

# The galaxy size – halo mass scaling relations and clustering properties of central and satellite galaxies

Facundo Rodriguez<sup>1,2\*</sup>, Antonio D. Montero-Dorta<sup>3,4</sup>, Raul E. Angulo<sup>5,6</sup>, M. Celeste Artale<sup>7</sup>, Manuel Merchán<sup>1,2</sup>

<sup>1</sup> Universidad Nacional de Córdoba. Observatorio Astronómico de Córdoba. Córdoba, Argentina

<sup>2</sup> CONICET. Instituto de Astronomía Teórica y Experimental. Laprida 854, X5000BGR, Córdoba, Argentina

<sup>3</sup> Departamento de Física, Universidad Técnica Federico Santa María, Casilla 110-V, Avda. España 1680, Valparaíso, Chile

<sup>4</sup> Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, Rua do Matão 1371, CEP 05508-090, São Paulo, Brazil

<sup>5</sup> Donostia International Physics Center (DIPC), Paseo Manuel de Lardizábal, 4, 20018 Donostia-San Sebastián, Spain

<sup>6</sup> IKERBASQUE, Basque Foundation for Science, 48013, Bilbao, Spain.

<sup>7</sup> Institut für Astro- und Teilchenphysik, Universität Innsbruck, Technikerstrasse 25/8, 6020 Innsbruck, Austria

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

In this work, we combine size and stellar mass measurements from the Sloan Digital Sky Server (SDSS) with the group finder algorithm of Rodriguez & Merchán in order to determine the stellar and halo mass – size relations of central and satellite galaxies separately. We show that, while central and satellite galaxies display similar stellar mass – size relations, their halo mass – size relations differ significantly. As expected, more massive haloes tend to host larger central galaxies. However, the size of satellite galaxies depends only slightly on halo virial mass. We show that these results are compatible with a remarkably simple model in which the size of central and satellite galaxies scales as the cubic root of their host halo mass, with the normalization for satellites being  $\sim 30\%$  smaller than that for central galaxies, which can be attributed to tidal stripping. We further check that our measurements are in excellent agreement with predictions from the IllustrisTNG hydrodynamical simulation. In the second part of this paper, we analyse how the clustering properties of central and satellite galaxies depend on their size. We demonstrate that, independently of the stellar mass threshold adopted, smaller galaxies are more tightly clustered than larger galaxies when either the entire sample or only satellites are considered. The opposite trend is observed on large scales when the size split is performed for the central galaxies alone. Our results place significant constraints for halo–galaxy connection models that link galaxy size with the properties of their hosting haloes.

**Key words:** large-scale structure of Universe – galaxies: statistics – galaxies: groups: general – galaxies: haloes

## 1 INTRODUCTION

In the standard hierarchical structure formation scenario, galaxies form by the condensation of baryons at the minima of the potential wells defined by the collisionless collapse of dark-matter (DM) haloes (White & Rees 1978; White & Frenk 1991). This highly nonlinear galaxy formation process involves a great diversity of astrophysical mechanisms over a wide range of scales. Despite this complexity, many observed properties of galaxies are known to exhibit tight correlations between each other. These *scaling relations* give us valuable statistical information about the galaxy populations and are therefore useful to constrain galaxy formation and evolution

models. One of these well-measured properties is the stellar mass – galaxy size relation. This relationship, which can be measured in terms of the half-mass or the half-light radius, has been intensively studied both in the local universe (e.g., Shen et al. 2003; Huang et al. 2013; Bernardi et al. 2014; Lange et al. 2015; Zhang & Yang 2019; Hearin et al. 2019; Zanisi et al. 2020) and at high redshift (e.g., Trujillo et al. 2004; Huertas-Company et al. 2013; van der Wel et al. 2014; Kawamata et al. 2015; Shibuya et al. 2015; Huang et al. 2017; Favole et al. 2018; Mowla et al. 2019). Through halo-galaxy connection techniques such as sub-halo abundance matching (SHAM, e.g., Trujillo-Gomez et al. 2011; Hearin et al. 2013), it is currently possible to relate the size of galaxies to the mass and virial radius of the halo in which they reside. It has been pointed out, in this context, that galaxy size is approximately proportional to the

\* E-mail: facundo.rodriguez@unc.edu.ar

virial radius of the hosting halo (e.g., Kravtsov 2013; Huang et al. 2017).

The mass – size relation is only one of the multiple correlations between galaxy size and other galaxy properties that have been measured, including also dependencies on redshift, colour, or star formation (Shen et al. 2003; Ferguson et al. 2004; Trujillo et al. 2006; Williams et al. 2010; Mosleh et al. 2012; Ono et al. 2013; van der Wel et al. 2014; Lange et al. 2015). Some works have also addressed the connection with several *evolutionary* properties, suggesting that size could ultimately be determined by the galaxy’s dynamical and assembly history (Mo et al. 1998; Naab et al. 2009; Somerville & Davé 2015; Robertson et al. 2006; Dekel & Burkert 2014; DeFelippis et al. 2017; Jiang et al. 2019).

As predicted by the hierarchical assembly of haloes and the merging of galaxies, groups generally present a main or *central galaxy*, which is typically more luminous and/or massive than the surrounding *satellite galaxies* (Larson et al. 1980; Balogh et al. 2000; McCarthy et al. 2008; Campbell et al. 2015; Hearin et al. 2019; Jiang et al. 2020). Many studies indicate that whether the galaxy occupies a “central” or “satellite” position in the halo that it inhabits determines, at least to some extent, its evolution (Kauffmann et al. 2004; Baldry et al. 2006; Weinmann et al. 2009; Wetzell et al. 2013; Lacerna et al. 2014). Studying satellite and central galaxies separately can therefore provide us with clues about their formation history (e.g., Spindler & Wake 2017).

In order to “tag” a galaxy as central or satellite, we need to be able to reliably identify the halo in which it resides. In observational data, the link between galaxies and their corresponding haloes is performed by means of a *group finder*. Several methods have been proposed to extract clusters from massive galaxy surveys, but the most widely used are the friends-of-friends (FOF; Huchra & Geller 1982) and halo-based methods (Yang et al. 2005). In a recent work, Rodríguez & Merchán (2020) present a technique that combines these two methods in order to efficiently identify groups in survey data.

Observational results can be contrasted with hydrodynamical simulations, which use sub-grid physical models to reproduce the central/satellite galaxy–halo connection. In this context, the recently released IllustrisTNG<sup>1</sup> suite of hydrodynamical simulations offers some additional advantages, such as the large size of some of their boxes (up to a side length of 205  $h^{-1}$ Mpc). By comparing the results obtained with observations with predictions from hydrodynamical simulations such as IllustrisTNG, the physical mechanisms responsible for the observed relations can be investigated. Using IllustrisTNG, in particular, Genel et al. (2018) analysed the size evolution of star-forming and quenched galaxies separately and were also able to reproduce their observed stellar mass – size relations.

The measurements of the mass – size relations can be combined with a clustering analysis in order to evaluate the dependence on the local and large-scale environment. In this framework, it is well known that halo mass is the main property that determines halo clustering. However, a number of secondary dependencies of halo clustering have been identified in recent years (on, e.g., halo age, concentration or spin, see, e.g., Sheth & Tormen 2004; Gao et al. 2005; Wechsler et al. 2006; Gao & White 2007; Angulo et al. 2008; Li et al. 2008; Lazeyras et al. 2017; Paranjape & Padmanabhan 2017; Salcedo et al. 2018; Han et al. 2018; Mao et al. 2018; Sato-Polito et al. 2019; Johnson et al. 2019; Ramakrishnan et al. 2019; Tucci et al. 2020). Whether these dependencies transmit to the galaxy

population or not is still a matter of intense debate (Zentner et al. 2016; Miyatake et al. 2016; Zu et al. 2016; Lin et al. 2016; Sunayama et al. 2016; Montero-Dorta et al. 2017; Artale et al. 2018; Niemiec et al. 2018; Zehavi et al. 2018; Walsh & Tinker 2019; Sunayama & More 2019; Montero-Dorta et al. 2020a,b; Obuljen et al. 2020; Salcedo et al. 2020), but the existence of *secondary halo bias* adds a important element to the modelling of the halo–galaxy connection (e.g. Hearin & Watson 2013; Hearin et al. 2014, 2016; Zentner et al. 2019; Wechsler & Tinker 2018; Contreras et al. 2020; Xu et al. 2020). For galaxy size, in particular, Hearin et al. (2019) show that smaller galaxies cluster more strongly than larger galaxies of the same stellar mass. The authors use these constraints to propose a model where galaxy size is proportional to the size of the virial radius at the time the halo reached its maximum mass.

The goal of this work is to provide a precise measurement of the relation between galaxy size and stellar and halo mass for central and satellite galaxies separately. These relations are measured both in SDSS at  $z = 0$  and in the IllustrisTNG hydrodynamical simulation. Also, we provide a high signal-to-noise measurement of clustering as a function of galaxy size and stellar mass. This analysis is first performed for the entire sample (without any membership classification) and subsequently compared with results obtained for central and satellite galaxies separately.

The structure of the paper is as follows. In Section 2 we describe the SDSS and TNG data that we use to measure the scaling relations and clustering properties of central and satellite galaxies. The group identification and galaxy classification (i.e., centrals and satellites) is briefly described in Section 3. Section 4 presents the main results of this paper. First, we show the scaling relations of centrals and satellites: the stellar-to-halo mass relation (SHMR), the stellar mass – size relation and the halo mass – size relation. Second, we present the clustering analysis both in observations and simulation data. Finally, Section 5 is devoted to discussing the implications of these results and providing a summary of the paper. Throughout this work we adopt the standard  $\Lambda$ CDM cosmology (Planck Collaboration et al. 2016), with parameters  $\Omega_m = 0.3089$ ,  $\Omega_b = 0.0486$ ,  $\Omega_\Lambda = 0.6911$ ,  $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$  with  $h = 0.6774$ ,  $\sigma_8 = 0.8159$  and  $n_s = 0.9667$ .

## 2 OBSERVATIONAL AND SIMULATION DATA

### 2.1 SDSS data

The main goal of this work is to study the relation between the size of galaxies and the mass of the haloes that they inhabit and to analyse the dependence of galaxy clustering on galaxy size. We use galaxy data from the Sloan Digital SkyServer DataRelease 7 (SDSS DR7 Abazajian et al. 2009) spectroscopic sample. Measurements of the half-light radius derived from galaxy profile decompositions are provided by Meert et al. (2015) and the stellar mass measurements, based on the population synthesis code of Bruzual & Charlot (2003), are taken from the MPA-JHU catalogue (Kauffmann et al. 2003). The catalogue contains 670,722 SDSS galaxies with improved photometry, which includes a better background subtraction (Vikram et al. 2010; Bernardi et al. 2013, 2014; Meert et al. 2013) and light profile fitting method. The improvement in the radius measurements comes from the fit with different models of two-dimensional  $r$ -band profiles.

For our analysis, we define a volume-limited sample taking into account the conditional probability density for stellar mass as a function of redshift in the SDSS. For this selection, and in order to

<sup>1</sup> <http://www.tng-project.org>

be consistent with previous analyses (e.g. Hearin et al. 2019), we use the same cut of Behroozi et al. (2015):

$$M_r < -0.25 - 1.9 \log_{10} \left( \frac{M_*}{M_\odot} \right), \quad (1)$$

where  $M_r$  is the galaxy's Petrosian  $r$ -band absolute magnitude and  $M_*$  is its stellar mass (for more details see Figure 2 of Behroozi et al. 2015).

