Clustering Constraints on the Relative Sizes of Central and Satellite Galaxies

Andrew Hearin[1], Peter Behroozi[2], Andrey Kravtsov[3], Benjamin Moster[4]

[1] *Argonne National Laboratory, Argonne, IL, USA 60439, USA*
[2] *Department of Physics, University of Arizona, 1118 E 4th St, Tucson, AZ 85721 USA*
[3] *Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637 USA*
[4] *Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 München, Germany*

10 October 2017

ABSTRACT
We place empirical constraints on the connection between dark matter halos and galaxy half-light radii, $R_{1/2}$. Using the SDSS, we show that smaller galaxies cluster much more strongly than larger galaxies at fixed stellar mass. Using Halotools to forward-model the observations, we find that the clustering signal generically requires satellite galaxies to be smaller than central galaxies of the same halo mass. We present a simple empirical model consistent with the clustering results, in which galaxy size is proportional to halo virial radius at the time of peak halo mass, and the scatter is strongly correlated with halo scale radius. Models in which galaxy size is related to stellar mass only or that have uncorrelated scatter in galaxy size both fail the clustering test, even considering effects of satellite stripping and orphan galaxies. These results suggest that the relative size of centrals and satellites is predetermined at the time of satellite infall, and supports the notion that L_* galaxy and halo profiles co-evolve across many Gyr of cosmic time.

1 INTRODUCTION

In the ΛCDM framework of cosmological structure formation, galaxies form at the centers of dark matter halos. Highly complex and nonlinear baryonic processes regulate galaxy formation, and the quest for a fine-grained understanding of these processes is one of the chief goals of theoretical astrophysics today.

Observationally, many properties of observed galaxies exhibit remarkably tight scaling relations. Among the most fundamental of these relations is the strong correlation between galaxy size and stellar mass and luminosity. The scaling of galaxy size with galaxy mass is well-measured in the local Universe (Shen et al. 2003; Guo et al. 2009; Huang et al. 2013; Zhang & Yang 2017) and at high-redshift (Trujillo et al. 2004; van der Wel et al. 2014; Kawamata et al. 2015; Shibuya et al. 2015; Huertas-Company et al. 2013; Lange et al. 2015; Huang et al. 2017).

These well-measured scaling relations are challenging to faithfully recover using ab initio galaxy formation methods such as hydrodynamical simulations and semi-analytic models, and provide useful boundary conditions for the calibration of such modeling efforts (Khochfar & Silk 2006; Dutton et al. 2011; Hopkins et al. 2010; Bottrell et al. 2017). The precision cosmology program also depends critically on accurate modeling of galaxies, so that cosmological parameter inference can be confidently conducted without undue interference from uncertainty

in baryonic physics (LSST Science Collaboration et al. 2009; Robertson et al. 2017).

2 DATA AND SIMULATIONS

Our galaxy sample comes from the catalog of SDSS galaxy profile decompositions provided by Meert et al. (2015). This catalog is based on Data Release 10 of the Sloan Digital Sky Survey (SDSS, Ahn et al. 2014), with improvements to the photometry pipeline and light profile fitting methods (Vikram et al. 2010; Bernardi et al. 2013, 2014; Meert et al. 2013). In the version of this catalog that we use, two-dimensional $r-$band profiles were fit with a two-component de Vaucouleurs + exponential profile to determine the half-light radius $R_{1/2}$. We apply the Bell et al. (2003) mass-to-light ratio to the $r-$band flux and $g-r$ colors in this catalog to obtain an estimate for the total stellar mass M_* of every galaxy. We calculate two-point clustering w_p of our SDSS galaxy sample using line-of-sight projection of $\pi_\mathrm{max} = 20$Mpc using the `correl` program in `UniverseMachine`.

As the bedrock of our modeling, we use the catalog of `Rockstar` subhalos identified at $z = 0$ in the Bolshoi-Planck simulation (Klypin et al. 2011; Behroozi et al. 2013,?; Riebe et al. 2013; Rodríguez-Puebla et al. 2016). The particular version of the catalog we use is made publicly available through `Halotools` (Hearin et al. 2016), with `version_name` = 'halotools_v0p4'.

We additionally explore the potential existence of satellite galaxies that reside in subhalos that are not identified by halo-finder to the present day, so-called "orphan galaxies" (see, e.g., Campbell et al. 2017). We use an extension of `Consistent Trees` that models the evolution of subhalos after disruption. The phase space evolution of orphans is approximated by following a point mass evolving in the host halo potential according to the orbital parameters of the subhalo at the time of disruption; the evolution of subhalo mass and circular velocity is approximated using the semi-analytic model presented in Jiang & van den Bosch (2014).

For mock galaxies, to compute galaxy clustering we employ the distant observer approximation by treating the simulation $z-$axis as the line-of-sight. We compute w_p using the `mock_observables.wp` function in `Halotools`, which is a python implementation of the algorithm in the `Corrfunc` C library (Sinha & Garrison 2017).

All numerical values of $R_{1/2}$ will be quoted in physical kpc, and all values of M_* and M_halo in $M_\odot$, assuming $H_0 = 67.8$ km/s $\equiv 100h$ km/s, the best-fit value from Planck Collaboration et al. (2016). To scale stellar masses to "$h = 1$ units" (Croton 2013), our numerically quoted values for M_* should be multiplied by a factor of h^2, while our halo masses and distances should be multiplied by a factor of h.

3   GALAXY-HALO MODEL

3.1   Stellar mass model

We map M_* onto subhalos with the best-fit stellar-to-halo mass relation from Moster et al. (2013):

$$\langle M_*/M_\mathrm{halo} \rangle = 2N \left[(M_\mathrm{halo}/M_1)^{-\beta} + (M_\mathrm{halo}/M_1)^{\gamma} \right]^{-1}. \tag{1}$$

For halo mass M_halo we use M_peak, the largest value of M_vir ever attained along the main progenitor branch of the subhalo.

