

Clustering Constraints on the Relative Sizes of Central and Satellite Galaxies

Andrew Hearin¹, Peter Behroozi², Andrey Kravtsov³, Benjamin Moster⁴

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*

1 November 2017

ABSTRACT

We place empirical constraints on the connection between dark matter halos and galaxy half-light radii, $R_{1/2}$. Low-redshift SDSS measurements 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. We use this model to predict how galaxy lensing, $\Delta\Sigma$, should depend on $R_{1/2}$ for M_* -complete samples. Other simple empirical models fail the clustering test, such as models in which galaxy size is related to stellar mass alone; these failures persist even when accounting for possible effects from satellite stripping and orphan galaxies. Our results suggest that the relative size of centrals and satellites is predetermined at the time of satellite infall, and that a remarkably simple galaxy–halo scaling relation emerges from the complex physics regulating galaxy size.

1 INTRODUCTION

In the Λ CDM 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 use stellar mass measurements made available through the MPA-JHU catalog (Kauffmann et al. 2003; Brinchmann et al. 2004), allowing us to define volume-limited samples of galaxies according to the same completeness cuts used in Behroozi et al. (2015) (see Figure 2).

We calculate two-point clustering w_p of our SDSS galaxy sample using line-of-sight projection of $\pi_{\max} = 20\text{Mpc}$ using the `correl` program in `UniverseMachine`. **PSB: Please fill in details about `correl` here.**

As the bedrock of our modeling, we use the pub-

licly available¹ 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).

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). **PSB: Make sure you agree with my description of your orphan treatment.**

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 $H_0 = 67.8 \text{ km/s} \equiv 100h \text{ km/s}$, 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 .

2.1 Classifying large vs. small galaxies

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 $w_p(r_p)$. 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 $N_{\text{gal}} = 1000$. Each galaxy is categorized as either “large” or “small” according to whether it is above or below the median 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. See Figure 1.

3 GALAXY-HALO MODEL

3.1 Stellar mass model

We map M_* onto subhalos using deconvolution abundance matching on the orphan-supplemented Bolshoi-Planck subhalo catalog, as described in detail in the Appendix.

¹ http://www.slac.stanford.edu/~behroozi/BPlanck_Hlists

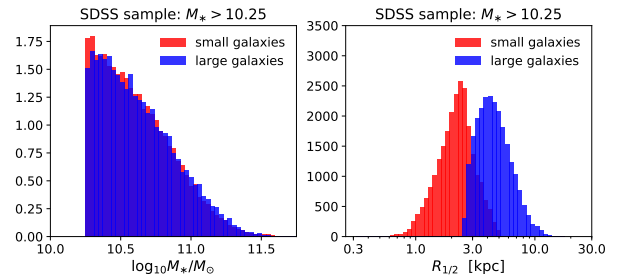


Figure 1. Definition of “small” and “large” galaxies. For a volume-limited SDSS galaxy sample defined by $M_* > 10^{10.25} M_\odot$, we visually demonstrate how we classify galaxies into “small” and “large” subsamples. As described in detail in §2.1, we compute the median value $\langle R_{1/2} | M_* \rangle_{\text{median}}$ using a sliding window with a width of 1000 galaxies at each value of M_* . The *left panel* shows a histogram of the stellar masses, confirming that our method yields identical stellar mass functions for the two subsamples. The *right panel* shows histograms of $R_{1/2}$ for the two subsamples, which partially overlap due to the finite range of M_* in the volume-limited sample.

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 $\langle R_{1/2} | M_* \rangle_{\text{median}}$. To implement this model, for simplicity we directly tabulate $\langle R_{1/2} | M_* \rangle_{\text{median}}$ directly from the data, rather than pursue a parametric form (see, e.g., Zhang & Yang 2017).

