The offsets between the light and the dark matter in the galaxy clusters of the Illustris simulation

Karen Y. Ng,[1] Annalisa P. Pillepich,[2] William A. Dawson,[3] D. Wittman,[1] Lars Hernquist,[2]

arXiv

ABSTRACT

Assuming that dark matter has a zero interacting cross section, how likely is it for us to see the offset values between dark matter and galaxies from real data? This paper formulates a test using cluster data in the cosmological simulation, the Illustris simulation, to examine that hypothesis. We examine the uncertainties of the different summary statistics, and see if the spatial distribution of galaxy closely follows those of the Dark Matter. TODO: result summary. We found that the uncertainty of the offset resulting from projection effects are non-negligible and vary in unpredictable ways.

Key words: galaxy clusters, dark matter, statistics

1 INTRODUCTION

During the latest stage of structure formation, the universe gave birth to non-linear, hierarchical structures known as galaxy clusters. These clusters, made up of dark matter, galaxies and hot gas, are constantly accreting mass, merging and evolving with their environments. Bright galaxies that belong to a galaxy cluster or group, in particular, highlight the overdensities of the underlying dark matter (DM) distribution.

In these dense regions of the clusters, the rates of particle interactions can be enhanced, including the long-suspected self-interaction of DM particles (hereafter, SIDM). Many papers have used the offsets between the summary statistics of the DM density and the galaxy density to give constraints on the self-interaction cross section, i.e. $\sigma_{\rm SIDM}$, of dark matter. A lot of observational studies focus on using merging galaxy clusters as they assume the high collisional velocity should increase the chance of detecting the effects of SIDM. By assuming galaxies being relatively collisionless $\sigma_{\rm gal} \approx 0$, any offset of the DM population from the galaxy provides $\sigma_{\rm SIDM}$ relative to $\sigma_{\rm gal} \approx 0 {\rm cm}^2$. These observational studies include Markevitch et al. (2004) and Bradač et al. (2006) reporting an offset of 25 kpc for the Bullet Cluster; Dawson (2013) reporting an offset of 129 kpc and 47 kpc for the southern and the northern subcluster respectively; Jee et al. (2015) reporting an offset of 190 kpc for MACSJ1752, and others that we list in detail in table 3. However, other studies using 129 X-ray selected relaxed galaxy groups, such as George et al. (2012) also report offsets of the same order of magnitude, at around $50 - 150$ kpc.

Simulators have also joined hands to study the velocity dependence of the effects of SIDM. There are many staged simulations of mergers of galaxy clusters that focused on detecting the signal from SIDM. The 'staging' step of these simulations usually involve prescribing parametric spatial distribution of galaxies (Randall et al. 2008, Kahlhoefer et al. 2014, Robertson et al. 2016), such as an

NFW profile, and may not simulate the galaxy morphology, nor include dynamical frictions. They try to show the level of offsets solely due to SIDM (Kahlhoefer et al. 2014) as by initializing the galaxy-DM offset to be zero at the beginning of their simulations. Furthermore, these staged simulations commonly use a much higher number of galaxy particles than the realistic observable number of galaxies. Randall et al. (2008) found an offset of only 1.8 kpc in the staged merger simulation with $\sigma_{\rm SIDM} = 0 {\rm cm}^2/$ g using 10^5 galaxy particles When assumed with zero impact parameter for mergers, Kim and Peter et al. (2016), using 5.7k or 57k galaxy tracer particles, also show null galaxy-DM offset during most periods of their control staged simulation with $\sigma_{\rm SIDM} = 0 {\rm cm}^2/$ g. While we provide a more in-depth comparison with these staged simulation in the discussion, we argue these staged simulations do not probe how statistical and observational uncertainties realistically contribute to the galaxy-DM offsets. As such, any offsets from aforementioned staged simulations when they increased $\sigma_{\rm SIDM}$, they can guarantee the offsets are maximally due to SIDM. When these simulations set the $\sigma_{\rm SIDM}$ to observationally motivated levels of $< 3 {\rm cm}^2/{\rm g}$, different authors have consistently reported offset signals on par with uncertainties estimated from individual observations (50 - 100 kpc). These simulations have raised questions about how strongly the galaxy-DM offsets can constrain the effects of SIDM. When Kahlhoefer et al. (2014) simulated SIDM with both low-momentum-transfer self-interaction and rare self-interactions of DM with high momentum transfer, they found maximum offsets that are < 30 kpc for $\sigma_{\rm SIDM}$ as high as 1.6 ${\rm cm}^2$ / g. The reported offset from Randall et al. (2008) for $\sigma_{\rm SIDM} = 1.24 {\rm cm}^2/{\rm g}$ is only 53.9 kpc. While Kim and Peter et al. (2016) found a maximum offset < 50 kpc for $\sigma_{\rm SIDM} = 3 {\rm cm}^2/{\rm g}$, and Robertson et al. (2016) also found a maximum offset $\lesssim 40$ kpc from a simulation suite of a Bullet Cluster analog with $\sigma_{\rm SIDM} = 1 {\rm cm}^2/$ g.

An alternative explanation for the observed galaxy-DM offsets

is due to statistical and observational uncertainties. Galaxies are sparse samples of the underlying DM overdensities, it is possible that the summary statistic of the sparse sample be different from those of the underlying distribution. It is not clear if there is any physical cause of the galaxy-DM offset in a CDM universe, but any statistical noise leading to an offset can influence this method of inferring σ_{SIDM}. Since the Illustris simulation assumes no SIDM, this study is complementary to staged simulation for understanding what can contribute to the offsets. Simply put, we perform a hypothesis test with the galaxy-DM offsets in the Illustris simulation directly corresponding to our null hypothesis $\mathcal{H}_0$, with:

$$\begin{cases} \text{the null hypothesis } \mathcal{H}_0 : \text{Cold Dark Matter (CDM)} \\ \text{the alternative hypothesis } \mathcal{H}_1 : \text{Self-interacting Dark Matter (SIDM)} \end{cases}$$
$$(1)$$

This exercise is further complicated by the fact that there is no theoretical foundation showing which observable would be the most sensitive to each possible type of SIDM. In fact, Kahlhoefer et al. (2014) have argued that SIDM does not cause significant offsets between the galaxy and DM peaks, and only cause an offset between the corresponding centroids for [TODO] a brief period of time after a merger. Popular choice for computing the offsets involves first inferring the summary statistic of the DM population and those of the galaxy population of a cluster independently before taking a difference. While there are well established procedures driven by lensing physics for inferring the DM spatial distribution, there is no standard procedure for mapping the sparse member galaxy distribution. We quantify the bias and uncertainty associated with the one-point summary statistic for summarizing member galaxy population.

Understanding the characteristics of different galaxy summary statistics of clusters are also important for probing the matter fluctuations in the universe. Galaxy clusters probe the large fluctuations in the matter power spectrum. As most large scale surveys are only in the optical, it is important to understand the characteristics of different summary statistics of galaxy populations of clusters. Performing lensing analyses on the stacked images of many small galaxy groups and clusters can give constraints to cosmological parameters. Stacking on the 'wrong' centers is a commonly cited source of uncertainties. By comparing the discrepancies of different galaxy summary, we can shed light to what can help maximize the lensing signal and the possible cause(s) of miscentering.

In this paper, we 1) extract realistic observables from the Illustris simulation for comparison with observations, 2) explore the pros and cons of the different statistic for summarizing *the member galaxy population* of a galaxy cluster, 3) give estimates for the offsets between the summary statistics of the galaxy population and the DM population under ΛCDM cosmology, which we call

$$\Delta \boldsymbol{s} \equiv \boldsymbol{s}_{\mathrm{gal}} - \boldsymbol{s}_{\mathrm{DM}}. \qquad (2)$$

where $\mathbf{s_{gal}}$ and $\mathbf{s_{DM}}$ are the two-dimensional (2D) spatial locations of the summary statistic of the galaxy population, and the density peak of DM respectively. This gives an estimate of the baseline scatter of offsets without any SIDM. And finally we 4) examine the properties of the clusters that give outliers in the offset distribution and 5) investigate the correlations between the physical properties of a cluster and the projected observables such as Δs.

The organization of this paper is as follows: In section 2, we will describe the physical properties of the data of the Illustris simulation (Vogelsberger et al. 2014a, Genel et al. 2014), and the selection criteria that we have employed to ensure that the quantities that we examine resemble observables but without noise and

systematics from observations. Then in section 3, we explain the methods for computing various one-point statistics of the spatial distribution of galaxies how we prepare our dark matter spatial data to resemble convergence maps. We show the statistical performance of the different summary statistics before we show the main results in section 4. In the discussion in section 5, we list the implications of our results and compare it to other simulations and observations. We also show how one may make use of the population offset statistical distribution from the Illustris data to construct a test with a null hypothesis of $\sigma_{\mathrm{SIDM}} = 0$ and discuss the caveats.

Our analysis makes use of the same flat Lambda Cold Dark Matter (ΛCDM) cosmology as the Illustris simulation. The relevant cosmological parameters are $\Omega_{\Lambda} = 0.7274, \Omega_m = 0.2726$, and $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2 THE ILLUSTRIS SIMULATION DATA

The Illustris simulation contains some of the most realistic, simulated galaxies to date, making it especially suitable for verifying the properties of galaxy clusters. We obtained our data from snapshot number 135 (cosmological $z = 0$) of the Illustris-1 simulation. The Illustris-1 simulation has the highest particle resolution and has incorporated the most comprehensive baryonic physics among the different Illustris simulation suites. The sophisticated galaxy formation model in Illustris-1 includes star formation rate, and stellar evolution due to environmental effects of the intracluster medium, such as ram pressure stripping and strangulation and feedback from Active Galactic Nuclei (AGN) etc. (Genel et al. 2014). The physics of stellar evolution were solved using a moving mesh code **AREPO** (Springel 2010). The observable properties of galaxies were statistically consistent with the Sloan Digital Sky Survey (SDSS) data (Vogelsberger et al. 2014a).

As the stellar population in Illustris were evolved from the initial condition, these makes the spatial distribution of galaxies in Illustris data more realistic than galaxies that are prescribed onto DM-only cosmological simulation data such as those used in Harvey et al. (2014). Gravitational effects in Illustris-1 have provided realistic dynamics and spatial distribution of subhalos. The simulated effects include tidal stripping, dynamical friction and merging. Since the profile of the galaxies clusters were not provided in symmetrical, parametric forms, we can study how asymmetry in the cluster profile affects the estimate of our summary statistic. This data allows us to examine cluster galaxies in a realistic, yet noise-free way. The softening length of the DM particles is 1.4 kpc and those of the stellar particles is 0.7 kpc, both in constant comoving units (Genel et al. 2014).

The two sets of data catalogs in use are obtained through two types of halo finders. The catalog that maps particles to the halo of a certain cluster was created by the **SUBFIND** algorithm. The friends-of-friends (FoF) finder (Davis et al. 1985) was further used to identify the affinity of galaxy-sized halos to a galaxy-cluster. These galaxy-size halos are referred to as *subhalos* and they are the dark matter hosts of what we refer to as galaxies in Illustris-1. Vogelsberger et al. (2014b) also extracted the absolute magnitude of each subhalo in the SDSS bands of g, r, i, z as part of the **SUBFIND** catalog using stellar population synthesis models.