Among the different size measurements provided in the Meert et al. (2015) catalogues, we use the half-light radius of the total fit,  $r_{\text{tot}}$ , and the axis ratio of the total fit,  $ba_{\text{tot}}$ . From these quantities, the half-light semi-major axis of each galaxy is simply computed as

$$R_a = \frac{r_{\text{tot}}}{\sqrt{ba_{\text{tot}}}}, \quad (2)$$

which provides us with a robust galaxy size estimate that is independent of morphology.

## 2.2 Simulation data

In this paper, we use the galaxy and dark-matter halo catalogues from *The Next Generation* Illustris (IllustrisTNG, Nelson et al. 2019) magneto-hydrodynamical cosmological simulations, which represent an updated version of the Illustris simulations (Vogelsberger et al. 2014a,b; Genel et al. 2014). The IllustrisTNG simulations are performed with the AREPO moving-mesh code (Springel 2010) and include sub-grid models that account for radiative metal-line gas cooling, star formation, chemical enrichment from SNI, SNIa and AGB stars, stellar feedback, supermassive-black-hole formation with multi-mode quasar, and kinetic black-hole feedback. The main updates with respect to the Illustris simulation are: a new implementation of black-hole kinetic feedback at low accretion rates, a revised scheme for galactic winds, and the inclusion of magneto-hydrodynamics (see Pillepich et al. 2018; Weinberger et al. 2017, for further details).

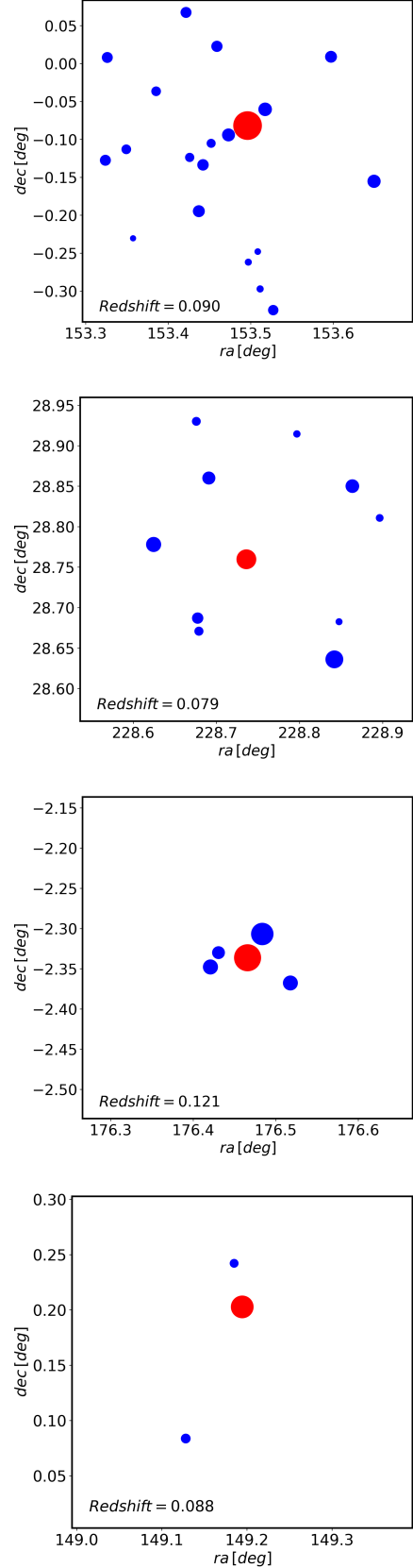
We analyse the IllustrisTNG300-1 run (hereafter TNG300 for simplicity), which is the largest simulated box from the IllustrisTNG suite featuring the highest resolution. This run adopts a cubic box of side  $205 h^{-1}$  Mpc with periodic boundary conditions. The TNG300 run follows the evolution of  $2500^3$  dark-matter particles of mass  $4.0 \times 10^7 h^{-1} M_\odot$ , and  $2500^3$  gas cells of mass  $7.6 \times 10^6 h^{-1} M_\odot$ .

In TNG300, dark-matter haloes are identified using a friends-of-friends (FOF) algorithm with a linking length of 0.2 times the mean inter-particle separation (Davis et al. 1985). The gravitationally bound substructures called subhaloes are subsequently identified using the SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009). Among all subhaloes, those containing a non-zero stellar component are considered galaxies. Each dark-matter halo typically contains multiple galaxies, including a central galaxy and several satellites, where the positions of centrals coincide with the FOF centres.

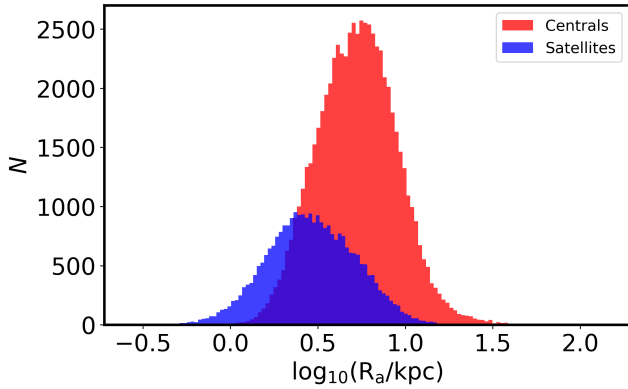
In our analysis, we use the virial mass of the haloes,  $M_{\text{vir}}$ , defined as the total mass enclosed within a sphere of radius  $R_{\text{vir}}$  (i.e., the radius at which the enclosed density equals 200 times the critical density). For the simulated galaxies (i.e., subhaloes with non-zero stellar components), the stellar mass,  $M_*$ , is defined as the total mass of all stellar particles bound to each subhalo.

We complement our analysis with post-processing TNG300 size data from Genel et al. (2018)<sup>2</sup>, which differ slightly from

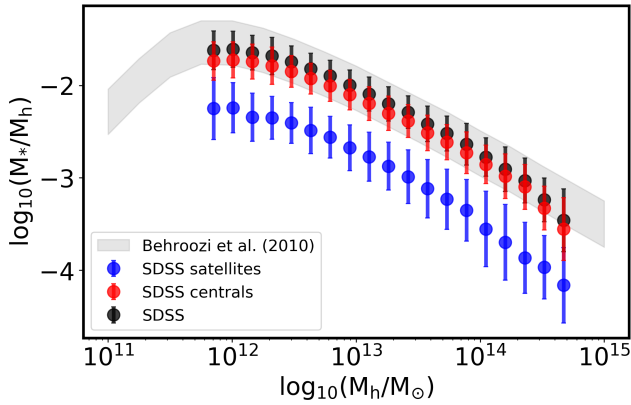
<sup>2</sup> Available on the IllustrisTNG database.



**Figure 1.** Examples of identified groups. The colour of the circles is given by our classification of central (red) and satellite (blue) galaxies, whereas the dot size indicates the stellar mass of each galaxy (largest in each group is the central galaxy).



**Figure 2.** The distribution of the size of satellite (blue) and central (red) galaxies in the SDSS sample.



**Figure 3.** The stellar-to-halo mass relation (SHMR) measured from the SDSS sample. The SHMR obtained using the entire stellar mass of the groups is represented in black dots, whereas the SHMR for central and satellite galaxies alone is shown in red and blue dots, respectively. Error bars correspond to the standard deviation of the measurements. The SHMR from Behroozi et al. 2010 is also shown in a light grey shaded region for comparison.

the standard TNG300 computations. For each galaxy, we use both the three-dimensional half-mass radius,  $R_{3D,1/2}$ , and the two-dimensional half-light radius in the r-band,  $R_{r,1/2}$ . The first quantity is based on the masses of all the stellar particles assigned to the galaxy, including burning stars and stellar remnants. The second measurement is obtained by projecting the simulation box along a random direction with respect to the orientation of each galaxy. For more information, we refer the reader to Genel et al. (2018).

One of the main goals of this work is to compare the scaling relations measured in observations with those predicted by IllustrisTNG. In addition to analyzing the results obtained directly from the TNG300 box, we build a mock catalogue that reproduces the SDSS DR12 properties. This TNG300 mock catalogue is used to evaluate the effect of potential observational biases in our simulation-data comparison.

In order to construct our mock catalogue, we first place the observer at the origin of the TNG300 box. Considering the box periodicity, the volume of the SDSS DR12 spectroscopic catalogue is simulated by adding the TNG300 volume repeatedly. The red-

shift of each galaxy is calculated by combining the cosmological distance and the distortion produced by proper motions. From these redshifts and the absolute magnitudes provided by the simulation, the apparent magnitudes of each galaxy is derived. To mimic the flux-limited selection of the SDSS DR12, the SDSS upper apparent magnitude threshold is imposed. Finally, to reproduce the angular selection function of the survey, a mask following the same procedure used in Rodríguez et al. (2015) is built to reproduce the SDSS DR12 selection and geometry. Note, finally, that one of the main uncertainties that exist for central and satellite galaxies in the SDSS is that this classification depends on group identification. For this reason, the group finder algorithm of Rodríguez & Merchán (2020) is also applied to the TNG300 mock catalogue.

### 3 GROUP IDENTIFICATION

In order to identify central and satellite galaxies in the SDSS sample described above, a galaxy group catalogue is built using the galaxy group finder presented in Rodríguez & Merchán (2020), which combines both the FOF and the halo-based methods. Assuming that a group is a gravitationally bound system containing at least one bright galaxy, the procedure begins by identifying groups using a proper FOF algorithm adapted to fit this definition. Subsequently, properties are assigned based on their characteristic luminosities. In a second step, an iterative halo-based method is incorporated to improve the reliability of the system identification. This iteration process is designed to add galaxies (or not) based on the properties of these FOF groups. If groups memberships change, the dark-matter properties are recalculated based on the new group luminosity. The process continues until no changes in group membership are needed. This method allows us to find systems with both low and high number of members, maintaining a high purity and completeness.

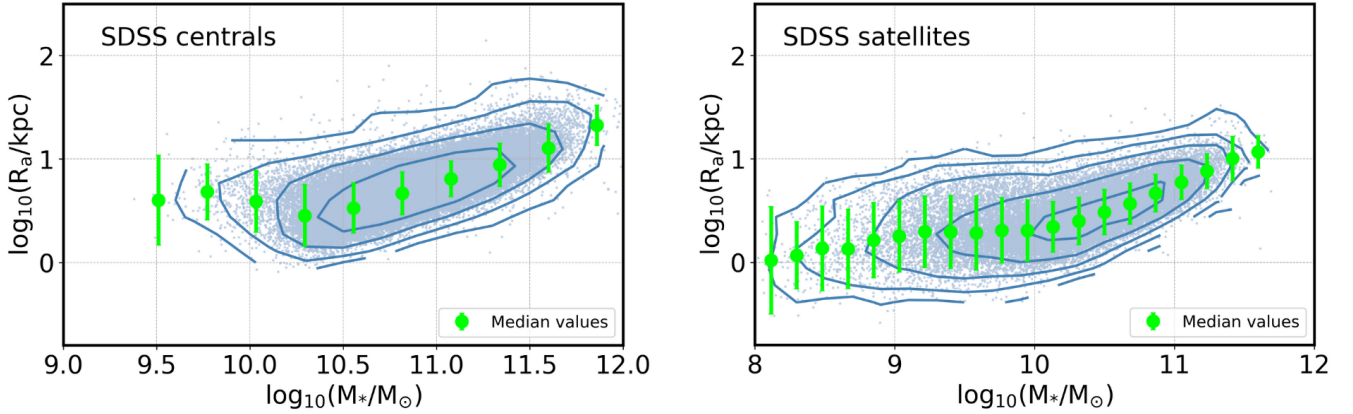
The performance of the method was assessed using a mock catalogue in Rodríguez & Merchán (2020). The results of this analysis show a very low percentage of interlopers in extracted groups along with a precise recovery of the galaxy membership in dark-matter haloes. In addition, the total number of identified clusters is found to be very similar to the total number of dark-matter haloes in the synthetic catalogue.