3.2.2 R_{vir} -only model

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

$$R_{1/2} = 0.01 R_{\text{vir}} \quad (1)$$

For the virial radius of halos and subhalos, we use $R_{M_{\text{peak}}}$, the value of R_{vir} in physical units of kpc measured at the time of peak subhalo mass, defined by

$$M_{\text{peak}} \equiv \frac{4\pi}{3} R_{M_{\text{peak}}}^3 \Delta_{\text{vir}}(z_{M_{\text{peak}}}) \rho_m(z_{M_{\text{peak}}}), \quad (2)$$

where for $\Delta_{\text{vir}}(z_{M_{\text{peak}}})$ we use the fitting function to the “virial” definition used in Bryan & Norman (1998). For the model we refer to as the “ R_{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 M_* -stripping model

As we will show in §4, the chief ingredient needed to recover the observed clustering properties of galaxies is that

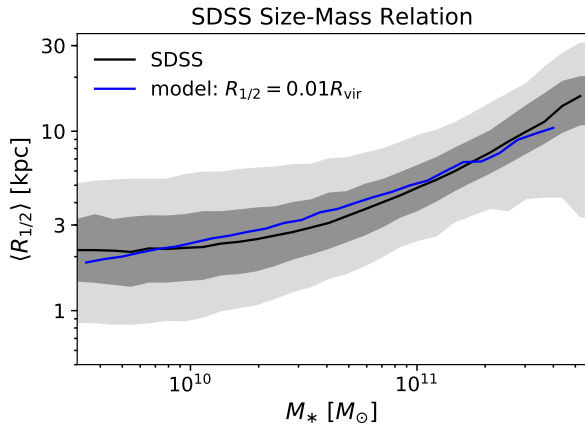


Figure 2. The black curve shows the median size-mass relation of SDSS galaxies as measured in Meert et al. (2015). The two gray bands enveloping the black curve show the 50% and 90% percentile regions. The blue curve shows R_{vir} -only model in which $\langle R_{1/2} | R_{\text{vir}} \rangle_{\text{median}} = 0.01 R_{\text{vir}}$. This figure confirms that a linear relationship between R_{vir} and $R_{1/2}$, convolved against the nonlinear relationships between R_{vir} , M_{halo} and M_* , predicts the characteristic curvature in the relation $\langle R_{1/2} | M_* \rangle_{\text{median}}$ over a wide range in mass.

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:

$$f_* = 1 - \exp(-14.2 f_{\text{DM}}) \quad (3)$$

For f_{DM} we use the ratio of present-day subhalo mass divided by the peak mass, $M_{\text{vir}}/M_{\text{peak}}$. If we denote the post-stripping stellar mass as M'_* , then we have $M'_* \equiv f_* M_*$, where M_* is given by Eq. ???. 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 R_{vir} -only model described in §3. This figure shows that models in which $R_{1/2} \propto R_{\text{vir}}$ 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.

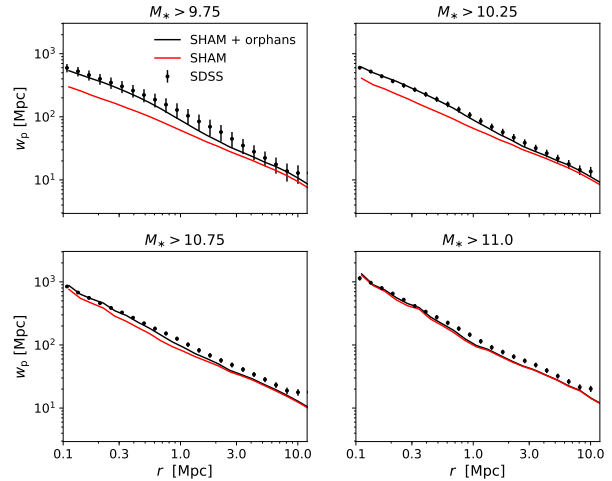


Figure 3. SHAM + orphan clustering predictions. Using the abundance matching methods described in §3.1, we compare the projected clustering of mock vs. SDSS galaxies. Each panel shows the comparison for a volume-limited sample defined by a different M_* -threshold. Black points with error bars show SDSS measurements; solid red (black) curves show the abundance matching prediction including (excluding) the effect of orphan subhalos (see Appendix A). Including orphans mitigates the discrepancy in the clustering predicted by traditional, M_{peak} -based SHAM based, though mild tension remains for $M_* \gtrsim 10^{10.75} M_\odot$.