For our analyses, we make use of galaxy clusters / groups with at least 50 member galaxies that are within a reasonable observational limit, i.e. apparent $i \leq 24.4$ which is the limiting magnitude of the DEIMOS spectrometer on the Keck telescope when we assume a cosmological redshift of $z = 0.3$ in the i band. The limiting

Figure 1. Left figure: Mass distribution of the group / cluster sized DM halos for different halo selection schemes. Mass estimates obtained by the FoF algorithm are labeled as M_{FoF}. Masses centered on the most bound particle within a radius those the average density is 200 or 500 times the critical density of the universe are labeled as M_{200c} and M_{500c} respectively. **Right figure:** Mass-richness relationship of galaxy clusters and groups with $M_{FoF} > 10^{13} M_{\odot}$ assuming different cosmological redshifts of the observed clusters.
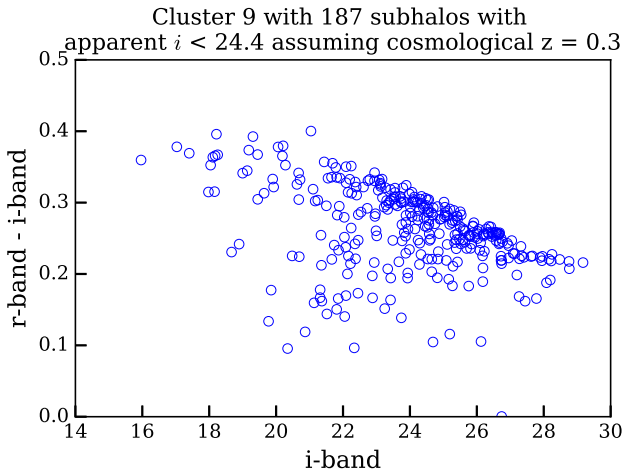

Figure 2. Color-magnitude diagram of one of the galaxy clusters that is selected for analysis. This cluster is the 9th most massive. The apparent magnitude is calculated assuming that the cosmological redshift (distance) is $z = 0.3$. We can see a clear overdense region that corresponds to a red-sequence. The color-magnitude diagrams of the other clusters can be found in the Jupyter notebook at https://goo.gl/TJmI6s.

magnitude of 24.2 in the F814W filter of the Hubble Space Telescope, and the limiting magnitude of the Canadian-Hawaii French Telescope of 24.5, are also close to our chosen limiting magnitude. There is relatively large statistical uncertainty if we try to analyze clusters with less than 50 member galaxies. As indicated by the right-hand panel of Fig. 1, a total of 43 clusters have survived this magnitude cut. These simulated galaxy clusters (or groups) have masses ranging from 10^{13} $M_{\odot}$ to 10^{14} $M_{\odot}$.

2.1 Cluster properties

2.1.1 Quantifying the dynamical states (relaxedness) of the galaxy clusters

Clusters undergo merger activities of a large range of physical scales and in the time scale of million of years. The dynamical history, or what we call "unrelaxedness" is hard to retrieve from simulations across different saved states and is missing from observations. We quantify the state of the cluster by providing several quantitative definitions of unrelaxedness and see how they correlate with Δs. Some possible definition of unrelaxedness referred by the simulation community include:

● unrelaxedness$_0$: the ratio of mass outside the dominant dark matter halo over the total mass of the galaxy cluster. The lower the ratio, the less substructures there are in the cluster.
● unrelaxedness$_1$: the distance between the most bound particle from the center of mass as a function of R_{200c}, the three-dimensional (3D) radius in which the average density is 200 times the critical density of the universe. The smaller the distance, there are less asymmetric substructures.

To relate these simulation quantities to observation, we compute more observation-oriented quantities in the method section 3.0.2.

2.2 Selection of the field-of-view

We make use of the **SUBFIND** member particle for the DM and the **FoF** subhalo identification as our default volume selection scheme for each cluster / group. We understand that this choice of volume selection can be more ideal than observational conditions. We make use of this volume selection scheme for baseline comparisons.

Assuming a conservative line-of-sight (los) distance , i.e. cosmological redshift, with [TODO] $z = 0.3$, the projected extent for most of the Illustris galaxy clusters and groups, fits inside the field of view of telescopes, such as the Subaru Suprime Camera,

Table 1. Selection criteria for stellar subhalos (member galaxies) for each cluster / group

Data	Selection strategy	Sensitivity	Relevant section
Field of view (FOV)	FoF halo finder	comparable to FOV of the Subaru Suprime camera	2.2
Observed filter	i-band	consistent among the redder r, i, z bands	2.3
Cluster richness	$i \leq 24.4$ and $z = 0.3$	sensitive to the assumed cosmological redshift of cluster and the assumed limiting magnitude of telescope	2
Two-dimensional projections	even HEALPix samples over half a sphere	discussed as results	2.2.1

which covers a physical area of [TODO] ~ 9 Mpc $\times 7$ Mpc (See https://goo.gl/ClZNvM for a Jupyter notebook showing the extent of the Dark Matter distribution of the most massive 129 clusters).

2.2.1 Spatial Projections

Unlike in staged simulation, picking out a particular projection for a cluster does not always make physical sense. For highly symmetrical clusters, most projections are similar. However, for mergers or asymmetrical clusters, there is no obvious choice for the plane of projection that can allow us to understand the cluster. There is no simplistic merger axis that should be projected along the plane of the field of view due to possible 3D spiraling trajectories of the merging components.

We therefore compute observables based on even sampling of angular orientation as our line-of-sight. As the order of projecting the data and estimating the summary statistic is non-commutative, we first project the data before estimating any projected observable. The computation of even angular orientation is done by using HealPy, which is a Python wrapper for **HEALPix** [1] (Gorski et al. 2005). Each line-of-sight centers on a **HEALPix** pixel. The number of projections that we employed is 768 for each cluster. With around 768 projections, the offset distributions of each cluster start to converge to a stable distribution. Even though there are at least 2 identical projections for each cluster due to one possible line-of-sight from the front and one from the back, it does not affect any summary statistic. We do not remove the duplication as it breaks the rotational symmetry in the 2D plane when we try to compute the 2D population distribution of offsets.

2.3 Properties of the galaxies in Illustris clusters

Different galaxies have different masses, so they should not be considered with equal importance for peak identification, which requires summing the mass proxies of different galaxies. One of the most common weighting schemes employed for galaxy data is to weight by the luminosity in a particular band. For some of the methods, we investigate the differences in peak identification with and without any luminosity weights. We pick the $i-$band magnitude associated with each subhalo as the weight. Since the $i-$band is one of the redder bands, the mass-to-light ratio is not skewed as much due to star formation activities. We further examined if the colors distribution of galaxies in Illustris-1 are similar to the observed color-magnitude diagrams for clusters. The Illustris cluster galaxies are realistic enough that it is easy to identify an overdense region of galaxies known as the red-sequence in the color-magnitude diagram such as Fig. 2. The red-sequence is prominent even if we use other colors formed by different combinations of the r, i, z bands.

---

[1] HEALPix is currently hosted at http://healpix.sourceforge.net

3 METHODS

A common and the most precise way of summarizing the DM distribution in a galaxy cluster is by finding the lensing peaks (Medezinski et al. 2013, Markevitch et al. 2004, Zitrin et al. 2013). Additionally, the peak region is physically interesting due to the higher particle density and interaction rates. The most direct analogous statistic for summarizing the member galaxy population in a cluster is therefore, also the peak. Comparing the DM peak with the summary statistics of the galaxy population that are not the peak therefore can have an *offset* purely due to the difference in the choice of the statistic for summarizing the two sets of identically distributed data. We will state the definition and implementation of five commonly used point statistic or location for summarizing the member galaxy population in a galaxy cluster.

We avoid any manual methods for comparison purposes, scalability and reproducibility. Since all the methods listed in this paper are automated with the source code openly available, it is possible for future studies to reuse our code for comparisons. Another major advantage for automation is that it allows us to apply the same methods across the different snapshots of the (Illustris) simulations to examine the variability of Δs across time in future studies.

3.0.1 Computing the weighted centroid

We follow the usual definition of the weighted centroid:

$$\bar{\mathbf{x}}_w = \frac{\sum_i w_i \boldsymbol{x}_i}{\sum_i w_i}, \tag{3}$$

with $\boldsymbol{x}_i$ being the positional vector of each subhalo and we use the i-band luminosity as the weight w_i for the i-th galaxy. Centroids can be biased by subcomponents from merging activities yet the centroid estimate do not provide explicit evidence for ongoing merger or accretion. These estimates are also sensitive to odd boundaries of the field of view.

3.0.2 Cross-validated Kernel Density Estimation (KDE) and the peak finder

Finding the exact peak of a sets of data points involves computing the density estimate of the data points and sorting through the density estimates. A specific version of this density estimation process is known as histogramming. During the making of histogram, each data point is given some weight using a tophat kernel and the weights are summed up at specific data locations (e.g. $\mathbf{x_i}$). Histogram is not good for peak estimate for *sparse* data for two reasons: 1) the choice of laying down the bin boundaries affects the count in each bin, 2) the choice of bin width also affects the count in the bin. Only when the available number of data points for binning is large, the estimates of histograms and smoothed density estimates are approximately the same. The number of member galaxies (< 500) is sparse enough for the uncertainty introduced by histogramming to bias our peak

estimate. For the density estimate of galaxy luminosity, we adopt a Gaussian kernel. The exact choice of the functional form of the smoothing kernel does not dominate the density estimate as long as the chosen kernel is smooth (Feigelson & Babu 2014).

For computing the density estimate, the most important parameter of computing is the bandwidth of the smoothing kernel, which takes the form of a matrix in the 2D case. We illustrate the choice of kernel width with Fig. 3. When the kernel width is too large (bottom left panel), the data is over-smoothed, resulting in a bias of the peak estimate. On the other hand, when the kernel width is too small, it results in high variances of the estimate and result in too many peaks due to noise. The decision of having to balance between creating high bias or high variance estimates is also known as the bias-variance tradeoff. Any other smoothing procedures, including interpolation with splines, polynomials, and filter convolutions, also face the same tradeoff.

A well-known way to minimize the fitting error from the density estimate is through a data-based approach called cross-validation to obtain the optimal 2D smoothing bandwidth matrix (H) of the 2D Gaussian kernel for the density estimate $\hat{f}$:

$$\hat{f}(\chi; H) = \frac{1}{n} \frac{1}{(2\pi)^{d/2} |H|^{1/2}} \sum_{i=1}^{n} w_i \exp((\chi - \mathbf{x}_i)^T H^{-1} (\chi - \mathbf{x}_i)), \tag{4}$$

where the dimensionality is $d = 2$ for our projected quantities, χ represents the uniform grid points for evaluation, and $\mathbf{x_i}$ contains the spatial coordinates for each of the identified member galaxies that survived our brightness cut and w_i is again the $i-$band luminosity weights for each galaxy. The idea behind cross-validation is to leave a small fraction of data point out as the test set, and use the rest of the data points as the training set for computing the estimated density. Then it is possible to estimate and minimize the asymptotic mean-integrated squared error (AMISE) by searching for the best set of bandwidth matrix values, eliminating any free parameter.

Specifically, we made use of the smoothed-cross validation (Hall et al. 1992) bandwidth selector in the statistical package `ks` (Duong 2007) in the `R` statistical computing environment (R Core Team 2014). Among all the different `R` packages, `ks` is the only package capable of handling the magnitude weights of the data points while inferring the density estimates (Deng & Wickham 2011). Although the particular implementation of KDE has a computational runtime of $O(n^2)$, the number of cluster galaxies is small enough for this method to finish quickly ($\lesssim 2$ second per projection per cluster).

The resulting KDE contains rich information about the spatial distribution of the clusters, and we focus on the peak regions. We employed both a first and second-order finite differencing algorithm to find the local maxima. The local maxima were then sorted according to the KDE density in a descending fashion before we perform peak matching and compute the offset. The exact procedure is discussed in section 3.3.

The number of distinct subclusters are also occupied by a bright peak. For each projection of each cluster, we normalize the density of all luminosity peaks to those of the brightest peak. Then we sum the density of all the galaxy peaks for a cluster and call this value ν. When the value of ν much gets bigger than 1, it indicates the presence of projected substructure(s). Even though ν is not expressed in terms of masses, it is a practical measurement for optical survey data where only galaxy magnitudes, but not spatial mass distribution from lensing estimates, are available.