As a result of the identification procedure, in addition to the galaxy group membership, the halo mass of each group ( $M_h$ ) is obtained. Group masses are allocated through an abundance matching technique (e.g., Kravtsov et al. 2004; Vale & Ostriker 2004; Conroy et al. 2006), based on luminosity. In this way, we can analyse both the galaxy properties and their dependence on halo mass. It is worth emphasising that the identification process is based on the same overdensity as that adopted in TNG300 ( $\Delta = 200$ ). Therefore, the SDSS group masses are directly comparable to those of TNG300.

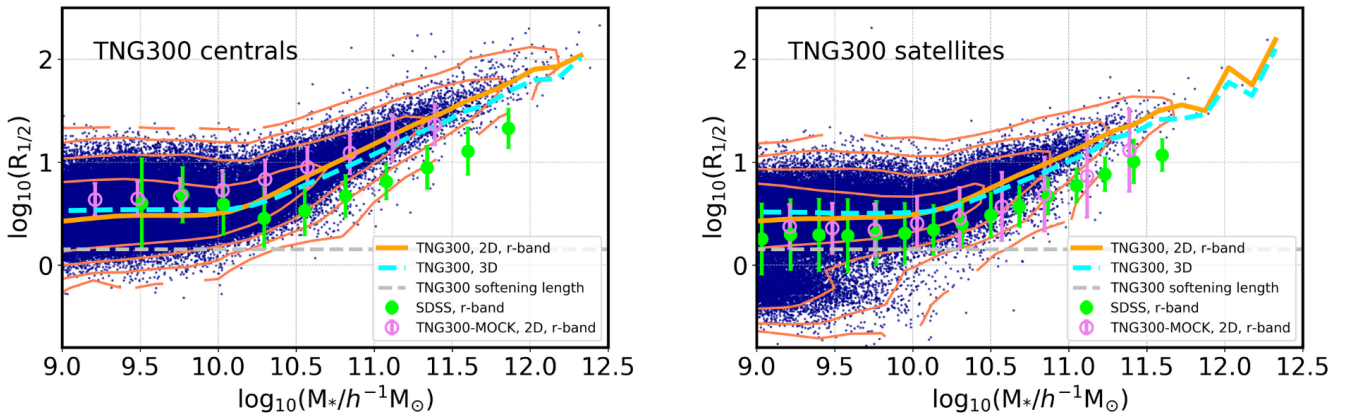
The Rodríguez & Merchán (2020) group finder, in its identification process, does not explicitly produce a central/satellite galaxy classification. However, we can use for this task a widely-extended criterion based on the luminosity and stellar mass of the galaxy group members: central galaxies are expected to have higher stellar mass and/or luminosity than satellite galaxies. We opt here to label the brightest and most massive galaxy in the group as the central galaxy and all the others as satellite galaxies.<sup>3</sup> In order to illustrate this process, in Figure 1 we show four randomly-chosen examples

<sup>3</sup> Note that to produce a better classification, those cases in which the brightest and most massive galaxy in the group do not coincide are discarded. These cases represent a small fraction of the total sample.





**Figure 4.** The stellar mass–size relation from the SDSS sample. The right-hand panel shows the semi-major-axis half-light radius for satellites galaxies as a function of stellar mass (blue dots and contours). The over-plotted green dots represent the median values in several stellar-mass bins. Error bars indicate the standard deviation of the measurements. The left-hand panel presents the same results for central galaxies.



**Figure 5.** The stellar mass–size relation from TNG300 in the same format as Figure 4 for central (left panel) and satellite galaxies (right panel). In both panels, blue dots, orange contours, and solid lines show results obtained with the 2D r-band half-mass radius, whereas the dashed blue line shows the relation based on the 3D size measurement and the open circles correspond to the TNG300 mock catalogue. The SDSS relations shown in Figure 4 are over-plotted for comparison. The horizontal dashed line marks the TNG300 softening length.

of identified groups with different numbers of members. The size of the dot representing each galaxy indicates its stellar mass (larger meaning more massive), whereas central and satellite galaxies are marked in red and blue colours, respectively.

As a result of our classification, and taking into account the mass–redshift cut shown in Section 2.1, a sample comprising 66,839 central galaxies and 28,680 satellites is obtained for the redshift range  $0.01 < z < 0.2$ . The resulting size distribution of central (red) and satellite (blue) galaxies is shown in Figure 2. As expected, the histograms for the two subsamples display some overlap, but central galaxies are typically larger. These galaxies have a median radius of 6.13 kpc with a standard deviation of 3.81 kpc, while the mean radius of satellites is 3.46 kpc and their standard deviation 2.22 kpc.

## 4 RESULTS

### 4.1 Scaling relations

The group finder allows us to link galaxy properties to those of the haloes that they inhabit. It also enables us to differentiate be-

tween central and satellite galaxies, which adds some important information for understanding the physical processes involved in galaxy growth. In this section, we analyse some observational constraints that are particularly relevant in the aforementioned context: the scaling relations between galaxy size and both the stellar and virial (halo) mass.

#### 4.1.1 Stellar-to-halo mass relation

We begin with the stellar-to-halo mass relation (SHMR), which provides the fraction of stellar mass as a function of halo mass. In order to measure this relation, which is one of the most fundamental halo–galaxy constraints, we make use of the stellar masses from the Meert et al. (2015) catalogue and the halo masses obtained through the abundance matching technique described above (Rodríguez & Merchán 2020).

Figure 3 displays the total SHMR, obtained using the combined stellar mass of the groups (i.e., the sum of the stellar masses of all group members, black dots), along with the SHMRs for central and satellite galaxies separately (red and blue dots, respectively). As expected, the combined and central-galaxy relations are quite simi-

lar, which is a consequence of satellite galaxies having, on average, significantly smaller stellar masses. It is, however, interesting that satellites follow a trend that resembles that of the combined/central-galaxy SHMRs. This is not necessarily expected, since the SHMR is computed following the same procedure as for the central galaxies, using the stellar mass of each satellite galaxy and the mass of the halo to which it belongs, averaged for each halo mass bin. We are, therefore, relating two different scales. This resemblance indicates a certain level of connection between the properties of satellite galaxies and those of their hosting haloes.

In Figure 3, our results are also compared with the central-galaxy SHMR of Behroozi et al. (2010) (light-grey shaded region). For the sake of simplicity, only the confidence intervals defined by the reported error bars are shown here. We find an excellent agreement between the aforementioned estimate and our measurements, both for the combined sample and for central galaxies alone. It is noteworthy that the Behroozi et al. (2010) relation is in good agreement with several other works (as shown in their Figure 11, e.g., Mandelbaum et al. 2006; Yang et al. 2009; Moster et al. 2010). The SHMR is a pivotal relation in halo-galaxy connection studies, so the level of agreement shown in Figure 3 gives us confidence in terms of addressing other scaling relations in the following sections.

#### 4.1.2 Stellar mass–size relation

To continue with our analysis, we focus now on the relation between the size and stellar mass of galaxies, i.e., the so-called stellar mass–size relation. Following our philosophy, the relations are measured for SDSS central and satellite galaxies separately, as defined by the criterion presented in Section 3. These results are shown in two separate panels (for centrals and satellites) in Figure 4. In each panel, the logarithm of galaxy size is shown as a function of stellar mass, including both the individual data points (blue dots and contours) and the median values of the radius in several mass bins (green dots). Error bars on the median values correspond to the standard deviation of the measurements. The stellar mass–size relation for central and satellite galaxies follow the expected trend, with more massive galaxies having larger radii (e.g., Shen et al. 2003; Shankar & Bernardi 2009; Ichikawa et al. 2012; Poggianti et al. 2012; Fernández Lorenzo et al. 2013; Hearin et al. 2019; Jiang et al. 2020).

It can be observed from Figure 4 that central galaxies cover a smaller stellar-mass range than satellite galaxies. This is to be expected, given that the former are selected as the brightest and most massive galaxies of their groups. Note that due to the identification procedure, central galaxies are by definition brighter than a given threshold, whereas satellite galaxies can span a wide range of luminosities (i.e., there is no cut for satellites). If we restrict ourselves to the stellar mass range where both central and satellite galaxies overlap, the median values of both populations are very similar, with centrals being only slightly larger. These results align well with those found by Huertas-Company et al. (2013) and Spindler & Wake (2017), which indicate that the stellar mass - size relation is largely the same for centrals and satellites. Another thing to notice from Figure 4 is the significant dispersion of sizes for both galaxy types, which seems to advise against central/satellite classification criteria based upon a simple size threshold (e.g., Hearin et al. 2019).

The TNG300 hydrodynamical simulation provides a means to contrast our observational constraints against a detailed galaxy formation model. Figure 5 displays, in the same format as Figure 4, the stellar mass – size relation for central (left) and satellite galaxies (right) in TNG300. Note that here the two different TNG300 half-

mass radius estimates (Section 2.2) are employed (with differences being quite small). Importantly, the TNG300 measurements confirm the idea discussed above: central and satellite galaxies display very similar stellar mass – size relations.

Although we have kept all objects in TNG300 for completeness, the small-size end of Figure 5 is, of course, expected to be affected by resolution. In fact, the TNG300 softening length for dark matter and stars is  $\epsilon_{\text{dm,stars}} = 1.0 h^{-1} \text{kpc}$ , which makes the size measurement unreliable for the smallest satellites observed in the right-hand panel of Figure 5. This resolution length is represented by a horizontal dashed line in Figure 5.

In order to evaluate the potential observational biases and those that can be introduced by the galaxy group identification algorithm, in Figure 5 we also show the results obtained with the TNG300 mock (open circles). We use the stellar mass and the two-dimensional half-light radius in the r-band. In agreement with previous results, a very similar mass–size relation is observed for central and satellites. The discrepancies found with respect to the mean values of the box (orange line) are a consequence of the misclassification of galaxies as central and satellite.

Over-plotted in Figure 5 are the median relations for the SDSS (Figure 4). When comparing SDSS results with those obtained for the TNG300, it can be seen that they both present similar sizes and the same qualitative trends for the central and satellite galaxies. Moreover, satellite and central galaxies in the TNG300 and in the SDSS, show that below  $\log_{10}(M_*/M_\odot) = 10.0$ , the increase in stellar mass is not clearly reflected in the increase in radius. Our results show that the galaxy size increases with stellar mass for central and satellite galaxies with  $\log_{10}(M_*/M_\odot) \gtrsim 10.0$  in both TNG300 and SDSS samples. On the other hand, galaxies with  $\log_{10}(M_*/M_\odot) \lesssim 10.0$  do display a roughly flat relation between the stellar mass and galaxy size for satellite and centrals in both samples. It is noteworthy that, despite the smaller agreement, the radii of SDSS galaxies are systematically smaller than those of the simulated galaxies at  $\log_{10}(M_*/M_\odot) \gtrsim 10.5$ .