4.1 Size-Dependent Clustering

In Figure 5 we present new measurements of the $R_{1/2}$ -dependence of projected galaxy clustering, $w_p(r_p)$.

We measure $w_p(r_p)$ 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_p^{all} , w_p^{large} , and w_p^{small} for each threshold sample. This allows us to compute the ratio $(w_p^{\text{large}} - w_p^{\text{small}})/w_p^{\text{all}}$, 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 5. 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 $R_{1/2} \propto R_{\text{vir}}$, halo mass $R_{\text{vir}} \propto M_{\text{halo}}^{1/3}$, 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 6, which compares the $R_{1/2}$ distributions of central, satellite, and splashback galaxies with the same halo

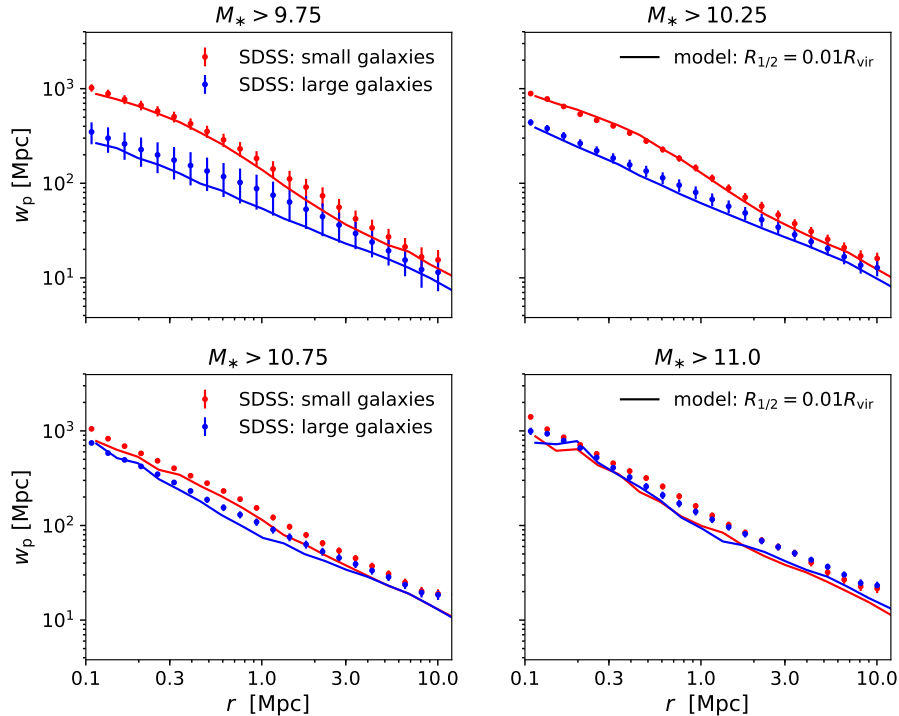


Figure 4. $R_{1/2}$ –dependence of galaxy clustering. Red and blue points with error bars show our SDSS measurements of the clustering of small and large galaxies, respectively. For each volume-limited sample of M_* –complete galaxies, the small and large subsamples have identical stellar mass functions, as shown in Figure 1. Solid curves show the clustering predictions of the R_{vir} –only model described in §3.2.2. The R_{vir} –only model inherits the shortcoming of ordinary abundance matching at $M_* \gtrsim 10^{10.75} M_\odot$, although the model faithfully captures the *relative* clustering of small vs. large galaxies, as shown in Figure 5.

mass $M_{\text{peak}} \approx 10^{12} M_\odot$. 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 other hand, we define a “true central” as a galaxy that has never been a satellite.

In the R_{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 $z_{M_{\text{peak}}}$. 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 5. 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 ?? shows that at fixed M_{peak} , larger galaxies will occupy halos with larger $R_{s, M_{\text{peak}}}$. Since halo concentration $c = R_{\text{vir}}/R_s$, 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_{vir} –only model.

