THE OVERTIDENSITY AND MASSES OF THE FRIENDS-OF-FRIENDS HALOS AND UNIVERSALITY OF HALO MASS FUNCTION

SURHUD MORE1, ANDREY V. KRAVTsov1,2,3, NEAL DALAL4, AND STEFAN GOTTLÖBER5

1 Kavli Institute for Cosmological Physics and Enrico Fermi Institute, The University of Chicago, Chicago, IL 60637, USA; surhud@kicp.uchicago.edu
2 Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637, USA
3 Enrico Fermi Institute, The University of Chicago, Chicago, IL 60637, USA
4 Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, Ontario M5S 3H8, Canada
5 Astrophysikalisches Institut Potsdam, An der Sternwarte 16, 14482 Potsdam, Germany

Received 2011 February 28; accepted 2011 May 12; published 2011 June 23

ABSTRACT

The friends-of-friends algorithm (hereafter FOF) is a percolation algorithm which is routinely used to identify dark matter halos from N-body simulations. We use results from percolation theory to show that the boundary of FOF halos does not correspond to a single density threshold but to a range of densities close to a critical value that depends upon the linking length parameter, \( b \). We show that for the commonly used choice of \( b = 0.2 \), this critical density is equal to 81.62 times the mean matter density. Consequently, halos identified by the FOF algorithm enclose an average overdensity which depends on their density profile (concentration) and therefore changes with halo mass, contrary to the popular belief that the average overdensity is \( \sim 180 \). We derive an analytical expression for the overdensity as a function of the linking length parameter \( b \) and the concentration of the halo. Results of tests carried out using simulated and actual FOF halos identified in cosmological simulations show excellent agreement with our analytical prediction. We also find that the mass of the halo that the FOF algorithm selects crucially depends upon mass resolution. We find a percolation-theory-motivated formula that is able to accurately correct for the dependence on number of particles for the mock realizations of spherical and triaxial Navarro–Frenk–White halos. However, we show that this correction breaks down when applied to the real cosmological FOF halos due to the presence of substructures. Given that abundance of substructure depends on redshift and cosmology, we expect that the resolution effects due to substructure on the FOF mass and halo mass function will also depend on redshift and cosmology and will be difficult to correct for in general. Finally, we discuss the implications of our results for the universality of the mass function.

Key words: cosmology: theory – dark matter – methods: numerical

Online-only material: color figures

1. INTRODUCTION

Over the last three decades, cosmological simulations have been playing an ever increasing role in testing cosmological structure formation models against observations using statistics that can be reliably measured in both. Given that most of the available observational information is about virialized peaks in the overall matter distribution, identification of corresponding virialized peaks, or halos, in simulations is of critical importance.

A number of automated halo finding algorithms have been developed over the years (e.g., Knebe et al, 2011, and references therein). One of the most popular of these is the “friends-of-friends” (hereafter FOF) algorithm which uniquely defines groups that contain all particles separated by distance less than a given linking length, \( b \ell \), where \( \ell \) is the mean interparticle separation in simulations (related to the mean number density \( \bar{n} \)) and \( b \) is a free parameter of the algorithm. The FOF algorithm is commonly applied both to identify groups of galaxies in redshift catalogs (Huchra & Geller 1982; Press & Davis 1982; Einasto et al. 1984; Eke et al. 2004; Berlind et al. 2006) and virialized halos in cosmological simulations (Einasto et al. 1984; Davis et al. 1985; Frenk et al. 1988; Lacey & Cole 1994; Klypin et al. 1999; Jenkins et al. 2001; Warren et al. 2006; Gottlöber & Yepes 2007).

An attractive feature of the FOF algorithm is its simplicity: the result depends solely on the linking length in units of the mean interparticle separation, \( b \). The FOF algorithm does not assume any particular halo shape and can therefore better match the generally triaxial mass distribution in halos forming in hierarchical structure formation models. In addition, studies over the last decade indicate that the appropriately parameterized mass function of FOF halos is universal for different redshifts and cosmologies at least to \( \sim 10\% \), although real systematic variations of \( \lesssim 10\% \) do exist (Jenkins et al. 2001; White 2002; Evrard et al. 2002; Hu & Kravtsov 2003; Warren et al. 2006; Reed et al. 2007; Lukić et al. 2007; Tinker et al. 2008; Bhattacharya et al. 2011; Crocce et al. 2010; Courteau et al. 2011). Mass functions of halos identified using the spherical overdensity (SO) algorithm, on the other hand, exhibit considerably larger differences for different cosmologies and redshifts (White 2002; Tinker et al. 2008). Given the importance of the halo mass function in interpreting observed counts of galaxies and clusters, it is interesting to understand the origin of deviations from universality, the role of mass definition, and differences between mass functions defined with the FOF and SO halo finders (e.g., Audit et al. 1998; Jenkins et al. 2001; White 2001, 2002; Tinker et al. 2008; Lukić et al. 2009). This, in turn, requires good understanding of properties of the FOF-identified groups. For example, a recent study by Courteau et al. (2011) shows that the degree of universality depends sensitively on the choice of the linking length parameter \( b \).

One could expect that for a given value of \( b \), the FOF algorithm defines the boundary of a halo as corresponding to a
certain isodensity surface, at least in the limit of large number of particles. Frenk et al. (1988) indicate that the overdensity (defined with respect to the mean density of the universe: \( \delta = \rho / \bar{\rho} - 1 \)) of this surface is \( \delta_{\text{fof}} \approx 2 h^{-3} \). Lacey & Cole (1994; see also Summers et al. 1995 and Audit et al. 1998) quote a value four times smaller, of \( \delta_{\text{fof}} = 3/(2\pi b^3) \approx 0.48 b^{-3} \), corresponding to the local overdensity of two particles within a sphere of radius \( b \). Clearly, such local overdensity is the absolute minimum overdensity that should be sampled by the particles of a FOF halo. For the most commonly used value of \( b = 0.2 \) this corresponds to a local overdensity of \( \delta_{\text{fof}} \approx 60 \), which for an isothermal density profile, \( \rho(r) \propto r^{-2} \), corresponds to an enclosed overdensity of \( \delta_{\text{fof}} \approx 180 \). This value is close to the virial overdensity predicted by the spherical collapse model in the Einstein–De Sitter cosmology and is usually regarded as a justification for using \( b = 0.2 \) in analyses of simulations.

More recently, Warren et al. (2006) have noted that their experiments on Poisson realizations of isothermal halos indicate that the FOF algorithm identifies the boundary at an overdensity \( \delta_{\text{fof}} \approx 74 \), which corresponds to an enclosed overdensity of \( \approx 280 \) rather than the canonical value of 180. Indeed, they report that direct measurements of internal overdensities of the FOF halos in their cosmological simulations identified with \( b = 0.2 \) range from \( \sim 200 \) for largest simulation boxes to \( \sim 400 \) for the smallest boxes. Given that small boxes resolve predominantly smaller mass halos compared to larger boxes, this result hints that the internal overdensity of the FOF halos is actually mass dependent.

Given that the FOF algorithm identifies boundary at a local overdensity and halos are described by an Navarro et al. (1997, hereafter NFW) profile with mass-dependent concentration, this result is not surprising. However, concentration also strongly depends on cosmology and redshift (e.g., Bullock et al. 2001; Zhao et al. 2003a, 2009), which immediately implies that the internal overdensity of FOF halos identified with a given value of \( b \) is also redshift and cosmology dependent. Interpretation of the FOF halo mass function and other statistics is therefore not trivial. For example, halo occupation distribution models typically assume that halos are defined within a spherical radius enclosing a well-defined overdensity. Also, creating mock galaxy catalogs by assigning galaxies to FOF halos requires knowledge of the internal halo overdensities in order to model the target galaxy bias properly.

In this study, we present a detailed analysis of the halo boundary and the corresponding overdensity selected by the FOF algorithm with a given linking length \( b \) based on random particle realizations of spherical NFW halos. We also present an analytical interpretation of the results of these experiments and compare its predictions to overdensities of FOF halos in cosmological simulations. We show that the boundary of the FOF halos does not correspond to a single local overdensity, but to a range of overdensities around a characteristic value that can be understood in terms of percolation theory. For the commonly used value of \( b = 0.2 \), the characteristic local overdensity is \( \delta \approx 81 \), a value higher than that quoted in previous studies. Correspondingly, the enclosed overdensity of the FOF halos is considerably higher than thought before and for \( b = 0.2 \) ranges from \( \sim 250 \) to \( \sim 600 \) for typical halo concentrations (overdensities for other values of \( b \) scale as \( \propto b^{-3} \)).

