



# How Do Galaxy Properties Affect Void Statistics?

Rushy R. Panchal<sup>1</sup>, Alice Pisani<sup>2</sup> , and David N. Spergel<sup>2,3</sup> 

<sup>1</sup>Princeton University, Department of Computer Science, 35 Olden Street, Princeton, NJ 08540, USA; [rpanchal@alumni.princeton.edu](mailto:rpanchal@alumni.princeton.edu)

<sup>2</sup>Princeton University, Department of Astrophysical Sciences, 4 Ivy Lane, Princeton, NJ 08540, USA; [apisani@astro.princeton.edu](mailto:apisani@astro.princeton.edu)

<sup>3</sup>Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA

Received 2020 April 3; revised 2020 July 21; accepted 2020 August 9; published 2020 September 24

## Abstract

Using mapping from dark matter halos to galaxy properties based on hydrodynamical simulations, we explore the impact of galaxy properties on the void size function and the void–galaxy correlation function. We replicate the properties of galaxies from *Illustris* on *MassiveNus* halos, to perform both luminosity and star formation rate cuts on *MassiveNus* halos. We compare the impact of such cuts on void properties with respect to cuts on halo mass (as usually performed on halo catalogs driven from  $N$ -body simulations). We find that void catalogs built from luminosity-selected galaxies and halos are consistent within errors, while void catalogs built from star formation rate-selected galaxies differ from void catalogs built on halos. We investigate the reason for this difference. Our work suggests that voids built on galaxy catalogs (selected through luminosity cuts) can be reliably studied using halos in dark matter simulations.

*Unified Astronomy Thesaurus concepts:* [Large-scale structure of the universe \(902\)](#); [Voids \(1779\)](#); [Catalogs \(205\)](#); [Surveys \(1671\)](#); [Cosmology \(343\)](#); [Hydrodynamical simulations \(767\)](#); [N-body simulations \(1083\)](#)

## 1. Introduction

Cosmic voids, the large underdense regions in the galaxy distribution, are a novel probe for cosmology (see Pisani et al. 2019 and references therein). Due to their low-density nature, they are particularly sensitive probes of dark energy, modified gravity, and neutrino properties (Lee & Park 2009; Odrzywołek 2009; Biswas et al. 2010; Lavaux & Wandelt 2010, 2012; Bos et al. 2012; Sutter et al. 2012, 2015; Clampitt et al. 2013; Spolyar et al. 2013; Carlesi et al. 2014; Barreira et al. 2015; Massara et al. 2015; Pisani et al. 2015; Zivick et al. 2015; Hamaus et al. 2016, 2017; Pollina et al. 2016; Voivodic et al. 2017; Cautun et al. 2018; Kreisch et al. 2019; Schuster et al. 2019; Perico et al. 2019; Pisani et al. 2019; Verza et al. 2019). Voids span sizes of tens to hundreds of  $h^{-1}$  Mpc. Their use for cosmology relies on averaged quantities such as their void size function or the stacked density profile (also known as the void–galaxy cross-correlation function)—needing large samples of voids.

In cosmology, theoretical models for voids are mainly built and tested on halo catalogs from large dark matter (DM)  $N$ -body simulations (see, e.g., Hamaus et al. 2014). In  $N$ -body simulations, halos are found using algorithms such as friends-of-friends (Springel et al. 2001; Dolag et al. 2009) and are then used as tracers to construct void catalogs. Ideally, model testing should be done on simulations able to mimic galaxy properties matching observations: mock galaxies would be ideal as tracers, since the ultimate goal is to apply those models to data (e.g., Hamaus et al. 2016, 2017; Hawken et al. 2017).

In the best case, however, large  $N$ -body simulations are populated with Halo Occupation Distribution (HOD) techniques (Zheng et al. 2005) to replicate the properties of a particular galaxy population—this is still far from optimal (Hadzhiyska et al. 2020), as it fails to include all the complicated effects impacting galaxy formation, and does not account for the impact of selection cuts performed on galaxies from observations. Typically, large simulations are unable to include the small-scale evolution of stars and gas and fail to reproduce the observed population ratios of spiral and elliptical

galaxies (Vogelsberger et al. 2014a). Indeed, modeling the baryonic component requires being able to mimic stars, gas, supermassive black holes, and their feedback. This is necessary to allow a robust prediction of galaxy properties (e.g., stellar content, morphology) on small scales and, consequently, to reliably reproduce galaxy populations over large scales. Such effects are only accounted for in hydrodynamical simulations.

Current hydrodynamical simulations model galaxy populations and features in great detail (including, among others, effects such as primordial and metal-line gas cooling, star formation, stellar feedback, supermassive black hole formation, growth, and feedback), but the simulation of all these effects is computationally expensive and can only be performed for relatively small simulation sizes (e.g., *Illustris*, Genel et al. 2014) up to a few hundred  $h^{-1}$  Mpc. Consequently, all recent works testing the extraction of cosmological information from voids rely on halo catalogs from DM simulations or HOD, implicitly assuming that the use of halos or HOD-populated fields as tracers of the cosmic web—instead of galaxies—to measure void observables is robust, and that the impact of such an assumption is small enough to be negligible (or, in any case, smaller than the error in the measurements of cosmological parameters).

As of today this assumption remains untested. With the increase in void numbers promised by upcoming surveys such as DESI (DESI Collaboration et al. 2016), Euclid (Laureijs et al. 2011), SPHEREx (Doré et al. 2018), and WFIRST (Spergel et al. 2015), errors in measurements of void related quantities will drop dramatically. The era of precision cosmology from voids steadily approaches: it becomes critical to test the impact of galaxy properties on void observables used for cosmology. Typically, when performing void finding in simulations we rely on halos above a certain mass.

The mass of halos is often modeled as correlated with galaxy luminosity or galaxy star formation rate (SFR). Most  $N$ -body simulations will assume a linear relationship between the halo mass and the galaxy luminosity (or galaxy SFR) to build halo catalogs. Nevertheless, it is well known that there is a

considerable scatter in this relationship (Guo et al. 2011; Greco et al. 2015). Very massive halos can correspond to low luminosity or low SFR galaxies. Voids are built using a tracer field; in observations the used tracers are galaxies, while in simulations we use halos. Void catalogs from simulations will then intrinsically be impacted by this scatter. To allow a proper understanding of applications to data, such scatter needs to be quantified.