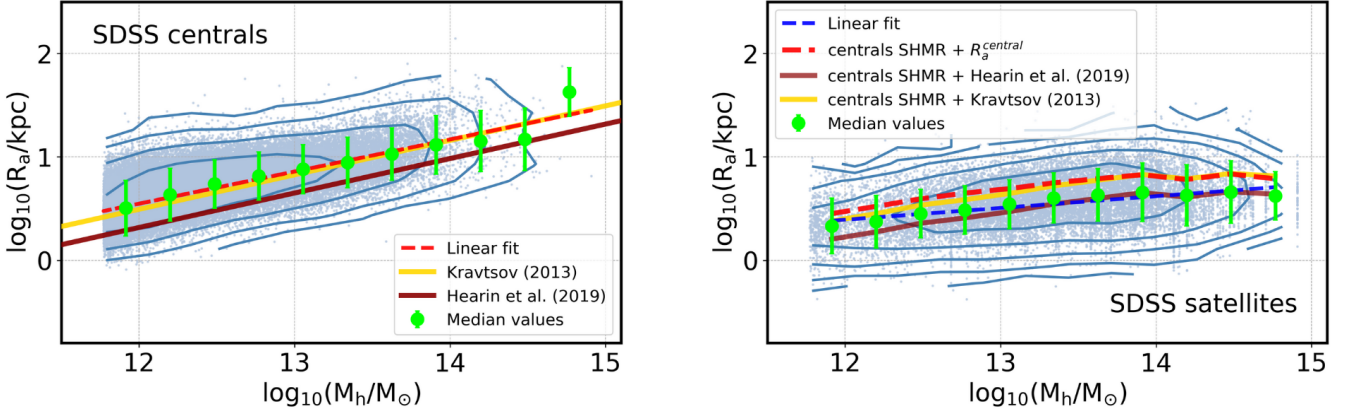
The discrepancies in the galaxy size predictions from hydrodynamical simulation and observational measurements are well documented in the literature. In the original Illustris simulation, galaxies with  $\log_{10}(M_*/M_\odot) \lesssim 11$  were larger than observed galaxies by roughly a factor of 2 (e.g., Bottrell et al. 2017). This problem was reportedly amended in TNG300 (as our Figure 5 confirms) by means of a combination of several modifications to the galactic winds model (Genel et al. 2014; Pillepich et al. 2018).

However, as can be seen in Figure 5, systematic differences with respect to the SDSS data are still present, mainly at the high-mass end, which could potentially be due to material losses by stripping processes, the applied AGN feedback processes, or other physical models implemented in TNG300. It is, nevertheless, outside the scope of this work to address these issues.

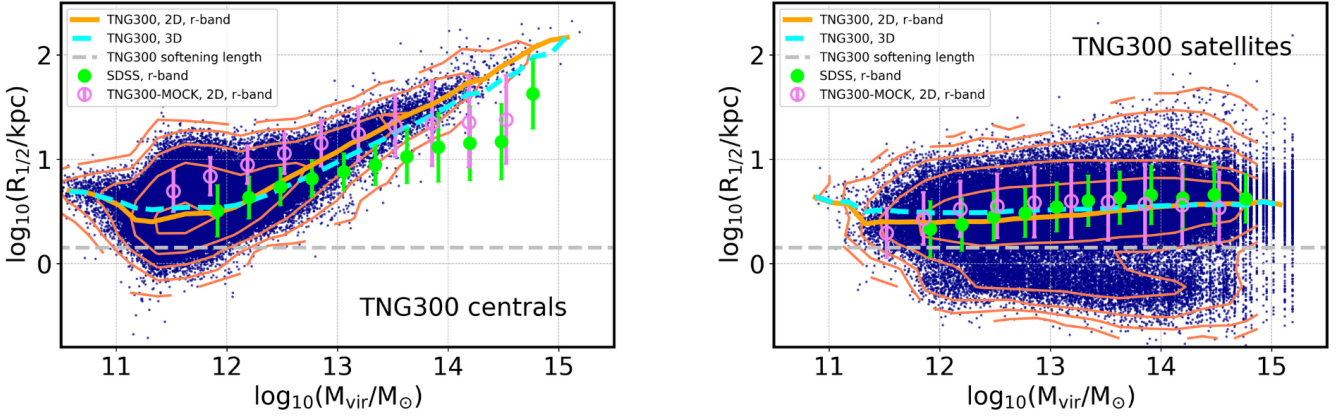
Despite the aforementioned differences, the qualitative agreement found between observational and simulation data reflects the ability of TNG300 to reproduce the main effect of galaxy formation and evolution processes. On the other hand, this result also demonstrates the effectiveness of our group identifier. Note that in observations, unlike in simulations, interlopers and/or misidentifications can easily blur the central/satellite galaxy classification, which would have an impact on our measured relations.

#### 4.1.3 Halo mass - size relation

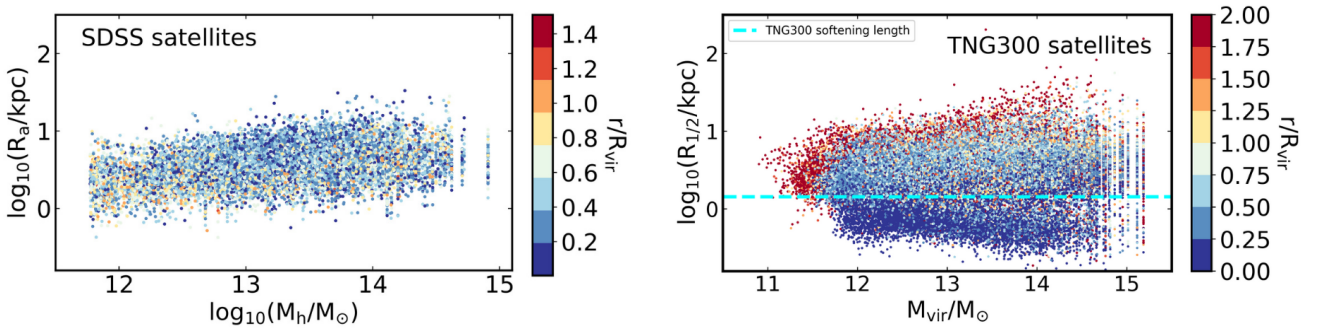
Galaxies form and evolve inside dark-matter haloes, which establishes an intrinsic connection that manifests itself in several halo



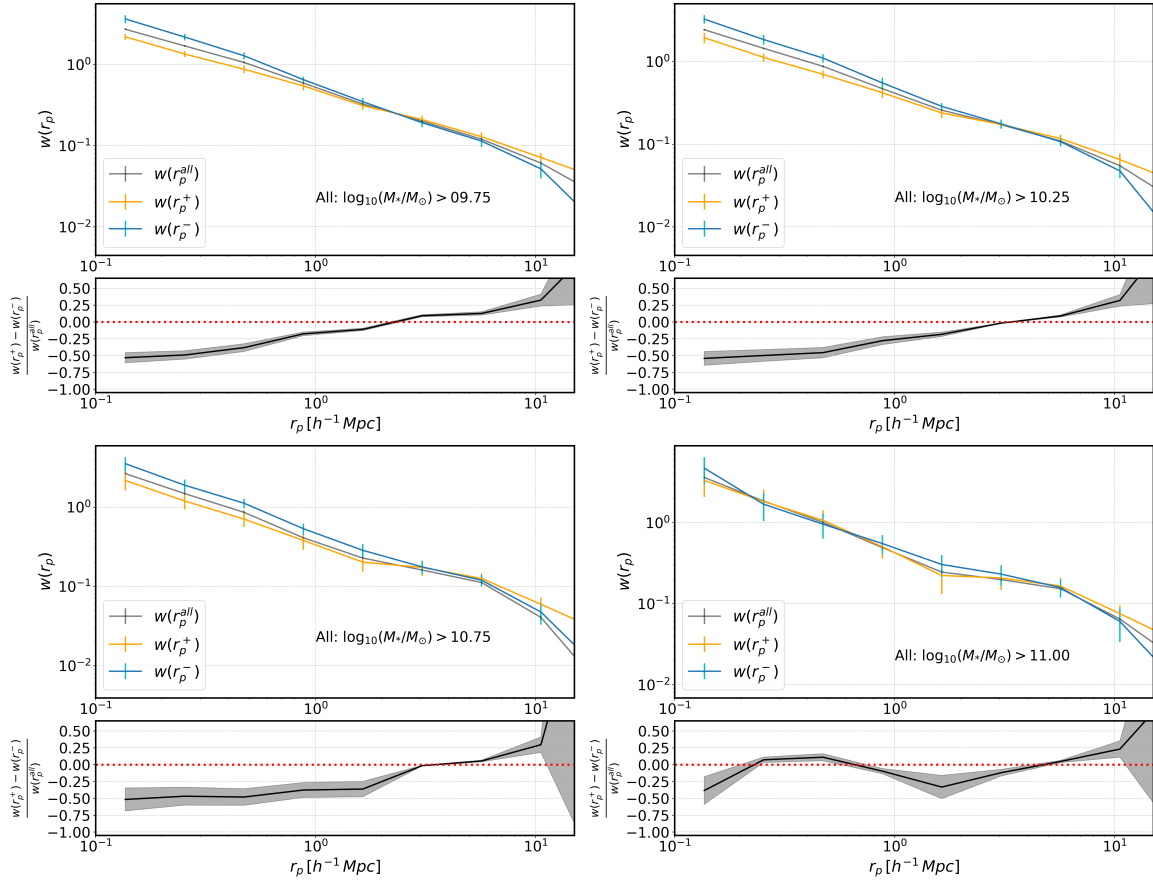
**Figure 6.** The halo mass–size relation from the SDSS sample in the same format as Figure 4, for central (left panel) and satellite galaxies (right panel) separately. A linear fit to these scaling relations and the models proposed by Kravtsov 2013 and Hearin et al. 2019 (adapted to our analysis, see text) have been included in both panels.



**Figure 7.** The halo mass–size relation from TNG300 in the same format as Figure 5, for central (left panel) and satellite galaxies (right panel) separately.



**Figure 8.** The dependence of the size of satellite galaxies on the halo-centric distance. Galaxies are colour-coded based on their projected distance to the centre of the halo, normalised by halo virial radius. The left-hand panel displays results for the SDSS whereas the right-hand panel shows the prediction from TNG300. The horizontal dashed line in the right-hand panel marks the TNG300 softening length.



**Figure 9.** The projected galaxy correlation function as a function of galaxy size in the combined sample, for four different stellar-mass thresholds. In each main panel, orange lines represent the correlation function for the 50 % subset containing the galaxies with larger radii,  $w(r_p^+)$ , whereas cyan lines show results for the remaining 50% smaller-radius subset,  $w(r_p^-)$ . Grey lines display the correlation functions for the entire sample above the corresponding stellar mass cut,  $w(r_p^{\text{all}})$ . Error bars correspond to the jackknife errors obtained from a set of 25 subsamples. In each subplot, the fractional difference between the clustering of larger and smaller galaxies is displayed (normalised to the total sample).

and galaxy properties. As shown in Section 4.1.1, the mass of central galaxies is tightly connected to the mass of the haloes that they inhabit. It is also well known that there is a relationship between the number of galaxies that occupy haloes and the halo mass (e.g., Berlind & Weinberg 2002; Zehavi et al. 2005; Zheng et al. 2005). However, the relationship between galaxy radius ( $R_a$ ) and halo mass has not been sufficiently explored. The size of galaxies can provide us with valuable information about their formation and evolution, whereas the mass of dark-matter haloes is related to the assembly history of their hosting haloes.

