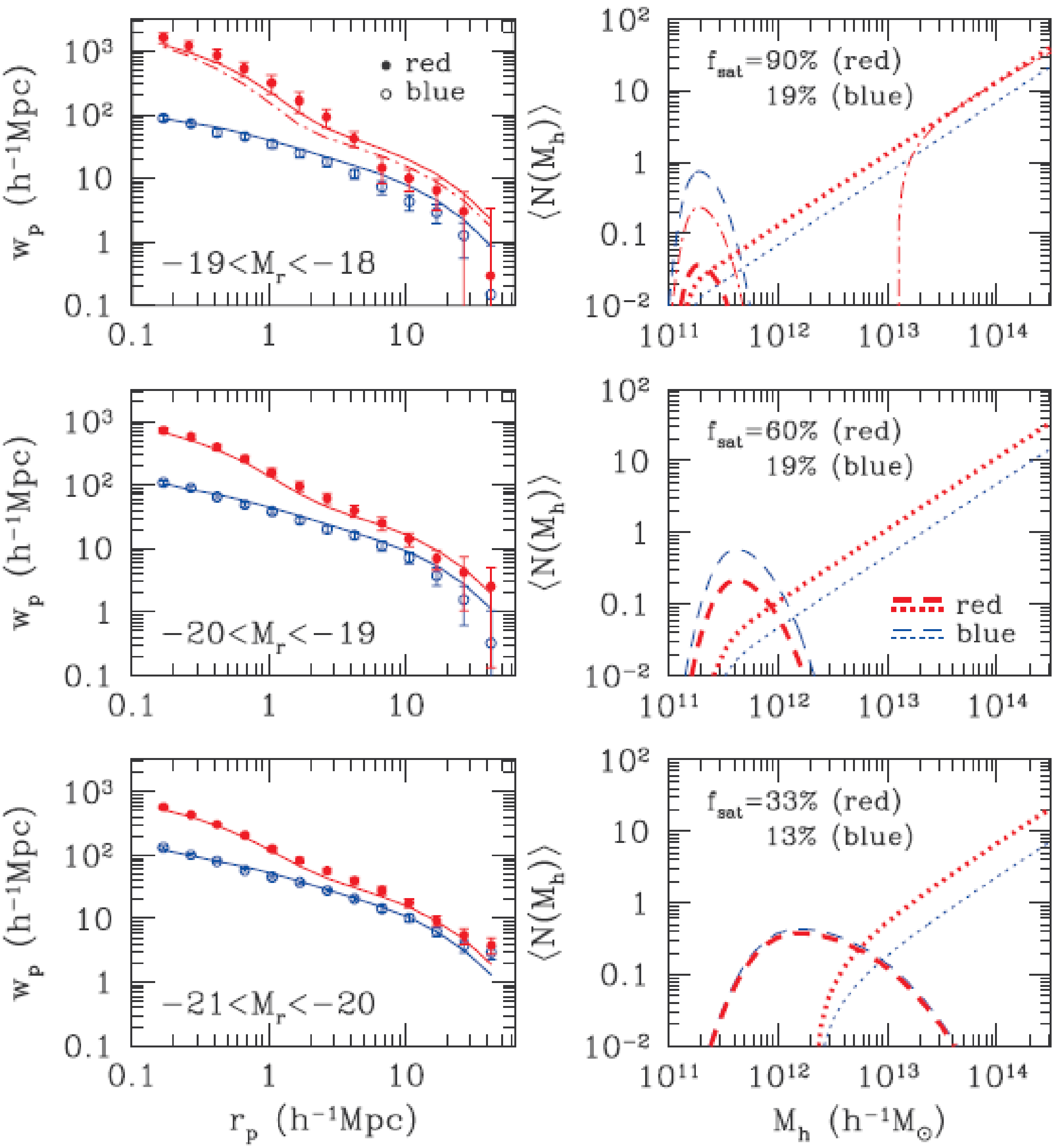

Figure 6: Projected correlation functions for blue and red subsamples of the Sloan Digital Sky Survey Seventh Release (Zehavi et al. 2011)