In this work we build a tool to replicate the properties of the *Illustris* galaxy population on a larger simulation, the *MassiveNus*  $\Lambda$ CDM. We then construct void catalogs from *MassiveNus*, using as tracers both the DM halos and the reconstructed galaxies. This allows us to analyze void properties in both the original halo catalog and in the galaxy catalog obtained by replicating the properties of *Illustris* on the larger *MassiveNus* simulation. The comparison quantifies the impact on void observables of the scatter in the mass-to-luminosity (or SFR) relationship.

The paper is organized as follows. Section 2 introduces the hydrodynamical and  $N$ -body simulations used for this work, as well as the void-finding algorithm. Section 3 presents the method used to replicate the properties of galaxies from *Illustris* in the large  $N$ -body *MassiveNus* simulation. Section 4 describes the void observables on which we test the impact of the tracer used to build void catalogs, and presents our results. We present our conclusions in Section 5.

## 2. Simulations and Void Finder

In this paper we rely on two simulations: the *Illustris* simulation provides us with a framework to reproduce galaxy properties relying on halos from the larger  $N$ -body simulation *MassiveNus*.

### 2.1. *Illustris*

The *Illustris* (Vogelsberger et al. 2014b) project is a set of hydrodynamical particle simulations aiming to simultaneously be complex enough to model galaxy formation, large enough to model evolution on cosmic scales, and resolved enough that small (as well as large) structures can be studied. The *Illustris* project grows out of previous work with the *Millennium* simulation and allows us to connect galactic scales to the large-scale structure of the universe.<sup>4</sup>

Cosmological parameters reflect a  $\Lambda$ CDM cosmology, with  $\Omega_m = 0.2726$ ,  $\Omega_b = 0.0456$ , and  $\Omega_\Lambda = 0.7274$ ,  $h = 0.704$ ,  $\sigma_8 = 0.809$ ,  $n_s = 0.963$ ; the parameters follow the Wilkinson Microwave Anisotropy Probe-9 measurements (Hinshaw et al. 2013). Minimal changes of cosmological parameters are not expected to affect the results of the paper. Initial conditions for the simulation are obtained with N-GenIC (Springel 2005).<sup>5</sup> The simulations consist of a total of  $1820^3$  dark matter particles and  $1820^3$  gas tracer particles, each with a particle mass of  $m_{\text{DM}} = 4.4 \times 10^6 M_\odot/h$  in a  $75 h^{-1}$  Mpc length cube ( $\simeq 106.5$  Mpc) (Genel et al. 2014; Vogelsberger et al. 2014a, 2014b; Nelson et al. 2015) built using the moving mesh code AREPO (Springel 2010). The *Illustris* simulation includes

<sup>4</sup> We note that a larger *IllustrisTNG* has been recently developed, with different sizes available (50, 100 and 300 Mpc), an extended mass range of galaxies and halos, and other improvements (Nelson et al. 2015). It is a valid option to consider for further work.

<sup>5</sup> More details can be found here: <http://www.illustris-project.org/data/downloads/Illustris-1/>.

modeling of subgrid physics, radiative gas cooling, star formation, galactic-scale winds from star formation feedback, supermassive black hole formation, accretion, and feedback. The simulation provides halo catalogs built with friend-of-friends (Davis et al. 1985). Redshift snapshots span from  $z = 0.0$  to  $z = 127.0$ . For this work we train our models on the  $z = 0.0$  halos data set, containing  $7.7 \times 10^6$  halos.

### 2.2. *MassiveNus*

*MassiveNus*<sup>6</sup> (Cosmological Massive Neutrino Simulations) (Liu et al. 2018) are a set of 101  $N$ -body simulations with sizes of  $512 h^{-1}$  Mpc each. The simulations have a  $\Lambda$ CDM version, as well as boxes including the effect of massive neutrinos. Massive neutrino modeling (for the boxes that we do not use in this work) is done by evolving neutrinos perturbatively and considering clustering with a nonlinear cold DM potential. Other parameters are varied in the 101 versions: aside from the sum of neutrinos masses, the matter density  $\Omega_m$  and the primordial curvature power spectrum  $A_s$  are changed. Each box has  $1024^3$  particles and the simulations span redshifts from 0 to 45. The publicly available simulated data include simulation snapshots and halo catalogs that are relevant for our work.

Initial conditions are obtained with an improved version of N-GenIC (Springel 2005), S-GenIC; simulations are built with the public tree-Particle Mesh code Gadget-2 (Springel 2005). The halo catalogs are generated with the halo finder code Rockstar (Behroozi et al. 2013), relying on a friends-of-friends algorithm that groups close particles and finds substructures within parent halos. The minimum halo mass in the simulations is roughly  $10^{11} M_\odot h^{-1}$ , while the number of high-mass halos ( $> 10^{14}$ ) is slightly below model predictions due to the box size.

Cosmological parameters for the massless neutrino  $\Lambda$ CDM simulation snapshot at  $z = 0$  used in this work are:  $A_s = 2.1 \times 10^{-9}$ ,  $\Omega_m = 0.3$ ,  $\Omega_{\text{rad}} = 0$ ,  $h = 0.7$ ,  $n_s = 0.97$ ,  $w = -1$ , and  $\Omega_b = 0.05$ . As for the *Illustris* simulation, we note that minimal changes of cosmological parameters are not expected to affect the results of the paper. The volume is large enough to observe a considerable statistics of voids, while still resolving relatively low-mass halos; this makes *MassiveNus* the most suitable simulation to test the impact of the halo mass-to-luminosity relationship scatter on void properties.

### 2.3. Void Finder

This project uses the *VIDE*<sup>7</sup> void finder (Sutter et al. 2015, embedding *ZOBOV* Neyrinck 2008), which operates in three steps: the tessellation of space, creation of a density field, and finally, a watershed transformation. The tessellation of space is performed as a Voronoi tessellation where each input particle, galaxy, or dark matter halo acts as a tracer of the large-scale structure.