4.3 Satellite Mass Loss

We conclude this section with a discussion of Figure 9, which provides an estimate of how satellite mass stripping and orphan galaxies impact size-dependent clustering ratios. In the M_* –stripping model described in §3.2.3, 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 9 show that produces the expected sign of the effect on the clustering ratios. The magnitude of the ef-

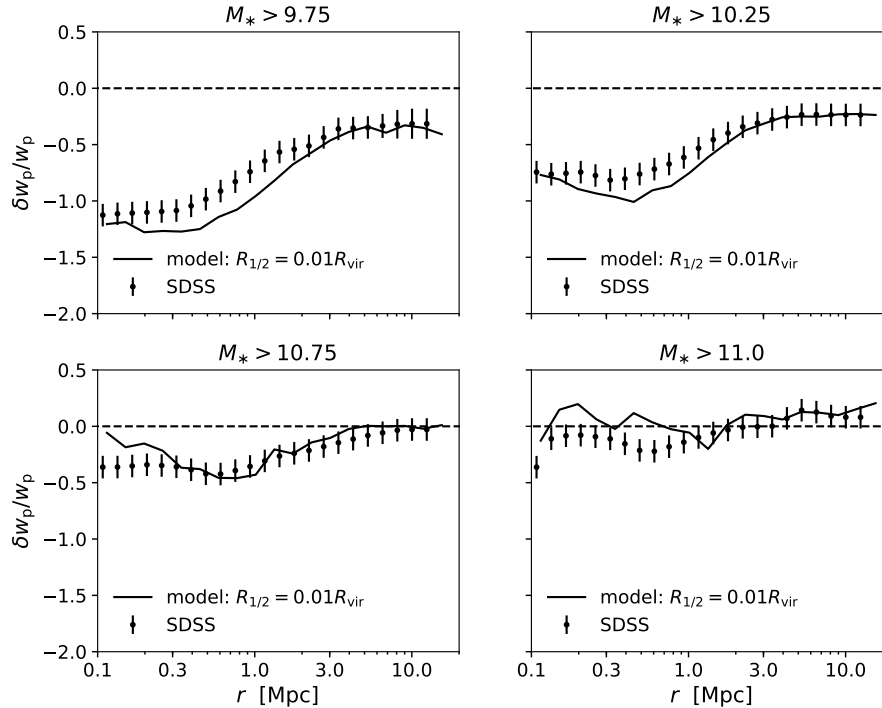


Figure 5. $R_{1/2}$ -dependence of galaxy clustering: clustering ratios. Closely related to Figure 4, the y-axes show *clustering strength ratios*, defined as $(w_p^{\text{large}} - w_p^{\text{small}})/w_p^{\text{all}}$. Thus a y-axis value of -0.5 corresponds to small galaxies being 50% more strongly clustered than large galaxies of the same stellar mass. Solid curves show the clustering ratio predictions of the R_{vir} -only model described in §3.2.2. Normalizing the measurements and predictions by w_p^{all} scales away the shortcoming of ordinary abundance matching at high stellar mass (see Figure 3), highlighting the successful prediction of the R_{vir} -only model for the $R_{1/2}$ -dependence of galaxy clustering.

fect, however is not strong enough to remedy the discrepancy of the clustering predictions of the R_{vir} -only model. We discuss the physical implications of this result in §5.3.

The dotted curves in Figure 9 show results of the R_{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 §3.2.3 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_{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 9 is expected. Orphan halos also typically have earlier-than-average values of $z_{M_{\text{peak}}}$ relative to ordinary subhalos, which also enhances differences between central and satellite sizes. See §5.3 for further discussion.

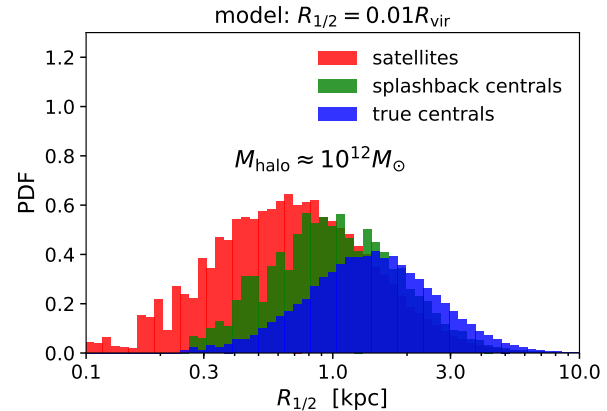


Figure 6. Relative sizes of centrals and satellites. In a narrow bin of halo mass $M_{\text{halo}} \equiv M_{\text{peak}} \approx 10^{12} M_{\odot}$, we show the distribution of model galaxy sizes for different subpopulations galaxies, as predicted by the R_{vir} -only model. The red histogram shows 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 centrals”). In the R_{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 splashback centrals relative to true centrals of the same M_{halo} .

