Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 19 September 2017 (MN LaTeX style file v2.2)

Empirical Modeling Galaxy Size Throughout the Cosmic Web

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*

19 September 2017

ABSTRACT

We derive empirical modeling constraints on the connection between dark matter halos and the half-light radius $R_{1/2}$ of galaxies. Using forward-modeling techniques based on `Halotools`, we confirm previous results in Kravtsov (2013) that $R_{1/2}$ is well-described by a linear scaling relation with halo virial radius. Novel to this work, we use new SDSS measurements of the $R_{1/2}-$dependence of galaxy clustering to test this modeling assumption. With no changes to the parameters, the model accurately predicts the observed two-point clustering on small- and large-scales over a wide range of stellar mass. This quantitative success is remarkable since the Kravtsov (2013) parameters were fit to the observed $\langle R_{1/2}|M_*\rangle$, and the $R_{1/2}-$dependence of SDSS galaxy clustering has heretofore never been measured. Moreover, this success is non-trivial, as we demonstrate that galaxy clustering is highly sensitive to the assembly history-dependent physics that shapes the relative size of centrals and satellites. Our results can be treated as a boundary condition for more complex and fine-grained models of galaxy size, and provide a simple means for cosmological surveys to generate synthetic galaxy populations with realistic sizes across the cosmic web.

1 INTRODUCTION

Some introduction goes here.

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 $w_{\rm p}$ of our SDSS galaxy sample using line-of-sight projection of $\pi_{\rm max} =$ 20Mpc using the `correl` program in `UniverseMachine`. Our results in § 4 will give special focus on the dependence of $w_{\rm p}$ upon $R_{1/2}$. We will quantify this dependence in terms of *clustering ratios* of "large" vs. "small" galaxies, defined according to whether composite galaxy size is above or below $\langle R_{1/2}|M_*\rangle$, computed as the median of a sliding stellar mass window with a width of $N_{\rm gal} = 1000$.

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'. 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 $w_{\rm p}$ 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 kpc, and all values of M_* and $M_{\rm halo}$ 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

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

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

For halo mass $M_{\rm halo}$ we use $M_{\rm peak}$, the largest value of $M_{\rm vir}$ 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 $M_*^{\rm MPA-JHU}$ 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 $\langle M_*^{\rm Meert+15}|M_*^{\rm MPA-JHU}\rangle$ in logarithmic bins spanning $9 < \log_{10} M_*^{\rm MPA-JHU}/M_\odot < 12$, 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 $M_*^{\rm MPA-JHU} \approx 10^{9.75}M_\odot$, and ~ 0.4 dex at $M_*^{\rm MPA-JHU} \approx 10^{11.5}M_\odot$.

In Kravtsov (2013), it was found that if a stellar-to-halo mass relation is inverted to map halo mass estimates $M_{\rm halo}$ onto SDSS galaxies, and then the $M_{\rm halo} - R_{\rm vir}$ relation is applied to map values of $R_{\rm vir}$ onto the galaxies, then the resulting $R_{1/2} - R_{\rm vir}$ relation of SDSS galaxies exhibits the following linear scaling across a wide range of stellar mass:

$$R_{1/2} = 0.0125R_{\rm vir} \quad (1)$$

Motivated by the simplicity of this scaling relation, we transform the Kravtsov (2013) into a forward model using `Halotools`. For the virial radius of halos and subhalos, we use $R_{\rm M_{peak}}$, the value of $R_{\rm vir}$ in physical units of kpc measured at the time of peak subhalo mass. When generating Monte Carlo realizations of our model galaxy sizes, we add log-normal scatter $\sigma_{\rm R_{1/2}} = 0.15$ dex.

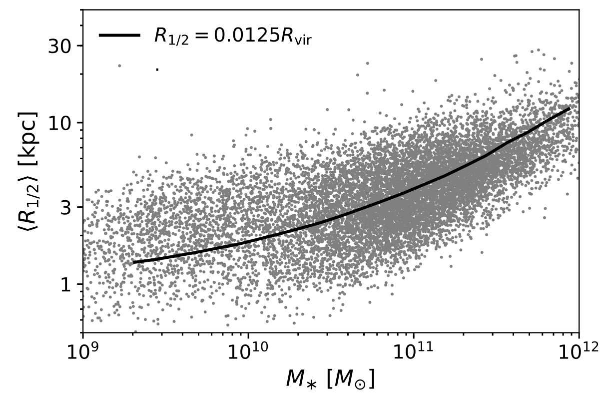

Figure 1. One-point data used to fit the fiducial model. Scattered points show the $R_{1/2} - M_*$ relation for SDSS galaxies as measured in Meert et al. (2015). The black curve shows the median $R_{1/2} - M_*$ relation implied by the model described in §3, in which $R_{1/2} = 0.0125R_{\rm vir}$. This figure confirms the findings in Kravtsov (2013) that a linear relationship between $R_{\rm vir}$ and $R_{1/2}$, convolved against the nonlinear relationships between $R_{\rm vir}$, $M_{\rm halo}$ and M_*, correctly predicts the characteristic curvature in the relation $\langle R_{1/2}|M_*\rangle$ over a wide range in stellar mass.