In Figure 6, we show, in the same format of Figure 4, the halo mass – size relation, i.e., galaxy size as a function of halo mass, for central (left) and satellite galaxies (right) separately. For central galaxies, as expected, there is a clear increasing trend by which more massive central galaxies are progressively larger as halo mass grows. To quantify the increase in radius, we fit a linear model, obtaining the following best-fitting relation:

$$\log_{10}(R_a^{\text{central}}) = a_c * \log_{10}(M_h) + b_c, \quad (3)$$

with  $a_c = 0.311 \pm 0.002$  and  $b_c = -3.182 \pm 0.009$ . This model is shown in a red dashed line in the left-hand panel of the Figure 6.

The above results are expected in the context of the hierarchical assembly framework. The most massive haloes are described

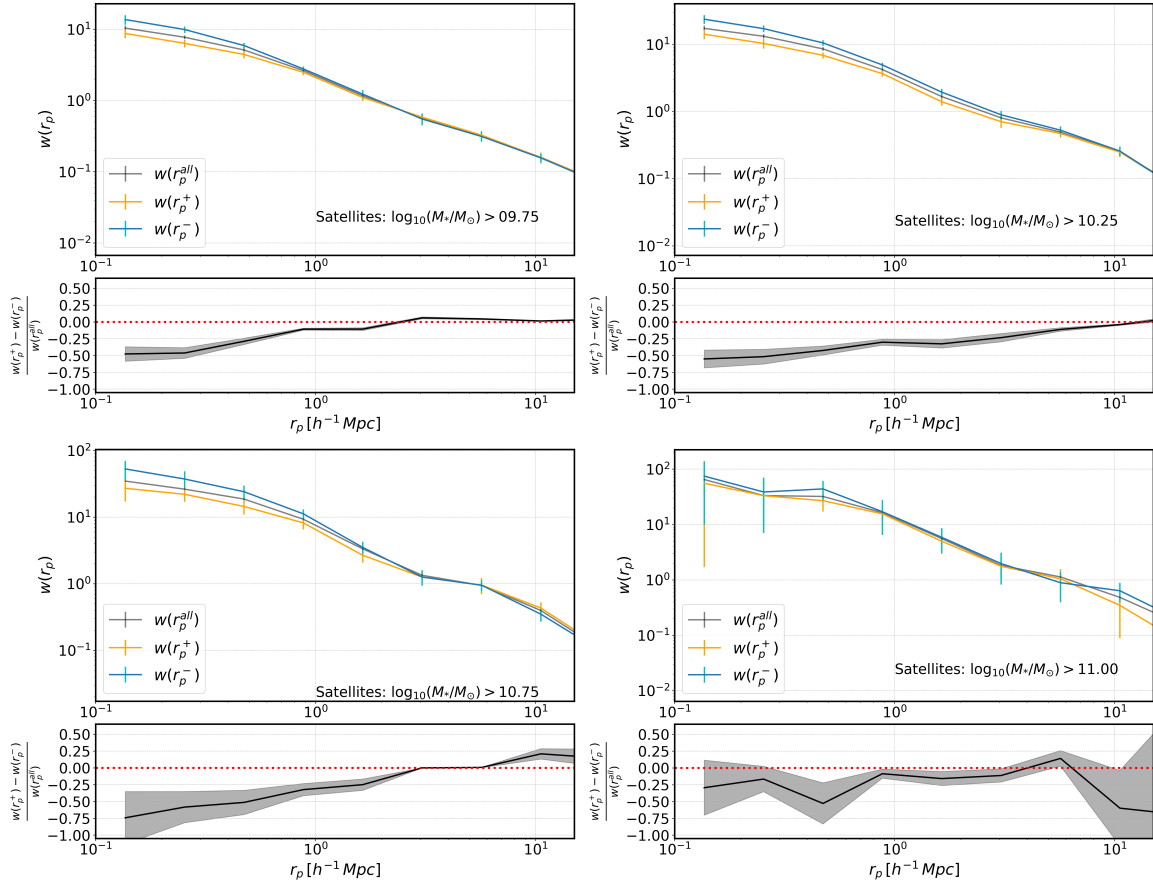
as the product of the fusion of many other smaller haloes and, as this process proceeds, the stellar content is expected to be affected. Central galaxies can accrete satellite galaxies in a process called galactic cannibalism (e.g., Ostriker & Tremaine 1975; White 1976; Malumuth & Richstone 1984; Merritt 1984; Aragón-Salamanca et al. 1998; De Lucia & Blaizot 2007) in which, due to dynamical friction, satellite galaxies lose energy and momentum, falling into the centre of the halo’s potential well and merging with the central galaxy. This simple mechanism could explain the simultaneous growth of the central galaxy and the halo mass in which they reside.

The size of the satellites, conversely, remains fairly constant throughout the mass range considered, with just a slight tendency for galaxies to be larger in more massive haloes. The blue dashed line in the right-hand panel of Figure 6 shows the linear fit to this trend. Namely:

$$\log_{10}(R_a^{\text{satellite}}) = a_s * \log_{10}(M_h) + b_s, \quad (4)$$

with  $a_s = 0.114 \pm 0.001$  and  $b_s = -0.972 \pm 0.006$ . The slope is, therefore, a factor  $\sim 3$  smaller for satellites. This result shows that satellite galaxies are affected by their host halo in a very different way than centrals are. This is expected, given the nature of satellites, which are external objects that are accreted by the halo and thus are not initially linked to it. The little correlation between satellite size and halo mass is also the result of satellites experiencing a





**Figure 10.** The projected galaxy correlation function as a function of galaxy size for satellite galaxies and four different stellar-mass thresholds. In each main panel, orange lines represent the correlation function for the 50 % subset containing the correlation function for the 50 % smaller-radius subset,  $w(r_p^+)$ , whereas cyan lines show results for the remaining 50% larger-radius subset,  $w(r_p^-)$ . Grey lines display the correlation functions for the entire satellite population above the corresponding stellar mass cut,  $w(r_p^{\text{all}})$ . Error bars correspond to the jackknife errors obtained from a set of 25 subsamples. In each subplot, the fractional difference between the clustering of larger and smaller satellite galaxies is displayed (normalised to the total satellite population).

combination of multiple physical processes inside haloes, including galaxy harassment, ram pressure, strangulation and/or tidal stripping (Gunn & Gott III 1972; Larson et al. 1980; Abadi et al. 1999; Van Gorkom 2004; Bosch et al. 2008; Pasquali & Nachname 2015).

Overall, our results indicate that the position that galaxies take in the group determines their growth. It is noteworthy that the properties of satellites, despite their secondary status within haloes, do appear to reflect, albeit slightly, those of their hosting haloes. This can also be understood in the context of hierarchical merging, since the largest haloes might be potentially capable of accreting larger haloes containing larger central galaxies (that are subsequently to become satellites).

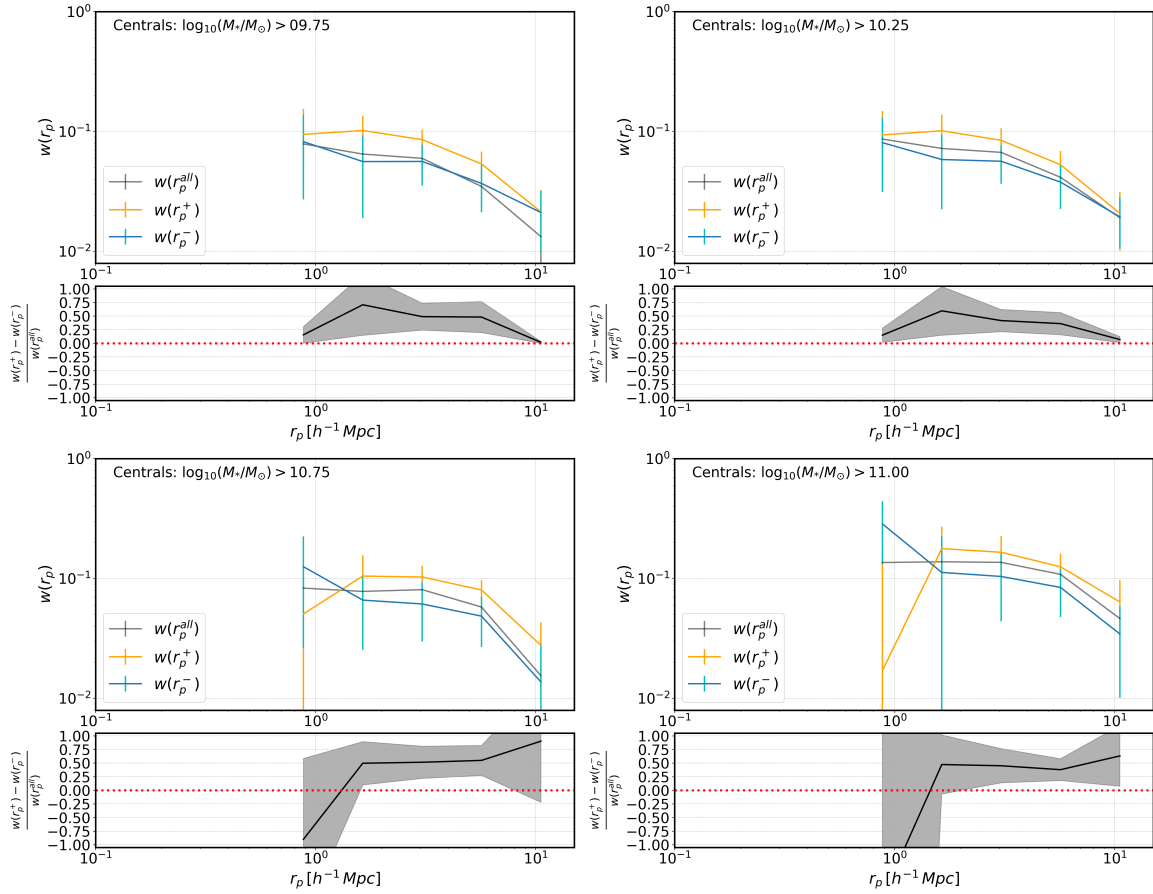
To compare our linear fits with models that link the galaxy size ( $R_{\text{gal}}$ ) to the virial radius of the halo ( $R_{\text{vir}}$ ), we show in Figure 6 the model proposed by Kravtsov (2013), i.e.,  $R_{\text{gal}} = 0.015 * R_{\text{vir}}$  (yellow line on the left-hand panel). We also compare our results with the similar model of Hearin et al. (2019):  $R_{\text{gal}} = 0.01 * R_{\text{vir}}$  (brown line). We find that our fit displays an excellent agreement with the Kravtsov (2013) model, which reinforces the idea that galaxy size is proportional to the virial radius. Figure 6 shows that, conversely, the Hearin et al. (2019) model underestimate the size of central galaxies.