The Voronoi tessellation (see, e.g., Barr & Schoenberg 2010) has been widely used in astrophysics (see, e.g., Ebeling & Wiedenmann 1993; van de Weygaert 2007; Cappellari 2009; van de Weygaert & Schaap 2009; Soares-Santos et al. 2010). It associates to each tracer a cell (called a Voronoi cell) that contains all of the points that are closer to the tracer itself than any other tracer. Each cell is assigned a uniform density,

<sup>6</sup> *MassiveNus* snapshots, halo catalogs, merger trees, and galaxy and CMB lensing convergence maps are publicly available at <http://ColumbiaLensing.org>.

<sup>7</sup> Publicly available at [https://bitbucket.org/cosmicvoids/vid\\_e\\_public](https://bitbucket.org/cosmicvoids/vid_e_public).

inversely proportional to its volume. This creates a density field across all of space. Cells corresponding to local density minima are merged with higher-density cells surrounding them, until local density maxima are reached. The watershed transform effectively merges smaller basins into larger ones. The merged basins are the detected voids.

VIDE has been used in many recent applications for cosmology (examples include Sutter et al. 2014a, 2015; Hamaus et al. 2015, 2016; Pisani et al. 2015; Pollina et al. 2017, 2019; Contarini et al. 2019; Kreisch et al. 2019; Schuster et al. 2019; Verza et al. 2019), both to test theoretical models and to constrain parameters with data—it will likely be used in upcoming years to analyze data from future surveys and for model testing with future state-of-the-art mocks and simulations.<sup>8</sup> It is therefore important to check the impact of tracer selection properties on void finding.<sup>9</sup>

### 3. Tracer Selection

This section describes the algorithm to reproduce the relationship between halos and galaxies modeled in *Illustris* on *MassiveNus* halos. *Illustris* properly takes into account the scatter in the relationship, thanks to the realistic implementation of galaxies’ properties. Rich data from *Illustris* allow us to bridge the computational gap between high-resolution, complex simulations, and larger simulations with relatively lower resolution and complexity—namely, *MassiveNus*. We build a predictive model mapping a halo’s mass to the properties of the galaxy within that halo. The model can then be utilized to probabilistically assign galaxy properties to a halo. It associates galaxy-like properties to pre-existing halos, by assigning galaxy features (luminosity and SFR), mapped from the halo mass, to existing halo positions.

In particular we focus on two galaxy properties that will impact selection when using real data: the luminosity of a galaxy and the SFR. *Illustris* allows us to link those properties to halo mass, reproducing the intrinsic scatter present in the relation.

We now describe how to build the model with *Illustris* and how to subsequently apply it to *MassiveNus*.

#### 3.1. Constructing the Model

Given the nature of the distribution of halo masses with luminosity and SFR, different groups of galaxies will have different probability distributions that accurately link halo masses to properties. To construct the model, we first split our sample of halo masses in even-width groups (the same

<sup>8</sup> Extensive tests of different techniques to find voids have been considered in the literature (see, e.g., Colberg et al. 2008; Cautun et al. 2018). Here we focus on the void finder VIDE: while we do expect that different void finders might be impacted differently, most recent applications using voids for cosmology rely on either VIDE or are based on ZOBOV—hence our results cover work from a broad range of groups.

<sup>9</sup> An interesting application for further work would be to focus on analyzing generalizations of the Voronoi smoothing technique (such as smoothers based on the empty space function  $F$ ; see, e.g., Baddeley et al. 2015), and in particular the impact of other techniques on the results of this work. Nevertheless, we note that if a different void finder were to be used, there would be no need to repeat the training, as the void finding procedure is performed after assigning galaxy properties to *MassiveNus* halos. If a different halo finder were to be used in the *Illustris* simulations, the mapping between galaxies and halos could change, but the scatter on the relationship between halos and galaxies should be preserved. In this case, rerunning the training would ensure an enhanced robustness against changes due to the halo definition. Globally we do not expect changes in either the halo definition or the void definition to impact our overall results.

number of halos in each). Each group will therefore have an adequate number of data points. We use five halos per group, but tested different combinations. Our goal is to find the optimal model for each group of galaxies. To do so, independently for each group, we find the discrete random variable that describes the distribution of the desired property in that group.

To find the discrete random variable for each group, we consider the dark matter halo masses for each galaxy in the group as a random variable and fit a discrete distribution. We compute a histogram with 100 bins over the data within the group itself. In other words, we use a normalized histogram of the dark matter masses. As an example, if a bin with average dark matter mass  $m$  contains half of the total halos in that group, then  $\text{PMF}(m) = 0.5$ . We repeat for each group to find the set of discrete random variables that best describes each case.

From the random variable associated with each group we can draw properties for each halo. Through this process, the two properties (luminosity and SFR) are independently assigned to the halos.

The final model is then the set of random variables and the boundaries defining the groups.

#### 3.2. Applying the Model

After building the model, we first test it on *Illustris* data. Figure 1 shows the probabilistic fits for luminosity and SFRs overlaid on top of original data from *Illustris*. The probabilistic distribution is reproducing the target distribution for the whole range of halo masses in both the luminosity case and the SFR case, with a mean squared error of, respectively,  $1.8 \times 10^9 h^{-1} M_{\odot}$  and  $0.021 (M_{\odot} \text{yr}^{-1})^2$ .