Radial Distribution of Galaxies and Subhalos within Host Halos

While the previous studies attempted to link galaxy properties to halo properties through clustering abundance, Nagai & Kravtsov (2005) demonstrate that within halos, galaxies do not strictly trace the dark matter distribution, resulting in biases when galaxies are used as an observable proxy for halo presence. Under dissipationless simulations and simulations which included gas cooling and stellar formation, Nagai & Kravtsov reveal a more gradual radial decay for matter than dark matter within a halo. Using simulations, they demonstrate that this is consistent with differential tidal stripping of stellar mass from the center of the halo to its fringes (about 70% for central galaxies, 30% for galaxies near the edge of the halo) as well as differential tidal stripping of the outlying dark matter as compared to the centrally located and much more strongly bound baryonic matter in the center of the halo (Figure Y). In turn, they demonstrate that using matter to trace halos can be reduced using properties at accretion time, rather than present time in simulations (Nagai & Kravtsov 2005).

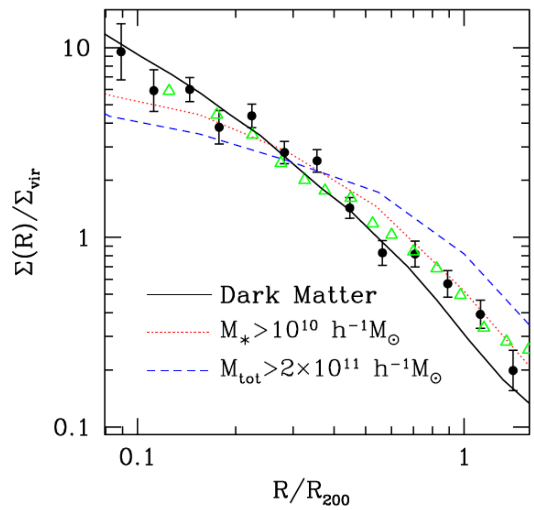

Figure 7: Dark and baryonic matter distributions as a function of distance from center of the halo.

Conditional Luminosity Function and Constraints on ΛCDM:

As desmontrated by Yang, Mo, & van den Bosch (2003), cluster analysis of galaxies also provides constraints on various cosmologies. The results of various N-body simulations is shown in Figure 8, indicating further support for the commonly accepted $\Omega_0 = .3$. As Yang, Mo, & van den Bosch (2003) detail further in the paper, the CLF framework also provides tests to break degeneracies in normalizations σ_8, settling on a value of σ_8~.87.