The paper is organized as follows. In Section 2, we present tests of the FOF algorithm on Monte Carlo realization of idealized spherical NFW halos and show explicitly that (1) the boundary of FOF halos does not correspond to a single local overdensity, but rather to a range of overdensities and (2) the enclosed overdensities of the FOF halos are significantly larger than commonly thought and depend on concentrations of halos and thus on mass, redshift, and cosmology. In Section 3, we develop a simple analytic model that encapsulates results of the Monte Carlo experiments of Section 2 (see also Appendix B for interpretation of these results in the context of percolation theory) and present tests of this model against results of cosmological simulations. In Section 4, we discuss implications of our results for the universality of halo mass function. In Section 5, we interpret results for idealized realizations of NFW halos in the context of percolation theory and present an accurate formula describing the dependence of the FOF mass on mass resolution based on this theory. In Section 5, we also consider real halos extracted from cosmological simulations of a ΛCDM cosmology and show that substructure present in real halos makes behavior of the FOF masses with resolution even more complicated. Finally, we summarize our results and conclusions in Section 6. In Appendix A, we review the basics of the percolation theory and in Appendix B, we demonstrate how the boundary of the FOF halos and their mass can be understood and predicted in its context.

2. TESTS WITH MONTE CARLO REALIZATIONS OF SPHERICAL NFW HALOS

To explore the boundary of the FOF halos and their enclosed overdensities, we follow the approach of Lukić et al. (2009) and consider Monte Carlo realizations of idealized spherical halos. We assume that the internal density distribution of the halos is described by the NFW density profile (Navarro et al. 1997):

\[
n(r) = \frac{A}{(r/r_s)(1+r/r_s)^2},
\]

which is a reasonable approximation to density profiles of halos formed in CDM cosmologies. Here, \( r_s \) denotes the scale radius. The boundary of a halo is usually defined with respect to the radius \( R_\Delta \) that encloses internal overdensity \( \Delta \) with respect to the mean density of the universe. The radii \( r_s \) and \( R_\Delta \) are related via the concentration parameter \( c_\Delta = R_\Delta/r_s \). The normalization, \( A \), is then given by

\[
A = \left( \frac{N_\Delta}{4\pi R^3_\Delta} \right) \frac{c^3_\Delta}{\mu(c_\Delta)}.
\]

where \( N_\Delta \) is the number of particles within \( R_\Delta \) and the function \( \mu(x) \) is given by

\[
\mu(x) = \ln(1+x) - \frac{x}{(1+x)}.
\]

For the Monte Carlo realizations presented in this section, we assume a concentration of \( c_\Delta = 10 \). We generalize our results for other concentrations in the following section. We generate such realizations with varying number of particles, \( N_p \), and mean interparticle separation, \( \bar{l} \). The latter can be expressed in terms of the radius \( R_\Delta \) and the number of particles \( N_\Delta \) as

\[
\bar{l} = \left[ \frac{4\pi R^3_\Delta}{3} \right]^{1/3} N_\Delta.
\]

As the boundary that the FOF algorithm will select is not known a priori, we conservatively generate particle distribution up to the radius of \( 2R_\Delta \).
Without any loss of generality, we use $\Delta = 180$, one of the most commonly used mass-defining overdensities and generate a series of halo realizations with $N_{180}$ varying from $10^7$ to 100 particles. To reduce Poisson noise, for small $N_{180}$ we generate multiple realizations and average over them. We use 10, 100, and 1000 realizations for halos with $10^4$, $10^5$, and 100 particles, respectively. As the particle distribution extends up to $2R_{180}$, the actual number of particles used in each of the realizations is larger than $N_{180}$ roughly by a factor of 1.4. We run the FOF halo finder on each of the halo realizations with a linking length equal to $0.2\bar{l}$. The algorithm links particles with each other if they are closer than the linking length. In what follows, we restrict our attention to the largest group that is found by the FOF algorithm.

Figures 1–3 show the fraction of particles in a Monte Carlo halo, $f_{\text{accept}}$, that are grouped into the central halo by the FOF algorithm at a given radius as a function of radius, local density, and enclosed overdensity, respectively. Although we generate realizations of spherically symmetric halos with no physical substructure, the figures show that the boundary of the FOF-identified halos is not sharp. The particles joined into the FOF group span a range of radii and overdensities. The “fuzziness” of the boundary increases dramatically for realizations with the smallest number of particles. Note, however, that even for realizations with millions of particles, $f_{\text{accept}}$ as a function of radius or overdensity does not converge to a step function, but rather converges to a well-defined shape spanning a range of radii. This implies that the boundary selected by the FOF algorithm is inherently fuzzy.

Figure 2 also clearly shows that the local overdensity of the majority of particles within the fuzzy FOF boundary is larger than $n_{180}$. Correspondingly, the mean enclosed overdensity within this boundary is also much larger than 180, contrary to what is usually assumed for $b = 0.2$ linking length (Figure 3).

The particles that are joined into an FOF group depend upon the percolation properties of the particle distribution. In Appendix B, we show that the behavior of $f_{\text{accept}}$ as a function of radius and overdensity demonstrated by Figures 1–3 can be understood in the framework of percolation theory. For example, percolation theory predicts that for a uniform particle distribution percolation (i.e., formation of a group spanning the entire region) should occur at the local number density equal to a critical value of

$$n_{\text{crit}} = n_c (b\bar{l})^3.$$  \hspace{1cm} (5)

This corresponds to the local overdensity (with respect to the mean density $\bar{n} = l^{-3}$) of

$$\delta_{\text{crit}} \equiv \frac{n_{\text{crit}}}{l^{-3}} - 1 = n_c b^{-3} - 1.$$  \hspace{1cm} (6)

Here, $n_c$ is a universal constant that arises in the percolation problem of spheres that follow a Poisson distribution. The value of this constant has been calibrated via extensive Monte Carlo experiments (Lorenz & Ziff 2001):

$$n_c = 0.652960 \pm 0.000005.$$  \hspace{1cm} (7)

We can expect that the boundary of FOF halos should approximately correspond to $n_{\text{crit}}$ because percolation across
a radial bin will be inhibited at smaller densities. For our choice of \( b = 0.2 \), this corresponds to \( n_{\text{crit}} = 81.62 \bar{I}^{-3} \), i.e., local overdensity \( \delta_{\text{crit}} = 80.62 \). This overdensity is shown by the vertical line in Figure 2, while vertical lines in Figures 1 and 3 show the corresponding radius and enclosed mean overdensity. The figures show that percolation threshold does indeed predict a characteristic overdensity and radius roughly in the middle of the FOF boundary range. In Appendix B, we show that percolation theory also explains the shape of the FOF halos. In the next section, we present a simple analytic model that describes this overdensity as a function of linking length \( b \) and halo concentration \( c \).

3. CONCENTRATION DEPENDENCE OF THE ENCLOSED FOF OVERDENSITY

3.1. Analytical Model

In the previous section, we showed that the boundary of the FOF algorithm corresponds to a wide range of local overdensities (with the width of the range dependent on the number of particles in a halo) around a characteristic local density \( n_{\text{crit}} = n_c (b \bar{I})^{-3} \) or the corresponding local overdensity \( \delta_{\text{crit}} \equiv n_{\text{crit}}/\bar{n} - 1 = n_c b^{-3} - 1 \). For the commonly used value of the linking length parameter \( b = 0.2 \), \( \delta_{\text{crit}} = 80.62 \). Given the characteristic local overdensity at the boundary, it is straightforward to derive an analytical expression for the average enclosed overdensity assuming that halos have NFW density profiles.

Let us denote the number of particles selected by the FOF algorithm as \( N_\Delta \), and the effective spherical radius enclosing these particles as \( R_\Delta \), where \( \Delta \) is the overdensity of the FOF halo which we wish to determine. Evaluating the number density at \( R_\Delta \) using Equations (1) and (2), and equating it to the critical number density, \( n_{\text{crit}} \) yields

\[
\left( \frac{N_\Delta}{4 \pi R_\Delta^3} \right) c_\Delta^3 \mu(c_\Delta) \frac{1}{c_\Delta^2(1 + c_\Delta)^2} = n_c (b \bar{I})^{-3}.
\]

(8)

Note that here \( c_\Delta \equiv R_\Delta/r_s \) is the concentration defined with respect to \( R_\Delta \).

The enclosed overdensity, \( \Delta \), of the halo is then given by

\[
\Delta = \left( \frac{3N_\Delta}{4 \pi R_\Delta^2 \bar{I}^{-3}} \right) - 1
\]

(9)

\[
= 3 n_c b^{-3} \frac{\mu(c_\Delta)(1 + c_\Delta)^2}{c_\Delta^3} - 1.
\]

(10)

This explicitly shows that the overdensity of an FOF halo depends not only upon the linking length parameter, \( b \), but also upon its concentration. In Figure 4, we show the average FOF halo overdensity as a function of the concentration, \( c_\Delta \), for three representative values of \( b \).

Note that one needs to know the concentration–mass relation to predict the overdensity of halos as a function of the FOF halo mass. The concentration of halos depends upon the radius of the halo (and hence the overdensity definition). The concentration and the average overdensity of FOF halos as a function of their mass can be calculated using the following steps. (1) As a first guess, we assume that FOF halos have a certain overdensity (say \( \Delta_c \)) with respect to the background. (2) We use the concentration–virial mass relation given by Zhao et al. (2009)\(^6\) and convert it to a concentration–mass relation for halos with overdensity \( \Delta_c \). (3) This concentration is used to find a new overdensity using Equation (10). We repeat steps (2) and (3) until we converge to a value of overdensity (and concentration).