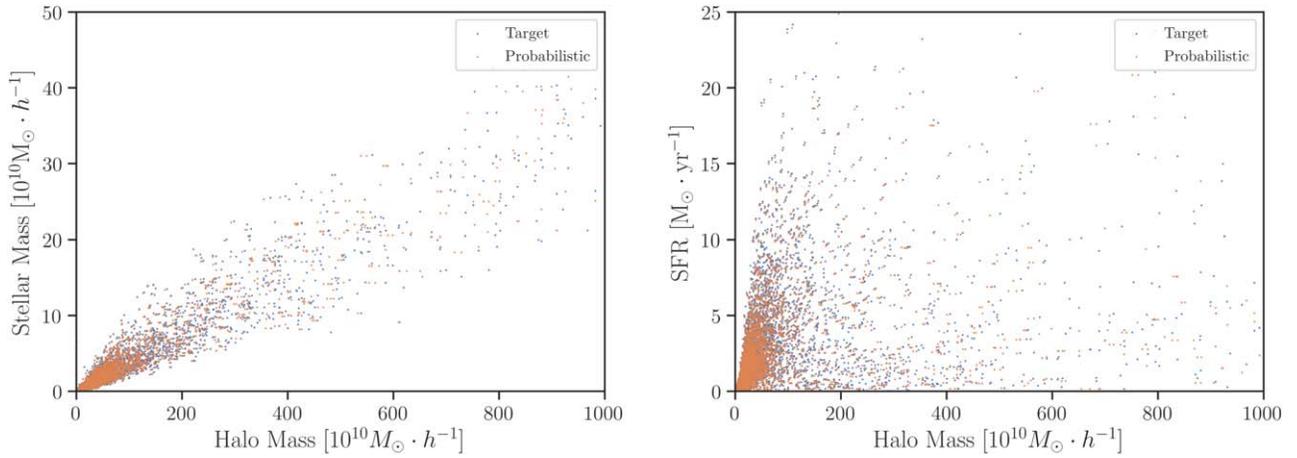
Our model is subsequently applied to *MassiveNus* by fitting each halo within the original slices: the probabilistic approach allows to replicate both galaxies luminosities, or SFRs on *MassiveNus* halos.

After assigning galaxy-like properties to halos, we now have three different sets of tracers from *MassiveNus* to perform void finding:

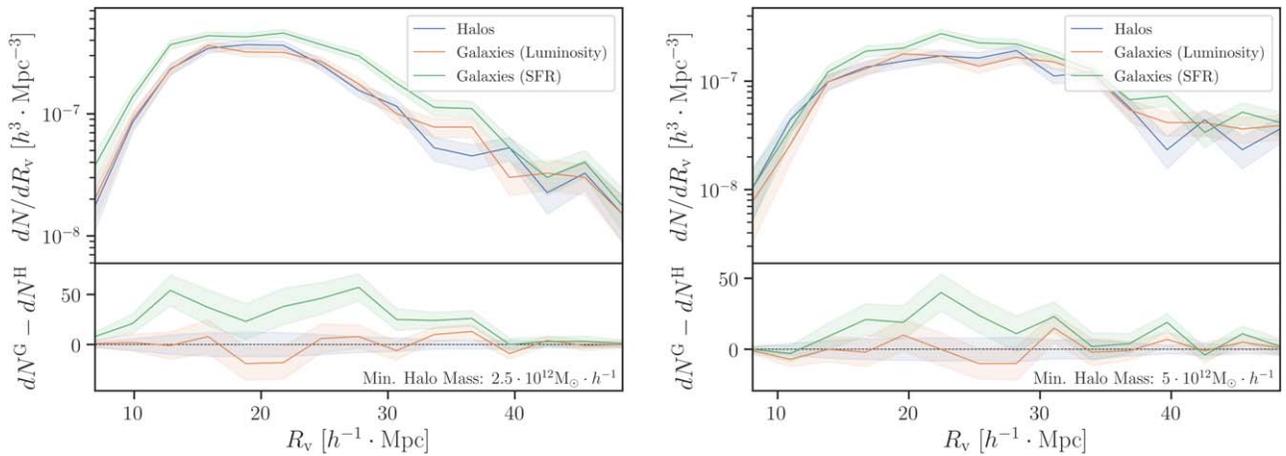
1. Halo catalog: the original output from the Rockstar halo finder provided as part of *MassiveNus*.
2. Luminosity-ordered galaxy catalog: built with our probabilistic approach from *MassiveNus* halos.
3. SFR-ordered galaxy catalog: built with our probabilistic approach from *MassiveNus* halos.

We wish to replicate what is done in simulations and observations, to test and subsequently apply models.

In simulations, a mass cut is performed on the original set of halos. To replicate what is done in observations, we apply a cut on galaxy catalogs chosen in terms of descending luminosity or SFR, keeping a number of resultant galaxies equivalent to the number obtained from the simulation mass cut. That is, if the simulation mass cut results in  $n$  halos, the galaxies with the  $n$  highest luminosity or SFR, respectively, are taken. The cuts on galaxy-like catalogs mimic real data treatment. We note the importance of keeping the same number of tracers, as void number density can be sensitive to tracer number density, both in terms of minimum mean particle separation (impacting smaller voids to be detected on a tracer catalog) and of box length (impacting the statistic of large voids and potentially their observed maximum size). We note that we perform two different mass cuts, and replicate tracer



**Figure 1.** Probabilistic fits for the luminosity (left) and SFR (right) overlaid on top of the original data.



**Figure 2.** Void abundances for the  $2.5 \times 10^{12} h^{-1} M_{\odot}$  set (left) and the  $5 \times 10^{12} h^{-1} M_{\odot}$  set (right). The bottom panels show the relative differences between halo-traced and galaxy-traced cases. The shaded region for each abundance is the Poisson error.

numbers in each case on the luminosity and SFR-ordered tracer catalogs. The mass cuts are  $M_{\min} = 2.5 \times 10^{12} h^{-1} M_{\odot}$  and  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$ .

We obtain six void catalogs from the tracers described in the items above (three kinds of tracers, considering two different mass cuts for each case). The comparison of void properties in the catalogs will allow us to estimate the impact of the scatter in the relationship between halo mass and luminosity or SFR-selected tracer catalogs.

In the next section we describe the two main void observables we focus our attention on, the void size function and the void density profile, and we present our results.

## 4. Results: Impact on Voids

### 4.1. Void Size Function

The void size function—the number of voids as a function of their radius, also known as void abundance—is a sensitive probe of cosmology. In particular it is sensitive to the properties of dark energy, modified gravity, and neutrinos (e.g., Massara et al. 2015; Pisani et al. 2015, 2019; Zivick et al. 2015; Sahlén et al. 2016; Kreisch et al. 2019; Sahlén 2019; Schuster et al. 2019; Verza et al. 2019).