The values of the best-fit parameters in Moster et al. (2013) were fit to a stellar mass function (SMF) with values $M_*^\mathrm{MPA-JHU}$ based on the MPA-JHU catalog (Kauffmann et al. 2003; Brinchmann et al. 2004), which differs from the SMF in our galaxy sample (see, e.g., Bernardi et al. 2014). We account for this difference by manually tabulating the median value $\langle M_*^\mathrm{Meert+15}|M_*^\mathrm{MPA-JHU} \rangle$ in logarithmic bins spanning $9 < \log_{10} M_*^\mathrm{MPA-JHU}/M_\odot < 12$, and applying the median correction to the Monte Carlo realization of the mock galaxy sample. This results in a typical boost of ~ 0.25 dex at $M_*^\mathrm{MPA-JHU} \approx 10^{9.75} M_\odot$, and ~ 0.4 dex at $M_*^\mathrm{MPA-JHU} \approx 10^{11.5} M_\odot$.

3.2   Galaxy size models

In §4, we calculate predictions for the $R_{1/2}-$dependence of galaxy clustering for several different kinds of empirical models, described in turn below.

3.2.1 M_*-only model

In the first class of models we explore, we suppose that stellar mass M_* is the statistical regulator of $R_{1/2}$, so that galaxy sizes are drawn from a log-normal distribution centered at $\langle R_{1/2}|M_* \rangle$, where M_* derives from the Moster et al. (2013) relation described above in §3.1. To implement this model, for simplicity we directly tabulate $\langle R_{1/2}|M_* \rangle$ from the data, rather than pursue a parametric form (see, e.g., Zhang & Yang 2017).

3.2.2 $R_\mathrm{vir}-$only model

Motivated by Kravtsov (2013), we explore a model in which $R_{1/2}$ is linearly proportional to halo virial radius:

$$R_{1/2} = 0.0125 R_\mathrm{vir} \tag{2}$$

For the virial radius of halos and subhalos, we use R_{M_peak}, the value of R_vir in physical units of kpc measured at the time of peak subhalo mass, defined by

$$M_\mathrm{peak} \equiv \frac{4\pi}{3} R_{M_\mathrm{peak}}^3 \Delta_\mathrm{vir}(z_{M_\mathrm{peak}}) \rho_\mathrm{m}(z_{M_\mathrm{peak}}), \tag{3}$$

where for $\Delta_\mathrm{vir}(z_{M_\mathrm{peak}})$ we use the fitting function to the "virial" definition used in Bryan & Norman (1998). For the model we refer to as the "$R_\mathrm{vir}-$only model", we add uncorrelated log-normal scatter of $\sigma_{R_{1/2}} = 0.2$ dex to generate a Monto Carlo realization of the model population.
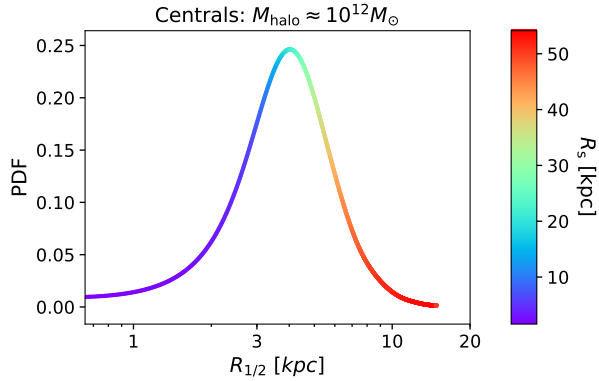
3.2.3   Profile co-evolution model

The "profile co-evolution" model is identical to the $R_\mathrm{vir}-$only model, but the scatter in $R_{1/2}$ at fixed R_{M_peak} is no longer purely stochastic, and is instead correlated with V_peak, defined as the value of V_max at z_{M_peak}. In this way, R_{M_peak} statistically determines the distribution of available sizes, but halos with extended dark matter profiles and large scale radii R_s host galaxies with above-average sizes, while halos with small R_s host smaller galaxies. See Figure 1 for a visual illustration of this residual correlation. We implement scatter correlations using the `halotools.empirical_models.conditional_abunmatch` function, which generalizes the Conditional Abundance Matching technique described in Hearin et al. (2014).

3.2.4 M_*-stripping model

As we will show in §4, the chief ingredient needed to recover the observed clustering properties of galaxies is that satellites need to be smaller than centrals of comparable halo mass. Thus it is natural to consider a class of models in which stellar mass is stripped from satellite galaxies after infall.

The basis of this class of models is the fitting function presented in Smith et al. (2016), which was calibrated by studying stellar mass loss in a suite of high-resolution hydrodynamical simulations. In this model, f_* quantifies the fraction of stellar mass lost as a function of f_DM, the amount of dark matter that has been stripped since infall:

Figure 1. Profile co-evolution model. In this model, the log-normal distribution of galaxy sizes is determined by $R_{\mathrm{M_{peak}}}$, the physical size of the halo virial radius at the time the halo reaches its peak mass. Thus log-normal the shape of this curve is determined by $\langle R_{1/2} | R_{\mathrm{M_{peak}}} \rangle = 0.0125 R_{\mathrm{M_{peak}}}$ with 0.2 dex of scatter. The profile co-evolution model further supposes that galaxy and halo profiles co-evolve across cosmic time. Thus within this log-normal distribution, for two halos with the same $R_{\mathrm{M_{peak}}}$, halos with larger scale radius at $z_{M_{peak}}$ will host galaxies with above-average $R_{1/2}$, and conversely for halos with smaller values of $R_{\mathrm{s, M_{peak}}}$.

$$f_* = 1 - \exp(-14.2 f_{\mathrm{DM}}) \qquad (4)$$

For f_{DM} we use the ratio of present-day subhalo mass divided by the peak mass, $M_{\mathrm{vir}}/M_{\mathrm{peak}}$. If we denote the post-stripping stellar mass as M_*', then we have $M_*' \equiv f_* M_*$, where M_* is given by Eq. 1. We then calculate the post-stripping radius by interpolating $\langle R'_{1/2} | M_*' \rangle$ directly from SDSS data.