Since the group finder does not provide an estimate of the masses of the subhaloes where the satellites reside, in order to

compare our results for satellites with the Kravtsov (2013) and Hearin et al. (2019) models, we produce an estimate based on the SHMR. We use the stellar mass of each satellite to estimate the mass of its subhalo and subsequently calculate its virial radius, the radius of each galaxy and the average radius in halo mass bins. It is important to note that this procedure may introduce slight differences since, for example, Hearin et al. (2019) uses the halo radius when the halo mass reached its maximum and it is not clear how good the approximation using SHMR is in this respect. The relation found by extrapolating our fit for central galaxies to the satellites is shown in red dashed line in the right-hand panel of Figure 6, whereas those for the Kravtsov (2013) and Hearin et al. (2019) models are represented by yellow and brown solid lines, respectively. As we can see, our fit and the Kravtsov (2013) model overestimate the radius of the satellite galaxies and, as expected, they agree with each other. On the other hand, the relation proposed by Hearin et al. (2019) predicts quite precisely the average radius of the satellite galaxies.

Our results on the halo mass – size relation are compared with predictions from TNG300 in Figure 7, which adopts the same format as Figure 5.

Again, the agreement with the TNG300 box and mock is qualitatively remarkable, both for central and for satellite galaxies. For central galaxies, the trend is again steeper for TNG300, but both



**Figure 11.** The projected galaxy correlation function as a function of galaxy size for central galaxies and four different stellar-mass thresholds. In each main panel, orange lines represent the correlation function for the subset containing 50 % of central galaxies with larger radii,  $w(r_p^+)$ , whereas cyan lines show results for the remaining 50% smaller-radius subset,  $w(r_p^-)$ . Grey lines display the correlation functions for the entire central population above the corresponding stellar mass cut,  $w(r_p^{\text{all}})$ . Error bars correspond to the jackknife errors obtained from a set of 25 subsamples. In each subplot, the fractional difference between the clustering of larger and smaller central galaxies is displayed (normalised to the full central-galaxy population).

sets of results have similar values. In the TNG300 mock, it is observed that the group finder introduces slight changes in the masses estimations and shows average sizes at high mass smaller than those of the simulation box. However, the behaviour is similar to that of the observations and do not affect the general tendency. For satellite galaxies, on the other hand, TNG300 predicts a subpopulation of small galaxies in low-mass haloes which are not observed in the SDSS. However, these objects have sizes comparable to the softening length of the simulation ( $\epsilon_{\text{dm,stars}} = 1.0 h^{-1}\text{kpc}$ ). In addition, we have checked that they all reside extremely close to the central galaxies of the corresponding haloes. Overall, the TNG300 display a flat halo mass – size relation, with mean size values that are almost identical to those measured in the SDSS. This agreement again reinforces our confidence in the robustness of the group member identification and halo-matching processes.

Finally, we have investigated potential secondary dependencies that can contribute to the scatter observed in the halo mass – size relations of Figure 7. For central galaxies, it seems that morphology could play a role. Disk-dominated central galaxies at fixed halo mass tend to be larger than bulge-dominated central galaxies, particularly towards the low-mass end. This dependence seems to align well with well-documented results on the stellar mass–size relation of early- and late-type galaxies (e.g. Shen et al. 2003; Kravtsov 2013; Lange et al. 2015).

The scatter in the halo mass – size relation of satellite galaxies could, in turn, be connected with physical processes such as tidal stripping, which are expected to be more severe in the proximity of the central galaxy. The right-hand panel of Figure 8 displays, in a colour code, the dependence on the halo-centric distance (here, the distance to the central galaxy in units of virial radii) for TNG300. Even though the smallest objects might be affected by resolution effects in the box, larger satellites seemingly tend to be located farther away from the central galaxies.

The above result from TNG300 would be consistent with the effect of tidal stripping, which would tend to remove stars from the outer regions of satellite galaxies, making them effectively smaller. However, this prediction is not confirmed in the SDSS, as the left-hand panel of Figure 8 demonstrates. This could be due to the uncertainties in the estimation of galaxy sizes from observations, which would tend to erase this small dependence. Alternatively, the level of tidal stripping implemented in TNG300 could be too high due to numerical artifacts and thus inconsistent with observations. Extending our analysis in the future using better photometric data will help alleviate this tension.

## 4.2 Clustering Analysis

In this section, we complement the measurements of scaling relations presented in Section 4.1 with a galaxy clustering analysis, which is also relevant for understanding the formation and evolution of galaxies and can provide clues about how central and satellite galaxies influence each other's formation. The dependence of clustering on galaxy size is addressed below for the SDSS and the TNG300 samples separately.

For the SDSS data, we measure clustering using the projected correlation function,  $w(r_p)$ , and the Landy & Szalay (1993) estimator. Therefore, we compute the following expression:  $W = (DD - DR - RD + RR)/RR$ , where DD is the number of galaxies, DR corresponds to the random pair counts, RR is the number of random centre-random tracer and RD is the random centre-galaxy pair counts. Several different subsets are analysed in this section (satellite/central, larger/smaller, mass thresholds) and for each of them, we compute a random catalogue that mimics its redshift distribution. Due to the complex selection function of the SDSS survey, we also make sure that the random distributions of points have the same selection function as the data, in order to normalise the galaxy pairs counts. This task is performed using the SDSS angle selection mask (described in Sec. 5.1 of Rodriguez et al. 2015).

To determine the uncertainties in the projected correlation function, we use a jackknife procedure. We generate 25 equal sub-samples and compute the correlation function after subtracting these subsamples. We also tested the results using 10, 50, 75, and 100 sub-samples. We checked that above 25 sub-samples, the variance of the measurement stabilises.

In order to analyse the dependence of galaxy clustering on galaxy size, the additional dependence on stellar mass must first be subtracted. This is performed following the simple procedure (detailed in Sec. 2.1 of Hearin et al. 2019). In essence, the galaxy population is binned in stellar mass, and each galaxy is classified as "large" or "small" based on whether they lie above or below the median size value at the corresponding stellar mass bin. Using this technique, for any  $M_*$  threshold sample, the stellar mass functions of the "large" and "small" subsamples are identical.

The first analysis that we perform is to split galaxies according to their size, independently of their membership status (central/satellite). This is equivalent to the analysis presented in Hearin et al. (2019). In several stellar-mass thresholds, we compute the median of the galaxy radius and create two subsets with the 50% of objects above and below this value (i.e., with larger and smaller radii, respectively). In Figure 9, we compare the clustering measured for the combined sample (grey), the larger galaxies (orange) and the smaller galaxies (cyan), for four different thresholds in stellar mass. Smaller galaxies are more tightly clustered than larger galaxies towards small scales for the  $\log_{10}(M_*/M_\odot) > 9.75, 10.25$  and  $10.75$  subsets, while no significant difference is detected for  $\log_{10}(M_*/M_\odot) > 11$ . On larger scales, the trend seems to invert. These results are consistent with those reported by Hearin et al. (2019), which support the notion that smaller galaxies cluster more strongly than larger galaxies on small scales and that this difference decreases with increasing stellar mass. As can be seen in the bottom panels, smaller galaxies can be up to 50% more strongly clustered than large galaxies of the same stellar mass below  $r \lesssim 3$  Mpc. At  $r \gtrsim 5$ , the trend inverts and larger galaxies begin to be more clustered.

The above results are in some sense misleading, since we are mixing central and satellite galaxies. In Section 4.1, we show significant differences in the scaling relations of central and satellite

galaxies when these are measured with respect to halo mass. We set out now to assess whether these differences have an impact on their clustering and whether the size dependence displayed in Figure 9 is altered when central and galaxies are treated independently. In this analysis, we generate the same subsets described above, but for central and satellite galaxies separately.

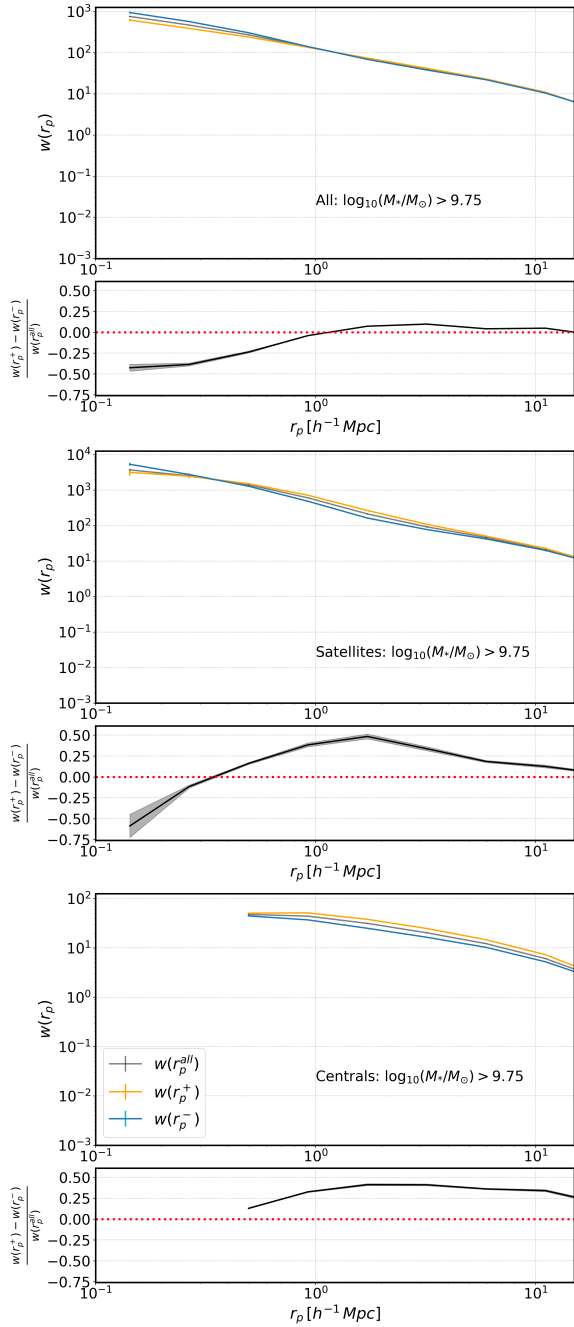
Figure 10 displays the projected correlation functions for satellite galaxies alone in the same format of Figure 9. As can be seen, there is a clear signal so that smaller galaxies are more tightly clustered than those with larger radii, although this effect decreases with stellar mass. A comparison with Figure 9 reveals that the clustering of satellites follows quite well that of the entire population on small scales.

We now focus on the clustering of central galaxies, shown in Figure 11. Note that here, only the 2-halo term of the correlation function is visible, which dictates that only scales above  $\sim 1$  Mpc are accessible. Within this range of scales, larger objects are more clustered than their smaller counterparts across the entire stellar mass range considered. Although the uncertainties are larger than those measured for the previous samples, we can say that larger central galaxies have typically 50 % higher clustering, across the stellar mass ranges considered. Again, a comparison with Figure 9 allows us to understand the measurement for the entire population: the inversion of the signal observed on larger scales is due to the increasing impact of central galaxies.

In summary, the results presented in this section provide a clarified picture for the dependence of galaxy clustering on galaxy size. When galaxies are separated exclusively by size (and not by group status), as shown in Figure 9, smaller objects appear more clustered than larger objects towards smaller scales, corresponding to the 1-halo term. This result is the consequence of satellites being dominant on these scales, since, as Figure 10 demonstrates, smaller satellites are more highly biased than larger satellites. On larger scales, conversely, the signal progressively inverts as we enter the 2-halo term, where the effect of centrals becomes more significant. Larger centrals have higher clustering amplitude than their smaller counterparts (Figure 11), which produces the inversion of the signal observed in Figure 9.