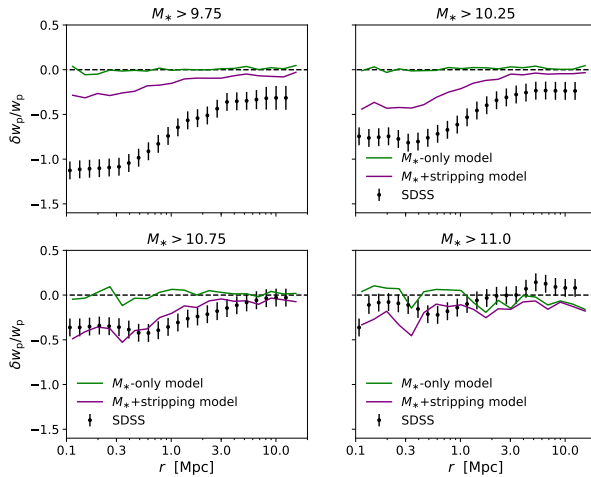


Figure 7. Impact of tidal stripping. In all panels, the axes and points with error bars are the same as in Figure 5. The solid green curves show the prediction of the M_* -only model, which is in gross tension with the data due to satellites having the same size as centrals of the same mass. The solid purple 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 §3.2.3. This produces satellites that are smaller than centrals, but the effect is too mild to correctly capture the observed clustering. Evidently, satellite-specific mass stripping plays a sub-dominant role in setting the relative size of centrals and satellites.

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 model quantities mapped onto galaxies are M_{halo} and R_{vir} ; another classic example of backwards modeling uses a group- or cluster-finding algorithm to assign M_{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_{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 5 shows how forward modeling allows us to exploit galaxy clustering measurements to quantitatively test whether M_{halo} or M_* is the statistical regulator of galaxy size. The ability disentangle coupled variables such as M_* and M_{halo} is just one example of this

advantage of forward modeling. Another example is illustrated in Figure 6, in which we explore the role of splash-back 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 functional 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