4 RESULTS

In Figure 2 we show the scaling of galaxy size $R_{1/2}$ with M_*. The black curve enveloped by the gray bands show the scaling relation for our SDSS galaxy sample, while the blue curve shows the median relation $\langle R_{1/2} | M_* \rangle$ implied by the profile co-evolution model described in §3. This figure shows that models in which $R_{1/2} \propto R_{\mathrm{vir}}$ can naturally give rise to the characteristic curvature in the $\langle R_{1/2} | M_* \rangle$ relation, confirming the results in Kravtsov (2013) in a forward modeling context.

4.1 Size-Dependent Clustering

In Figure 3 we present new measurements of the $R_{1/2}$−dependence of projected galaxy clustering, $w_{\mathrm{p}}(r_{\mathrm{p}})$. Because galaxy clustering has well-known dependence upon M_* that is not the subject of this work, we wish to remove this influence and focus purely on the relationship between $R_{1/2}$ and $w_{\mathrm{p}}(r_{\mathrm{p}})$. To do so, we determine the value $\langle R_{1/2} | M_* \rangle$ by computing a sliding median of $R_{1/2}$, calculated using a window of width $N_{\mathrm{gal}} = 1000$. Each galaxy is categorized as either "large" or "small" according to whether it is above or below the median
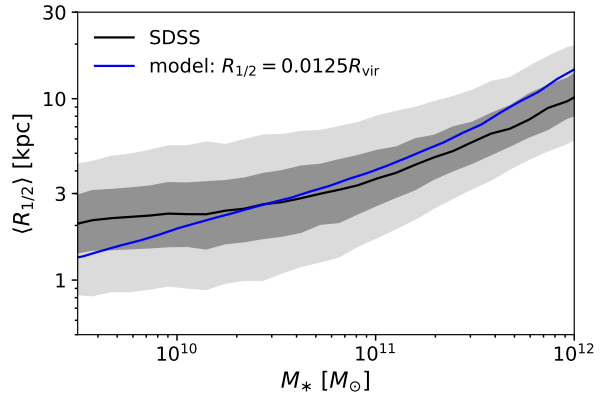
Figure 2. The black curve shows the median $R_{1/2} - M_*$ relation SDSS galaxies as measured in Meert et al. (2015). The two gray bands enveloping the black curve show the 50% and 90% percentile regions. The blue curve shows profile co-evolution model in which $\langle R_{1/2} | R_{\mathrm{vir}} \rangle = 0.0125 R_{\mathrm{vir}}$, as in Kravtsov (2013). This figure confirms that a linear relationship between R_{vir} and $R_{1/2}$, convolved against the nonlinear relationships between $R_{\mathrm{vir}}, M_{\mathrm{halo}}$ and M_*, predicts the characteristic curvature in the relation $\langle R_{1/2} | M_* \rangle$ over a wide range in mass.

value appropriate for its stellar mass. Using this technique, we stress that for any M_*−threshold sample, the SMF of the "large" and "small" subsamples are identical, by construction.

We measure $w_{\mathrm{p}}(r_{\mathrm{p}})$ separately for large and small subsamples for four different M_* thresholds, $M_* > 10^{9.75} M_\odot$, $M_* > 10^{10.25} M_\odot$, $M_* > 10^{10.75} M_\odot$, and $M_* > 10^{11.25} M_\odot$. We make the same measurements for each volume-limited M_*−threshold sample *without* splitting on size, giving us measurements $w_{\mathrm{p}}^{\mathrm{all}}, w_{\mathrm{p}}^{\mathrm{large}}$, and $w_{\mathrm{p}}^{\mathrm{small}}$ for each threshold sample. This allow us to compute the ratio $(w_{\mathrm{p}}^{\mathrm{large}} - w_{\mathrm{p}}^{\mathrm{small}})/w_{\mathrm{p}}^{\mathrm{all}}$, which we refer to as *the $R_{1/2}$ clustering ratio*. These ratios are the measurements appearing on the y-axis in each panel of Figure 3. Points with jackknife-estimated error bars show SDSS measurements, solid curves show the clustering ratios of model galaxies as predicted by the models described in §3.

The salient feature of the clustering ratio measurements is that they are negative: small galaxies cluster more strongly than large galaxies of the same stellar mass, a new result. This feature also holds true for galaxies predicted by the R_{vir}−only model. This result may be surprising, since $R_{1/2} \propto R_{\mathrm{vir}}$, halo mass $R_{\mathrm{vir}} \propto M_{\mathrm{halo}}^{1/3}$, and clustering strength increases with M_{vir}. Based on this simple argument, one would expect the opposite trend to the measurements shown here.