Following the same philosophy of Section 4.1, we have compared our SDSS clustering measurements with predictions from the TNG300 hydrodynamical simulation box. As mentioned above, TNG300 is one of the largest hydrodynamical boxes available to the community and one of the first where statistically-significant measurements of clustering can be performed (for certain scales and mass ranges, see, e.g., Montero-Dorta et al. 2020a). Figure 12 displays, in the same format of Figures 9, 10, and 11, the size dependence of the correlation function for the entire, the central, and the satellite galaxy samples, respectively, in TNG300. For the sake of simplicity, only the less restrictive mass threshold is shown here, i.e.,  $\log_{10}(M_*/M_\odot) = 9.75$ . Following Montero-Dorta et al. (2020a), the computation of errors is based on a standard jackknife technique, where the TNG300 box is divided in 8 sub-boxes ( $L_{\text{sub-box}} = L_{\text{box}}/2 = 102.5 h^{-1}\text{Mpc}$ ). The correlation function for each subset is measured in 8 different configurations of equal volume, obtained by subtracting one sub-box at a time. The errors correspond to the standard deviation of all individual configurations.

Figure 12 qualitatively reproduces the observational SDSS-based trends (note that the two measurements are, strictly speaking, different). For the entire sample, where central and satellite galaxies are treated indistinctly, smaller galaxies are found to be more tightly clustered than larger galaxies below  $r \approx 2$  Mpc. The difference on the smallest scales considered reaches  $\sim 75\%$ . Above this scale, the



**Figure 12.** The TNG300 projected galaxy correlation function, for the combined sample (upper panel), satellites (middle panel) and central galaxies (bottom panel), in the same format of Figures 9, 10 and 11. Errors are computed as the standard deviation obtained from a set of 8 jackknife configurations (see text).

dependence on galaxy size is weak. This is qualitatively similar to the results found for the SDSS (note that in the SDSS, the bias trend actually inverts).

When analysed separately, smaller TNG300 satellites cluster more tightly than their larger counterparts across the scale domain of the 1-halo term, with differences reaching over  $\sim 60\%$ . On larger scales, the signal inverts, which is something that is also observed (albeit to a lesser extent) for some mass thresholds in the SDSS. Central TNG300 galaxies, on the other hand, show more signal over

the 2-halo term, where larger centrals are up to  $\sim 20\%$  more biased than their smaller counterparts. This general agreement between observations and simulation is also recovered for the rest of the mass thresholds considered. Overall, this comparison makes us confident about the robustness of our results and reinforces, even more, the importance of treating satellite and central galaxies separately. We have also checked that similar results are obtained when the TNG300 mock instead of the box is employed.

## 5 DISCUSSION AND CONCLUSIONS

In this paper, we investigate the scaling relations and clustering of central and satellite galaxies separately, focusing on the dependence on galaxy size. We use data from the SDSS at  $z \sim 0.1$  that include accurate measurements of galaxy size and stellar mass (Meert et al. 2015; Kauffmann et al. 2004) and we further identify galaxy groups following the method proposed by Rodríguez & Merchán (2020). To establish a central/satellite galaxy classification scheme within each group, we use a simple criterion where the most massive and luminous galaxy is chosen as the central, with the remaining members being labelled satellites. We compare our results with measurements from the TNG300 simulation, which serves as a means of confirming the robustness of our central/satellite identification.

Regarding the size-related scaling relations, we show for the SDSS and the TNG300 simulation that central and satellite galaxies display significantly different behaviours. For the stellar mass-size relation, the sizes of central galaxies are only slightly larger than those of satellites. However, these differences amplify drastically when the halo mass - size relation is analysed. Our linear fits to the SDSS measurement indicate that the slope is a factor  $\sim 3$  smaller for satellites as compared to centrals. These relations also allow us to demonstrate that our results are in excellent agreement with the model proposed by Kravtsov (2013), in which galaxy radius is proportional to the virial radius of the hosting halo times a constant 0.015. For satellites, on the other hand, our results are more consistent with the model proposed by Hearin et al. (2019), where the proportionality constant is 0.01.

Our results illustrate the fact that the evolution of galaxies is different depending on their location and status within the halo, with central galaxies displaying a much more pronounced increase in size as compared to satellites (as halo mass grows). Importantly our findings are consistent with the TNG300 predictions, which demonstrates the robustness of our measurements. The disparity in the halo mass - size relations reflects the fact that centrals are intrinsically more connected to the halo than satellites, which are accreted over time. It also reflects the different physical mechanisms that underpin their evolution. Central galaxies accrete material from the surrounding galaxies that pertain to the group and could even absorb entire galaxies in a process called galactic cannibalism (e.g., Ostriker & Tremaine 1975; White 1976; Malumuth & Richstone 1984; Merritt 1984; Aragón-Salamanca et al. 1998; De Lucia & Blaizot 2007). Satellite galaxies, on the other hand, suffer from various physical processes that will prevent their growth and cause them to lose material, namely, i.e., galaxy harassment, ram pressure, strangulation and/or tidal stripping (Gunn & Gott III 1972; Larson et al. 1980; Abadi et al. 1999; Van Gorkom 2004; Bosch et al. 2008; Pasquali & Nachname 2015).

The above arguments are enriched by the fact that satellite galaxies also seem to be on average slightly larger in more massive haloes. In the context of hierarchical merging, it would be conceivable that larger haloes were more likely to accrete other large



haloes, which would tend to host larger central galaxies. This is an interesting aspect that will be addressed in more detail in follow up works.

We have also investigated the potential secondary dependencies of the halo-centric distance for satellite galaxies. Our results in the SDSS do not show a clear dependence of size on the distance to the centre. However, in IllustrisTNG300 we see a slight trend where it seems that the smaller satellite galaxies are closer to the centre, which would be in favour of tidal stripping. Although this result is not very conclusive considering the smoothing length, it would be interesting to explore this further in both simulations and observations.

By studying the dependence of clustering on galaxy size, we have provided a clarified picture that helps to understand previous measurements. When this dependence is studied independently of the central/satellite galaxy classification (i.e., for the entire sample) smaller galaxies tend to be more tightly clustered on small scales, with the trend inverting at the large-scale end. These results are qualitatively in agreement with the measurements of [Hearin et al. \(2019\)](#). However, when central and satellite galaxies are treated independently, the intrinsic trends are revealed. For satellite galaxies, smaller objects are indeed more highly biased than larger objects within the entire scale range considered. For central galaxies, conversely, the larger objects are the ones that are more clustered, within the 2-halo term scale range accessible (large scales). We have checked that these behaviours are also predicted by TNG300, from a qualitative standpoint.

The general agreement we found throughout this work between the groups identified in the SDSS and the haloes of TNG300, in addition to ensuring the robustness of our scaling relations and clustering results, adds another level of confidence to the group identification method of [Rodríguez & Merchán \(2020\)](#). On the other hand, the scaling relations and the differences between the clustering of central and satellite galaxies shown could be useful in terms of improving the mass estimates and the member assignment in group finders.

Finally, although we have not addressed it here, our well-calibrated machinery can be used to perform clustering measurements at fixed halo mass for both central and satellite galaxies separately. This type of measurements are relevant in the context of galaxy assembly bias (see, e.g., [Zentner et al. 2016](#); [Miyatake et al. 2016](#); [Zu et al. 2016](#); [Lin et al. 2016](#); [Sunayama et al. 2016](#); [Montero-Dorta et al. 2017](#); [Artale et al. 2018](#); [Niemic et al. 2018](#); [Zehavi et al. 2018](#); [Walsh & Tinker 2019](#); [Sunayama & More 2019](#); [Obuljen et al. 2020](#); [Montero-Dorta et al. 2020a,b](#)). The excellent classification of central and satellite galaxies that we have achieved with the SDSS can also be employed to investigate the manifestation of physical processes that drive the evolution of galaxy populations, particularly with the improved data quality of upcoming surveys.

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding authors.

## ACKNOWLEDGEMENTS

We thank the referee for the helpful comments that helped improve this work. We also appreciate Mariangela Bernardi responses re-

garding the use of the data. FR gratefully acknowledges comments from Mario Abadi.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No 734374. ADMD thanks FAPESP for financial support. MCA acknowledges financial support from the Austrian National Science Foundation through FWF stand-alone grant P31154-N27. REA acknowledges the support of the ERC-StG number 716151 (BACCO). FR and MM thanks the support by Agencia Nacional de Promoción Científica y Tecnológica (PICT 2015-3098), the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) and the Secretaría de Ciencia y Tecnología de la Universidad Nacional de Córdoba (SeCyT-UNC, Argentina).

## REFERENCES

- Abadi M. G., Moore B., Bower R. G., 1999, *Monthly Notices of the Royal Astronomical Society*, 308, 947
- Abazajian K. N., et al., 2009, *The Astrophysical Journal Supplement Series*, 182, 543
- Angulo R. E., Baugh C. M., Lacey C. G., 2008, *MNRAS*, 387, 921
- Aragón-Salamanca A., Baugh C. M., Kauffmann G., 1998, *Monthly Notices of the Royal Astronomical Society*, 297, 427
- Artale M. C., Zehavi I., Contreras S., Norberg P., 2018, *Monthly Notices of the Royal Astronomical Society*, 480, 3978
- Baldry I. K., Balogh M. L., Bower R., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, *Monthly Notices of the Royal Astronomical Society*, 373, 469
- Balogh M. L., Navarro J. F., Morris S. L., 2000, *The Astrophysical Journal*, 540, 113
- Behroozi P. S., Conroy C., Wechsler R. H., 2010, *The Astrophysical Journal*, 717, 379
- Behroozi P. S., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 1546
- Berlind A. A., Weinberg D. H., 2002, *The Astrophysical Journal*, 575, 587
- Bernardi M., Meert A., Sheth R., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 697
- Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R., 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 874
- Bosch F. C., Pasquali A., Yang X., Mo H., Weinmann S., McIntosh D. H., Aquino D., 2008, arXiv preprint arXiv:0805.0002
- Bottrell C., Torrey P., Simard L., Ellison S. L., 2017, *MNRAS*, 467, 2879
- Bruzual G., Charlot S., 2003, *Monthly Notices of the Royal Astronomical Society*, 344, 1000
- Campbell D., van den Bosch F. C., Hearin A., Padmanabhan N., Berlind A., Mo H., Tinker J., Yang X., 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 444
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *The Astrophysical Journal*, 647, 201
- Contreras S., Angulo R., Zennaro M., 2020, arXiv e-prints, [p. arXiv:2005.03672](#)
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- De Lucia G., Blaizot J., 2007, *Monthly Notices of the Royal Astronomical Society*, 375, 2
- DeFelippis D., Genel S., Bryan G. L., Fall S. M., 2017, *The Astrophysical Journal*, 841, 16
- Dekel A., Burkert A., 2014, *Monthly Notices of the Royal Astronomical Society*, 438, 1870
- Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, 399, 497
- Favole G., Montero-Dorta A. D., Prada F., Rodríguez-Torres S. A., Schlegel D. J., 2018, *MNRAS*, 480, 1415
- Ferguson H. C., et al., 2004, *The Astrophysical Journal Letters*, 600, L107