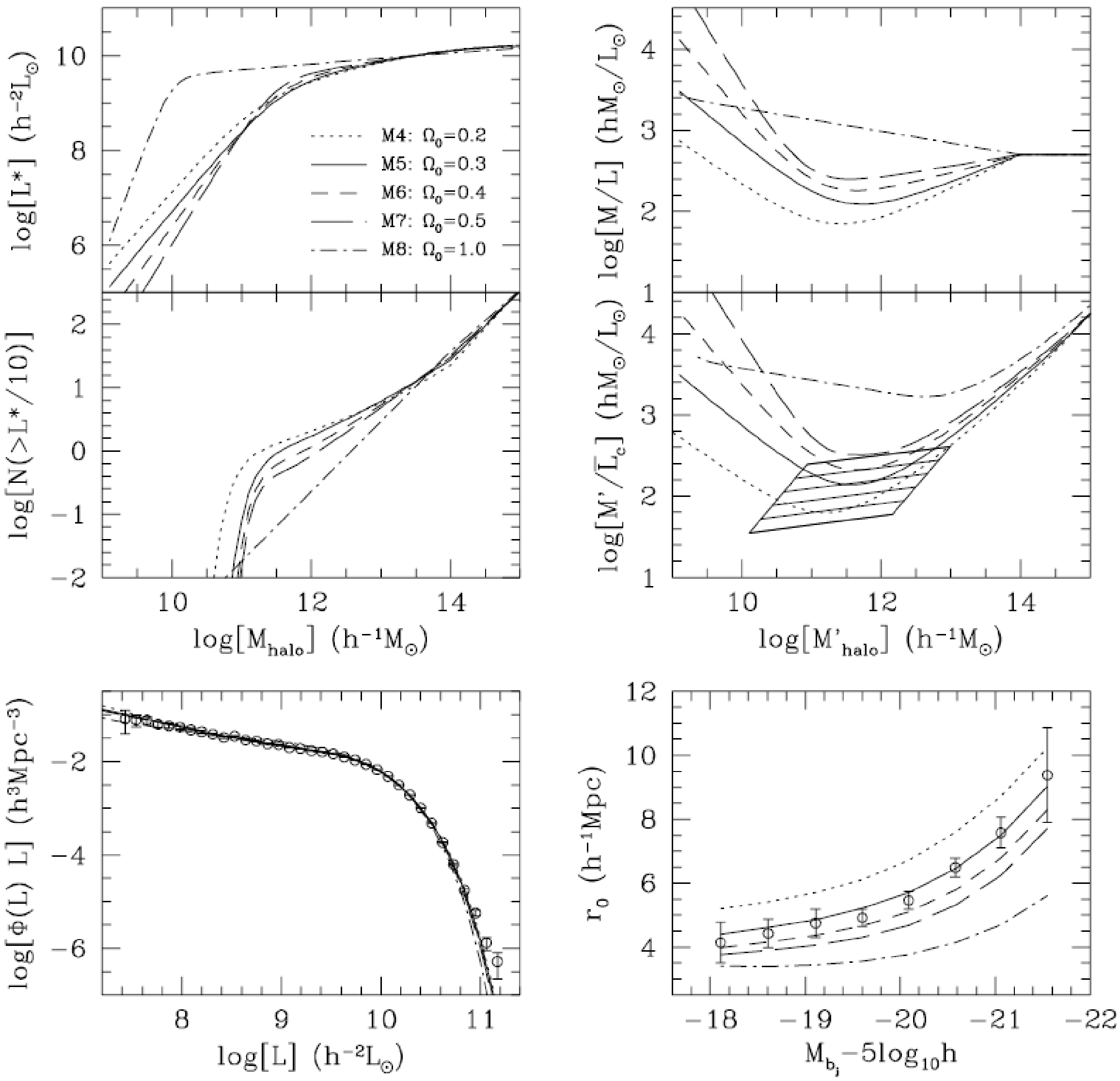

Figure 8: Clustering results of N-body simulations for various cosmogonies.

Orphan Galaxies

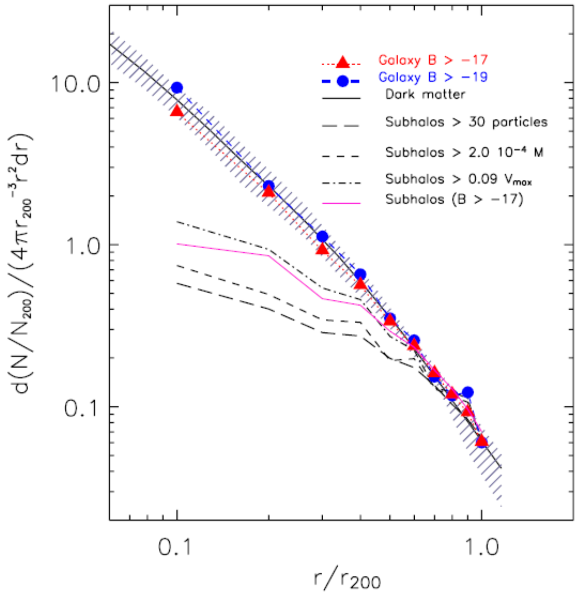

Figure 9: distribution of dark matter versus sub-halos within 10 high-resolution N-body simualtions