4.2 Central vs. Satellite Sizes

A straightforward resolution to the above puzzle is shown in Figure 4, which compares the $R_{1/2}$ distributions of central, satellite, and splashback galaxies with the same halo mass $M_{\mathrm{peak}} \approx 10^{12} M_\odot$. A "splashback" galaxy is defined as a present-day central that used to be a satellite, i.e., its main progenitor halo passed inside the virial radius of a larger halo at some point in its past history. On the
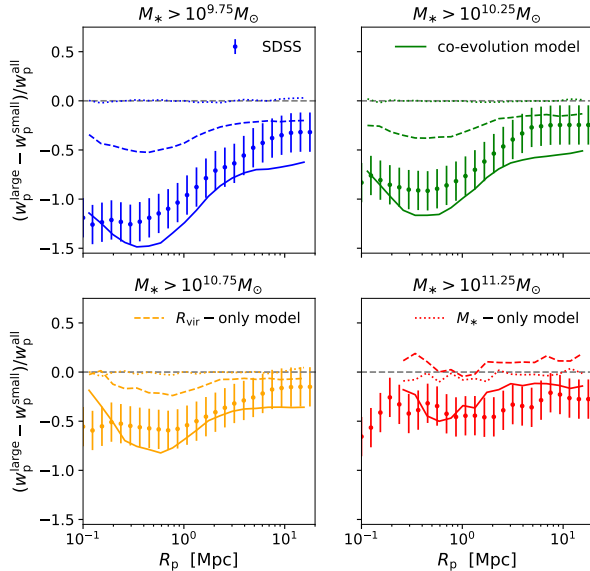
Figure 3. $R_{1/2}-$**dependence of galaxy clustering.** Points with error bars show new SDSS measurements of the $R_{1/2}-$dependence of projected galaxy clustering, $w_{\rm p}$. We define a galaxy as "large" or "small" according to whether it is above or below the median size for its stellar mass, so that in each panel, the stellar mass functions of the "large" and "small" subsamples are identical, as described in the text. The y-axis shows *clustering strength ratios,* so that, for example, a y-axis value of -0.5 corresponds to small galaxies being 50% more strongly clustered than large galaxies of comparable stellar mass. Each panel shows results separately for different volume-limited M_*-threshold samples; predictions of three different models are shown in each panel. See §3.2 for a description of each model.

other hand, we define a "true central" as a galaxy that has never been a satellite.

In the $R_{\rm vir}-$only model, satellite and splashback galaxies are smaller than centrals of the same halo mass due to the physical size of their halo being smaller at earlier times $z_{M_{\rm peak}}$. There are two distinct reasons why this feature results in small galaxies being more strongly clustered relative to larger galaxies of the same mass. First, satellite galaxies statistically occupy higher mass host halos that are more strongly clustered. In models where satellites are smaller than centrals, for any given M_*-threshold the "small" subsample will naturally have a higher satellite fraction, resulting in a negative clustering ratio as seen in SDSS data. Second, at fixed mass, halos of L_* galaxies that form earlier are more strongly clustered, a phenomenon commonly known as *halo assembly bias.* Since splashback halos are typically earlier forming than true centrals, then models where splashback halos host smaller-than-average galaxies will naturally predict negative clustering ratios.

In the M_*-only model, in which $R_{1/2}$ is statistically set by present-day stellar mass with no other dependencies. Thus neither of the above features are present: satellites and centrals of the same mass have no differences in size, and there is essentially no $R_{1/2}-$dependence to

galaxy clustering, in gross disagreement with our SDSS observations.

Much more successful is the profile co-evolution model shown with the solid curves in Figure 3. The strongly negative clustering ratios predicted by the assumption of profile co-evolution can be understood in terms of the same features discussed above. Figure 1 shows that at fixed $M_{\rm peak}$, larger galaxies will occupy halos with larger $R_{\rm s, M_{\rm peak}}$. Since halo concentration $c = R_{\rm vir}/R_{\rm s}$, this implies that high-concentration galaxies will tend to host smaller galaxies. Both satellites and splashback centrals have higher concentration than true centrals, another manifestation of assembly bias, and so the assumption of profile co-evolution naturally results in an enhancement of the clustering ratios beyond the $R_{\rm vir}-$only model.

4.3   Tidal Stripping and Orphan Satellites

We conclude this section with a discussion of Figure 5, which provides an estimate of how satellite mass stripping and orphan galaxies impact size-dependent clustering ratios. In the M_*-stripping model described in §3.2.4, satellites lose stellar mass in a way that mimics what is seen in high-resolution hydrodynamical simulations. This naturally results in an enhancement of size differences between satellites and centrals, and the dot-dashed curves in Figure 5 show that produces the expected sign of the effect on the clustering ratios. The magnitude of the effect, however is not strong enough to remedy the discrepancy of the clustering predictions of the $R_{\rm vir}-$only model. We discuss the physical implications of this result in §5.3.

The dotted curves in Figure 5 show results of the $R_{\rm vir}-$only model applied to a subhalo catalog that includes a orphan halos. As described in §**??**, orphan halos are subhalos that are no longer resolved by `Rockstar`, but whose evolution is tracked in a post-processing phase. We apply the same mass-stripping model described in §3.2.4 to this subhalo catalog that includes orphan halos, keep all model galaxies that retain more than half of their stellar mass, and show results for the $R_{\rm vir}-$only model applied to the resulting catalog.

As the orphan catalog has a higher satellite fraction than the standard subhalo catalog, the negative boost shown the dotted curves in Figure 5 is expected. Orphan halos also typically have earlier-than-average values of $z_{M_{\rm peak}}$ relative to ordinary subhalos, which also enhances differences between central and satellite sizes. See §5.3 for further discussion.