- Fernández Lorenzo M., Sulentic J., Verdes-Montenegro L., Argudo-Fernández M., 2013, *Monthly Notices of the Royal Astronomical Society*, 434, 325
- Gao L., White S. D. M., 2007, *MNRAS*, 377, L5
- Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
- Genel S., et al., 2014, *MNRAS*, 445, 175
- Genel S., et al., 2018, *MNRAS*, 474, 3976
- Gunn J. E., Gott III J. R., 1972, *The Astrophysical Journal*, 176, 1
- Han J., Li Y., Jing Y., Nishimichi T., Wang W., Jiang C., 2018, preprint, ([arXiv:1802.09177](https://arxiv.org/abs/1802.09177))
- Hearin A. P., Watson D. F., 2013, *MNRAS*, 435, 1313
- Hearin A. P., Zentner A. R., Berlind A. A., Newman J. A., 2013, *MNRAS*, 433, 659
- Hearin A. P., Watson D. F., Becker M. R., Reyes R., Berlind A. A., Zentner A. R., 2014, *MNRAS*, 444, 729
- Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016, *MNRAS*, 460, 2552
- Hearin A., Behroozi P., Kravtsov A., Moster B., 2019, *Monthly Notices of the Royal Astronomical Society*, 489, 1805
- Huang S., Ho L. C., Peng C. Y., Li Z.-Y., Barth A. J., 2013, *The Astrophysical Journal*, 766, 47
- Huang K.-H., et al., 2017, *The Astrophysical Journal*, 838, 6
- Huchra J., Geller M., 1982, *The Astrophysical Journal*, 257, 423
- Huertas-Company M., Shankar F., Mei S., Bernardi M., Aguerri J. A. L., Meert A., Vikram V., 2013, *ApJ*, 779, 29
- Ichikawa T., Kajisawa M., Akhlaghi M., 2012, *Monthly Notices of the Royal Astronomical Society*, 422, 1014
- Jiang F., et al., 2019, *Monthly Notices of the Royal Astronomical Society*, 488, 4801
- Jiang F., Dekel A., Freundlich J., Bosch F. C., Green S. B., Hopkins P. F., Benson A., Du X., 2020, *arXiv preprint arXiv:2005.05974*
- Johnson J. W., Maller A. H., Berlind A. A., Sinha M., Holley-Bockelmann J. K., 2019, *MNRAS*, 486, 1156
- Kauffmann G., et al., 2003, *MNRAS*, 341, 33
- Kauffmann G., White S. D., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, *Monthly Notices of the Royal Astronomical Society*, 353, 713
- Kawamata R., Ishigaki M., Shimasaku K., Oguri M., Ouchi M., 2015, *The Astrophysical Journal*, 804, 103
- Kravtsov A. V., 2013, *The Astrophysical Journal Letters*, 764, L31
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottloeber S., Allgood B., Primack J. R., 2004, *The Astrophysical Journal*, 609, 35
- Lacerna I., Rodríguez-Puebla A., Avila-Reese V., Hernández-Toledo H. M., 2014, *The Astrophysical Journal*, 788, 29
- Landy S. D., Szalay A. S., 1993, *The Astrophysical Journal*, 412, 64
- Lange R., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 447, 2603
- Larson R., Tinsley B., Caldwell C. N., 1980, *The Astrophysical Journal*, 237, 692
- Lazeyras T., Musso M., Schmidt F., 2017, *J. Cosmology Astropart. Phys.*, 2017, 059
- Li Y., Mo H. J., Gao L., 2008, *MNRAS*, 389, 1419
- Lin Y.-T., Mandelbaum R., Huang Y.-H., Huang H.-J., Dalal N., Diemer B., Jian H.-Y., Kravtsov A., 2016, *ApJ*, 819, 119
- Malumuth E., Richstone D., 1984, *The Astrophysical Journal*, 276, 413
- Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, *MNRAS*, 368, 715
- Mao Y.-Y., Zentner A. R., Wechsler R. H., 2018, *MNRAS*, 474, 5143
- McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, *Monthly Notices of the Royal Astronomical Society*, 383, 593
- Meert A., Vikram V., Bernardi M., 2013, *Monthly Notices of the Royal Astronomical Society*, 433, 1344
- Meert A., Vikram V., Bernardi M., 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 3943
- Merritt D., 1984, *The Astrophysical Journal*, 276, 26
- Miyatake H., More S., Takada M., Spergel D. N., Mandelbaum R., Rykoff E. S., Rozo E., 2016, *Physical Review Letters*, 116, 041301
- Mo H., Mao S., White S. D., 1998, *Monthly Notices of the Royal Astronomical Society*, 295, 319
- Montero-Dorta A. D., et al., 2017, *ApJ*, 848, L2
- Montero-Dorta A. D., Artale M. C., Abramo L. R., Tucci B., 2020a, *arXiv e-prints*, p. [arXiv:2008.08607](https://arxiv.org/abs/2008.08607)
- Montero-Dorta A. D., et al., 2020b, *MNRAS*, 496, 1182
- Mosleh M., et al., 2012, *The Astrophysical Journal Letters*, 756, L12
- Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, *The Astrophysical Journal*, 710, 903
- Mowla L., van der Wel A., van Dokkum P., Miller T. B., 2019, *The Astrophysical Journal Letters*, 872, L13
- Naab T., Johansson P. H., Ostriker J. P., 2009, *The Astrophysical Journal Letters*, 699, L178
- Nelson D., et al., 2019, *Computational Astrophysics and Cosmology*, 6, 2
- Niemiec A., et al., 2018, *MNRAS*,
- Obuljen A., Percival W. J., Dalal N., 2020, *arXiv e-prints*, p. [arXiv:2004.07240](https://arxiv.org/abs/2004.07240)
- Ono Y., et al., 2013, *The Astrophysical Journal*, 777, 155
- Ostriker J., Tremaine S., 1975, *The Astrophysical Journal*, 202, L113
- Paranjape A., Padmanabhan N., 2017, *MNRAS*, 468, 2984
- Pasquali A., Nachname V., 2015, *Astronomische Nachrichten*, 336, 505
- Pillepich A., et al., 2018, *MNRAS*, 473, 4077
- Planck Collaboration Ade P. A., et al., 2016, *Astronomy & Astrophysics*, 594, A13
- Poggianti B., et al., 2012, *The Astrophysical Journal*, 762, 77
- Ramakrishnan S., Paranjape A., Hahn O., Sheth R. K., 2019, *MNRAS*, 489, 2977
- Robertson B., Bullock J. S., Cox T. J., Di Matteo T., Hernquist L., Springel V., Yoshida N., 2006, *The Astrophysical Journal*, 645, 986
- Rodríguez F., Merchán M., 2020, *A&A*, 636, A61
- Rodríguez F., Merchán M., Sgró M. A., 2015, *A&A*, 580, A86
- Salcedo A. N., Maller A. H., Berlind A. A., Sinha M., McBride C. K., Behroozi P. S., Wechsler R. H., Weinberg D. H., 2018, *MNRAS*, 475, 4411
- Salcedo A. N., et al., 2020, *arXiv e-prints*, p. [arXiv:2010.04176](https://arxiv.org/abs/2010.04176)
- Sato-Polito G., Montero-Dorta A. D., Abramo L. R., Prada F., Klypin A., 2019, *MNRAS*, 487, 1570
- Shankar F., Bernardi M., 2009, *Monthly Notices of the Royal Astronomical Society: Letters*, 396, L76
- Shen S., Mo H., White S. D., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *Monthly Notices of the Royal Astronomical Society*, 343, 978
- Sheth R. K., Tormen G., 2004, *MNRAS*, 350, 1385
- Shibuya T., Ouchi M., Harikane Y., 2015, *The Astrophysical Journal Supplement Series*, 219, 15
- Somerville R. S., Davé R., 2015, *Annual Review of Astronomy and Astrophysics*, 53, 51
- Spindler A., Wake D., 2017, *Monthly Notices of the Royal Astronomical Society*, 468, 333
- Springel V., 2010, *MNRAS*, 401, 791
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Sunayama T., More S., 2019, *MNRAS*, 490, 4945
- Sunayama T., Hearin A. P., Padmanabhan N., Leauthaud A., 2016, *MNRAS*, 458, 1510
- Trujillo-Gomez S., Klypin A., Primack J., Romanowsky A. J., 2011, *ApJ*, 742, 16
- Trujillo I., et al., 2004, *The Astrophysical Journal*, 604, 521
- Trujillo I., et al., 2006, *The Astrophysical Journal*, 650, 18
- Tucci B., Montero-Dorta A. D., Abramo L. R., Sato-Polito G., Artale M. C., 2020, *arXiv e-prints*, p. [arXiv:2007.10366](https://arxiv.org/abs/2007.10366)
- Vale A., Ostriker J., 2004, *Monthly Notices of the Royal Astronomical Society*, 353, 189
- Van Gorkom J. H., 2004, *Clusters of Galaxies: Probes of Cosmological Structure and Galaxy Evolution*, p. 305
- Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G., 2010, *Monthly Notices of the Royal Astronomical Society*, 409, 1379
- Vogelsberger M., et al., 2014a, *MNRAS*, 444, 1518

- Vogelsberger M., et al., 2014b, *Nature*, 509, 177
- Walsh K., Tinker J., 2019, *MNRAS*, 488, 470
- Wechsler R. H., Tinker J. L., 2018, *ARA&A*, 56, 435
- Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *ApJ*, 652, 71
- Weinberger R., et al., 2017, *MNRAS*, 465, 3291
- Weinmann S. M., Kauffmann G., Van Den Bosch F. C., Pasquali A., McIntosh D. H., Mo H., Yang X., Guo Y., 2009, *Monthly Notices of the Royal Astronomical Society*, 394, 1213
- Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, *Monthly Notices of the Royal Astronomical Society*, 432, 336
- White S. D., 1976, *Monthly Notices of the Royal Astronomical Society*, 174, 19
- White S. D. M., Frenk C. S., 1991, *ApJ*, 379, 52
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Williams R. J., Quadri R. F., Franx M., Van Dokkum P., Toft S., Kriek M., Labbé I., 2010, *The Astrophysical Journal*, 713, 738
- Xu X., Zehavi I., Contreras S., 2020, arXiv e-prints, p. [arXiv:2007.05545](https://arxiv.org/abs/2007.05545)
- Yang X., Mo H., Van Den Bosch F. C., Jing Y., 2005, *Monthly Notices of the Royal Astronomical Society*, 356, 1293
- Yang X., Mo H. J., van den Bosch F. C., 2009, *The Astrophysical Journal*, 693, 830
- Zanisi L., et al., 2020, *Monthly Notices of the Royal Astronomical Society*, 492, 1671
- Zehavi I., et al., 2005, *The Astrophysical Journal*, 630, 1
- Zehavi I., Contreras S., Padilla N., Smith N. J., Baugh C. M., Norberg P., 2018, *The Astrophysical Journal*, 853, 84
- Zentner A. R., Hearin A., van den Bosch F. C., Lange J. U., Villarreal A., 2016, preprint, ([arXiv:1606.07817](https://arxiv.org/abs/1606.07817))
- Zentner A. R., Hearin A., van den Bosch F. C., Lange J. U., Villarreal A., 2019, *MNRAS*, 485, 1196
- Zhang Y.-C., Yang X.-H., 2019, *Research in Astronomy and Astrophysics*, 19, 006
- Zheng Z., et al., 2005, *The Astrophysical Journal*, 633, 791
- Zu Y., Mandelbaum R., Simet M., Rozo E., Rykoff E. S., 2016, preprint, ([arXiv:1611.00366](https://arxiv.org/abs/1611.00366))
- van der Wel A., et al., 2014, *The Astrophysical Journal*, 788, 28