3.0.3 Shrinking aperture estimates

Another popular method among astronomers for finding the peak of a spatial distribution is what we call the shrinking aperture method. While we do not endorse this method, we test if the shrinking aperture method is able to reliably recover the peak of the luminosity map. This method is dependent on the initial diameter and the initial center location of the aperture. This method does not evaluate if the cluster is made up of several components. The estimate using the shrinking aperture algorithm can be biased by substructures. The only way to inform the algorithm about substructures would be to introduce another parameter to restrict the extent of the aperture, or to partition the data with another (statistical) algorithm. More to the point, the convergence of results of this method is unstable. We use a convergence criteria of having the aperture distance not change more than 2% between successive iterations as a reference. The actual implementation in Python can be found at https://goo.gl/nqxJl8 while the pseudo-code can be find in Appendix A.

3.0.4 Brightest Cluster Galaxies (BCG)

The BCGs are formed by the merger of many smaller galaxies. The galaxy-cannibalism makes BCGs typically brighter than the rest of the cluster galaxy population by several orders of magnitude. However, star formation can cause less massive galaxies to be brighter in the bluer photometric bands. To avoid star formation from biasing our algorithm for identifying the BCG, we find the brightest galaxies in redder bands i.e. the r, i, z bands and found that they give consistent results for all selected clusters. We used the $i-$band to pick the BCG for computing the plots and the final results.

3.1 Comparison of the methods from Gaussian mixture data

In order to examine the statistical properties of commonly used point-estimates of the distribution of the galaxy data, we test them on data drawn from Gaussian mixtures with known mean and variance. (See Fig. 4). The main factors that affect the performance of the methods are sensitive to the statistical fluctuations of the drawn data, e.g. the spatial distribution of the data, including 1) the density profile and 2) the location(s) of subdominant mixtures, , and 3) the number of data points that we draw. It is also not enough to just compare the performance by applying each method for one realization of the data. We provide the 68% and the 95% confidence regions by applying the each method for many Monte Carlo realizations. In general, the peaks identified from the KDE density is closer to the peak of the dominant mixture (more accurate) than both the weighted centroid method and the shrinking aperture method. For example, in the bottom middle panel of fig. 4, it is clear that the green contours that represents the confidence region for the shrinking aperture peak is biased due to the substructure, whereas the confidence region for the centroid is so biased that it is outside the field of view of that panel. For the bottom right panel of fig. 4, we present how the confidence region shrinks as the number of data points increases. From that here is also a catastrophic outlier for the shrinking aperture method for 500 data points. The outlier shows how the shrinking aperture method can have radical behavior when there are subclusters in the data.

Figure 4. Comparison of peak finding performances of different methods by drawing data points (i.e. 20, 50, 100, 500) from known number of Gaussian mixtures. Panels from the top row contain data drawn from a single Gaussian mixture. The panels from the middle row contain data from two Gaussian mixtures with weight ratio = 7:3. The panels from the bottom row contain data drawn from three Gaussian mixtures with weight ratio = 55:35:10. The left column shows how 50 data points drawn from the fixed number of Gaussian mixtures look like. Due to the statistical nature of this exercise, we sampled the data and performed the analyses [TODO: state how many times] many times to create the 68% and 95% Monte Carlo confidence contours of the estimates in the zoomed-in view of the data in the middle column. The rightmost column shows how the size (median contour radius) of the confidence regions vary as a function of the number of drawn data points from the Gaussian mixtures. From the middle and the rightmost column, we can tell that the KDE peak estimate is the most accurate. [TODO]

3.2 Modeling the DM map in Illustris-1 and the lensing kernel

The most well established method of inferring the projected dark matter spatial distribution from observations is through gravitational lensing. It works by detecting subtle image distortions of background galaxies due to the foreground dark matter. The resolution of the inferred map therefore depends on the properties of the source galaxies that are being lensed, such as the projected number density, the intrinsic ellipticities and morphology etc. To achieve a sufficient signal-to-noise ratio for lensing, Hoag et al. (2016) has performed simulation for inferring the optimal bandwidth for a Gaussian smoothing kernel for the cluster MACSJ0416. In the strong lensing regime, Hoag et al. (2016) found a resolution of 11 arcseconds can best fit the MACSJ0416 data. A kernel bandwidth (this is the standard deviation) of 11 arcseconds translates to an angular diameter distance of 50 kpc assuming a cosmological redshift of $z \approx 0.3$. Clowe et al. (2012) also described using a 60 kpc radial Gaussian kernel to smooth out spikes in the weak lensing map of Abell 520. In order to match the resolution of lensing data, we also employed a smoothing kernel of a similar physical size of 50 kpc.

To compute the DM spatial distribution from our data, we first make histogram with 2 kpc × 2 kpc bin size which is slightly larger than the DM softening length of 1.4 kpc. After that, we use a (50, 50) kpc 2D radial Gaussian kernel to smooth the DM histogram Illustris DM particle data. There are resolution differences between the smoothed and unsmoothed DM maps. The unsmoothed histograms tend to show many more local maxima around the major density peaks (i.e. show high variance). The number of DM particles for each cluster is of [TODO double check] the order of millions and densely packed in the region of interest. Physically, the smoothed histograms of the dark matter of each cluster is analogous to a convergence map from a lensing analysis.

3.3 Finding the offsets

It is possible to have several peak estimates from the KDE of the member galaxy population of a cluster. From the density estimate at each peak, we can sort the peaks according to their densities. We only match luminosity peaks that are at least 20% as dense as the brightest galaxy-luminosity peak to avoid computing the offsets of spurious substructures, such as the peaks due to small number of galaxies that are located far away from the main concentration of mass.

In general, there are many more DM peaks because there are many more dark subhalos than galaxies for each cluster and the resolution of the DM data is much higher. To find the nearest DM peak to the significant galaxy peaks just as any astronomers would do manually, we construct a k-dimensional tree (KD-Tree; in our case, k = 2) using the densest n_{DM} number of DM peaks:

$$n_{\mathrm{DM}} = \begin{cases} 3 \times (n_{gal} + 1) & \text{if } n_{gal} < 3 \\ 3 \times n_{gal} & \text{if } n_{gal} \geq 3. \end{cases} \quad (5)$$

where n_{gal} is the number of significant galaxy peak, and n_{DM} is the number of peaks that went into the construction of the KD-tree. When there are more than one dense galaxy peaks located far away from one another, the top few densest DM peaks (subhalos) can be located around the same galaxy peak. i.e. there is no one-to-one matching between the luminosity of galaxies and the density of detected DM peaks. Matching purely based on density and luminosity leads to larger offsets. From inspection, using eq. (5) works well to match the appropriate peaks. We also show in the result section that

most of the chosen DM peaks do not have significant deviation from the most gravitationally bound particles. After identifying the DM peaks, we also compute the offsets between the DM peaks, and the following spatial estimates, including

- the most (gravitationally) bound particle
- the shrinking aperture peaks, the corresponding offset is $\Delta s_{\mathrm{shrink}}$,
- the number density peaks, the corresponding offset is $\Delta s_{\mathrm{num.dens}}$,
- the BCGs, the corresponding offset is Δs_{BCG} and
- the luminosity weighted centroid, the corresponding offset is Δs_{BCG}.

Since there can be more than one number density peak from the corresponding KDE map, we also use a KD-tree to location the closest number density peak to the identified DM peak.

[TODO] Figure out how the Illustris simulation figure out the most bound particle. The most bound particle is the location with minimum gravitational potential of the **Subfind** identified cluster. Due to substructures, it is possible for there to be several minima of similar gravitational potential level.

3.4 Constructing the hypothesis test

After matching the peaks, we use the offsets as the basis of our hypothesis test. We compute the p-value as the narrowest density interval of simulated offsets that are below observed values of offsets in the literature. The distribution of offsets in the Illustris simulation represents the possible ways that offsets would be observed in a CDM universe. This gives us a rough estimation of the probability of seeing the offset from observations under the null hypothesis of CDM being true. We also provide the biweight location and the density interval characterizing the distribution of offsets computed from each of the listed methods. For instance, the 95% interval is computed as the interval that encompasses 95% of total density (2.5% of density mass at each end of the tail is excluded). In case of degeneracy, the interval is also required to cover the location estimate for the distribution.

The different representations of the distributions of Δs have different statistical power for the hypothesis test, i.e. there are discrepancies between the spread of the offset distributions, so it affects the significance of an observation. The most faithful representation of the offsets without any information loss is:

$$\Delta \mathbf{s} = (\mathbf{x}_{\mathrm{gal}} - \mathbf{x}_{\mathrm{DM}}, \mathbf{y}_{\mathrm{gal}} - \mathbf{y}_{\mathrm{DM}}). \quad (6)$$

The PDF of $\Delta \mathbf{s}$ in eq. 6 peaks at $(0, 0)$ when there is no real offset. However, when one takes the magnitude of $\Delta \mathbf{s}$, i.e.:

$$|\Delta \mathbf{s}| = \sqrt{(\mathbf{x}_{\mathrm{gal}} - \mathbf{x}_{\mathrm{DM}})^2 + (\mathbf{y}_{\mathrm{gal}} - \mathbf{y}_{\mathrm{DM}})^2}, \quad (7)$$

the resulting 1D distribution of $|\Delta \mathbf{s}|$, those support being $[0, \infty)$, will not peak at zero even if the original distribution of $\Delta \mathbf{s}$ peaks at $(0, 0)$. To illustrate how difficult it is to interpret the magnitude values $|\Delta s|$, we can imagine the following transformation. The values drawn from:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2, 0 \\ 0, \sigma^2 \end{pmatrix} \right), \quad (8)$$

will result in $|\Delta s|$ values that follow the Rayleigh distribution:

$$f(\Delta s | \sigma) = \Delta s / \sigma^2 \exp(-\Delta s^2 / 2\sigma^2) \quad (9)$$

those peak is at $|\Delta s| = \sigma$, the standard deviation of the 2D Gaussian and ruling out any probability mass at $|\Delta s| = 0$ by construction. The dependency of $|\Delta s|$ on the parameters of the 2D distribution

is even more complicated when the the 2D distribution does not approximate a Gaussian or when there is more than one peak in the 2D space. The shifting of the peak location due to variable transformation is seen in the distribution of $|\Delta s|$ recorded in table C3. For an asymmetrical 1D distribution due to taking the magnitude, finding the narrowest 68% and 95% interval is not a standard statistical procedure and can be prone to error. In fact, several studies have reported poor single 2D Gaussian fits to 2D offset data due to the long tails (Zitrin et al. 2012, Oguri et al. 2010).

On the other hand, the 1D distributions of offsets along a particular spatial axis, e.g. Δx and Δy, each with a support of $\mathbb{R}$, will not exhibit a discontinuity at zero. Any shift in the 2D peak location is still obvious. The distributions represented by Δx or Δy can be symmetrical without any sharp cutoff at zero so the highest density interval is easier to find. Since we have enough samples for there to be rotational symmetry for the distribution of $(\Delta x, \Delta y)$, we will show that it does not matter much if we picked Δx or Δy for the 1D representation. We compute the hypothesis test significance level with the offset Δx along one of the spatial axes. To report the statistics, we also make use of estimates that do not make any underlying assumption of the shape of the distributions that may skew the statistical parameter estimate. We report the percentile for the offset distributions, and also the biweight statistical estimates such as the biweight location (analogous to the median) and the midvariance (robust standard deviation estimate). The biweight statistics are less susceptible to the effects of outliers (Beers et al. 1990). We compute the robust statistics using the implementation from Robitaille et al. (2013) as part of `astropy`.

In following sections of this paper, we use Δs to represent the two-dimensional offsets, $|\Delta s|$ for the magnitude of the offset as calculated according to the Euclidean distance, and Δx or Δy to denote the one-dimensional offset along one of the spatial dimensions. To compare with observed data, we estimate the 1D spatial components of the offsets from the merging cluster observations from various sources. We make our best attempt to measure Δy_{obs}, the spatial component of the observed offset along the axis connecting the subclusters if subcomponents exist. In our observed samples, Abell 3827 is the only exception to have no subclusters but only four bright galaxy peaks in the central region. We also show Δx_{obs} components, the offset perpendicular to Δy_{obs} for comparison. For most of the observed offsets, we obtain the estimates from the contour plots and descriptions of the corresponding papers. For the offsets that are roughly in line with the axis connecting the two subclusters, we let $\Delta y = |\Delta s|$. If Δs_{obs} is not aligned along the line joining the subclusters, using $|\Delta s_{\mathrm{obs}}|$ instead of Δy_{obs} to come up with a p-value from the distribution of Δy will lead to an spurious increase in significance.