4 RESULTS

In §4.1 we show comparisons between the galaxy size model described in §3 and our SDSS sample. We identify the key ingredients that determine the characteristic $R_{1/2}-$dependence of galaxy clustering in §**??**. In so doing, we demonstrate the sensitivity of galaxy clustering measurements to the underlying model assumptions, establishing the success of our model as non-trivial.

4.1 Testing Model Predictions

In Figure 1 we show the scaling of galaxy size $R_{1/2}$ with M_*. Scattered gray points show the scaling relation for our SDSS galaxy sample, while the black curve shows the median relation $\langle R_{1/2}|M_*\rangle$ implied by the model described in §3.

In Figure 2 we present new measurements of the $R_{1/2}-$dependence of projected galaxy clustering, $w_{\rm p}(r_{\rm p})$. 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 $w_{\rm p}(r_{\rm p})$. 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 $N_{\rm gal} = 1000$. 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. For any M_*-threshold sample, the SMF of the "large" and "small" subsamples are identical, by construction.

We measure $w_{\rm p}(r_{\rm p})$ 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 $w_{\rm p}^{\rm all}, w_{\rm p}^{\rm large}$, and $w_{\rm p}^{\rm small}$ for each threshold sample. This allow us to compute the ratio $(w_{\rm p}^{\rm large} - w_{\rm p}^{\rm small})/w_{\rm p}^{\rm all}$, 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 2. Points with error bars show SDSS measurements, solid curves show the clustering ratios of model galaxies as predicted by the model described in §3. Before unpacking the information contained in these clustering measurements, we stress that the good agreement shown between model and data in Figure 2 is a genuine model prediction, since the model parameters were taken directly from Kravtsov (2013), which were fit to the one-point measurements in Fig. 1, whereas the two-point measurements appearing in Figure 2 have heretofore never been measured.

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 model galaxies. This result may be surprising, since $R_{1/2} \propto R_{\rm vir}$, halo mass $R_{\rm vir} \propto M_{\rm halo}^{1/3}$, and clustering strength increases with $M_{\rm vir}$. Based on this simple argument, one would ex-

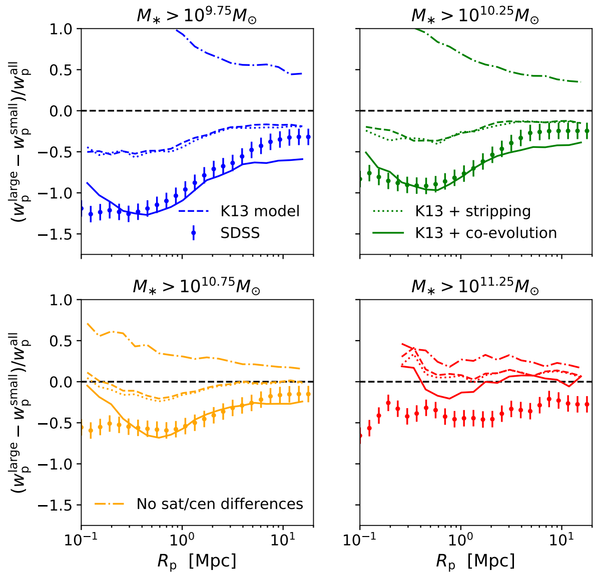

Figure 2. $R_{1/2}$**−dependence of galaxy clustering.** Points with error bars show new SDSS measurements of the $R_{1/2}$−dependence of projected galaxy clustering, w_p, compared to predictions by the model tuned to the measurements shown in Fig. 1. We define a galaxy as "large" or "small" according to whether it is above or below the median size for its stellar mass, so that in each panel, the SMF 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 a different volume-limited M_*−threshold samples. Fiducial model predictions appear as the solid curve in each panel.

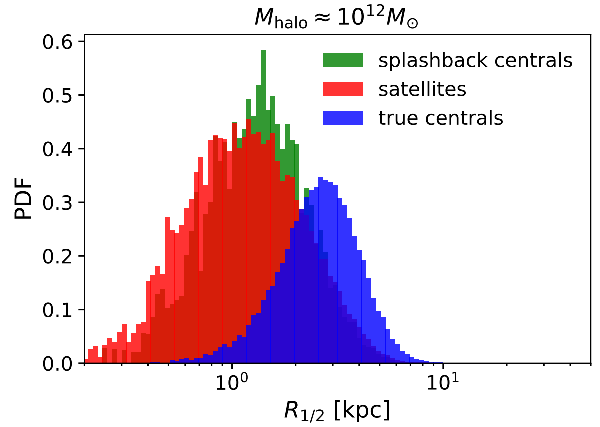

Figure 3. Relative sizes of centrals and satellites. In a narrow bin of halo mass $M_{\rm halo} = M_{\rm peak} \approx 10^{12} M_\odot$, we show the distribution of model galaxy sizes for different subpopulations galaxies. The red histogram shows the sizes of satellites; the blue histogram shows host 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 fiducial model, galaxy size is set by the physical size of the virial radius at the time the (sub)halo attains its peak mass, naturally resulting in smaller sizes for satellites and backsplash centrals relative to true centrals of the same $M_{\rm peak}$.

pect the opposite trend to the measurements shown here. We present a resolution to this puzzle in the following section.