5   DISCUSSION

5.1   Progression from Backwards to Forward Modeling

Our results give an archetypal demonstration of the natural scientific progression from backwards to forward modeling. In backwards modeling, some mapping is applied to observed galaxies to estimate the values of model quantities such as halo mass. In Kravtsov (2013), the
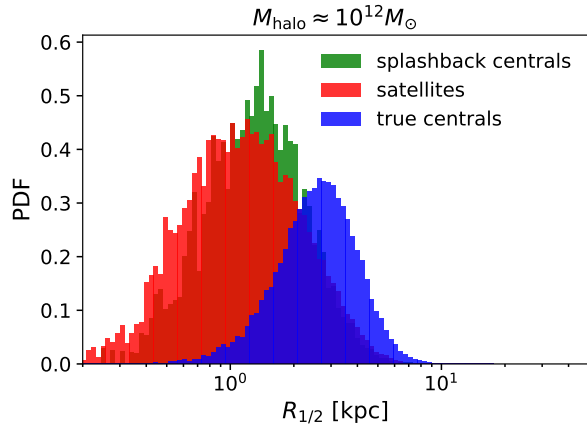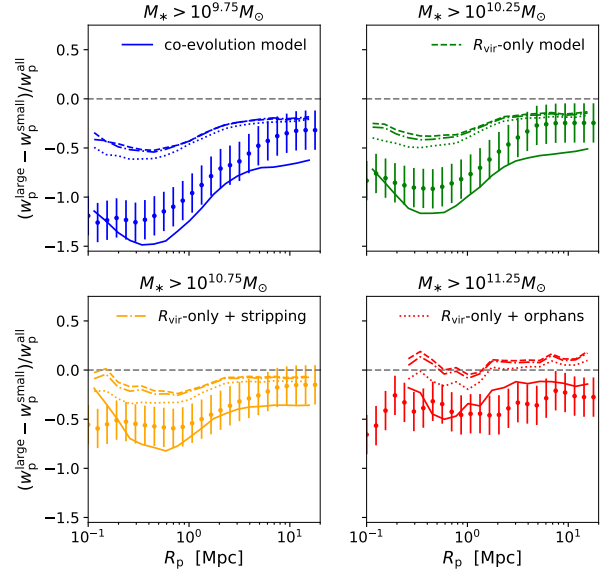
Figure 4. Relative sizes of centrals and satellites. In a narrow bin of halo mass $M_{\rm halo} = M_{\rm peak} \approx 10^{12} M_\odot$, we show the distribution of model galaxy sizes for different subpopulations galaxies, as predicted by the $R_{\rm vir}$−only model. The red histogram shows the sizes of satellites; the blue histogram shows host halos that have never passed inside the virial radius of a larger halo ("true centrals"); the green histogram host halos that were subhalos inside a larger at some point in their past history ("splashback halos"). In the $R_{\rm vir}$−only model, galaxy size is set by the *physical* size of the virial radius at the time the halo attains its peak mass, naturally resulting in smaller sizes for satellites and backsplash centrals relative to true centrals of the same $M_{\rm peak}$. The central/satellite size differences shown here are not enough to correctly predict the observed clustering ratios (see Figure 3), motivating the M_*−stripping and profile co-evolution models.

Figure 5. Impact of tidal stripping and orphan satellites. In all panels, the axes and points with error bars are the same as in Figure 3. The dashed curves ("$R_{\rm vir}$−only model") and solid curves ("profile co-evolution model") from Figure 3 are also shown for convenience. The dot-dashed curves show results for a model in which satellites lose mass after infall in a manner similar to what is seen in high-resolution hydrodynamical simulations, as described in §3.2.4. The dotted curve shows results for the same model, but including the effect of "orphan" galaxies that live in subhalos no longer resolved by `Rockstar`.

model quantities mapped onto galaxies are $M_{\rm halo}$ and $R_{\rm vir}$; another classic example of backwards modeling uses a group- or cluster-finding algorithm to assign $M_{\rm halo}$ to observed galaxies (e.g., Berlind et al. 2006; Yang et al. 2005; Rykoff et al. 2014). Once the observed galaxies have been supplemented with model variables, then the relations of the galaxy-halo connection can be inferred, for example, by calculating quantities such as the mean stellar mass or quiescent fraction as a function of halo mass (e.g., Yang et al. 2005; Weinmann et al. 2006).

In forward modeling, the direction of inference is turned around: a mapping is instead applied to the model quantities such as $M_{\rm halo}$. In the case of `Halotools`, this transforms a cosmological simulation into a synthetic galaxy catalog that can be directly compared with observations. This enables a richer quantitative study of modeling hypotheses relative to backwards modeling. For example, Figure 3 shows how forward modeling allows us to exploit galaxy clustering measurements to quantitatively test whether $M_{\rm halo}$ or M_* is the statistical regulator of galaxy size. The ability disentangle coupled variables such as M_* and $M_{\rm halo}$ is just one example of this advantage of forward modeling. Another example is illustrated in Figure 4, in which we explore the role of splashback halos in setting galaxy size. In our forward modeling approach, this is entirely straightforward; in backwards modeling, such an investigation would not even be possible without introducing additional modeling ingredients.

Backwards modeling the galaxy-halo connection is useful for generating hypotheses and motivating func-

tional forms. Forward modeling becomes necessary when the problem at hand becomes encumbered by multiple relevant variables, as is the case with galaxy size. Forward modeling also makes it possible to conduct rigorous Bayesian inference, which we consider to be the next natural step in the progression described here (see §5.4 for further discussion).

5.2 Relation to Previous Work

Backwards modeling methods have been used extensively in the literature to gain insight into the relationship between galaxy mass, size, and environment. Broadly speaking, such studies proceed by using a galaxy group catalog to classify observed galaxies as centrals or satellites, and estimate their halo mass. Employing such methods, several analyses of the Yang et al. (2005) group catalog have found that the $M_* − R_{1/2}$ relation of early-type galaxies exhibits weak, if any, environmental dependence (Huertas-Company et al. 2013; Shankar et al. 2014).