4 RESULTS

4.1 The dynamical states (relaxedness) of the clusters

Out of the $43 \times 768 = 33024$ projections, $\sim 45\%$ of the projections have one dominant luminosity peak and negligible substructures, with the total peak density of the projection being $\nu \leq 1.2$. Another $\sim 50\%$ of the projections have more than one dominant luminosity peak with $1.2 < \nu < 2.2$. From Fig. C1 that shows the spread of ν estimate over different projections, most clusters show signs of other subdominant peaks.

Visually, the spread of the ν distribution is indicated by the horizontal length of the blue box. the median of ν per cluster are

indicated by the red central vertical line inside each box in Fig. C1. Only 7 clusters (with ID = 15, 16, 17, 22, 31, 35, 51) out of 43 clusters have $\nu \lesssim 1.2$ for most of the projections. Clusters with median values of $\nu > 2.2$ usually have multiple subclusters. Cluster with ID = 7, for instance, is made up of around 4 disconnected clusters that span several Mpc. There is also a strong correlation of ~ 0.8 between each of the two unrelaxedness quantities defined in section 2.1.1 and the median of ν over different projections of each cluster.

4.2 Offset between the matched DM peaks and the corresponding most bound particle

There is no significant differences between the matched DM peaks and the most gravitationally bound particle (hereafter most bound particle). The median of the offset between the DM peak and the gravitationally bound particle is (0, 0) kpc. The 75-th percentile of the offsets are at ($\pm 2, \pm 2$) kpc. Most of the other offset values occur below ($\pm 9, \pm 9$) kpc. Large offsets are only seen for clusters with $\nu > 1.2$. The densest DM peak in 3D where the most bound particle is located, does not necessarily correspond the densest projected peak in 2D in the presence of significant DM substructures.

4.3 Offset between galaxy summary statistic and the most bound particle

As another sanity check, we computed the offsets between different galaxy summary statistic and the most bound particle. Interested reader can refer to table C2 for the different percentile and robust estimates of the distribution of offsets from the most bound particle. The ranking in terms of increasing distance to the most bound particle computed by different method is as follows:

- BCG
- densest peak of the luminosity map created by weighted the KDE
- shrinking aperture center from the luminosity weighted galaxy data
- densest peak of the number density map created by the unweighted KDE
- centroid estimate using luminosity weights, which is a proxy for the center of mass

In fact, most of the BCG offsets are very small except for two clusters with ID 13 and 33. Both clusters have all the values of $\nu > 1.5$ over different projections. From the projected density map, we further confirm that both clusters have significant substructures. It is therefore possible for the most bound particle to have similar gravitational potential level as another substructure where the BCG is located. In general, the offset distributions between the galaxy summary statistics and the most bound particle have approximately the same level of variance but more extreme outliers (at the 99%) than the offset distribution between the DM peaks and the corresponding galaxy summary statistics.

4.4 Galaxy-DM Offset in Illustris

4.4.1 The two-dimensional distribution and distribution of Δy

The 2D distribution of Δs from most methods peak at around zero ($\lesssim 4$ kpc) with high rotational symmetry. The luminosity weighted centroid offset, i.e. median($\Delta x'_{\mathrm{centroid}}$) = -37 kpc, has the highest asymmetry along x-axis for the peak. These offsets distributes are recorded in detail in table C3. The estimates of the offsets from
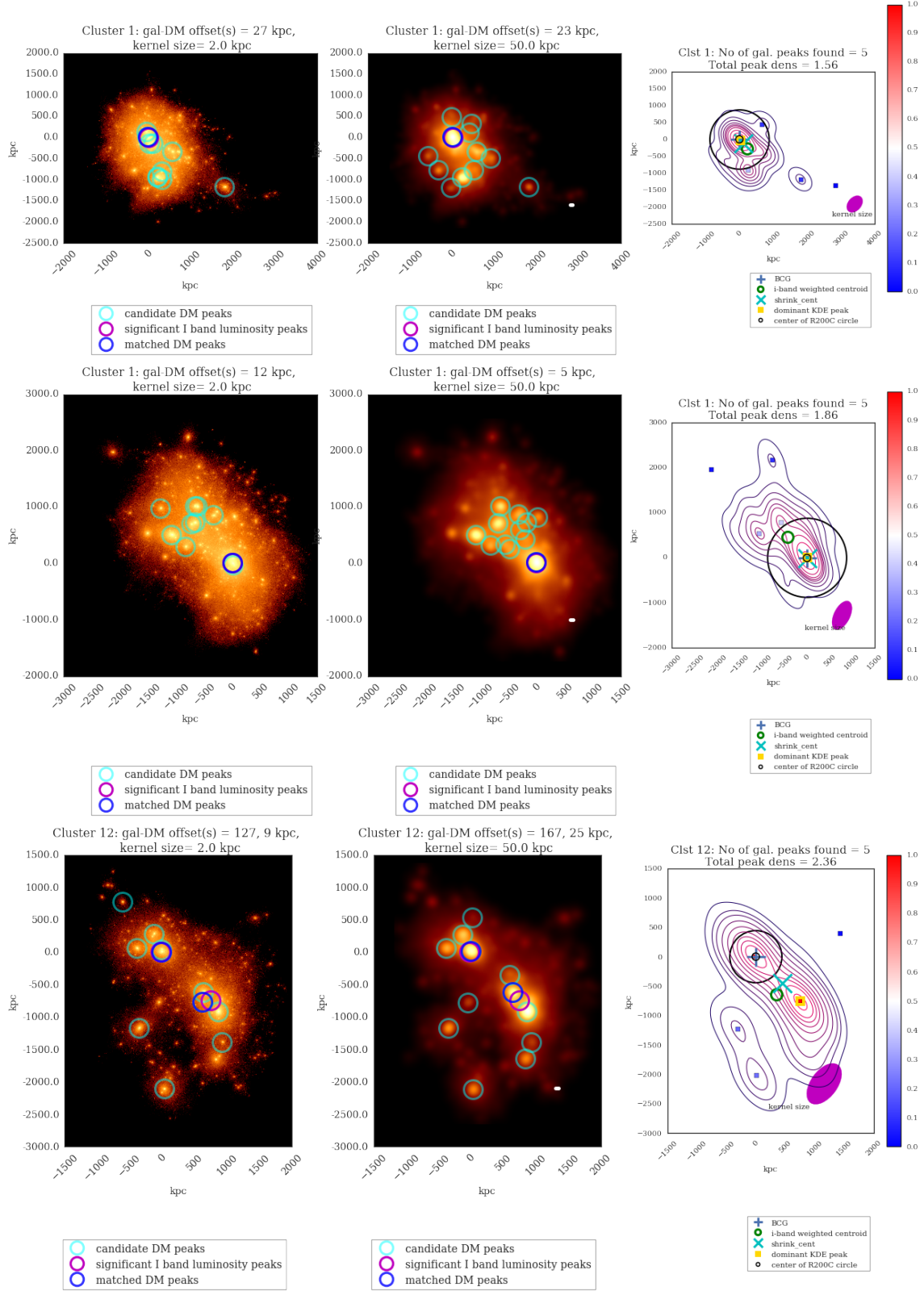
Figure 5. Visualization of clusters (each row is for the same projection of the same cluster). **Left column:** Projected density distribution of DM particle data is shown in orange, with the dense regions in yellow. The identified density peaks are indicated by colored circles. **Middle column:** The same DM projection after smoothing with a 50 kpc smoothing kernel (kernel size is indicated by a white dot on lower right of the panel. The thickness of the dot may be larger than 2 kpc for the plots on left hand column). **Right column:** Projected galaxy kernel density estimates (KDE) of the i-band luminosity map for the member galaxies of the same clusters. Each colored contour denotes a 10% drop in density mass starting from the highest level in red. Each of the magenta ellipse on the bottom right corner of each plot show the Gaussian kernel matrix H from eq. (4). The big black circle is centered on the most bound particle as identified by **SUBFIND** and the radius of the circle indicates the R_{200C}. The luminosity peaks (square markers) are colored by the relative density to the densest peak, the relative density is shown by the color bar. See http://goo.gl/WiDijQ and http://goo.gl/89edcM for the visualization of the selected clusters inside two Jupyter notebooks.

© 2015 RAS, MNRAS **000**, 1–23

Table 2. Robust estimates and the distribution of offsets along the y-axis (This is different from the magnitude which has discontinuity at zero).

sample	offset (kpc)	location	lower 68%	lower 95%	lower 99%	upper 68%	upper 95%	upper 99%
all ν	Δy_{BCG}	0	-3	-22	-496	3	456	1449
all ν	$\Delta y'_{\text{KDE}}$	0	-25	-79	-127	25	79	126
all ν	$\Delta y_{\text{num.dens}}$	0	-84	-303	-693	84	302	691
all ν	$\Delta y'_{\text{shrink}}$	0	-65	-295	-652	65	295	655
$\nu < 1.2$	Δy_{BCG}	0	-3	-10	-19	2	9	19
$\nu < 1.2$	$\Delta y'_{\text{KDE}}$	0	-18	-48	-82	18	48	83
$\nu < 1.2$	$\Delta y'_{\text{centroid}}$	0	-108	-255	-395	108	254	394
$\nu < 1.2$	$\Delta y_{\text{num.dens}}$	0	-73	-195	-303	73	195	302
$\nu < 1.2$	$\Delta y'_{\text{shrink}}$	0	-51	-187	-285	51	187	285
$1.2 < \nu < 2.2$	Δy_{BCG}	0	-3	-160	-684	4	807	1570
$1.2 < \nu < 2.2$	Δy_{KDE}	0	-32	-89	-125	32	89	124
$1.2 < \nu < 2.2$	$\Delta y'_{\text{centroid}}$	0	-262	-663	-905	262	663	904
$1.2 < \nu < 2.2$	$\Delta y_{\text{num.dens}}$	0	-87	-299	-739	87	298	738
$1.2 < \nu < 2.2$	$\Delta y'_{\text{shrink}}$	0	-85	-386	-777	85	386	779

Table 3. Observed offsets from clusters with reported evidence of mergers along line connecting two subclusters (Δy) and the approximate perpendicular offset (Δx). The table mainly contains clusters that have been used to constrain σ_{SIDM} using the reported offsets. Any approximate error estimates are the corresponding 68% lensing peak uncertainty in the figure(s) of the references, this is due to the lack of uncertainty estimates from the galaxy summary statistics from most literature. Error estimates are omitted when they are not reported by the authors. All p-value lower bounds are reported by matching to the corresponding method for estimating galaxy summary statistic in table 2.