Aside from being a relevant cosmological observable, the void size function provides an estimate of the number of voids

to be observed by an hypothetical survey. Since the number of voids will impact error bars on measurements of the density profiles or other void features sensitive to cosmology, it is important that the void size function is modeled correctly. Recent developments have established a robust framework to predict void abundance based on an improved version of the popular Sheth and Van de Weygaert model (an excursion set model, Sheth & van de Weygaert 2004): the volume conserving model (Vdn, Jennings et al. 2013). This model has proven successful in predicting void numbers from dark matter simulations, using both dark matter and halos (see recent results in Contarini et al. 2019; Ronconi et al. 2019; Verza et al. 2019), and accounting for halo bias in standard  $\Lambda$ CDM and different dark energy models.<sup>10</sup>

The impact on void abundance of the scatter in the relationship between halo mass and luminosity or SFR-selected tracer catalogs has yet to be estimated—a relevant point to consider as the void size function is becoming an established tool to use with upcoming data.

We run the void finder on all the tracer catalogs described in Section 3 and compare the results. We show in Figure 2 the comparison of the void size function. Error bars are obtained considering Poisson error. Within error bars, voids obtained

<sup>10</sup> Interestingly, the match with theoretical predictions is done with the same void finder `VIDE` that we use in our work.

from the halo distribution have the same void size function as voids selected from galaxies using a luminosity cut. Consistency remains for the two different mass cuts used, showing that testing theory and models on voids from halo catalogs can be used as a reliable proxy of data application, where galaxies are selected through their luminosity. We conclude that *void abundances are robust against the scatter present in the relationship between the dark matter halo mass and the luminosity of galaxies.*

Given the volume of the MassiveNus simulation, the tracer number densities for the  $M_{\min} = 2.5 \times 10^{12} h^{-1} M_{\odot}$  tracer catalog and the  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$  correspond, respectively, to a tracer number density of  $\simeq 1.9 \times 10^{-3} h^3 \text{Mpc}^{-3}$  and  $\simeq 9.7 \times 10^{-4} h^3 \text{Mpc}^{-3}$ . These number densities roughly span the ranges of tracer densities to be observed by upcoming surveys (including, for example, DESI, PFS, Euclid or WFIRST—for some of the redshift bins, since the latter is expected to exceed these number densities for the most populated bins). This work thus mimics the level of precision that can be expected, for a given redshift bin, with future surveys, and shows that void abundance measurements will be robust against the scatter present in the halo mass-to-galaxy luminosity relationship.

On the other hand, the selection of galaxies using SFR provides a void size function with higher values compared to voids from halos. While the global trend is similar for void abundance from halos and void abundance from SFR-selected galaxies, the overall value is different, and is due to an overall higher void number in the SFR galaxies case.

The larger number of voids in the SFR catalogs can be attributed to a bias in how tracers are selected for this type of threshold (see Section 4.3). The galaxies in this catalog are chosen by taking the  $n$  galaxies with the highest SFRs.

Galaxies within clusters have low SFRs, compared to those that are not in clusters, because they have less gas available from which to form stars. Thus, the galaxies with the highest SFRs, which are precisely the ones chosen for these catalogs, will be those outside of clusters; the SFR-cut tracer catalogs are biased toward galaxies that are distributed more sparsely. This is confirmed, for example, by the lower value of the two-point correlation function for the probabilistic SFR galaxy catalogs with respect to the two-point correlation function of halos (which we do not show).

Selecting a more sparsely distributed set of tracers will result in splitting up larger voids by adding galaxies inside them, and consequently increase the overall void statistics. Since the SFR-cut tracer catalogs are biased toward galaxies that are distributed more sparsely, the splitting of large voids can be linked to the mass of the host halos. Galaxies within clusters have low SFRs, therefore SFR-ordered galaxies (with higher SFR) are associated with halos with lower masses that have lower bias.

We note that for the  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$  tracer catalog, the increase in void abundance is lower, confirming the fact that it is due to a bias tracer selection effect. Indeed, with the  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$  tracer catalog, we are selecting more highly biased tracers, making the splitting of voids harder, and hence resulting in a lower difference in void abundance. Since current void size function models take into account the bias of tracers to build the theoretical prediction, we expect this shift to be correctly modeled—provided that the tracer bias can be either measured or theoretically predicted.

## 4.2. Density Profiles

We now estimate the impact of the scatter in the relationship between halo mass to luminosity and SFR of tracers for void density profiles. Density profiles of voids are mathematically equivalent to the cross-correlation between void centers and tracers (galaxies, halos, dark matter particles) (see, e.g., Cai et al. 2016; Hamaus et al. 2017; Massara & Sheth 2018), representing the variation of density contrast as a function of the distance from the void center. The void-tracer cross-correlation function is then underdense in the center, and the density increases toward void edges, up to the overdense shell that surrounds the void (Sheth & van de Weygaert 2004; Paz et al. 2013; Hamaus et al. 2014; Pisani et al. 2014).