Note that since the concentration of a halo depends on cosmology, redshift, and halo mass, the enclosed overdensity of halos selected by the FOF algorithm also depends upon cosmology, redshift, and mass. Furthermore, even for a fixed cosmology, redshift, and mass, halo concentrations exhibit substantial scatter and we can therefore expect a corresponding scatter in enclosed overdensities. We will consider these dependencies and scatter in the next section, where we compare the predictions of Equation (10) to overdensities of FOF halos identified in cosmological simulations.

3.2. Comparison with Cosmological Simulations

To test the simple model presented in the previous section, we compare predictions of Equation (10) with actual overdensities of halos identified in dissipationless cosmological simulations of the LCDM model. Halos have been identified using the FOF algorithm with different linking lengths \( b \) and at different redshifts in two cosmological simulations of the same flat LCDM cosmology: the matter and baryon density in units of the critical density \( \Omega_m = 1 - \Omega_\Lambda = 0.27 \) and \( \Omega_\Lambda = 0.0469 \), the Hubble constant \( h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.70 \), the rms amplitude of linear fluctuations in spheres of radius \( 8 h^{-1} \text{ Mpc} \sigma_8 = 0.82 \), and the power-law slope of the initial power spectrum, \( n_s = 0.95 \).

The first is the Bolshoi simulation of a cubic volume of size \( L_B = 250 h^{-1} \text{ Mpc} \), described in detail in Klypin et al. (2010), while the second is the MultiDark simulation of volume size \( L_{\text{MD}} = 1 h^{-1} \text{ Gpc} \) (Prada et al. 2011).\(^7\) Both simulations

---

\(^6\) Zhao et al. (2009) calibrate concentration–mass relation for concentration and masses defined with respect to the radius enclosing virial overdensity, \( \Delta_{\text{vir}} \).

\(^7\) Data from both simulations are publicly available at http://www.multidark.org/MultiDark/.
followed the evolution of 2048^3 particles, which corresponds to particle masses of $1.36 \times 10^8 h^{-1} M_{\odot}$ and $8.72 \times 10^9 h^{-1} M_{\odot}$ for the Bolshoi and MultiDark simulations, respectively. The peak spatial resolution was $1 h^{-1}$ kpc and $7 h^{-1}$ kpc in these simulations, respectively.

The FOF algorithm used to identify halos in these simulations is based on the minimal spanning tree and is described in Knebe et al. (2011). Given that the shape of the FOF halos can be arbitrary and rather complicated, measurement of their volume is not trivial. We estimate the volume employing the following procedure. For each FOF halo, the three-dimensional (3D) distribution of particles is projected onto a two-dimensional plane perpendicular to one of the coordinate axis (e.g., the $x$-axis in the following). A grid of cells of size $s = b I$ is then overlaid on this plane. The volume occupied by particles in each individual cell $i$ is estimated as

$$V_i = s^2 \times (x_{\text{max}} - x_{\text{min}}),$$  \hspace{1cm} (11)

where $x_{\text{min}}$ and $x_{\text{max}}$ are the minimum and maximum $x$ coordinates of particles in the cell and $x_{\text{max}} - x_{\text{min}}$ is the extent of the particle distribution along $x$. The total volume of the halo, $V$, is calculated as a sum over all cells containing particles $V = \sum V_i$. This procedure is repeated for the other two axes and the final halo volume is assumed to be the maximum of $V_x$, $V_y$, and $V_z$.

The procedure used for estimating the volume roughly approximates the convex hull algorithm.\footnote{http://en.wikipedia.org/wiki/Convex_hull} It is designed to avoid the pitfall of estimating volume using 3D grid as a sum of cells containing particles. Such estimate leaves many empty cells within the halo unaccounted for. Moreover, such method does not converge to a well-defined volume value as the 3D grid cell size is varied. Note that for two spherical halos of the same size that just touch each other (equal mass bridged halos), the convex hull algorithm shall give an answer which is 25% larger than the sum of the volumes of the two spheres. Our method, on the other hand, will lead to such an overestimate of the volume only when the separation vector between the two spheres lies roughly along one of the axis. Such situations are expected to be fairly rare.

Figure 5 shows overdensities of individual FOF halos selected from simulations as a function of the FOF halo mass selected using different linking length parameters. The three panels show results for FOF with linking lengths $b = 0.085$, $b = 0.17$, and $b = 0.2$. In each panel, the dashed lines show the median overdensity as a function of halo mass, while the dotted lines show the 16 and 84 percentiles of the distribution. The blue (short-dashed) and purple (long-dashed) lines correspond to the results of the Bolshoi and MultiDark simulations, respectively. The red solid lines show the prediction for the overdensity of FOF halos as a function of halo mass given by Equation (10) and concentration–mass relation of Zhao et al. (2009). The red dotted lines show the rms scatter predicted by the model, if we assume a scatter of 0.14 dex of concentrations at a given mass.

Figure 6 shows overdensities of the FOF halos identified with linking lengths $b = 0.085$, 0.17, and 0.20 in the Bolshoi and MultiDark simulations. In each panel, the dashed lines show the median overdensity, while the dotted lines show the 16 and 84 percentiles of the distribution. The blue and purple lines correspond to the results of the Bolshoi and MultiDark simulations, respectively. The gray points show halos from the Bolshoi simulation (the MultiDark halos are not shown for clarity). The red solid lines show the prediction of our model given by Equation (10) and concentration–mass relation of Zhao et al. (2009). The red dotted lines show the rms scatter predicted by the model, if we assume a scatter of 0.14 dex of concentrations at a given mass.

(A color version of this figure is available in the online journal.)

Figure 5. Enclosed overdensities of the FOF halos identified with linking lengths $b = 0.085$, 0.17, and 0.20 in the Bolshoi and MultiDark simulations. In each panel, the dashed lines show the median overdensity, while the dotted lines show the 16 and 84 percentiles of the distribution. The blue and purple lines correspond to the results of the Bolshoi and MultiDark simulations, respectively. The gray points show halos from the Bolshoi simulation (the MultiDark halos are not shown for clarity). The red solid lines show the prediction of our model given by Equation (10) and concentration–mass relation of Zhao et al. (2009). The red dotted lines show the rms scatter predicted by the model, if we assume a scatter of 0.14 dex of concentrations at a given mass.

(A color version of this figure is available in the online journal.)

except perhaps at the smallest and largest masses. At small masses, overdensities of simulated halos exhibit a downturn in both the Bolshoi and MultiDark simulations. The masses at which the downturn occurs are different in the two simulations. This downturn is due to the percolation properties of halos represented by small particle numbers, as we discuss in more detail in Section 5 below and in Appendix B. Note, for example, that the downturn shifts to smaller masses for smaller values of $b$ (i.e., larger local particle densities at the boundary) and almost entirely outside the shown mass region for $b = 0.085$. The overdensities of simulated halos also exhibit a somewhat weaker trend with mass than predicted by our model for masses $> 5 \times 10^{13} h^{-1}$ Mpc. It is not clear what is the source of this discrepancy, but we note that it is quite small and amounts to less than 10%.

Figure 6 shows overdensities of the FOF halos identified with $b = 0.17$ at redshifts $z = 0.0$, 1.0, and 2.5. The evolution of overdensity predicted by the model due to the redshift evolution of concentrations, predicted using the model of Zhao et al. (2009), matches the redshift trend observed in the simulations.
remarkably well. The scatter of overdensities is also well reproduced by the scatter of concentrations at all redshifts. Note that enclosed overdensity for this \( b \) in the mass range probed by the simulations reaches the floor value of \( \approx 400 - 450 \) by \( z = 2.5 \), as virial concentration of halos reaches a floor of \( c_{\text{vir}} \approx 4 \) (Zhao et al. 2003b, 2009).

We note that although the Einasto profile is a better description of the density profile of dark matter halos identified from numerical simulations (Navarro et al. 2004; Merritt et al. 2005), the differences between the Einasto profile and the NFW profile that we have assumed in this paper are less than 5% for radii larger than 0.2 times the virial radius (Gao et al. 2008). The differences between the ratio of the mass enclosed within a given radius to the mass enclosed within the virial radius predicted by the Einasto and the NFW profiles are further smaller. This justifies our use of the NFW profile for calculating the overdensities of halos in the numerical simulations.

4. IMPLICATIONS FOR UNIVERSALITY OF HALO MASS FUNCTIONS

Our results on the enclosed overdensity of the FOF-identified halos have important and interesting implications for the interpretation of recent results on the universality of the halo mass function. A number of studies have found that the halo mass function can be expressed in a cosmology- and redshift-independent way as a universal function of the peak height, \( \delta_c/\sigma(M) \), where \( \delta_c(z) \) is the linearly evolved overdensity of a peak at the time of collapse in the spherical collapse model and \( \sigma(M) \) is the rms fluctuation of perturbations of mass \( M \) (Sheth et al. 2001; Jenkins et al. 2001; Warren et al. 2006; Tinker et al. 2008).