A direct comparison to this finding is not possible because the conclusions drawn here pertain to M_*−limited galaxy samples. As discussed in Spindler & Wake (2017), it is entirely possible that size differences between centrals and satellites of the same mass can be accounted for by mutual covariance with an additional variable such as star formation rate or morphological type (see also Lilly & Carollo 2016, for an explicit demonstration of this scenario). In particular, suppose that centrals and

satellites of the same mass have different early-type fractions as indicated in Weinmann et al. (2006), and that early- and late-type galaxies exhibit universal, but distinct, $M_* - R_{1/2}$ relations. In such a case, centrals would be larger than satellites of the same mass, even though *early-type centrals* would have the same sizes as *early-type satellites*.

As discussed in §5.4, we are currently pursuing follow-up work in which we jointly model morphological type together with galaxy size. We consider forward modeling methods to be a requisite for progress on this issue, not only to properly handle the multi-dimensional nature of the problem, but also to rigorously treat systematic errors that plague inference based on galaxy group catalogs (see Campbell et al. 2015, for a thorough discussion).

Our approach is closely aligned with the methods used in Somerville et al. (2017), who studied the empirical modeling features that are necessary to recover the tight scatter in the observed $\langle R_{1/2}|M_* \rangle$ relation. By building models where $R_{1/2}$ is set by halo spin λ_{halo}, the authors in Somerville et al. (2017) found that the level of intrinsic scatter about $\langle \lambda_{\text{halo}}|M_{\text{halo}} \rangle$ in dark matter halos is at least as large as the scatter about $\langle R_{1/2}|M_* \rangle$ seen in observed galaxies. Since the latter necessarily receives an additional contribution from measurement error, this implies some tension with the common semi-analytic modeling assumption that λ_{halo} scales with $R_{1/2}$. As noted in Somerville et al. (2017), tension in the level of scatter cannot be used to directly test the $\lambda_{\text{halo}} \propto R_{1/2}$ assumption because the physical motivation for this correlation is largely limited to disk galaxies (Mo et al. 1998). This tension is not present in our approach for the reason that the level of scatter is simply a modeling parameter in our approach, and we make no attempt to uncover the physical origin of this scatter. However, our fiducial value choice was motivated by Somerville et al. (2017), and in the ongoing follow-up work discussed in §5.4 we will systematically test the large-scale structure implications of the assumption that $\lambda_{\text{halo}} \propto R_{1/2}^{\text{disk}}$.

5.3   Implications for Satellite Evolution

Explanations for differences between the sizes of centrals and satellites fall into two categories, depending on whether or not the processes involved are specific to post-infall physics. One obvious possibility that could lead to satellites being smaller than centrals of the same M_{halo} is tidal stripping: satellite M_* could decrease after infall, reducing the size of the galaxy. We have quantitatively tested this possibility with the M_*-stripping model described in §3.2.4. Comparing the clustering ratios of the "$R_{\text{vir}}-$only" and "$R_{\text{vir}}+$stripping" models in Figure 5, we find that the impact of physically motivated levels of satellite mass loss cannot explain the $R_{1/2}-$dependent clustering seen in SDSS. As an additional test, we reach the same conclusion when implementing an even more extreme M_*-stripping model in which stellar mass loss is linearly proportional to halo mass loss (Watson et al. 2012, Model 1).

Both the $R_{\text{vir}}-$only model and the co-evolution model (see sections §3.2.2 and §3.2.3, respectively) rep-

resent explanations falling into the second category, in which centrals and satellites of the same M_{halo} have different sizes due to their different evolutionary paths *prior* to satellite infall. For these models the story is very different: the influence of pre-infall co-evolution has a dramatic impact on $R_{1/2}-$clustering ratios, with a scale- and M_*-dependence that closely resembles the SDSS signal. Quantitative constraints on the true level of this co-evolution effect will be a natural outcome of the program described in §5.4, but it is already clear that the long term co-evolution of galaxy and halo profiles should be considered a critical ingredient to models for the distribution of galaxy sizes across the cosmic web.

5.4   Future Directions for Empirical Modeling of Galaxy Size

5.4.1   *Jointly modeling $\langle M_*|M_{\text{halo}} \rangle$*

The scope of this work is to serve as a pilot study in which we identify the chief ingredients that can influence the $R_{1/2}-$dependence of galaxy clustering. While our results suggest that our profile co-evolution model is giving the correct qualitative picture, our implementation cannot be correct in quantitative detail: the clustering ratios of this model shown in Figure 3 are too strong relative to SDSS. This likely indicates that the central/satellite size differences depicted in Figure **??** are too strong, and that the correlation strength in the real Universe is intermediate, rather than maximal. Our publicly available code that reproduces our results already contains the machinery to implement continuously variable levels of $R_{1/2} - R_{\text{s}, \text{M}_{\text{peak}}}$ correlation strength, but for present purposes we have chosen to focus on exploring bracketing cases of qualitative ingredients, rather than fine-tuning the model.

In addition to varying the $R_{1/2} - R_{\text{s}, \text{M}_{\text{peak}}}$ correlation strength, there are other factors that should be included in any proper likelihood analysis of galaxy size models. In particular, since the clustering signal is strongly influenced by differences between centrals and satellites, then the satellite fraction $F_{\text{sat}}(M_*)$ plays an important role in $R_{1/2}-$dependent clustering. For fixed scaling relations $\langle R_{1/2}^{\text{cens}}|M_{\text{halo}} \rangle$ and $\langle R_{1/2}^{\text{sats}}|M_{\text{halo}} \rangle$, models with different satellite fractions will exhibit different $R_{1/2}-$clustering ratios because $F_{\text{sat}}(M_*)$ controls the relative weighting of the two populations. Our results based on the catalog that includes orphan subhalos (Figure 5, dotted curves) give an explicit demonstration of the impact of $F_{\text{sat}}(M_*)$ on the clustering signal.