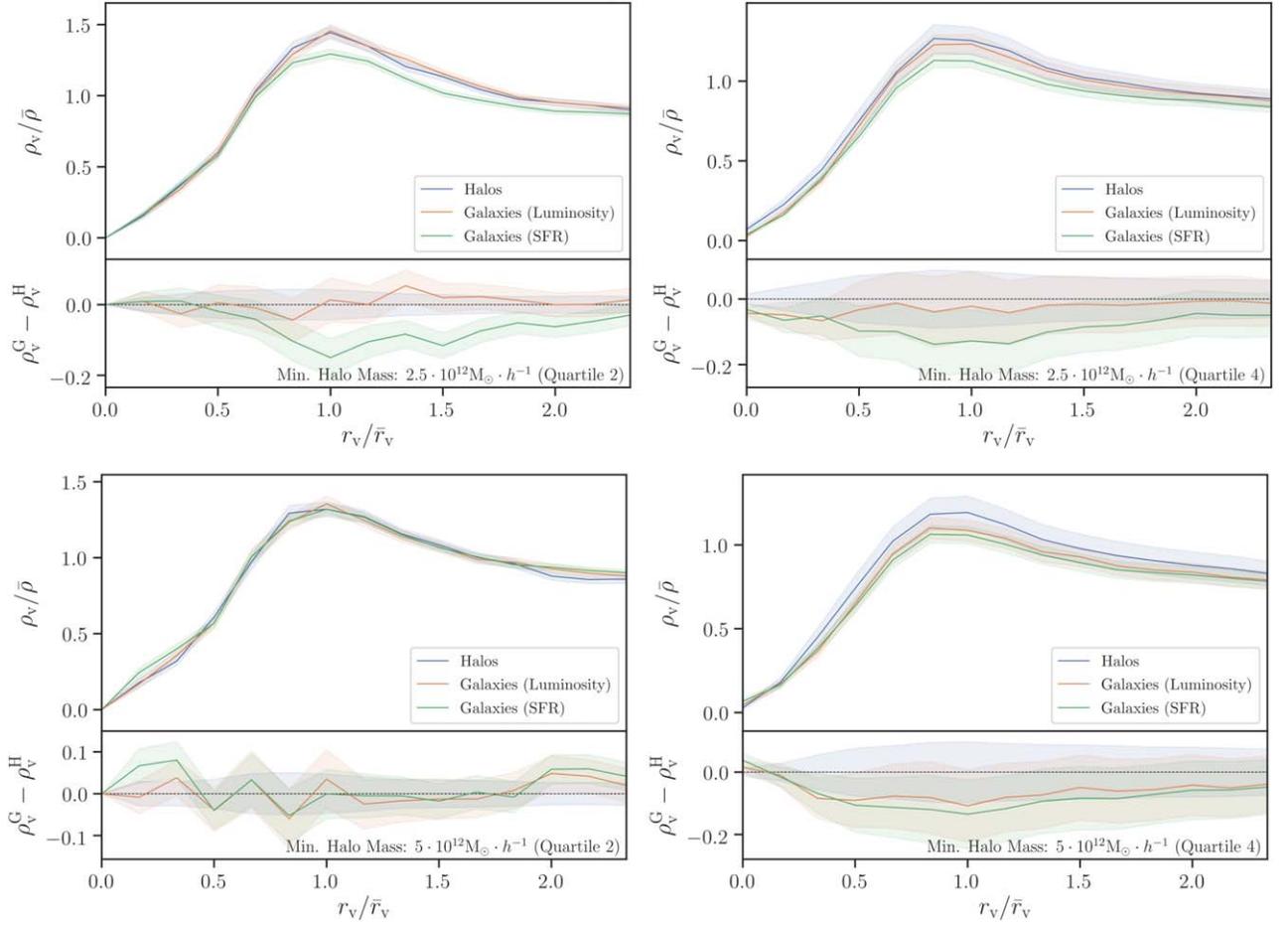
Void density profiles are powerful tools to constrain cosmology: the shape of void stacks is used as a standard sphere for the Alcock–Paczynski test (Lavaux & Wandelt 2012; Sutter et al. 2014b; Hamaus et al. 2016), or to analyze the pattern of redshift-space distortions around voids and constrain the growth rate of structure (Cai et al. 2016; Hamaus et al. 2017). The measure of the void-galaxy cross-correlation function is, as of today, the only void observable that provided constraints on cosmological parameters from data (e.g., Hamaus et al. 2016, 2017; Achitouv et al. 2017; Hawken et al. 2017; Achitouv 2019).

We measure the density profiles from the different void catalogs described in Section 3. Figure 3 shows the results for different radius bins.<sup>11</sup> Error bars are obtained considering Poisson error. We note that profiles from different tracers are mostly consistent within error bars, with some exception for the  $M_{\min} = 2.5 \times 10^{12} h^{-1} M_{\odot}$  case, where the SFR-galaxy based catalog results in a lower wall. This is consistent with the interpretation provided for the void size function: with a lower mass cut on halos, the SFR-galaxy catalog will be biased toward galaxies residing closer to void centers, hence selecting a different population of voids, defined by less biased tracers and hence with lower walls. For this reason the SFR void density profile has a lower ridge.

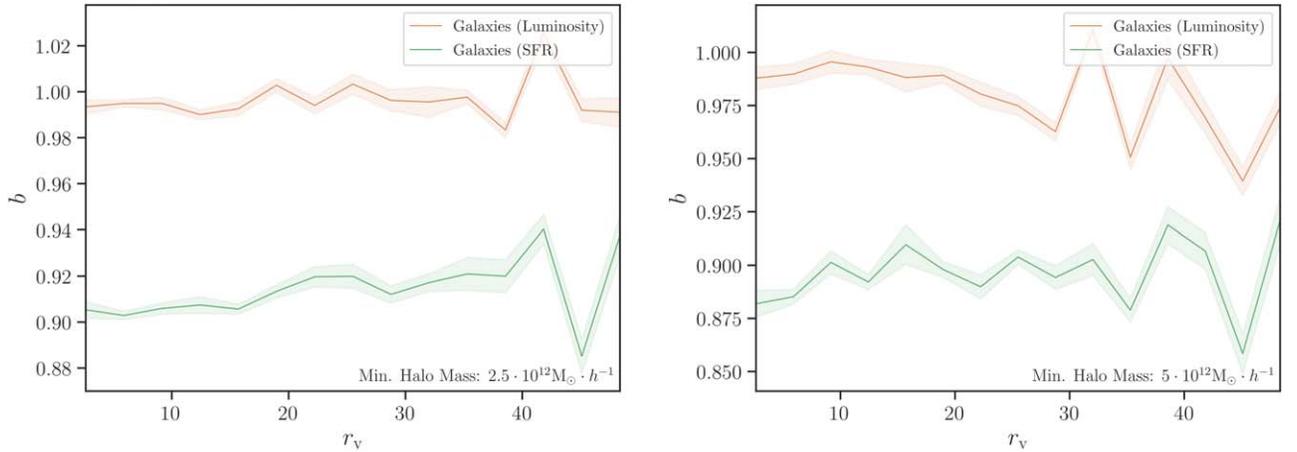
This difference disappears when using the  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$ : in this case profiles from the different cases overlap within error bars, as expected. Two reasons can explain the better match in this case. First, the  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$  case has less voids, and hence larger error bars, than the  $M_{\min} = 2.5 \times 10^{12} h^{-1} M_{\odot}$  case, mitigating the differences between the halos and galaxy cases. Second, since the  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$  case selects more strongly biased galaxies as a starting tracer catalog (than the  $M_{\min} = 2.5 \times 10^{12} h^{-1} M_{\odot}$  case) the impact of the SFR selection—which selects less biased objects among the available ones—is lower.