4.2 Origin of the size-dependence of galaxy clustering

We can understand the origin of the shape of the clustering strength ratios shown in Figure 2 through a series of straightforward "shuffle tests". The points with error bars and solid curves in each panel of Fig.4 are identical to those appearing in Fig.2, although the dynamic range of the y-axis has changed to accommodate the additional curves.

We generated each additional curve by variations on the following exercise. In logarithmic bins of $M_{\rm peak}$ with width of 0.1dex, we randomly shuffle the modeled $R_{1/2}$ values of some particular subpopulation of model galaxies that reside in subhalos in the $M_{\rm peak}$ bin. After repeating this shuffle over the entire range of $M_{\rm peak}$, we then re-divide model galaxies into "large" and "small" populations, and remeasure the clustering strength ratios. By shuffling sizes amongst different subpopulations, we can gain an understanding of the subhalo properties that contribute to the characteristic shape of the clustering ratios.

In the *brown, dotted* curves labeled "all-gal scramble", we shuffle the sizes of *all* model galaxies in the $M_{\rm peak}$ bin, so that in the resulting mock, galaxy size has no correlation with any subhalo property beyond $M_{\rm peak}$. In particular, the "all-gal scramble" mock erases the possible influence of a correlation between galaxy size and the redshift at which the subhalo attains peak mass. The resulting clustering strength ratios become exactly as predicted by the naive expectation sketched at the conclusion of §4.1: $R_{1/2} \propto R_{\rm vir} \propto M_{\rm halo}^{1/3}$, and so large galaxies clustering more strongly relative to small galaxies of comparable stellar mass.

Statistically, there are actually two distinct effects at work in the "all-gal scramble". Central and satellite galaxies of similar $M_{\rm peak}$ have their sizes shuffled amongst each other, and the potential influence of central galaxy assembly bias is also erased. We parse these separate effects by exploring three variations on the shuffle tests, described below.

For the *cyan dashed* curve labeled "all-cen-scramble", we only shuffle sizes amongst present-day host halos, i.e., amongst subhalos for which Rockstar `upid` equals -1. For the *black dashed* curve labeled "true-cen-scramble", we only shuffle amongst host halos that have never passed within the virial radius of a larger halo throughout their entire assembly history, i.e., we only shuffle amongst non-backsplash host halos. Finally, the *purple dot-dashed* curve labeled "sat-scramble", we only shuffle sizes amongst satellite galaxies, i.e., subhalos that are presently located inside the virial radius of a larger halo.

We conclude this section by estimating how satellite mass stripping manifests in $w_p(r_p)$. To do so, we con-

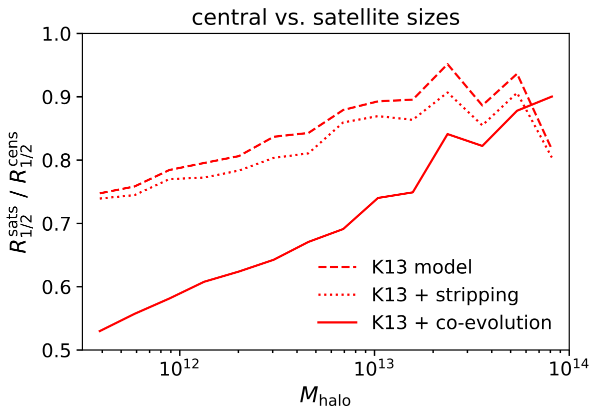

Figure 4. Origin of the $R_{1/2}-$dependence of clustering. Here we compare our fiducial model, in which satellite galaxy size is set by $R_{\rm vir}$ at the time of infall, to a set of alternative models created by shuffling the sizes of various subsamples of the fiducial mock.

struct an extension our fiducial model with a simple additional ingredient for post-infall evolution of satellites mass and size.

5 DISCUSSION

5.1 Progression from Backwards to Forwards Modeling

5.2 Implications for Satellite Mass Loss

5.3 Future Directions for Empirical Modeling of Morphology

6 CONCLUSIONS

6.1 Summary

ACKNOWLEDGMENTS

APH thanks John Baker for the *Toejam & Earl* soundtrack.

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
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
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
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS , 351, 1151
Croton D. J., 2013, PASA, 30, e052
Hearin A., Campbell D., Tollerud E., et al., 2016, ArXiv e-prints
Kauffmann G., Heckman T. M., White S. D. M., et al., 2003, MNRAS , 341, 33
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ , 740, 102
Kravtsov A. V., 2013, ApJL , 764, L31
Meert A., Vikram V., Bernardi M., 2013, MNRAS , 433, 1344
Meert A., Vikram V., Bernardi M., 2015, MNRAS , 446, 3943
Moster B. P., Naab T., White S. D. M., 2013, MNRAS , 428, 3121
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
Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS , 462, 893
Sinha M., Garrison L., , 2017, Corrfunc: Blazing fast correlation functions on the CPU, Astrophysics Source Code Library
Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, MNRAS , 409, 1379