On the one hand, this degeneracy with the satellite fraction is unfortunate, because it means galaxy sizes does not leave a pure and unique signature on $R_{1/2}-$clustering ratios. However, this can also be viewed as an opportunity to extract tighter constraints on $F_{\text{sat}}(M_*)$, which are sorely needed to discriminate between competing models. For this problem, the path forward is clear. Traditional galaxy clustering is already being used to validate and/or fit models of the stellar-to-halo-mass relation (e.g., Moster et al. 2013; Behroozi et al. 2013; Lehmann et al. 2017). Based on our results, we advocate that empirical models for $\langle M_*|M_{\text{halo}} \rangle$ be supplemented with additional model ingredients for $\langle R_{1/2}|M_* \rangle$,

and that the parameters of composite model are jointly constrained by measurements of the stellar mass function, M_*−dependent clustering, and $R_{1/2}$−dependent clustering. Careful attention will need to be paid to the orphan subhalo population, which is well-known to have a significant impact on small-scale clustering (Guo & White 2014; Campbell et al. 2017).

5.4.2   *Jointly modeling morphology*

Previous studies of large-scale structure measured by SDSS indicate that morphological type has a significant influence on galaxy clustering (Skibba et al. 2009). Because morphology is covariant with star-formation rate (SFR) and $R_{1/2}$, accounting for this dependency will require simultaneously modeling each of these variables.

To date, such a task has been beyond the purview of empirical models. However, recent progress has substantially improved the richness of the predictions that galaxy-halo models are capable of (Becker 2015; Cohn 2017; Moster et al. 2017). This new class of models has accomplished this by empirically modeling star-formation rate, and then integrating SFR along each halo's history to differentially build up stellar mass.[1] We are currently extending this new class of models to differentially build up the disk and bulge across cosmic time. In addition to treating the multi-dimensional nature of the problem, this also allows us to use data at higher redshift in a self-consistent manner.

6   CONCLUSIONS

We have presented new measurements of the dependence of galaxy clustering upon galaxy size, and used `Halotools` to identify the basic ingredients that influence the signal. We conclude with a brief summary of our primary findings:

(i) Small galaxies cluster more strongly than large galaxies of the same stellar mass, a new result. Differences between the clustering of small and large galaxies increase on small scales $R \lesssim 1\mathrm{Mpc}$, and decrease with stellar mass.

(ii) The most important ingredient influencing this signal is the relative size of central and satellite galaxies. The magnitude, scale-dependence, and M_*−dependence of $R_{1/2}$−dependent clustering provides strong evidence that satellite galaxies are smaller than central galaxies of the same halo mass.

(iii) Simple empirical models in which $R_{1/2}$ is regulated by M_*, rather than M_{halo}, are grossly discrepant with the observed clustering signal.

(iv) The "co-evolution model" shown with solid curves in Figure 3 exhibits a clustering signal that is strikingly similar to that seen in SDSS. In this model, $R_{1/2}$ is statistically regulated by $R_{\mathrm{M_{peak}}}$, the halo virial radius R_{vir} at

---

[1] This approach towards predicting M_* is thus mathematically akin to traditional semi-analytic models, although Becker (2015) have retained the empirical modeling prioritization of simplicity and computational efficiency.

the time of peak halo mass, with strong residual correlation with $R_{\mathrm{s,M_{peak}}}$, the halo scale radius at that time. The success of this model supports the notion that galaxy and halo profiles co-evolve across many Gyr of cosmic time.

Our results can be treated as a boundary condition for more complex and fine-grained models of galaxy size, such as semi-analytic models and hydrodynamical simulations. We view the present work as a pilot study that motivates a Bayesian inference program to tightly constrain the galaxy size-halo connection with forward modeling techniques, in direct analogy to the literature on the stellar-to-halo-mass relation. Our publicly available code provides a simple means for cosmological surveys to generate synthetic galaxy populations with realistic sizes across the cosmic web.

REFERENCES

Ahn C. P., Alexandroff R., Allende Prieto C., Anders F., Anderson S. F., Anderton T., Andrews B. H., Aubourg É., Bailey S., Bastien F. A., et al. 2014, ApJS , 211, 17

Astropy Collaboration Robitaille T. P., Tollerud E. J., Greenfield P., Droettboom M., Bray E., Aldcroft T., et al., 2013, AAP, 558, A33

Becker M. R., 2015, ArXiv:1507.03605

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ , 770, 57

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ , 762, 109

Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013, ApJ , 763, 18

Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS , 149, 289

Berlind A. A., Frieman J., Weinberg D. H., et al., 2006, ApJS , 167, 1

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MN-RAS , 436, 697

Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2014, MNRAS , 443, 874

Bottrell C., Torrey P., Simard L., Ellison S. L., 2017, MNRAS , 467, 2879

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS , 351, 1151

Bryan G. L., Norman M. L., 1998, ApJ , 495, 80

Campbell D., van den Bosch F. C., Hearin A., Padmanabhan N., Berlind A., Mo H. J., Tinker J., Yang X., 2015, MNRAS , 452, 444

Campbell D., van den Bosch F. C., Padmanabhan N., Mao Y.-Y., Zentner A. R., Lange J. U., Jiang F., Villarreal A., 2017, ArXiv:1705.06347

Cohn J. D., 2017, MNRAS , 466, 2718

Croton D. J., 2013, PASA, 30, e052

Dutton A. A., van den Bosch F. C., Faber S. M., Simard L., Kassin S. A., Koo D. C., Bundy K., Huang J., Weiner B. J., Cooper M. C., Newman J. A., Mozena M., Koekemoer A. M., 2011, MNRAS , 410, 1660