We conclude that *void density profiles are robust against the scatter present in the relationship between the dark matter halo mass and the luminosity of galaxies.* For SFR-selected galaxies, void density profiles will also be robust for upcoming surveys with tracer number densities lower than  $\simeq 1 \times 10^{-3} h^3 \text{Mpc}^{-3}$  (corresponding to the  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$  case), while void profiles from denser surveys will only be robust if galaxies are luminosity or flux-selected.

<sup>11</sup> Note that we split in quartiles, for the  $M_{\min} = 2.5 \times 10^{12} h^{-1} M_{\odot}$  case; quartile 2 includes voids from roughly 16–21  $h^{-1} \text{Mpc}$  and quartile 4 includes voids above 26  $h^{-1} \text{Mpc}$ ; for the  $M_{\min} = 5 \times 10^{12} h^{-1} M_{\odot}$  case, quartile 2 includes voids from roughly 20–26  $h^{-1} \text{Mpc}$  and quartile 4 includes voids above 32  $h^{-1} \text{Mpc}$ .



**Figure 3.** Density profiles, of the second (left) and fourth (right) quartiles of void radii, for the  $2.5 \times 10^{12} M_{\odot} h^{-1}$  set (top panels) and  $5 \times 10^{12} h^{-1} M_{\odot}$  set (bottom panels). The bottom subplots for each panel show the relative differences between halo-traced and galaxy-traced cases. The shaded region for each profile is the error as the normalized standard deviation.



**Figure 4.** Relative tracer bias for the  $2.5 \times 10^{12} h^{-1} M_{\odot}$  set (left) and  $5 \times 10^{12} h^{-1} M_{\odot}$  set (right).

In the next section we wish to verify that bias is indeed different for galaxies selected with respect to luminosity or SFR, to confirm our interpretation of the void size function and density profile differences.

#### 4.3. Tracer Bias

We calculate the relative bias of the galaxies selected by luminosity and by SFR with respect to halos (error bars are

obtained through jackknife resampling). As expected Figure 4 shows that SFR-selected galaxies have a lower bias than the luminous mass-selected galaxies, justifying the observed higher abundance and the density profiles with lower walls.

## 5. Conclusion and Discussion

In this work we have tested the assumption that the void size function and the void-galaxy cross-correlation function are

robust against the scatter in the halo mass-to-luminosity, or halo mass-to-SFR relationship for tracers used to build void catalogs. Mimicking the large  $N$ -body simulation *MassiveNus* the scatter observed in the *Illustris* simulation, we find that both the abundance and the void density profiles remain consistent within error bars for galaxies selected through luminosity. This result sets the ground to use void properties in upcoming data and robustly apply models developed with halos from  $N$ -body simulations. We note that using larger simulation volumes would reduce error bars, but also likely lead to a lower number density of tracers due to the decrease in resolution; this would mimic a harsher luminosity cut, and hence a stronger bias: the effects of the SFR selection should be increasingly mitigated.

Future applications could involve machine-learning techniques such as those used by He et al. (2019) to improve the population of large  $N$ -body simulations with galaxies, as well as testing our results on larger simulations—provided that a sufficiently low halo mass is reached even with simulations spanning larger volumes (that is, maintaining a high resolution). While the void cross-correlation function is already currently used to extract competitive cosmological constraints from voids (e.g., relying on SDSS data and VIPERS data), steps that we leave for future work (such as the impact of survey mask, and of peculiar velocities) need to be investigated and thoroughly modeled to reliably use the void size function.

The authors are grateful to Elena Massara, Mélanie Habouzit, Nico Hamaus, Jia Liu, and Shy Genel for useful discussions. The authors thank the anonymous referees for their helpful comments. A.P. is supported by NASA grant 15-WFIRST15-0008 to the Nancy Grace Roman Space Telescope Science Investigation Team “Cosmology with the High Latitude Survey.” For using *MassiveNus* in our work, we thank the Columbia Lensing group (<http://columbialensing.org>) for making their suite of simulations available.