Cluster	Δy (kpc)	Δx (kpc)	$|\Delta s|$ (kpc)	galaxy peak	DM peak	p-value	subcluster	mass (10^{14} M$_\odot$)	reference
Bullet	9	-23	25 $\pm$ 29	num. or lum.	SL & WL	$\gtrsim 0.32$	northwest	1.5	Randall et al. 2008
Baby Bullet	-40	0	$\sim 40 \pm \sim 50$	lum.	SL & WL	$\gtrsim 0.05$	northwest	2.6	Bradač et al. 2008:Fig.4
Baby Bullet	30	0	$\sim 30 \pm \sim 75$	lum.		$\gtrsim 0.32$	southeast	2.5	Bradač et al. 2008:Fig.4
Musketball	129	0	$129 \pm \sim 63$	num.	WL	> 0.05	southern	3.1	Dawson 2013:Fig.4.7
Musketball	-47	0	$47 \pm \sim 50$	num.		> 0.32	northern	1.7	Dawson 2013:Fig.4.7
Abell 3827	6	0	6	BCG	SL	> 0.05	central		Williams & Saha 2011
Abell 520	0	50	$\sim 50 \pm \sim 50$	lum.	WL	$\gtrsim 0.32$	blue	5.7	Clowe et al. 2012:Fig. 4
El Gordo	58	0	$\sim 58 \pm \sim 100$	lum.	WL	> 0.05	northwest	11	Jee et al. 2014:Fig.7,8
El Gordo	30	110	$115 \pm \sim 60$	num.		> 0.32	northwest		Jee et al. 2014:Fig.7,8
El Gordo	6	25	$\sim 26 \pm \sim 50$	lum.		> 0.32	northwest	7.9	Jee et al. 2014:Fig.7, 8
El Gordo	280	280	$400 \pm \sim 40$	num.		> 0.05	southeast		Jee et al. 2014:Fig.7, 8
Sausage	160	100	$\sim 190 \pm \sim 150$	num.	WL	$\gtrsim 0.05$	north	11.	Jee et al. 2015:Fig.10
Sausage	160	160	$\sim 190 \pm \sim 150$	num.		$\gtrsim 0.05$	south	9.8	Jee et al. 2015:Fig.10
Sausage	320	130	$\sim 340 \pm \sim 150$	lum.		$\lesssim 0.01$	north	11.	Jee et al. 2015:Fig.10
Sausage	160	160	$\sim 230 \pm \sim 150$	lum.		$\lesssim 0.01$	south	9.8	Jee et al. 2015:Fig.10

num. is a short hand for the peak estimate from the number density map.
lum. is a short hand for the peak estimate from the luminosity density map, or KDE' in the method description.
SL is a short hand for strong lensing.
WL is a short hand for weak lensing.

each method, are summarized with the biweight location statistic. It is noteworthy that the population spread of Δs computed by each method differ enough that one needs to know which offset method an astronomical study used for a fair comparison.

The 2D offset Δs_{BCG} has most of its density located near zero but contains outliers. Having outlier is possible because the DM peak is chosen as the closest DM peak to match the brightest luminosity peak in a particular projection. See bottom right panel of 5, the BCG is identified to coincide with the most bound particle. However, the luminosity peak of the cluster is located at the other mass substructure. When there are distantly separated subclusters of similar masses, the brightest projected luminosity peak may shift from one subcluster to another subcluster between different projections, while the BCG identification is unchanged between projections. The aforementioned bias from substructure can be seen when we compare the offset estimates between the relatively relaxed sample of $\nu < 1.2$ and the unrelaxed samples $1.2 < \nu < 2.2$. The 99-th percentile increased drastically from ± 19 kpc to an asymmetrical extreme

estimates of $(-684, +1570)$ kpc. Again, these values are possible because there can be several DM peaks of similar density due to subclusters located far apart from one another. The finite number of projections, combined with the substructures, have caused the 95-th and 99-th percentile tails of Δy_{BCG} of both the full sample and the unrelaxed sample exhibit noticeable asymmetry, but not the relaxed samples. Some field of views of the same cluster can be similar and happen to project the cluster along its longest extent. [TODO verify] It is likely that elongated clusters, such as cluster 7, and cluster 33, are responsible such asymmetry. We do not cut off the outliers nor fit any distribution to the Δy_{BCG}. This is because the tight peak of 3 kpc is probably due to numerical noise, (the softening length is around 1.4 kpc). There may not be a physical reason why the BCG center is offset from the DM peak. While we trust our 68% limit for the BCG for the relaxed sample, the percentile estimates for the 95% and 99% seem noisy for the unrelaxed sample.

The other offset estimates also exhibit difference variance. For the full sample in table 2 and fig. 6, the offsets computed by the

peak from the luminosity weighted KDE has the second smallest variance. The 68-th percentile of $\Delta y'_{\mathrm{KDE}}$ is at ± 25 kpc. Using shrinking aperture to estimate the peak location from the luminosity map increased the 68-th percentile of the offset to more than double those of $\Delta y'_{\mathrm{KDE}}$ at ± 65 kpc. The peak estimate from the number density map has even larger variance, with its 68-th percentile being ± 84 kpc.

Most of the percentile intervals of the unrelaxed samples $1.2 < \nu < 2.2$, when compared to the relaxed samples, are around a factor of 2 larger. With the relatively relaxed samples, the variance of the inferred offsets from different methods still show significant discrepancies. The variance of the offset computed from the shrinking aperture method, the number density map, and the weighted centroid are still at least a factor of 1.5 larger than those computed using the luminosity-weighted KDE. In particular, the 68% percentile of the centroid method is ± 108 kpc. This is around one-fourth the typical core radius of massive clusters (Allen 1998).

The spread of the offsets inferred by each method affects their ability for constraining σ_{SIDM}. We will further elaborate on this point when we compare our results with staged simulations of SIDM in section 5.3.

4.5 Offset projection uncertainty of each cluster

When we gather the offsets $\Delta s'_{\mathrm{KDE}}$ of the 768 projections for each cluster, we can find the offset uncertainty due to projection effects. The distributed are illustrated in the box plot of Fig. 7. The values of the biweight mid- variance of $\Delta y'_{\mathrm{KDE}}$ for half of the clusters are < 23 kpc. Of the ten clusters (ID = 3, 7, 12, 20, 21, 32, 33, 37, 40 and 46) that have mid-variance > 40 kpc, all of them have the median of $\nu > 1.2$.

4.6 Correlations between different variables and the offsets

Here we investigate a list of physical quantities that have significant to little correlation with the offsets. We use the Pearson product-moment correlation coefficient to quantify linear relationship between the pairs of variables. (aka Pearson's r, hereafter ρ) We describe the significance of the correlation based on the p-value reported by `Scipy` of seeing the level of correlation by chance assuming the pair of quantities have no correlation. If the p-value is greater than 0.1, we call the correlation as insignificant. As a reference, the correlation of between the unrelaxedness criteria defined in section 2.1.1 for the 43 selected clusters is as high as 0.82.

Each of the unrelaxedness have significant a positive correlation of ~ 0.70 with the maximum of $\Delta s'_{\mathrm{KDE}}$ of each cluster (hereafter $\max(\Delta s'_{\mathrm{KDE}})$). The offset $\max(\Delta s_{\mathrm{KDE}})'$ per cluster also show a high correlation of 0.77 with the median ν per cluster (hereafter, median(ν)). The FoF mass of each cluster shows only slight correlation of 0.28 with median(ν).

There is weak correlation between the richness of the clusters with $\max(\Delta s_{\mathrm{KDE}})$ ($\rho = 0.21$). This weak correlation may be due to the fact that the peak estimate is only affected strongly by a few bright galaxies near the peak. The richness is slightly more strongly correlated with the median of ν with $\rho = 0.33$.

There is insignificant negative correlation ($\rho = -0.20$) between the different mass measured within a certain density threshold, such as M_{200C}, M_{500C}, and $\max(\Delta s_{\mathrm{KDE}})$. The quantities M_{200C} and M_{500C}, which are computed within a shell centered on the most bound particles, have symmetry assumptions that may not capture the total mass well if there are substructures. The FoF mass, which

captures the total mass without any symmetry assumption, correlates positively and weakly ($\rho = 0.13$) with $\max(\Delta s_{\mathrm{KDE}})$.

5 DISCUSSION

5.1 Other findings from the visual inspection of the simulated galaxy clusters

We inspected both the luminosity maps and the number density maps of the member galaxy populations. With the same selection of bright galaxies of apparent i-band< 24.4 at $z = 0.3$, the luminosity maps in general resemble the DM maps more closely than the number density maps. A comparison of the projected DM map, the luminosity map and the number density map of 129 clusters can be found at https://goo.gl/kZUWrg, https://goo.gl/R7VNi9 and https://goo.gl/lmQUPd respectively. We encourage our readers to view the luminosity map to see how it is possible to create scientifically accurate contours of the DM distribution if the member galaxy data is of high completeness and purity. The KDE is not just a fancy way of identifying possible BCG candidates close to the luminosity peaks. (As the high resolution figures need to be downloaded, the corresponding Jupyter notebooks may take some time and several refresh of the web page before they are rendered properly.)

In real observations, missing selection of member galaxies, or foreground objects can both affect the inference of the galaxy spatial distribution. The number density map can be less susceptible to bias from bright foreground objects. It is less clear about the effect of missing member galaxies for computing the luminosity map because there is a selection bias favoring bright member galaxies.

5.2 Comparison to merging cluster observations

The offset distributions in table 2 represent an estimation of the spread of Δs in a ΛCDM universe. While the compilation of the offset distribution suffers from the small number of high mass clusters in the Illustris simulation, the distribution probes projection uncertainties comprehensively. Possible discrepancies that affect our comparison can arise due to the high purity and completeness of the Illustris galaxy data. Most observations of massive galaxy clusters with total mass of $M_{\odot} \approx 10^{14}$ have richness $\lesssim 300$, while the most massive Illustris cluster ($M_{\mathrm{FoF}} = 3.23 \times 10^{14} M_{\odot}$) has a richness of 483. Furthermore, the Illustris clusters are less massive than the observed clusters with huge offsets. The Illustris clusters may also provide sufficient samples of possible spatial configurations of observed clusters. Despite the possible differences, we try to do a fair comparison to observation by matching the method of offset inference.

Other complications can arise from how one choose to subset the offset estimate based on the cluster properties. The full sample in table 2 preserves the cluster mass-abundance relationship, i.e. each cluster has the same number of projections in the full sample. However, it also underestimates the offset spread because the full sample includes $\sim 45\%$ of relatively relaxed projections that only have one primary luminosity component. These clusters would have been excluded for comparison with bimodal mergers. Subsetting with $1.2 < \nu < 2.2$ picks out cluster projections that are in more similar dynamical states as the observed merging cluster. However, some simulated clusters may have more projections included in this sample than the other clusters. From inspection of the mass abundance relations in Fig. B1, we found that subsampling with $1.2 < \nu < 2.2$ includes a higher proportion of projections from
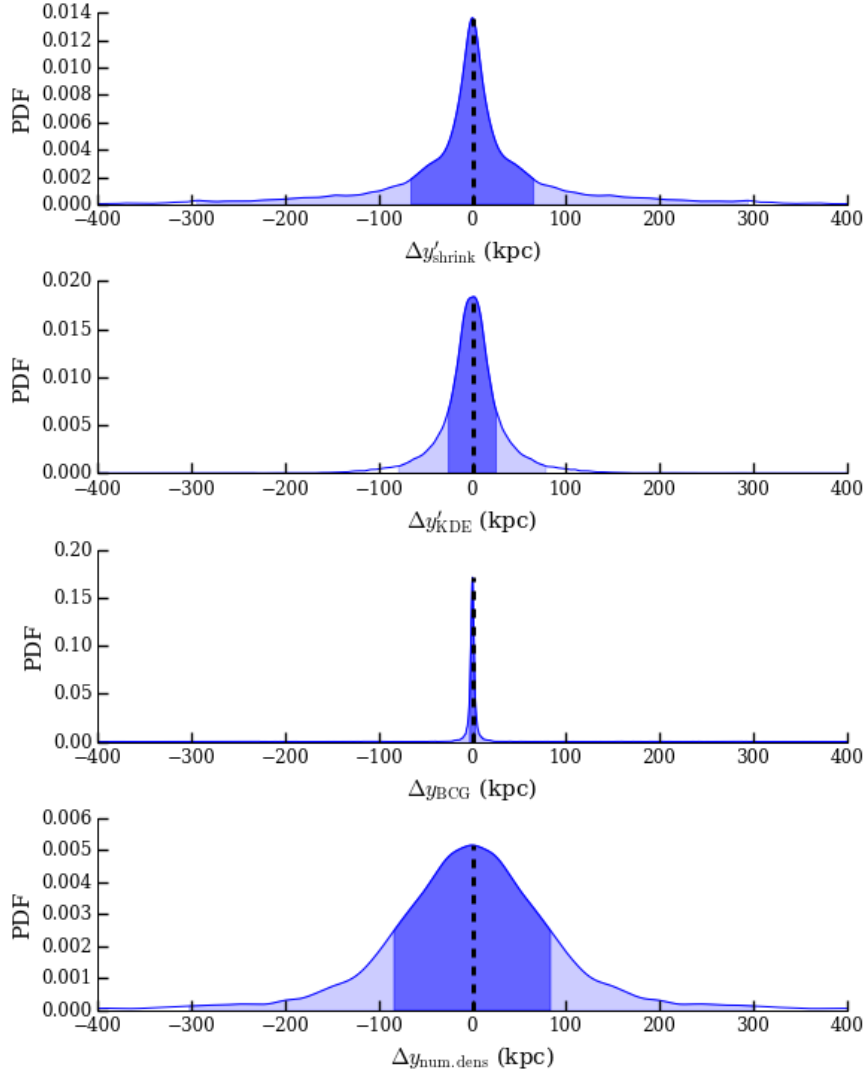
Figure 6. The smoothed distribution of different offsets of 43 clusters with all 768 projections. The smoothing bandwidth is determined by Scott's rule for visualization. For estimates where several peaks of galaxy data are possible, only the densest peak is matched to the DM peak for computing the offsets in this figure. The dark blue area indicates the 68% density interval while the light blue area shows the 95% density interval. The table summarizing the statistic of each distribution is available in table

massive clusters ($\sim 20\%$ more) than the full sample. This sampling should not introduce significant bias. The observed merging clusters are still more massive than the unrelaxed samples.