Although deviations from universal behavior have been found for both the FOF- and SO-identified halos, these deviations are markedly smaller for the FOF mass functions (e.g., Lukič et al. 2007; Tinker et al. 2008; Courtin et al. 2011). Courtin et al. (2011) showed that deviations from universality are not random but are correlated with the nonlinear virialization overdensity, \( \Delta_{\text{vir}} \), expected from the spherical collapse model for a given cosmology and redshift. In particular, they showed that the linking length, \( b_{\text{univ}} \), required to minimize deviations of the FOF mass function from universal form for a given cosmology and redshift is correlated with the corresponding \( \Delta_{\text{vir}} \) as

\[
\left( \frac{b_{\text{univ}}}{0.2} \right)^{-3} = 0.24 \left( \frac{\Delta_{\text{vir}}}{178} \right) + 0.68. \tag{12}
\]

This is an interesting and important result, as it indicates that deviations from universality can be minimized if one takes into account cosmology dependence of virialization parameters properly. However, as noted by Courtin et al. (2011), the form of Equation (12) is different from \( (b/0.2)^{-3} = \Delta_{\text{vir}}/178 \), which one would expect if the FOF algorithm with \( b = 0.2 \) would identify halos with a constant internal overdensity of \( \approx 178 \). This form thus begs for a physical explanation. Our results presented in the previous sections can help explain this empirical correlation, at least partially. First, we showed that the typical overdensity of FOF halos identified with \( b = 0.2 \) at \( z = 0 \) is significantly larger than 178. Second, we showed that overdensity of FOF halos depends not only on \( b \) but also on halo concentrations (Equation (10)), and thus on mass, cosmology, and redshift. In light of these results we expect that the linking length required to identify halos enclosing a certain overdensity \( \Delta \) is given by (see Equation (10))

\[
\left( \frac{b}{0.2} \right)^{-3} = \frac{\Delta + 1}{244.86} \psi(c_{\Delta}), \tag{13}
\]

where the function \( \psi(c_{\Delta}) \) is given by

\[
\psi(c) = \frac{c^2}{\mu(c)(1 + c)^2}. \tag{14}
\]

Equation (13) can thus be used to predict what linking length is needed to identify a halo boundary enclosing virial overdensity \( \Delta_{\text{vir}} \).

Figure 7 shows simulation results of Courtin et al. (2011) for values of \( b_{\text{univ}} \) as a function of \( \Delta_{\text{vir}} \) (squares with error bars) and the best fit to their results (dot-dashed line). It also shows the \( b_{\text{univ}} - \Delta_{\text{vir}} \) dependence given by Equation (13) (solid blue line). This line is computed assuming \( \sigma_{\text{vir}} - M \) relation for a flat ΛCDM cosmology consistent with WMAP5 results given by the model of Zhao et al. (2009) for the redshift range from \( z > 2 \) (where \( \Omega_m \approx 1.0 \) and \( \Delta_{\text{vir}} \approx 178 \)) to negative redshifts into the future to sample low-\( \Omega_m \), high-\( \Delta_{\text{vir}} \) regime. For all redshifts the model is computed for a fixed halo mass \( M_{\text{vir}} = 10^{14} M_\odot \), a value representative of the mass range probed by Courtin et al.’s simulations. As can be seen from the figure, prediction of Equation (13) is much closer to the results of Courtin et al. (2011) than the commonly assumed \( (b/0.2)^{-3} = \Delta_{\text{vir}}/178 \). Note that the slope is also different due to the dependence on concentrations via the function \( \psi(c) \).

This implies that results of Courtin et al. (2011) indeed indicate that deviations from universality are largely due to the
use of halo parameters not adjusted for different virialization overdensities in different cosmologies and redshifts. Note, however, that agreement between our model and their results is not perfect. This could be due to several factors. First, the virialization overdensities of halos may be somewhat different from those expected in the spherical collapse model, given that most halos form out of triaxial perturbations via a complicated sequence of mergers and smooth accretion. Second, the well-known bridging effect of the FOF halo finder may play a role at smaller values of \( \Delta_{\text{vir}} \) (i.e., larger values of \( b \)). For the commonly used value of \( b = 0.2 \), the FOF algorithm joins \( \approx 10\% \)–\( 15\% \) of neighboring halos by bridging at \( z = 0 \) (e.g., Davis et al. 1985; Bertschinger & Gelb 1991; Cole & Lacey 1996; Lukić et al. 2009), although this fraction is likely to be higher at larger redshifts (e.g., Cohn & White 2008). Figure 7, on the other hand, shows that our model predicts that the linking length should increase to \( b \approx 0.24 \) to reach \( \Delta_{\text{vir}} \). We can expect that bridging will become severe for such large linking length and would definitely affect FOF halo mass function. The weak dependence of \( b_{\text{uni}} \) on \( \Delta_{\text{vir}} \) for virial overdensities of \( \approx 180 \)–\( 300 \) may therefore reflect the fact that universality of the FOF mass function is compromised by bridging, which prevents \( b_{\text{uni}} \) from reaching lower values.

9 A dramatic effect of bridging on \( z = 10 \) halo mass function can be observed in Figure 3 of Cohn & White (2008), which shows abundance of FOF halos as a function of FOF mass with \( b = 0.2 \) and mass counted around centers of the same halos in spheres enclosing overdensity \( \Delta = 180 \). Although the FOF halos for \( b = 0.2 \) should have mean overdensities considerably larger than 180, and hence FOF mass smaller than SO(180) mass \( M_{\text{FOF}}/M_{180} \) expected from Equation (10) for halo masses between \( 10^7 \) \( h^{-1} \) \( M_\odot \) and \( 10^{10} \) \( h^{-1} \) \( M_\odot \) at \( z = 10.0 \) \( \approx \) \( 0.84 \), that figure shows that the average FOF mass of halos of a given abundance is actually about two times larger than their SO mass with \( \Delta = 180 \).

Some of the discrepancy between Equation (13) and Courtin et al. simulation results could also be due to the fact that their points comprised simulations of different cosmologies all using the same power spectrum and normalization \( \sigma_8 \) at \( z = 0 \), while our prediction is made for a single cosmology as a function of redshift. Given that concentrations of halos in a given cosmology depend not only on \( \Omega_m \), but also on \( \sigma_8 \), results of Courtin et al. (2011) for \( b_{\text{uni}} \) scaling are likely not universal. For example, for cosmology with the same \( \Omega_m \) and \( \Omega_b \) but different values of \( \sigma_8 \), halo concentrations, and hence value of \( b_{\text{uni}} \), will be different but \( \Delta_{\text{vir}} \) will be the same.

Incidentally, the dependence of enclosed overdensity of FOF halos on concentration could also explain why deviations of the halo mass function from universality at different redshifts have been found to be considerably smaller for the FOF halos identified with constant \( b \) than for the SO mass function with masses defined using constant overdensity (White 2002; Lukić et al. 2007; Tinker et al. 2008; Courtin et al. 2011). This more universal behavior could, in principle, be an indication that the FOF somehow identifies halos better related to the initial density field or assigns mass to halos more correctly than the SO algorithm. This would, of course, be interesting for understanding the physical origin of the universality of the mass function.

However, given the significant bridging effect for \( b \approx 0.2 \) discussed above, one should already be skeptical that some deep physics underlies a more universal behavior of the \( b = 0.2 \) FOF mass functions. In addition, our results imply that smaller deviations of the FOF halo mass function from universality are also due to a partial cancellation of some of the redshift evolution of the halo mass function by redshift evolution of halo concentrations. Indeed, for ΛCDM models for which these deviations with redshift have been studied, the enclosed overdensities for high-mass FOF halos at \( z = 0 \), when halo concentrations are relatively high, are \( \sim 300–400 \). These overdensities are close to the virial overdensity of halos in the ΛCDM cosmology. At higher redshifts, however, halo concentrations decrease as \( c(M, z) \propto (1 + z)^{-1} \) (Bullock et al. 2001) until they reach a floor value of \( \approx 4 \) (Zhao et al. 2003a, 2009). For \( c \sim 4 \), the overdensity of FOF halos should approach \( \sim 250 \) (see Figure 4), which is close to the virial overdensity at high redshifts where \( \Omega_m(z) \) is closer to unity. The FOF overdensity thus roughly tracks the virial overdensity in the concordance ΛCDM cosmology. However, we stress that this rough tracking is coincidental. This is because halo concentrations depend on the halo formation times (e.g., Wechsler et al. 2002; Neto et al. 2007; Zhao et al. 2009), which in turn depend on power spectrum normalization among other things. Thus, concentrations would still evolve with redshift in the Einstein–de Sitter \( \Omega_m = 1 \) cosmology, even though virial overdensity would not. The deviations of the FOF mass function from universality would therefore also be affected by power spectrum normalization, or any other parameter that affects concentrations.