Guo Q., White S., 2014, MNRAS , 437, 3228

Guo Y., McIntosh D. H., Mo H. J., Katz N., Van Den Bosch F. C., Weinberg M., Weinmann S. M., Pasquali A., Yang X., 2009, MNRAS , 398, 1129

Hearin A., Campbell D., Tollerud E., et al., 2016, ArXiv e-prints

Hearin A. P., Watson D. F., Becker M. R., Reyes R., Berlind A. A., Zentner A. R., 2014, MNRAS , 444, 729

Hopkins P. F., Bundy K., Croton D., Hernquist L., Keres D., Khochfar S., Stewart K., Wetzel A., Younger J. D., 2010, ApJ , 715, 202

Huang K.-H., Fall S. M., Ferguson H. C., van der Wel A., Grogin N., Koekemoer A., Lee S.-K., Pérez-González P. G., Wuyts S., 2017, ApJ , 838, 6

Huang S., Ho L. C., Peng C. Y., Li Z.-Y., Barth A. J., 2013, ApJ , 766, 47

Huertas-Company M., Mei S., Shankar F., Delaye L., Raichoor A., Covone G., Finoguenov A., Kneib J. P., Le F. O., Povic M., 2013, MNRAS , 428, 1715

Huertas-Company M., Shankar F., Mei S., Bernardi M., Aguerri J. A. L., Meert A., Vikram V., 2013, ApJ , 779, 29

Jiang F., van den Bosch F. C., 2014, ArXiv e-prints

Jones E., Oliphant T., Peterson P., et al., 2001-2016, http://www.scipy.org

Kauffmann G., Heckman T. M., White S. D. M., et al., 2003, MNRAS , 341, 33

Kawamata R., Ishigaki M., Shimasaku K., Oguri M., Ouchi M., 2015, ApJ , 804, 103

Khochfar S., Silk J., 2006, ApJL , 648, L21

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ , 740, 102

Kravtsov A. V., 2013, ApJL , 764, L31

Lange R., Driver S. P., Robotham A. S. G., et al., 2015, MNRAS , 447, 2603

Lehmann B. V., Mao Y.-Y., Becker M. R., Skillman S. W., Wechsler R. H., 2017, ApJ , 834, 37

Lilly S. J., Carollo C. M., 2016, ApJ , 833, 1

LSST Science Collaboration Abell P. A., Allison J., Anderson S. F., Andrew J. R., Angel J. R. P., Armus L., Arnett D., Asztalos S. J., Axelrod T. S., et al. 2009, ArXiv e-prints

Meert A., Vikram V., Bernardi M., 2013, MNRAS , 433, 1344

Meert A., Vikram V., Bernardi M., 2015, MNRAS , 446, 3943

Mo H. J., Mao S., White S. D. M., 1998, MNRAS , 295, 319

Moster B. P., Naab T., White S. D. M., 2013, MNRAS , 428, 3121

Moster B. P., Naab T., White S. D. M., 2017, ArXiv e-prints

Planck Collaboration Ade P. A. R., Aghanim N., Arnaud M., Ashdown M., Aumont J., Baccigalupi C., Banday A. J., Barreiro R. B., Bartlett J. G., et al. 2016, AAP, 594, A13

Riebe K., Partl A. M., Enke H., Forero-Romero J., Gottlöber S., Klypin A., Lemson G., Prada F., Primack J. R., Steinmetz M., Turchaninov V., 2013, Astronomische Nachrichten, 334, 691

Robertson B. E., Banerji M., Cooper M. C., Members of the LSST Galaxies Science Collaboration 2017, ArXiv e-prints

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS , 462, 893

Rykoff E. S., Rozo E., Busha M., et al., 2014, ApJ , 785, 104

Shankar F., Mei S., Huertas-Company M., Moreno J., Fontanot F., Monaco P., Bernardi M., Cattaneo A., Sheth R., Licitra R., Delaye L., Raichoor A., 2014, MNRAS , 439, 3189

Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS , 343, 978

Shibuya T., Ouchi M., Harikane Y., 2015, ApJS , 219, 15

Sinha M., Garrison L., , 2017, Corrfunc: Blazing fast correlation functions on the CPU, Astrophysics Source Code Library

Skibba R. A., Bamford S. P., Nichol R. C., Lintott C. J., Andreescu D., Edmondson E. M., Murray P., Raddick M. J., Schawinski K., Slosar A., Szalay A. S., Thomas D., Vandenberg J., 2009, MNRAS , 399, 966

Smith R., Choi H., Lee J., Rhee J., Sanchez-Janssen R., Yi S. K., 2016, ApJ , 833, 109

Somerville R. S., Behroozi P., Pandya V., et al., 2017, ArXiv e-prints

Spindler A., Wake D., 2017, MNRAS , 468, 333

Trujillo I., Rudnick G., Rix H.-W., Labbé I., Franx M., Daddi E., van Dokkum P. G., Förster Schreiber N. M., Kuijken K., Moorwood A., Röttgering H., van der Wel A., van der Werf P., van Starkenburg L., 2004, ApJ , 604, 521

Van Der Walt S., Colbert S. C., Varoquaux G., 2011, ArXiv:1102.1523

van der Wel A., Franx M., van Dokkum P. G., et al., 2014, ApJ , 788, 28

Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, MNRAS , 409, 1379

Watson D. F., Berlind A. A., Zentner A. R., 2012, ApJ , 754, 90

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS , 366, 2

Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005, MNRAS , 358, 217

Yang X., Mo H. J., van den Bosch F. C., Jing Y. P.,

2005, MNRAS , 356, 1293
Zhang Y., Yang X., 2017, ArXiv e-prints