We compile table 3 with the corresponding lower bounds for the p-values using the Δy distributions of the unrelaxed samples ($1.2 < \nu < 2.2$). The p-value lower bounds are reported base on the Δy distribution from the same methods that the observed values were computed. The majority of the p-values from table 3, 13 out of 15, are > 0.05, only 2 p-values are below 0.05. This means the observations are mainly consistent with the null hypothesis: It is possible to see offset values as extreme as reported by observations in a CDM universe. However, this does not mean that the CDM model is more probable than the SIDM model. The full physical implications of this result is rather involved so we defer the discussion only for interested readers in section [TODO]. We continue to discuss the possible use of 2 and the p-values in tables 3.

To avoid any report from this study directly quoting a p-value, we did not combine the p-values from different observations. This

is because the computation of p-value does not fully takes the true uncertainties of the observations into account. Most of the quoted uncertainties from 3 are only from the lensing estimate, but not the error estimates of the galaxy summary statistic. Any future studies that wish to claim significance or rejection based on comparison to a simulation will need to carefully track down the contribution of uncertainties from each aspect of the cluster analysis. It may be reasonable to inversely weight offsets by the reported bootstrapped uncertainty $|\Delta s_{\mathrm{KDE}}|$ for each cluster.

As warned by the American Statistical Association in a statement [TODO], p-value can be misinterpreted easily, so we must take caution in the procedure of computing and interpreting the p-values. For example, the high energy physics community deliberately set the detection threshold to be $5\,\sigma$ to account for how systematic and bootstrapped uncertainties can be underestimated or unaccounted for. If the reliability of each observation is approximately the same, averaging the p-values can be one way of combining the p-values but there is no consensus of the best practice of combining the p-
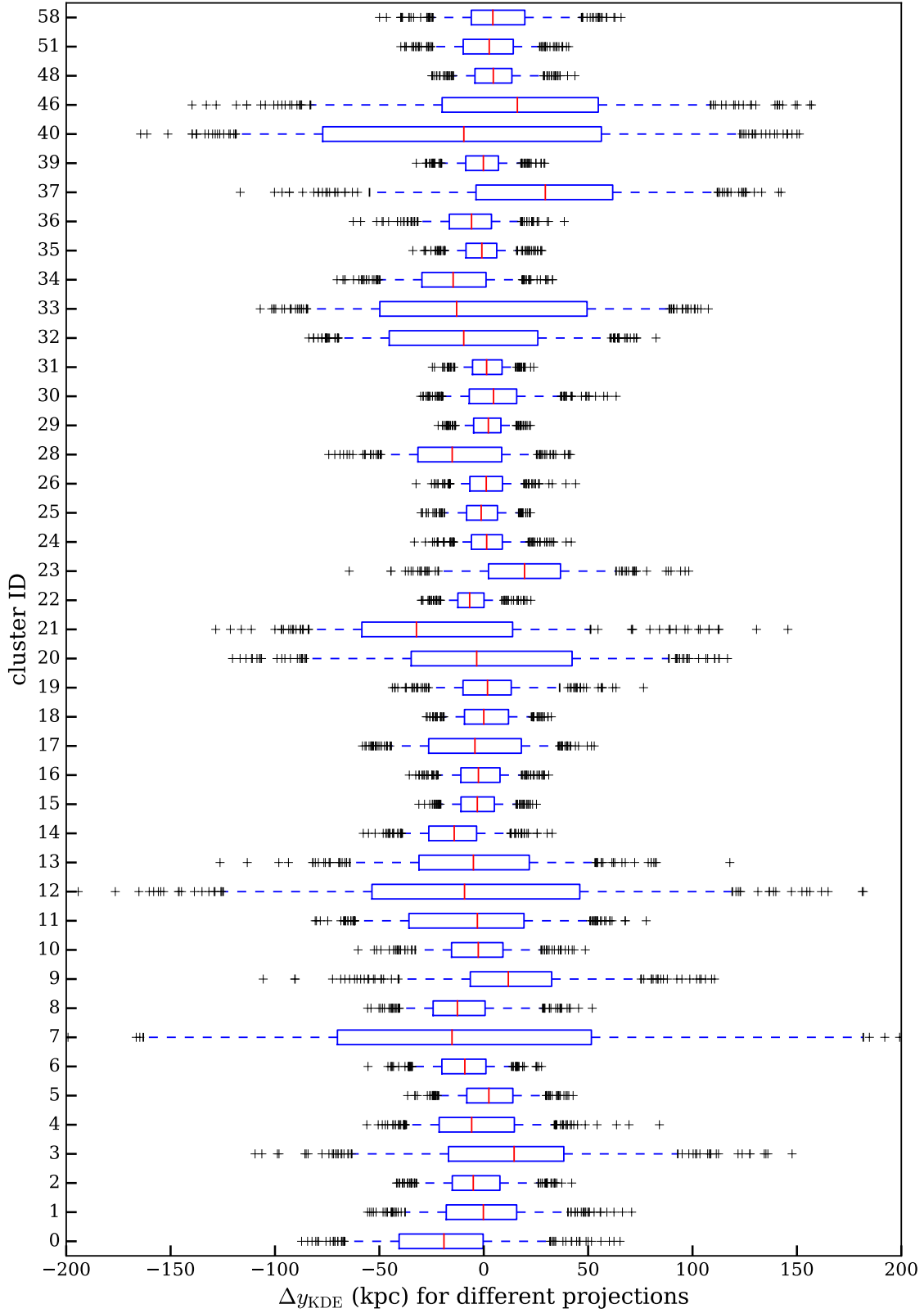
Figure 7. A box plot showing the distribution of Δy_{KDE} for each cluster based on 768 projections. The red line shows the median of the projections, the box encompasses the 25-th and 75-th percentile of the distribution while the whiskers mark the 5-th and the 95-th percentile. The other black crosses are data points with extreme values beyond the 5-th and 95-th percentile. The offsets were computed between the closest DM peak to the brightest luminosity peak of each cluster.

values. The rule used to combine the p-values needs to be carefully designed according to the goal of the experiment. When one keeps on including more observations, there will be some observations with extreme p-values that greatly influences the overall p-value. It is necessary to adjust for the sample size for such running experiments. The particle physics community, for example, discusses how to set stopping rules [TODO add reference] to account for the sample size to determine what p-value level should indicate meaningful discrepancies from the expectation of the null hypothesis *before* analyzing the data. How to count the sample size, i.e. whether to count the offsets from the same system as the same sample, is a subject up for debate. Regardless, computing a combined p-value for all the listed observations alone does not help understand how to best constrain SIDM. We instead focus on discussing possible ways to build on top of the results of this study and improve the design of the analysis in section [TODO].

5.2.1 Comparison of the offset results to studies of clusters and groups

Our results showing offset distribution are highly relevant to two other types of observational studies: The first type tries to map the DM distribution of galaxy clusters using lensing techniques. The second type focuses on finding the best way to stack galaxy groups for inferring cosmological parameters from the galaxy cluster mass function. In the following section, we will show that cosmological simulations generally show the BCGs has the tightest offsets from the DM peaks, which is consistent with our finding. Observational studies, however, in general do not find offset as tight as the simulations. We discuss some factors (other than SIDM) quoted in the literature that have shown to affect the observed offsets, including the lensing resolution and the choice of summary statistic for the member galaxy population. In particular, in the comparison of the luminosity peak to other galaxy summary statistic, the luminosity peak shows promise to give the second least amount of bias, after the BCG.

To establish the baseline of the tightest Δs, we first discuss and compare the $\Delta s_{\rm BCG}$ constraints. Cui et al. (2016), using 184 galaxy clusters with M$> 10^{14}$ M$_\odot$ in a N-body and hydrodynamical cosmological simulation suite powered by GADGET-3, also identified the maximum smoothed particle hydrodynamic (SPH) density peak to summarize the galaxy population in their cluster samples. Cui et al. (2016) found most offsets between BCGs and the most gravitationally bound particle to be mostly below 10 h^{-1} kpc. They also reported some extreme outliers spanning up to several hundred h^{-1} kpc due to the disturbed cluster morphology. Our tight 68-th percentile of $\Delta y_{\rm BCG}$ at ± 3 kpc gives some confidence that we have identified most of the BCGs correctly in the Illustris simulation.

The distributions of $\Delta s_{\rm BCG}$ derived from simulations, in general, are less spread out than those computed in most observational studies. For example, Oguri et al. (2010) have analyzed 25 X-ray luminous massive galaxy clusters of the LoCuSS survey. Observations were performed with a large FOV ($\sim 3h^{-1}$ Mpc) using the Subaru Suprime Camera. By fitting elliptical NFW models to the weak lensing data, Oguri et al. (2010) showed a long tail distribution for $\Delta s_{\rm BCG}$, which they fit with two 2D Gaussians. The first tighter 2D Gaussian had a standard deviation being 90 h^{-1} kpc for describing the offset for most clusters, the long tail spans around 1 Mpc was fit by a second 2D Gaussian with a standard deviation of 420 h^{-1} kpc. This second component in the tail region contains $\sim$ 10% of the clusters in the study and is consistent with the number of extreme outliers that we have. Note that by subsetting to only

consider relatively relaxed clusters, there is a tight maximum bound ± 19 kpc for $\Delta y_{\rm BCG}$ in our study.

One major source of uncertainties for computing $\Delta s_{\rm BCG}$ is the misidentification of BCG. To see the effects of misidentifications of BCGs, some studies have made use of N-body hydrodynamical cosmological simulations to compute the 2D distances between the BCG and the second most massive galaxies. Johnston et al. (2007) and Hilbert & White (2010) (using the Millenium simulation) found the one sigma level offsets for misidentified BCGs at 380 h^{-1} kpc, and 410 h^{-1} kpc respectively. This is consistent with the 95-th percentile of the unrelaxed and the full sample of the BCG in our study and also the tail of the $|\Delta s_{\rm BCG}|$ for Cui et al. (2016). If one wishes to use BCG with high confidence, it may be necessary to set stringent standard of the morphological characteristics of the BCG for classification, e.g. a minimum half light radius.