5. Masses of FOF Halos

5.1. Masses of the Idealized FOF Halos in the Context of Percolation Theory

Using Monte Carlo simulations of isothermal halos with varying numerical resolution, Warren et al. (2006) were the first to demonstrate that the mass of halos selected by the FOF algorithm depends upon the resolution with which the halo is
sampled. They found that at lower resolutions the FOF algorithm assigns systematically larger masses to halos. They devised an empirical formula to correct the effects of such systematic bias on the halo mass function. More recently, Lukić et al. (2009) carried out Monte Carlo simulations of NFW halos and found a qualitatively similar effect (see also Bhattacharyya et al. 2011). They also devised an empirical formula to correct for the resolution-dependent mass bias for the specific case of \( b = 0.2 \) and idealized spherical NFW halos that they studied. Lukić et al. (2009) showed that this correction depends not only on the number of particles but also upon the concentration of the halo.

As can be seen from Figure 1, our experiments also reveal a qualitatively similar effect. The boundary identified by the FOF algorithm significantly widens with decreasing number of halo particles. Therefore, the mass selected by the FOF algorithm also increases with decreasing number of particles. In Figure 8, we show the mass of the halo identified by FOF for each of our spherical Monte Carlo halos normalized by \( M_\Delta \), the mass expected within the overdensity predicted by using Equation (10). We plot this quantity as a function of \( L_\text{size} \) given by

\[
L_\text{size} = \frac{2R_\Delta}{b l} = \frac{2}{b} \left( \frac{3N_\Delta}{4\pi\Delta} \right)^{1/3}.
\]

Note that by definition \( L_\text{size} \) approximately corresponds to the inverse of the fractional accuracy with which a halo boundary can ever be identified by the FOF algorithm and it depends upon the resolution of the halo via \( N_\Delta \). As described in the appendices, \( L_\text{size} \) is thus the appropriate parameter to use from the standpoint of percolation theory to parameterize the dependence of FOF mass for a given halo on the numerical resolution.

Figure 8 shows that FOF mass can be systematically biased high by \( \approx 10\% \)–20\% for \( L_\text{size} \lesssim 10 \). Most of the modern state-of-the-art simulations are in this regime. For example, the Bolshoi and MultiDark simulations used in the previous section, followed evolution of \( 2048^3 \approx 8.59 \times 10^9 \) particles in boxes of \( 250 h^{-1} \) Mpc and \( 1000 h^{-1} \) Mpc, respectively. For \( b = 0.2 \), these simulations have \( b l \) of \( \approx 24.4 h^{-1} \) kpc and \( \approx 97 h^{-1} \) kpc, respectively. Thus, \( L_\text{size} \lesssim 10 \) corresponds to halos with virial radii \( R_\Delta \lesssim 122 h^{-1} \) kpc and \( R_\Delta \lesssim 488 h^{-1} \) kpc, respectively, both well within the range of halos resolved by these simulations. A wider range of masses would be affected for lower resolution simulations. Dependence of \( L_\text{size} \) on the number of particles in a halo for the choice of \( b = 0.2 \) and typical halo concentration is presented in Figure 16 in Appendix B, which shows that \( L_\text{size} \lesssim 10 \) for \( N_\Delta \lesssim 10^3 \).

In Appendix B, we show that the extra mass identified by the FOF algorithm at a given resolution (i.e., a given \( L_\text{size} \)) can be accurately corrected by the following formula motivated by percolation theory:

\[
M_\text{fof} = M_\text{fof} \left( 1 + 0.22 \alpha \frac{1}{\sqrt{L_\text{size}}} \left| \frac{\partial \ln M_\Delta}{\partial p} \right| \right)^{-1}.
\]

Here, \( M_\text{fof} \) denotes the mass of the halo that FOF would identify at infinite resolution, \( \nu \) is a critical exponent from percolation theory and is \( \approx 1.33 \) in our case (see Appendix B for details), \( \alpha \) denotes the logarithmic slope of the halo density profile at the percolation-theory-predicted boundary, \( R_\Delta \). For an NFW density profile, \( \alpha \) is given by

\[
\alpha = 1 + \frac{2c_\Delta}{1 + c_\Delta}.
\]

The probability \( p(r) \) (see Appendix A for the connection to percolation theory) at a given radius depends upon the number density of particles at that radius, \( n(r) \), via

\[
p(r) = \exp \left\{ -\frac{\pi}{6}(b\Delta)^3n(r) \right\},
\]

and \( \partial \ln M_\Delta/\partial p \) denotes the derivative of the logarithm of the mass with respect to \( p \) at the percolation-threshold-predicted boundary, \( R_\Delta \). Larger values of \( L_\text{size} \) correspond to higher resolution and the mass measured by the FOF algorithm tends to \( M_\text{fof} \) asymptotically. Note that our correction formula depends upon the number of halo particles, \( N_\Delta \), the linking length parameter, \( b \), and the concentration parameter, \( c_\Delta \). The circles in Figure 8 show the result of this correction. The figure shows that the mass corrected by this formula is independent of \( L_\text{size} \). The triangles, on the other hand, show the empirical correction of Warren et al. (2006), which clearly fails to correct the effect fully. This is not surprising as this formula was devised to correct resolution bias in the halo mass function, rather than the mass of individual-idealized NFW halos. As we show below, other resolution effects affect masses of real CDM halos, thereby affecting the halo mass function. The presented exercise simply indicates that the formula of Warren et al. (2006) does not describe the mass bias of idealized halos considered here.

Also note that even at infinite resolution the FOF algorithm selects a mass which is smaller than \( M_\Delta \) by \( \approx 2\% \). This is because the boundary of FOF halos is not a step function even at infinite resolution (see Figure 1). We defer detailed discussion of this effect to Appendix B and show that this small additional correction can also be calculated from percolation theory. The bold circles in Figure 8 show the result of correcting the masses taking into account this additional small effect. As the figure shows, the full correction brings the value of the FOF halo masses in good agreement with the true mass \( M_\Delta \).

Figure 9 shows the results of the Monte Carlo realizations of spherical NFW halos of differing concentrations carried out by
Lukić et al. (2009, shown by squares) and predictions of our model (shown by solid lines). These authors applied the FOF algorithm with $b = 0.2$ to identify halos from the realizations of spherical NFW halos by Lukić et al. (2009) are shown by squares, while predictions of our model for each concentration given by Equation (16) and after applying the additional FOF boundary correction are shown by solid lines.

(A color version of this figure is available in the online journal.)

We would like to note that the correction formula presented by Lukić et al. (2009) is a numerical fit to their results and is only valid for different values of linking length parameter, concentrations, and values of the numerical resolution ($L_{\text{size}}$).

In Appendix B, we also test our correction against simulated halos with varying slopes of the number density profile and show that it works remarkably well for different slopes. We also show that we are able to explain the empirical results for isothermal halos\(^{10}\) found by Warren et al. (2006).

Given that the density of CDM halos decreases rapidly near the outer virialized regions, an overestimate of mass for small $L_{\text{size}}$ and $N_A$ corresponds to an underestimate of the enclosed overdensities of FOF halos. This underestimate can be seen in the form of downturn of overdensity for halos from ACDM simulations observed in Figures 5 and 6. For a fixed mass and fixed value of $b$, the Bolshoi simulation has a larger value of $L_{\text{size}}$ than the MultiDark simulation. This explains why the downturn occurs at lower halo masses for the Bolshoi than for the MultiDark simulation. It is also clear from Equation (16), that $L_{\text{size}} \propto b^{-1}$, and therefore the downturn in overdensity shifts to smaller masses for decreasing values of $b$.

5.2. Resolution Dependence of the FOF Mass for Real $\Lambda$CDM Halos

In the previous subsection, we showed that the mass of halos selected by the FOF algorithm depends upon $L_{\text{size}}$. The mass $M$ selected by FOF at finite $L_{\text{size}}$ can be larger than $M^\infty$ by as much as 5%-20% for small values of $L_{\text{size}}$. This effect, if not corrected for, can potentially introduce systematic errors in the determination of the mass function using halos selected by FOF. We have also shown that the percolation-theory-motivated formula given by Equation (16) is able to correct this dependence of the mass on $L_{\text{size}}$ for spherical NFW halos (or for spherical halos with a power-law density profile). Real halos, however, are not spherical and contain substructure. In this section, we therefore test the correction formula derived for idealized halos against undersampled versions of real halos selected from cosmological simulations.

For this purpose, we make use of the L1000W simulation of size $L_H = 1$ $h^{-1}$ Gpc, described in detail in Tinker et al. (2008). The simulation follows the evolution of dark matter particles in a $\Lambda$CDM cosmology with parameters that are slightly different from the Bolshoi and the MultiDark simulation: the matter density and the baryon density in units of the critical density, $\Omega_m = 0.27$ and $\Omega_b = 0.044$, the Hubble constant $H = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.70$, the rms amplitude of linear fluctuations in spheres of radius $8 h^{-1}$ Mpc, $\sigma_8 = 0.79$, and the power-law slope of the initial power spectrum, $n_s = 0.95$. We run the FOF algorithm with a linking length parameter $b = 0.2$ on the redshift zero snapshot of the simulation. For the purpose of our tests, we focus our attention to the 25 most massive halos selected by FOF.