While previous studies fit observations of luminous galaxies with halo distributions, Gao et al. (2004) showed that a nontrivial percentage of galaxies do not reside in dark matter halos after a certain period of evolution due to disruption. Under a set of 10 N-body resimulations of a massive galaxy cluster in the ΛCDM cosmology ($\Omega_0 = .3, \Omega_\Lambda = .7, h = .7, \sigma_8 = .9$) along with a comparison against the semi-analytic model by De Lucia et al. (2004b) which tracked baryonic interactions such as chemical evolution, chemical enrichment, and feedback, Gao et al. (2004) found nearly 55% of galaxies within galaxy clusters and larger halos had no associated sub-halo at observation time, which accounts for the substantial difference between galaxy and sub-halo distribution, as evidenced by the discrepancy between subhalos and dark matter distributions (Figure 9)

V. Summary and Further Considerations

Although the ΛCDM hierarchical models predict that galaxies sit inside dark-matter halos, cluster analysis reveals a greater details of how observable galaxies might be distributed amongst dark-matter halos. Most importantly, the study of deviations in the power law form of a correlation functions support the idea of nested sub-halos within larger halos, as the “shoulder” shape can be reproduced in simulation as a transition between clustering dominated by galaxies occupying the same distinct halo and galaxies occupying different distinct halos (Zehavi et al. 2005). An examination of luminosity and color dependence illustrates that populations of galaxies exhibit different correlation functions and bias, indicating that more than just stellar mass may be influenced by galaxies’ surrounding dark matter halos (Zehavi et al. 2011). Within halos, the distribution of galaxies does not trace the underlying distribution of dark matter (Zheng, Coil, & Zehavi 2008). In particular, satellite galaxies near the edges of halos lose much less mass to dynamical friction and tidal stripping than galaxies near the central galaxy of the halo. While the Navarro, Frenk, and White (NFW) profile is one attempt to describe this distribution, this results in the “cuspy halo” problem: a spike in concentration of dark matter is expected at the center of the halo, but is not supported by rotational velocity observations (Zheng, Coil, & Zehavi). Using the conditional luminosity function framework, galaxy clustering also provides a test of cosmological parameters. As demonstrated by Yang, Mo, & van den Bosch (2002), current clustering statistics constrain ΛCDM parameters in agreement with values obtained from microlensing experiments: $\Omega_0 = .3, \sigma_8 = .89$. Finally, tidal disruption is effective in removing dark matter from halos, but not more tightly bound baryonic matter, which tends to concentrate in the central galaxy. As a result, dark matter-only simulations may neglect significant populations of “orphan galaxies” if population is done strictly on the basis of threshold mass (Gao et al. 2004). Future work may involve either inclusion of more baryonic physics in simulations, as Gao et al. suggest, although Nagai & Kravtsov (2004) as well as Conroy, Wechsler, & Kravtsov (2006) demonstrate that disspationless simulations are sufficient for thermal interactions. As Yang suggests, a treatment using the CLF framework over the HOD and SHAM frameworks, may avoid this problem altogether.

References

Conroy, C., Wechsler, R. H., Kravtsov, A.V. 2006, ApJ, 647, 201

De Lucia G., Kauffmann G., White S. D. M., 2004b, MNRAS, 349, 1101

Gao, L., De Lucia, G., White, S.D.M., Jenkins, A. 2004, MNRAS, 352, L1

Nagai, D., & Kravtsov, A. V. 2005, ApJ, 618, 557

Peebles, P. J. E. 1974, A&A, 32, 197

Spergel D.N. et al. 2003, ApJS, 148, 175

Yang, X., Mo, H. J., van den Bosch, F. C. 2003, MNRAS, 339, 1057

Zehavi et al. 2005, ApJ, 608, 16

Zehavi et al. 2011, ApJ, 736, 59

Zheng, Z., Coil, A. L., Zehavi, I. 2008, ApJ, 667, 760