Another source of uncertainty for $\Delta s_{\rm BCG}$ was from lensing. Dietrich et al. (2012) performed an analogous analysis of the work of Oguri et al. (2010) using the N-body Millenium Run (MR) simulation. They showed that a combination of shape noise and modeling choices alone can lead to hundred-kpc-level offsets between the most bound particle (proxy of the BCG) and the lensing peak for cluster-sized DM halos (M $> 10^{14} M_\odot$). Dietrich et al. (2012) ray-traced through the DM substructures in the MR simulation as mock lensing observation of 512 clusters. Without any smoothing and shape noise, Dietrich et al. (2012) showed 90% of the lensing peak and the most bound particle, which is a proxy for the BCG, agree to 2.0 h^{-1} kpc. With shape noise, even at the source galaxy density of space-based quality optical data of n = 80 arcmin^{-1}, fitting NFW halos and using the center as the DM gave a distribution of $|\Delta s_{\rm BCG}|$ those mode is at around 9 arcsecs (this is approximately the 1 standard deviation estimate in 2D). Lowering the source galaxy density to 30 arcmin^{-1} increased the mode to 22 arcsec with a 95-th percentile at 85 arcsec ($\sim$ 90 kpc and $\sim$ 400 kpc at z = 0.3, comparable to Oguri et al. 2010). Smoothing, in the presence of shape noise, resulted in an offset distribution with the mode at 15 arcsec ($\sim$ 60 kpc at z = 0.3). The offsets do not simply depend on the smoothing bandwidth, but also the number density of the source galaxies that are lensed. While the uncertainty from smoothing the DM map is relatively unimportant in the Illustris analysis due to the much higher resolution of the DM particles than the sparser source galaxies used for lensing, it highlights why the bootstrapped uncertainties from the observed DM peak need to be accounted for during the comparison between the Illustris results and the observations.

It is noteworthy from the work of Dietrich et al. (2012) that smoothing alone can caused the peak offset to shift from several arcsecs to around 1 arcmin. Any other morphological features from the smoothed DM map will be subject to uncertainty of similar order of magnitude as the peak estimate, but with lower signal. There is also a risk of seeing substructures that mimic the sought-after morphological patterns. After all, other cosmological experiments that have shown that it is possible to detect patterns once scientists look hard enough, such as the initials of Stephen Hawking in the seventh-year Wilkinson Microwave Anisotropy Probe (WMAP) Cosmic Microwave Background data (Bennett et al. 2011).

Now we turn to exploring complementary methods for summarizing the galaxy population of a cluster, and show that the luminosity peak is the second best choice than the BCG for summarizing the galaxy statistic. A unique BCG does not always exist for a cluster (or the subcluster). Bright galaxies in the dense region of the cluster are the possible progenitors of the BCG and therefore a reasonable choice when there is no unique BCG. George et al. (2012), for example, examined 129 X-ray selected non-merging galaxy groups in the

COSMOS field. They found that around 20% to 30% of groups have non-negligible discrepancies between different galaxy centroids. By stacking on a bright galaxy near the X-ray centroid, they found the resulting lensing strength is higher than the stacked lensing signal based on other galaxy centroids, including the BCG. For groups with clear BCG candidate, George et al. (2012) gave the range of offset between the BCG and the assumed halo center as $\lesssim 75$ kpc. The KDE peaks from the luminosity maps of the Illustris samples show a much tighter offset to the lensing center than any other centroids that George et al. (2012) investigated. The weighted or unweighted centroid measurement from George et al. (2012) has a $|\Delta s|$ with standard deviation at 50 - 150 kpc from the lensing center with long tails (of around several hundred kpc). In comparison, the median (26 kpc), mean (37 kpc), standard deviation (35 kpc) and 75-th percentile (49 kpc) of $|\Delta s_{\mathrm{KDE}}|$ from all the Illustris samples are below 50 kpc.

We attribute the small amount of population bias of Δs_{KDE} in our analysis due to cross validation, a well-tested procedure accepted by all the top journals from the statistics and the machine learning communities. Not only does the algorithm help determine the eigenvalues, but also the optimal eigenvector direction of the bandwidth matrix. Most literature do not discuss how the bandwidth of the kernel for the smoothed maps are determined. It is unclear if such results will enjoy the same accuracy level as what we demonstrated. If scientists hand tune the smoothing bandwidth, it is hard to avoid setting the bandwidth to fit the preconception of how the density contours of the cluster should look like, and inadvertently biasing Δs.

[TODO revise this paragraph]

5.3 Comparison to staged simulations with SIDM

Staged simulation are controlled experiments for probing the contribution of SIDM to offsets in mergers of cluster components. The non deterministic nature of particle interactions means that it is not easy to predict the offsets analytically. There is no consensus of what type of clusters might best show the effects of SIDM. This is because the effects are dependent on the model of SIDM. Currently, the studies of SIDM have been restricted to those with isotropic scattering and velocity-independent cross sections. The two main classes of SIDM that are studied include those with frequent, short-range interactions that can be broadly be modeled by an effective drag force, and those with rare, longer range interactions those effects are not well described by a drag-force approximation (Kahlhoefer et al. 2014). In addition having a dependence on the merger configurations, the offset due to SIDM also has a time-dependence.

At SIDM cross section level favored by current literature $\sigma_{\mathrm{SIDM}} \lesssim 1\mathrm{cm}\ \mathrm{g}^{-1}$, a list of SIDM simulation studies (Kim & Peter 2016, Robertson et al. 2016, Kahlhoefer et al. 2014, Randall et al. 2008) have reported that, during the offset is observable, offset generally scales linearly with σ_{SIDM}. However, most of the upper limit of the offsets due to SIDM to are of the order of magnitude of 50 kpc. For simulations of major mergers of two subclusters, it takes time for the self-interaction of DM to manifest and lag behind the galaxies. Kim & Peter (2016), for example, showed that this lag starts to be observable approximately after the subclusters reached apocenter. While the offset may persist when the subclusters are returning for a second collision, the magnitude of the offset can fluctuate over this period. In general, the offset is affected by the phase of major mergers that cannot be accurately estimated.

This small level of offset due to SIDM rules out the use of most summary method of galaxy statistics for computing SIDM

constraints. The uncertainties overwhelm the signal. A offset level of 50 kpc is within the one-sigma level of $\Delta s_{\mathrm{num.dens}}$, $\Delta s_{\mathrm{shrink}}$, and $\Delta s_{\mathrm{centroid}}$, the two-sigma level of the Δs_{KDE}. Despite Δs_{BCG}. has tight distribution for relaxed clusters, misidentification of the BCG may increase the tail of the distribution to render it unsuitable to be used for secure constraints. And to sample the tight distribution at higher resolution than provided here, a simulation at galactic scales with realistic baryonic feedback will be needed.

6 THE ROLE OF CDM COSMOLOGICAL SIMULATIONS FOR STATISTICALLY INFERRING SIDM FROM A POPULATION OF GALAXY CLUSTERS

So far the analysis has focused on comparing the offset uncertainties from the inference methods and the SIDM offset signal, but we must emphasize that the high level of uncertainty from CDM simulation matching the observed level is insufficient to rule out small SIDM cross section. We have not shown if realistic mock observations of cosmological simulations of SIDM with small cross section would match observations in better or worse ways. It is also possible that the SIDM results would be largely indistinguishable from our results. This is because the uncertainties from the offset inference method is not guaranteed to bias the offset level high. Also having an small SIDM cross section only produces offset under restricted set of merger configurations.

Assume we have chosen a certain offset estimation method and to study n clusters of a certain mass range and dynamical states (i.e. merger with $1.2 < \nu < 2.2$), the distribution of population observed offset signal has two main contributions that we assume to be independent:

$$\Delta s_{\mathrm{obs}} = \Delta s_{\mathrm{SIDM}} + \Delta s_{\mathrm{noise}} \qquad (10)$$

From our analysis, we showed that for most of the observation methods, $\Delta s_{\mathrm{noise}}$ has an approximately zero mean and a certain covariance matrix. We need to evaluate a few things before deciding whether the contribution from Δs_{SIDM} with a certain cross-section will create a second peak excess for Δs_{obs} away from zero. The distribution of Δs_{SIDM} depends heavily on the fraction of merger configurations that can give non-zero offset signal of SIDM given a cross section, the SIDM offset distribution of such mergers and the chance of picking such merger configurations to observe over the total observed population. As we know that the mean offset signal of SIDM, within the cross section range favored by current literature, is close to the peak of the $\Delta s_{\mathrm{noise}}$ distribution, if Δs_{SIDM} from the chosen mergers have high enough variance, studying more mergers will not result in an excess of population offset signal of SIDM at any characteristic length scale, when compared to $\Delta s_{\mathrm{noise}}$. Although, even in the optimistic case that one figures out how to pick mergers to create a population offset signal excess at a certain length scale, there are practical limits on how well one can find them in observations.

7 FUTURE RESEARCH DIRECTION

the data in the Illustris simulation has proven to be an excellent source of understanding observational uncertainties. further work can be done to make mock observations more realistic, including:

- adding foreground structures

• adding more realistic peak detection threshold from the KDE luminosity map based on the flux per area at the peak location and compare that to observational limitations

• examining the effects of incompleteness removing some member galaxies randomly

• computing the velocity dispersion of the galaxies along different line-of-sight to see how the velocity dispersion correlates the different measures of unrelaxedness

• testing for identifying member galaxies using the red sequence

• zoomed-in simulations of clusters with SIDM physics

8   SUMMARY

We have shown that mock observations from simulations are important when there are no theory mapping observables to σ_{SIDM}. Kim & Peter 2016 have shown that the SIDM offsets depend on the time of the merger, impact parameters, collisional velocity and mass concentration. There is yet to be an analytical model that includes all these merger parameters. We cannot propagate observational uncertainty analytically when we compare between theories and have to rely on simulations. we showed that

• it is possible to see Δs as extreme as those in observed merging galaxy clusters assuming that ΛCDM is the true underlying physical model

• the contribution of statistical uncertainty to the galaxy-DM offsets for ΛCDM clusters is not negligible when compared to the reported levels of offset from staged simulations (~ 50 kpc). Any observational study that uses galaxy-DM offsets to constrain σ_{SIDM} has to account for this contribution or else σ_{SIDM} will be biased.

• while the location estimates of the 2D spatial distribution of offsets and the 1D spatial distributions of the offsets (Δy) in the Illustris simulation are approximately zero. The root-mean-square of the magnitude of the offsets is ~ 30 kpc from the Illustris sample. If one uses the magnitude of the offset to constrain σ_{SIDM}, $\sigma_{\mathrm{SIDM}} = 0$ does not map to zero magnitude of offset. The converse may also not be true but needs to be verify by a cosmological simulation with realistic galaxy distribution and SIDM.

• with high completeness and purity of member galaxy data, the luminosity map resembles the DM spatial distribution more closely than the number density map of member galaxies.

• the locations of the most gravitationally bound particle are consistent with the BCG for systems with little substructures.

• the BCG has the smallest offset to the dominant DM peak; and the KDE peak of the luminosity map gives the second tightest offsets from the DM peak.

• there are offset observations in the Illustris simulation that are as extreme as the offset observed in more massive merging galaxy clusters.

• Lensing analysis that stack clusters based on the center of mass will see miscentering due to non-symmetrical and / or clumpy spatial distributions.

furthermore, we have provided a set of python functions for making accurate contour levels for spatial maps and inferring density peaks at https://goo.gl/MNrSQV.