We selected all particles within a radius $R_{\text{max}} = 10 h^{-1}$ Mpc of the center of mass of each of these halos. We have verified that all the particles of each halo selected by FOF lie well within $R_{\text{max}}$. We created 1000 subsamples each of particles around every halo by using only a fraction $f \in \{0.2, 0.4, 0.6, 0.8\}$ of the particles. We then run FOF on each of these subsamples using a linking length parameter $b = 0.2 f^{-1/3}$. We use the symbol $\mu_f$ to denote the ratio of the mass selected by FOF when run on a subsample with a fraction $f$ of the original particles to the mass of the FOF halo when using all the particles.

In the left-hand panel of Figure 10, we show the distribution of $\mu_f$ for different values of $f$ using different line types. Note that the peak of the distribution shifts toward larger values of $\mu_f$ for smaller values of $f$. This is qualitatively similar to the behavior of FOF discussed in Section 5. However, we also notice that the distribution of $\mu_f$ has a significant tail toward smaller values of $\mu_f$. In roughly one-third of the cases (9 out of 25), the FOF algorithm often fails to bridge a structure in the outer parts of the halo with the main halo. The effect appears less severe because we have plotted the combined distribution of $\mu_f$ values for the 25 halos. However, in the case of halos for which bridging is an issue, the distribution of $\mu_f$ clearly shows a bimodal distribution.

The right-hand panel of Figure 10 shows the cumulative distribution of $\mu_f$. Note that smaller values of $f$ have a slightly larger tendency to avoid bridging. This counteracts the tendency

---

\(^{10}\) We note that the empirical formula given by Warren et al. (2006) does not explain the results of their isothermal halos.
to select larger masses at smaller values of $f$. If we assign a mass for each halo for a given value of $f$ as the average of the FOF mass over the 1000 subsamples, we often find that this average FOF mass increases as $f$ increases contrary to our idealized NFW halos. Clearly using the average is sensitive to the tails of the distribution. Therefore, we used the median of the FOF masses of the 1000 subsamples to test our correction formula.

We denote the median mass selected by the FOF algorithm when run on a fraction $f$ of the particles by $M_f$ and the median mass after correcting for the finite size effect using Equation (16) by $M_f^\infty$. The top panel of Figure 12 shows the ratio of $M_f^\infty/M_{1,0}^\infty$ for the 25 most massive halos. Our correction formula, which worked extremely well for the idealized spherical NFW halos, seems to systematically overcorrect for the finite size effect for small values of $f$ by $\approx 3\%$–$5\%$.

The two plausible causes for this behavior are (1) the non-sphericity of real halos and (2) the presence of substructure in real halos. We carried out another set of Monte Carlo simulations of idealized triaxial halos where the number density of particles is given by a NFW-like profile with the radius $r$ replaced by $\xi$ such that

$$\xi^2 = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}. \quad (19)$$

We used values of $a/c = 0.6$ and $b/c = 0.8$, typical for halos found in numerical simulations of dark matter. We have verified that the correction formula given by Equation (16) works perfectly well even if our triaxial halos are incorrectly assumed to be spherical. Our use of the spherically averaged number density distribution to determine the correction does not introduce any systematic errors. We also experimented with particles whose number density distribution follows a power law in radius and found an identical result.

To investigate the effects of substructure, we carried out the following test. We first obtained the smoothed particle hydrodynamic estimate of the density at the location of all particles in each of the halos using 128 nearest neighbor particles. We used the position of the particle with the largest density as the center of the halo. We then randomly reassigned the angular coordinates of each of the particles within a 10 $h^{-1}$ Mpc sphere with respect to the center of the halo. In this manner, we were able to disperse the substructure over a wider range of angular coordinates while still preserving the radially averaged density profile. We then repeated our exercise of running FOF on subsampled versions of this set of particles.

We show the results of this exercise in Figure 11, which shows the distribution of values of $\mu_f$ thus obtained. In contrast to Figure 10, the distribution of $\mu_f$ is much more symmetric with no significant presence of tails. The peak of the distribution occurs at larger values of $\mu_f$ as $f$ is decreased. The lower panel of Figure 12 shows the ratio $M_f^\infty/M_{1,0}^\infty$ for halos where the substructure has been dispersed. Contrary to the results in the top panel, in this case our correction formula corrects masses accurately. This shows that failure of the correction formulae derived for idealized halos is due to substructure present in real CDM halos simulated with sufficiently high resolution.

The results of this exercise show that the masses selected by FOF for realistic halos cannot be corrected for finite size effects in a straightforward manner. Although percolation-motivated correction formula we derived for halos without substructure (Equation (16)) is highly accurate, it cannot be blindly applied to correct halo masses selected by the FOF algorithm. Substructure introduces strong resolution-dependent effects. The amount of substructure depends on resolution of simulations in a non-trivial way and will vary for halos of different mass within a simulation. It will also vary with redshift for a given halo mass. This indicates that any empirical formula designed to correct masses of halo mass function for resolution effects will also depend in a non-trivial way on resolution, cosmology, and redshift. We thus caution against the use of empirical formulae that depend just upon the number of particles in a halo calibrated for a single cosmology and redshift, as these will likely be inaccurate for other cosmologies and redshifts.

6. DISCUSSION AND CONCLUSIONS

In this paper, we have explored properties of halos identified by the FOF algorithm focusing on the halo boundary. Using idealized Monte Carlo realizations of spherical NFW halos we showed that boundary of the FOF halos spans a range of local overdensities and is inherently “fuzzy.” The fuzziness of the boundary increases with decreasing number of halo particles. We demonstrate that these results can be interpreted in terms of the percolation theory, which we discuss in detail in Appendix B. The value of characteristic local overdensity within
Figure 11. Same as Figure 10, except when the angular coordinates of the particles around the center of the FOF halo are shuffled to disperse substructure (see the text for details).

(A color version of this figure is available in the online journal.)

Figure 12. Ratio $M_{\infty}^f / M_{\infty}^{1.0}$ for the 25 most massive halos selected from the simulation. Here, $M_{\infty}^f$ denotes the median of the distribution of masses selected by the FOF algorithm when run on a fraction $f$ of the particles after correcting for the finite size effect using Equation (16). The top panel shows the result of the real halos, while the bottom panel shows the results when the angular coordinates of the particles around the center of the FOF halo are shuffled to disperse substructure (see the text for details). As indicated in the legend, different line types are used to indicate different values of the fraction $f$. The error bars are used to indicate the 16 and 84 percentile of the distribution. The error bars for different values of $f$ are shifted in the x-direction for clarity.

(A color version of this figure is available in the online journal.)
the FOF boundary derived from our Monte Carlo realizations and predicted by percolation theory is given by (Equation (6)): $\delta_{\text{fof}} = 0.65296b^{-3} - 1$, which gives $\delta_{\text{fof}} = 80.62$ for the commonly used value of $b = 0.2$. This is significantly larger than the local overdensity of $\approx 60$ usually assumed for this value of linking length. Correspondingly, the enclosed overdensity of typical FOF halos is significantly larger than 180 and ranges from $\sim 250$ to $\sim 600$. Specific value of the enclosed overdensity is determined by the concentration of halo (density distribution) and therefore depends on cosmology, halo mass, and redshift. We predict this dependence using a simple analytic model based on NFW density profile and show that this model reproduces results of cosmological simulations of $\Lambda$CDM cosmology at different halo masses, redshifts, and values of the linking length $b$.

For a given linking length $b$, the range of overdensities (i.e., the fuzziness) in the boundary of FOF halos increases with decreasing number of halo particles due to changing properties of percolation for smaller values of parameter $L_{\text{size}} \equiv 2 R_\Lambda/(b l)$, where $R_\Lambda$ is the effective radius of the FOF boundary. For a given simulation, this results in a systematic and increasing overestimate of the FOF mass with decreasing halo mass. This effect has been found empirically by Warren et al. (2006) and Lukic et al. (2009).

We demonstrate how it can be understood qualitatively on the basis of percolation theory. We also present an accurate formula for correcting this systematic FOF mass bias for idealized halos without substructure. This formula is accurate for different values of linking lengths $b$, halo concentrations, and values of parameter $L_{\text{size}}$. We note, however, that this accurate correction requires knowledge of the halo concentration–mass relation, which itself would need to be accurately calibrated for different cosmologies. Moreover, as we demonstrated in Section 5.2, substructure in real halos introduces additional substantial resolution-dependent biases into masses of FOF halos. Given that amount of substructure depends on resolution of simulations and simulation cosmology and redshift in a non-trivial way, any empirical mass correction formula should also depend in a non-trivial way on resolution, cosmology, and redshift.