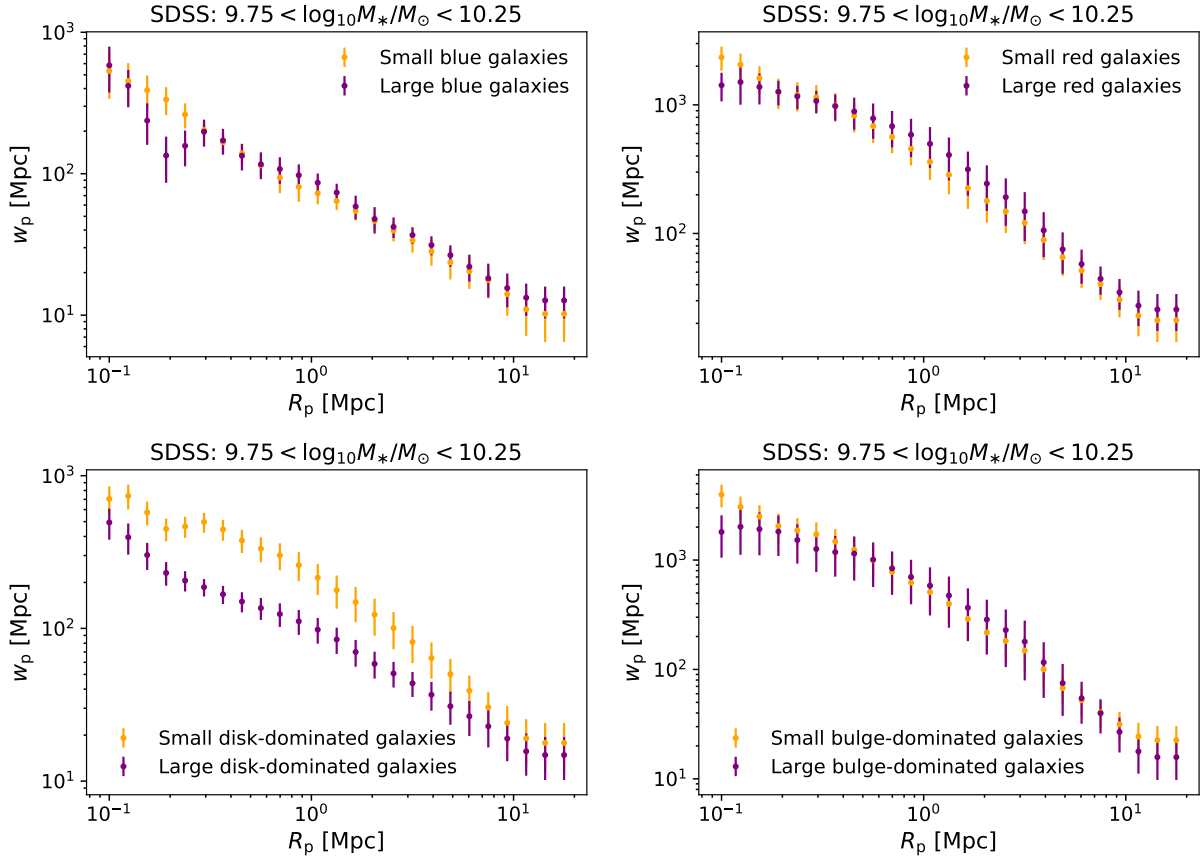
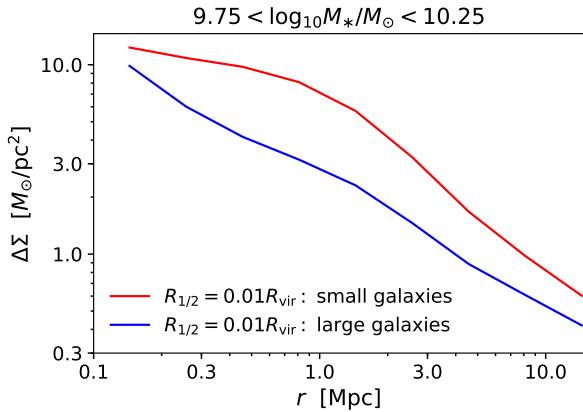


Figure 8. Some caption.


 Figure 9. Prediction for $R_{1/2}$ -dependence of galaxy lensing.

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

5.4 Future Directions for Empirical Modeling of Galaxy Size

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

5.4.2 Jointly modeling morphology

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 $R \lesssim 1\text{Mpc}$, and decrease with stellar mass.
- (ii) The most important ingredient influencing this sig-

nal 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.

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.

ACKNOWLEDGMENTS

APH thanks John Baker for the *Toejam & Earl* soundtrack. Thanks also to Frank van den Bosch, Doug Watson, and Risa Wechsler for thoughtful feedback at various stages of the development of this work, and to Faustin Carter and Sebastian Bocquet for sharing their python expertise.

We thank the *Astropy* developers for the package-template (Astropy Collaboration et al. 2013), and *NumPy* (Van Der Walt et al. 2011), *SciPy* (Jones et al. 2016), *IPython*, *Matplotlib*, and *GitHub* for their extremely useful free software.

This research was supported in part by the National Science Foundation under Grant No. NSF PHY11-25915. Work done at Argonne National Laboratory was supported under the DOE contract DE-AC02-06CH11357.

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
- 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
- Behroozi P. S., Zhu G., Ferguson H. C., Hearin A. P., Lotz J., Silk J., Kassin S., Lu Y., Croton D., Somerville R. S., Watson D. F., 2015, *MNRAS*, 450, 1546
- 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, *MNRAS*, 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*
- 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 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
- 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
- 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
- Planck Collaboration Ade P. A. R., Aghanim N., Ar-

- naud 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
- 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
- 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

APPENDIX: TREATMENT OF ORPHAN SUBHALOS