### ORCID iDs

Alice Pisani  <https://orcid.org/0000-0002-6146-4437>

David N. Spergel  <https://orcid.org/0000-0002-5151-0006>

### References

- Achitouv, I. 2019, *PhRvD*, **100**, 123513
- Achitouv, I., Blake, C., Carter, P., Koda, J., & Beutler, F. 2017, *PhRvD*, **95**, 083502
- Baddeley, A., Rubak, E., & Turner, R. 2015, *Spatial Point Patterns: Methodology and Applications with R* (Boca Raton, FL: CRC Press)
- Barr, C. D., & Schoenberg, F. P. 2010, *Biometrika*, **97**, 977
- Barreira, A., Cautun, M., Li, B., Baugh, C. M., & Pascoli, S. 2015, *JCAP*, **2015**, 028
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, *ApJ*, **762**, 109
- Biswas, R., Alizadeh, E., & Wandelt, B. D. 2010, *PhRvD*, **82**, 023002
- Bos, E. G. P., van de Weygaert, R., Dolag, K., & Pettorino, V. 2012, *MNRAS*, **426**, 440
- Cai, Y.-C., Taylor, A., Peacock, J. A., & Padilla, N. 2016, *MNRAS*, **462**, 2465
- Cappellari, M. 2009, arXiv:0912.1303
- Carlesi, E., Knebe, A., Lewis, G. F., Wales, S., & Yepes, G. 2014, *MNRAS*, **439**, 2943
- Cautun, M., Paillas, E., Cai, Y.-C., et al. 2018, *MNRAS*, **476**, 3195
- Clampitt, J., Cai, Y.-C., & Li, B. 2013, *MNRAS*, **431**, 749
- Colberg, J. M., Pearce, F., Foster, C., et al. 2008, *MNRAS*, **387**, 933
- Contarini, S., Ronconi, T., Marulli, F., et al. 2019, *MNRAS*, **488**, 3526
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, **292**, 371
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
- Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, *MNRAS*, **399**, 497
- Doré, O., Werner, M. W., Ashby, M. L. N., et al. 2018, arXiv:1805.05489
- Ebeling, H., & Wiedenmann, G. 1993, *PhRvE*, **47**, 704
- Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, **445**, 175
- Greco, J. P., Hill, J. C., Spergel, D. N., & Battaglia, N. 2015, *ApJ*, **808**, 151
- Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, *MNRAS*, **413**, 101
- Hadzhiyska, B., Bose, S., Eisenstein, D., Hernquist, L., & Spergel, D. N. 2020, *MNRAS*, **493**, 5506
- Hamaus, N., Cousinou, M.-C., Pisani, A., et al. 2017, *JCAP*, **2017**, 014
- Hamaus, N., Pisani, A., Sutter, P. M., et al. 2016, *PhRvL*, **117**, 091302
- Hamaus, N., Sutter, P. M., Lavaux, G., & Wandelt, B. D. 2014, *JCAP*, **2014**, 013
- Hamaus, N., Sutter, P. M., Lavaux, G., & Wandelt, B. D. 2015, *JCAP*, **2015**, 036
- Hamaus, N., Sutter, P. M., & Wandelt, B. D. 2014, *PhRvL*, **112**, 251302
- Hawken, A. J., Granett, B., Iovino, A., et al. 2017, *A&A*, **607**, A54
- He, S., Li, Y., Feng, Y., et al. 2019, *PNAS*, **116**, 13825
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, **208**, 19
- Jennings, E., Li, Y., & Hu, W. 2013, *MNRAS*, **434**, 2167
- Kreisch, C. D., Pisani, A., Carbone, C., et al. 2019, *MNRAS*, **488**, 4413
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
- Lavaux, G., & Wandelt, B. D. 2010, *MNRAS*, **403**, 1392
- Lavaux, G., & Wandelt, B. D. 2012, *ApJ*, **754**, 109
- Lee, J., & Park, D. 2009, *ApJL*, **696**, L10
- Liu, J., Bird, S., Zorrilla Matilla, J. M., et al. 2018, *JCAP*, **2018**, 049
- Massara, E., & Sheth, R. K. 2018, arXiv:1811.03132
- Massara, E., Villaescusa-Navarro, F., Viel, M., & Sutter, P. 2015, *JCAP*, **2015**, 018
- Nelson, D., Pillepich, A., Genel, S., et al. 2015, *A&C*, **13**, 12
- Neyrinck, M. C. 2008, *MNRAS*, **386**, 2101
- Odrzywołek, A. 2009, *PhRvD*, **80**, 103515
- Paz, D., Lares, M., Ceccarelli, L., Padilla, N., & Lambas, D. G. 2013, *MNRAS*, **436**, 3480
- Perico, E. L. D., Voivodic, R., Lima, M., & Mota, D. F. 2019, *A&A*, **632**, A52
- Pisani, A., Lavaux, G., Sutter, P. M., & Wandelt, B. D. 2014, *MNRAS*, **443**, 3238
- Pisani, A., Massara, E., Spergel, D. N., et al. 2019, *BAAS*, **51**, 40
- Pisani, A., Sutter, P. M., Hamaus, N., et al. 2015, *PhRvD*, **92**, 083531
- Pollina, G., Baldi, M., Marulli, F., & Moscardini, L. 2016, *MNRAS*, **455**, 3075
- Pollina, G., Hamaus, N., Dolag, K., et al. 2017, *MNRAS*, **469**, 787
- Pollina, G., Hamaus, N., Paech, K., et al. 2019, *MNRAS*, **487**, 2836
- Ronconi, T., Contarini, S., Marulli, F., Baldi, M., & Moscardini, L. 2019, *MNRAS*, **488**, 5075
- Sahlén, M. 2019, *PhRvD*, **99**, 063525
- Sahlén, M., Zubeldía, I., & Silk, J. 2016, *ApJL*, **820**, L7
- Schuster, N., Hamaus, N., Pisani, A., et al. 2019, *JCAP*, **2019**, 055
- Sheth, R. K., & van de Weygaert, R. 2004, *MNRAS*, **350**, 517
- Soares-Santos, M., de Carvalho, R. R., Annis, J., et al. 2010, *ApJ*, **727**, 45
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
- Spolyar, D., Sahlén, M., & Silk, J. 2013, *PhRvL*, **111**, 241103
- Springel, V. 2005, *MNRAS*, **364**, 1105
- Springel, V. 2010, *MNRAS*, **401**, 791
- Springel, V., Yoshida, N., & White, S. D. M. 2001, *NewA*, **6**, 79
- Sutter, P. M., Lavaux, G., Hamaus, N., et al. 2014a, *MNRAS*, **442**, 462
- Sutter, P. M., Lavaux, G., Hamaus, N., et al. 2015, *A&C*, **9**, 1
- Sutter, P. M., Lavaux, G., Wandelt, B. D., & Weinberg, D. H. 2012, *ApJ*, **761**, 187
- Sutter, P. M., Pisani, A., Wandelt, B. D., & Weinberg, D. H. 2014b, *MNRAS*, **443**, 2983
- van de Weygaert, R. 2007, arXiv:0707.2877
- van de Weygaert, R., & Schaap, W. 2009, in *Data Analysis in Cosmology, Lecture Notes in Physics*, Vol. 665, ed. V. J. Martínez et al. (Berlin: Springer), 291
- Verza, G., Pisani, A., Carbone, C., Hamaus, N., & Guzzo, L. 2019, *JCAP*, **2019**, 040
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, *Natur*, **509**, 177
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014b, *MNRAS*, **444**, 1518
- Voivodic, R., Lima, M., Llinares, C., & Mota, D. F. 2017, *PhRvD*, **95**, 024018
- Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, *ApJ*, **633**, 791
- Zivick, P., Sutter, P. M., Wandelt, B. D., Li, B., & Lam, T. Y. 2015, *MNRAS*, **451**, 4215