The concentration and non-trivial resolution dependence of enclosed overdensities and masses of the FOF halos make it difficult to interpret their raw mass function and its universality physically in terms of an underlying model of nonlinear collapse. For instance, as we note in Section 4, concentration dependence of FOF overdensity is likely behind smaller deviations of the FOF halo mass function from universality, as some of the real redshift evolution of the halo mass function is partially cancelled by redshift evolution of halo concentrations. Although such partial cancellation may work for a single $\Lambda$CDM cosmology, it will not work in general as halo concentrations do depend on cosmological parameters. All this also makes it more complicated to connect FOF halo masses to observational estimates of masses, which are typically made within spherical apertures enclosing a fixed (and fairly high) overdensity, with concentration of density profile not known a priori.

Nevertheless, results of Courtin et al. (2011) do indicate that universality of the halo mass function can be improved if cosmology dependence of nonlinear virialization is properly taken into account in the definition of halo mass. In Section 4, we show that their empirical findings can be understood better in terms of our results and model. Further exploration of this issue is definitely warranted. Overall, even though interpretation of FOF halo statistics is more complicated in light of our results, improved understanding of the FOF-identified halos makes any interpretation more robust.

Our results should be also useful in constructing mock catalogs of galaxies based on FOF halo catalogs. To reproduce galaxy clustering properly this procedure requires good knowledge of internal overdensity of identified halos. Model and percolation theory results presented in this paper can be used to accurately estimate this overdensity even for halos with small numbers of particles.

S.M. and A.V.K. are supported by the Kavli Institute for Cosmological Physics at the University of Chicago through the NSF grant PHY-0551142 and an endowment from the Kavli Foundation. A.V.K. is also supported by the NSF grants AST-0507596 and AST-0708154. S.M. and A.V.K. are grateful to the members of structure formation group at the University of Chicago for many useful discussions. A.V.K. is also grateful to Aleksand Kravtsov for patience during completion of this paper. S.G. acknowledges support of DAAD through the PPP program. The MultiDark simulation used in this paper were performed and analyzed at the NAS Ames Research Center. We thank A. Klypin (NMSU) and J. Primack (UCSC) for making these simulations available to us. These simulations will be available to the community via the MULTIDARK database http://www.multidark.org/MultiDark/ as part of the activities of the German Astrophysical Virtual Observatory. We thank Zarija Lukic (LANL) for providing data from their paper in electronic form and we are grateful to him and Frank van den Bosch for comments on an early draft of this paper. S.M. thanks Robert Ziff for interesting discussions on percolation theory. We also thank the anonymous referee for constructive comments. This work made extensive use of the NASA Astrophysics Data System and arXiv.org preprint server.

**APPENDIX A**

**BRIEF REVIEW OF THE RELEVANT ASPECTS OF PERCOLATION THEORY**

Consider a point process that generates a set of points on an $N$-dimensional manifold. Percolation theory deals with the statistics of clusters (or groups of friends in FOF terminology) formed by grouping together neighboring points on the manifold. Traditionally, the percolation problem is defined on a lattice where the occupation of each lattice cell is determined by a random process (Stauffer & Aharony 1994). However, the continuum percolation (Swiss-cheese) model is more relevant to our discussion of the FOF algorithm (Roberts & Storey 1968; Domb 1972; Lorenz & Ziff 2001). In this appendix, we briefly describe this model and how the profile of the boundary of a FOF halo can be understood in more detail.

The Swiss-cheese percolation model considers a set of spheres of equal radius, $R$, whose centers are distributed by a random Poisson process with a constant average number density $n(x)$ in a $L \times L \times L$ volume, where $L \gg R$. The spheres can be thought of as spheres carved in a slab of cheese, from which the model derives its name. Groups of overlapping spheres form clusters of varying sizes. The largest cluster that forms in the system is of particular importance, and for a fixed value of $R$, its size depends upon the average number density of spheres in the system. As the number density of spheres is increased, the size of the largest cluster increases until at a critical number density the largest cluster size becomes $\approx L$.  

---

The Astrophysical Journal Supplement Series, 195:4 (17pp), 2011 July

More et al.
This event is called percolation, the smallest number density at which it happens is called the critical percolation threshold and the corresponding cluster is called the infinite cluster. The critical density, \( n_c \) in units of \( 1/(2R)^3 \) is a universal constant and has been accurately measured by extensive Monte Carlo simulations: \( n_c = 0.652960 \pm 0.000005 \) (Lorenz & Ziff 2001).

The linking length of the FOF algorithm, \( b_l \), corresponds to the diameter \( 2R \) of the spheres in the Swiss-cheese percolation model. The centers of overlapping spheres correspond to “friend” particles in the FOF algorithm as the distance between the centers is less than the linking length. In the FOF language, the critical density threshold is therefore \( n_{\text{crit}} \equiv n_c/(2R)^3 = n_c b^{-3} l^{-3} \), which corresponds to an overdensity of \( \delta = n_{\text{crit}} / \bar{n} - 1 = n_c b^{-3} - 1 \).

For the Swiss-cheese model, the probability for any given point \( x \) in the \( L \times L \times L \) volume to belong to a non-zero number of spheres is given by

\[
p(x) = 1 - \exp\left\{-\frac{4}{3} \pi R^3 n(x)\right\} = 1 - \exp\left\{-\frac{1}{6} \pi (2R)^3 n(x)\right\}.
\]

(A1)

It is conventional to define the percolation problem in terms of this probability instead of the number density \( n(x) \), in which case the critical threshold for percolation \( p_c \) is related to \( n_c \) via

\[
p_c = 1 - \exp\left(-\frac{\pi}{6} n_c\right).
\]

(A2)

Close to the percolation threshold, the probability that any point \( x \) belongs to the infinite cluster, \( P_\infty \), also called the strength of the infinite cluster, follows the scaling relation

\[
P_\infty \approx (p - p_c)\beta.
\]

(A3)

where \( \beta \) is a constant which depends upon the dimensionality of the problem. Only few problems in percolation have exact analytical solutions. Hence, the constant \( \beta \) has to be determined by Monte Carlo experiments and it has been found to approximately equal to 0.42 for percolation in three dimensions (see, e.g., Stauffer & Aharony 1994). Another quantity of interest is the correlation or the connectivity length, denoted by \( \xi \), and defined as the average distance between two points that belong to the same cluster. As \( p \) approaches \( p_c \), \( \xi \) follows the scaling relation given by

\[
\xi \propto |p - p_c|^{-\nu},
\]

where the constant \( \nu \) again depends upon the dimensionality of the problem and is approximately equal to 0.88 in three dimensions and 4/3 in two dimensions.

How do these basics of the percolation theory relate to the halos identified by the FOF algorithm? In the context of the Monte Carlo realizations of spherical NFW halos considered in Section 2, the particle distribution of a given realization is a set of points distributed in a spherical volume of radius \( 2R_{180} \). The FOF algorithm with linking length \( b \) applied to these points treats particles as a set of spheres of radius \( R = b l / 2 \). Those particles whose spheres overlap are considered friends. The difference from a simple uniform density example considered above is that our halos have non-uniform density distribution. Thus, instead of considering percolation in a uniform distribution for different particle number densities, we are considering percolation as we decrease the number density of particles as a function of increasing radius. For a given \( b \), there will be a certain radius at which the critical number density for percolation, \( n_c \) (and corresponding probability \( p_c \)) is reached. Particles around this radius will have a high probability \( P_\infty \) to be a part of the infinite cluster—i.e., to be joined into FOF halo. It is these particles that form the boundary of an FOF halo. Below we consider the properties of this boundary in the context of the percolation theory.

APPENDIX B
DETAILED ANALYSIS OF THE FOF
BOUNDARY OF NFW HALOS

In the left panel of Figure 13, we show the probability \( p \) for a point to be within a distance \( b l / 2 \) from any particle as a function of its position \( x = r / r_s \) for the Monte Carlo realizations of spherical NFW halos analyzed in Section 2. In percolation theory, for point distributions with non-uniform density the infinite cluster is defined as the cluster connected to spheres that lie in the region where the probability \( p \rightarrow 1 \). In our case, this is equivalent to the group that consists of particles at the
center of the halo and is the largest group found by the FOF algorithm.

We denote the fraction of spheres at any given radius that belong to the infinite cluster by \( f_{\text{accept}} \). This fraction is simply the ratio of the strength of the infinite cluster to the probability for any point to belong to any sphere:

\[
 f_{\text{accept}} = \frac{P_\infty}{p}. \tag{B1}
\]

In the right panel of Figure 13, we show \( P_\infty \) as a function of \( p \) for the NFW halo realizations. The line types and colors are the same as in Figures 1–4. For \( p \gg p_c \), \( f_{\text{accept}} \approx 1 \) and \( p f_{\text{accept}} = p \). Near the percolation threshold \( p_c \), the fraction \( f_{\text{accept}} \) falls steadily from one to zero in a way that depends upon the mean interparticle separation in the halo relative to the linking length.

We first investigate the strength of the infinite cluster, \( P_\infty \), for \( p > p_c \). In the left panel of Figure 14, we show the dependence of \( P_\infty \) on \( p - p_c \). For \( p > p_c \), \( P_\infty \) exhibits critical scaling behavior, \( \xi \propto |p_c - p|^{-\nu} \).

In the more general case, other scales like the system size \( L_{\text{size}} \) or local scale length \( s = p/|\nabla p| \) can be important as well. For example, in finite volumes percolation occurs when the connectivity length becomes of order the system size, \( \xi \propto O(L_{\text{size}}) \), which occurs at a lower density than infinite percolation. The percolation threshold, therefore, decreases as the system size decreases, and we can easily see that setting \( \xi \approx L_{\text{size}} \) in Equation (A4) shows the finite size threshold \( p_c \) scales as (Stauffer & Aharony 1994)

\[
 \tilde{p}_c - p_c \propto L_{\text{size}}^{-1/\nu}. \tag{B2}
\]

Similarly, density gradients also modify the percolation transition. Regions where the density is below the naive critical threshold, \( p < p_c \), can still be linked to regions above threshold, if the connectivity length is of order the distance to the super-critical region. In other words, gradients will smear out the percolation transition, by an amount that is straightforward to estimate. If we Taylor expand about the location where \( p = p_c \), writing \( p(x) = p_c + (\nabla p)x + \ldots \), then setting \( x \approx \xi \) shows that the transition is smeared by a distance of roughly

\[
 \xi \propto |p_c - p(\xi)|^{-\nu} = |p_c - p_c - (\nabla p)\xi|^{-\nu} \Rightarrow \xi \propto |\nabla p|^{-\nu(1+\nu)}. \tag{B3}
\]

This corresponds to a width \( \sigma_p \) in \( p(x) \) such that

\[
 \sigma_p \propto |\nabla p|^{-\nu(1+\nu)}. \tag{B4}
\]

Thus, for non-uniform distributions, the density gradient results in a much more gradual transition of \( P_\infty \) to zero, which extends to \( p < p_c \) (Rosso et al. 1986), as illustrated in Figure 14.

For realistic halos, both of the above effects (finite size and density gradient) could be significant, but their importance must diminish as the particle number in the halo increases. To judge the importance of these effects for finite particle numbers, the quantity of interest is \( L_{\text{size}} = 2R_\Lambda/(b\bar{v}) \), where \( R_\Lambda \) is the

\[
 L_{\text{size}} = 2R_\Lambda/(b\bar{v}) \tag{14}
\]

The volume of the system enclosed by the boundary \( R_\Lambda \) is equal to \( 4/3\pi R_\Lambda^3 \) and the number of spheres of radius \( (b\bar{v})/2 \) that can fit in this volume is equal to \( L_{\text{size}}/2 = 8R_\Lambda^3/(b\bar{v})^3 \), which gives \( L_{\text{size}} = 2R_\Lambda/(b\bar{v}) \).
where \( N \) is the number of particles within \( R = 1 \), to identify halos. We generated halos with \( \alpha \in (1.5, 1.75, 2.0, 2.25, 2.5, 2.75) \). For each \( \alpha \), we generated 10⁵ realizations each consisting of 100, 500, and 1250 particles, 100 realizations each consisting of 10,000 and 80,000 particles, 10 realizations of 6.4 × 10⁶ particles, two realizations of 6.4 × 10⁸, and one realization with 10¹⁷ particles. The value of the radius \( R_\Delta \) predicted using Equation (5) for these halos is given by

\[
R_\Delta = \left( \frac{4\pi n_c}{1.25(3-\alpha)} \right)^{-1/\alpha}. \tag{B9}
\]

Note that \( R_\Delta \neq R = 1 \) is the effective radius of the FOF boundary and we used the fact that \( R = 1 \) in our model in the derivation of above equation. The corresponding value of \( L_{\text{size}} \) depends upon \( \alpha \) and is given by

\[
L_{\text{size}} = \frac{2R_\Delta}{b_l} = 2 \left( \frac{N}{1.25} \right)^{1/3} \left( \frac{4\pi n_c}{1.25(3-\alpha)} \right)^{-1/\alpha}. \tag{B10}
\]

Note that for increasing \( \alpha \), the same number of particles, thus correspond to a smaller value of \( L_{\text{size}} \). We would also like to point out that the form of the density profile we chose in Equation (B7) above requires \( \alpha < 3 \) to avoid the divergence in mass at \( r = 0 \). This does not imply that our formalism to correct the masses of low-resolution halos breaks down for \( \alpha > 3 \). As long as \( L_{\text{size}} \), \( \partial M/\partial \rho \) and \( \alpha \) are calculated appropriately at the boundary of the percolation threshold, our formalism should work.

In each panel of Figure 16, square symbols show the halo mass of the main FOF halo as a function of \( L_{\text{size}} \) for \( \alpha = 2.0, 2.25, 2.5 \) and 2.75. Other values of \( \alpha \) give similar results. The mass of the FOF halo asymptotes to its true value as the number of particles with which the halo is sampled is increased. This effect was first identified empirically by Warren et al. (2006) and triangles show their proposed empirical correction. The figure shows, however, that this correction does not account for the entire effect. The circles show the FOF masses corrected using Equation (B6) with a proportionality constant of 0.22 \( \alpha \) and \( \nu = 4/3 \):12

\[
M_\text{tol}^\infty = M_\text{tol} \left( 1 + 0.22 \alpha \left( L_{\text{size}} \right)^{-1/3} \frac{\partial \ln \Delta M}{\partial \rho} \right)^{-1}. \tag{B11}
\]

This correction almost entirely eliminates the \( L_{\text{size}} \) dependence of the FOF-identified halo mass. The circles thus represent the mass, \( M_\text{tol}^\infty \) that would be selected by the FOF algorithm if it were run on a realization with infinite number of particles. We note that for steeper density profiles (i.e., larger values of \( \alpha \)) a larger number of particles is required to converge to \( M_\text{tol}^\infty \).

As was pointed out in Section 5 and is clearly shown in Figure 16, the mass \( M_\text{tol}^\infty \) is smaller than the mass enclosed within an overdensity \( \Delta \) given by Equation (10) by a few percent. This is because the boundary profile of the FOF halos is not a step function but has a specific shape that can be approximately described by Equation (A3) (see Figure 17). This allows us to calculate an estimate of the fraction \( M_\text{tol}^\infty /\Delta M \) as

\[
\frac{M_\text{tol}^\infty}{\Delta M} = \frac{1}{\mu(c_\Delta)} \int_{0}^{c_\Delta} f_{\text{accept}} n(x) x^2 dx. \tag{B12}
\]

12 We have verified with simple three-dimensional gradient percolation experiments similar to Rosso et al. (1986) that \( \nu = 4/3 \) in contrast to \( \nu = 0.88 \) found for three dimensions in case of uniform continuum percolation experiments.
Figure 16. Mass of the FOF halos characterized by different $L_{\text{size}}$ for halos with power-law density profiles $n(r) \propto r^{-\alpha}$. Different panels correspond to different logarithmic slopes $\alpha$, as indicated in the legends. Squares show the mass selected by the FOF algorithm ran on Monte Carlo realizations of halos, while triangles show masses corrected using empirical correction of Warren et al. (2006). Open circles correspond to the FOF masses corrected using Equation (B11). The horizontal solid lines show the true mass $M_{\Delta}$ for each halo model.

Figure 17. Fraction of particles that are joined by the FOF algorithm (with $b = 0.2$) into the main halo as a function of the radius in units of $R_{180}$ for our Monte Carlo realizations of spherical NFW halos. Bold solid line shows the percolation theory prediction for uniform particle density, which can be compared to the results of our simulations shown with lines of different style and color. Number of particles in each halo realization is indicated in the legend. (A color version of this figure is available in the online journal.)

Here, the fraction $f_{\text{accept}}$ and $P_{\infty}$ are given by Equations (B1) and (A3), respectively. As can be seen in Figure 8, this boundary effect correction leads to values of the masses that are very close to true mass $M_{\Delta}$.

In this appendix, we have presented a thorough analysis of the boundary of the FOF halos in the context of percolation theory. We have shown that percolation theory accurately predicts the shape of the boundary of the FOF halos close to the density threshold for percolation, at least for halos without significant amounts of substructure (see Section 5). We have also discussed how the finite number of particles with which a halo is sampled affect this boundary and have found a percolation-theory-motivated formula to correct for this dependence. Finally, we have also shown how the fraction of mass identified by FOF in an infinite resolution halo relates to the mass within a SO given by Equation (10). These results provide a basis and theoretical interpretation for the empirical results presented in the main text of the paper.

REFERENCES

Audit, E., Teyssier, R., & Alimi, J. 1998, A&A, 333, 779
Berlind, A. A., et al. 2006, ApJS, 167, 1
Bertschinger, E., & Gelb, J. M. 1991, Comput. Phys., 5, 164
Bhattacharya, S., Heitmann, K., White, M., Lukić, Z., Wagner, C., & Habib, S. 2011, ApJ, 732, 122