REFERENCES

Allen S. W., 1998, MNRAS, 296, 392
Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
Bennett C. L., et al., 2011, Astrophys. J. Suppl. Ser., 192, 17
Bradač M., et al., 2006, ApJ, 652, 937
Bradač M., Allen S. W., Treu T., Ebeling H., Massey R., Morris R. G., von der Linden A., Applegate D., 2008, ApJ, 687, 959
Clowe D., Markevitch M., Bradač M., Gonzalez A. H., Chung S. M., Massey R., Zaritsky D., 2012, ApJ, 758, 128
Cui W., et al., 2016, MNRAS, 456, 2566
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dawson W. A., 2013, PhD thesis, University of California, Davis
Deng H., Wickham H., 2011, Technical report, Density estimation in R. had.co.nz
Dietrich J. P., Böhnert A., Lombardi M., Hilbert S., Hartlap J., 2012, MNRAS, 419, 3547
Duong T., 2007, J. Stat. Softw., 21, 1
Feigelson E. D., Babu G. J., 2014, Contemp. Phys., 55, 126
Genel S., et al., 2014, MNRAS, 445, 175
George M. R., et al., 2012, ApJ, 757, 2
Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hall P., Marron J. S., Park B. U., 1992, Probab. Theory Relat. Fields, 92, 1
Harvey D., et al., 2014, MNRAS, 441, 404
Hilbert S., White S. D. M., 2010, MNRAS, 404, 486
Hoag A., et al., 2016, arXiv Prepr., p. 48
Jee M. J., Hughes J. P., Menanteau F., Sifón C., Mandelbaum R., Barrientos L. F., Infante L., Ng K. Y., 2014, ApJ, 785, 20
Jee M. J., et al., 2015, ApJ, 802, 46
Johnston D. E., et al., 2007, arXiv Prepr.
Kahlhoefer F., Schmidt-Hoberg K., Frandsen M. T., Sarkar S., 2014, MNRAS, 437, 2865
Markevitch M., Gonzalez a. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S., Tucker W., 2004, ApJ, 606, 819
Medezinski E., et al., 2013, ApJ, 777, 43
Oguri M., Takada M., Okabe N., Smith G. P., 2010, MNRAS, 405, no
R Core Team 2014, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/
Randall S. W., Markevitch M., Clowe D., Gonzalez A. H., Bradač M., Bradac M., 2008, ApJ, 679, 1173
Robertson A., Massey R., Eke V., 2016, arXiv Prepr., p. 20
Robitaille T. P., et al., 2013, A&A, 558, A33
Springel V., 2010, MNRAS, 401, 791
VanderPlas J., Connolly A. J., Ivezic Z., Gray A., 2012, in 2012 Conf. Intell. Data Underst.. IEEE, pp 47–54, doi:10.1109/CIDU.2012.6382200, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6382200

Vogelsberger M., et al., 2014a, MNRAS, 444, 1518
Vogelsberger M., et al., 2014b, Nature, 509, 177
Williams L. L. R., Saha P., 2011, MNRAS, 415, 448
Zitrin A., Bartelmann M., Umetsu K., Oguri M., Broadhurst T., 2012, MNRAS, 426, 2944
Zitrin A., Menanteau F., Hughes J. P., Coe D., Barrientos L. F., Infante L., Mandelbaum R., 2013, ApJ, 770, L15

APPENDIX A: ALGORITHM OF THE SHRINKING APERTURE ESTIMATES

Data: subhalo that satisfy cuts as a galaxy

---

initial aperture centroid = weighted mean galaxy location in each spatial dimension
distance array = euclidean distances between initial aperture center and each galaxy location
aperture radius = 90th percentile of the weighted distance array
while *(newCenterDist - oldCenterDist) / oldCenterDist ≥ 2e-2* **do**
  | new data array = old data array within aperture
  | newCenter = weighted mean value of new data along each spatial dimension
end

Algorithm 1: Shrinking aperture algorithm with luminosity weights

APPENDIX B: STATISTICAL ASPECT OF REPRESENTING THE DISTRIBUTION OF OFFSETS

Another statistical aspect that needs attention for computing the significance level is the interpretation of the projections. We interpret the projections as the possible ways for observing cluster of the Illustris simulation. However, we also need to preserve the mass abundance of clusters, i.e. number of clusters in a certain cluster mass range. When we try to draw conclusions from the full sample of the clusters with all the projections, the mass abundance is consistent with the original simulation. However, when we subset the samples, the mass abundance will deviate.

If we assume the chosen projection of each cluster to have negligible dependence of the projection of another cluster, the ways to choose all the projections proceed as follows: For each of the 43 clusters, there are 768 projections to draw from, giving 768^{43} combinations of possible observations of the Illustris simulation, which is a daunting number. In fact, the number of possible observations is daunting even if we reduce the number of projections per cluster to only 12, giving 12^{43} possible observations. We therefore pick a few special cases to discuss while keeping in mind that the mass abundance should be preserved.
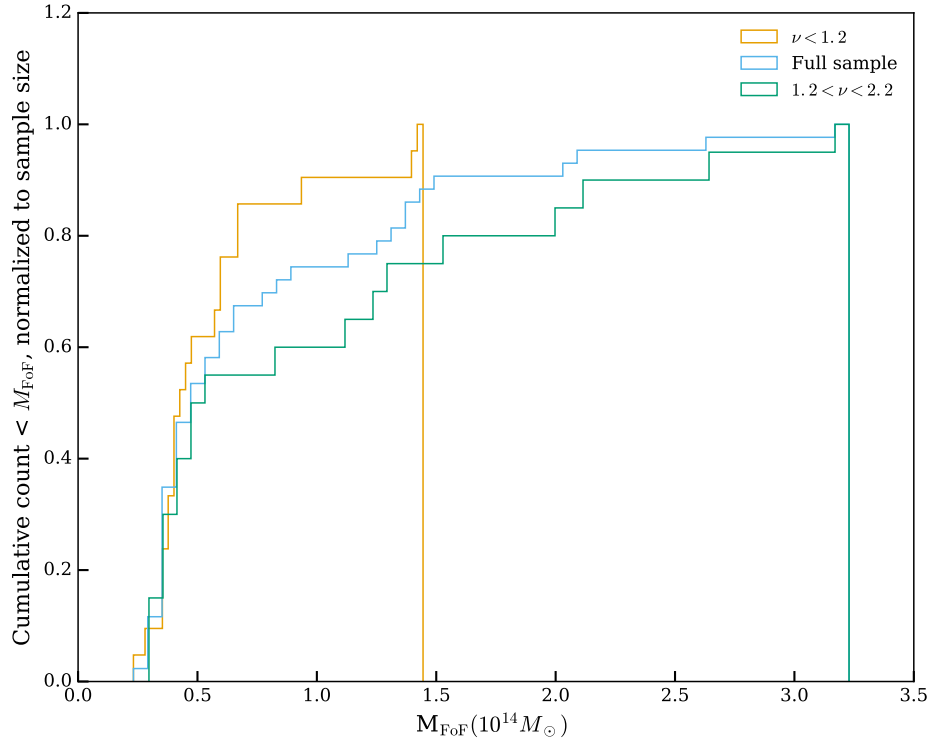
APPENDIX C: TABLE OF RESULTS

Figure B1. Cumulative distribution of clusters above a certain mass threshold for different samples. Each distribution is normalized to the sample size. The unrelaxed samples $1.2 < \nu < 2.2$ If the subsets have the same cluster mass abundance as the full sample, the three plots should lie on top of one another.

Table C1. Properties of the clusters used in the analysis. Richness is computed based on $i-$band < 24.4 assuming $z = 0.3$.

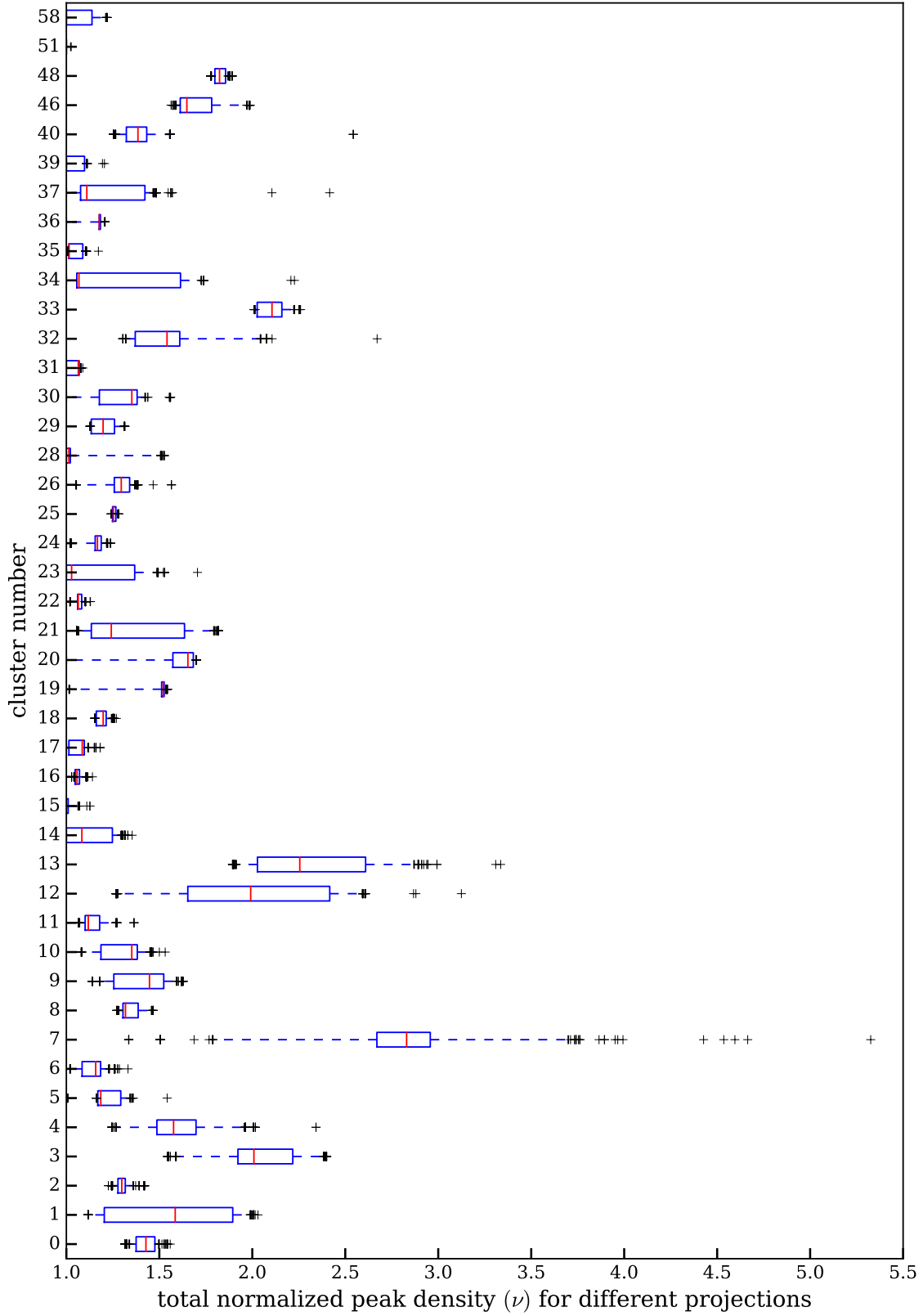| ID	richness	M_{200C} $(10^{14} M_\odot)$	M_{500C} $(10^{14} M_\odot)$	M_{FoF} $(10^{14} M_\odot)$	unrelaxedness$_0$	unrelaxedness$_1$	midvar(Δy_{KDE}) (kpc)	max(Δy_{KDE}) (kpc)	median(ν)
0	483	1.64	1.09	3.23	29	33	31	65	1.43
1	338	1.57	0.62	2.68	20	16	25	71	1.59
2	267	1.53	0.87	2.12	17	3	18	42	1.30
3	343	0.82	0.56	2.03	37	59	44	148	2.01
4	213	1.19	0.66	1.54	21	4	24	84	1.58
5	212	0.90	0.56	1.44	20	27	16	43	1.19
6	225	0.96	0.60	1.40	18	7	15	28	1.16
7	230	0.31	0.17	1.41	54	280	101	379	2.83
8	148	0.83	0.54	1.34	24	26	20	52	1.32
9	187	0.79	0.50	1.29	23	12	33	111	1.45
10	158	0.73	0.53	1.15	19	8	19	49	1.35
11	134	0.57	0.33	0.95	20	9	36	78	1.12
12	164	0.20	0.09	0.87	64	142	77	218	1.99
13	115	0.22	0.14	0.79	63	143	38	118	2.26
14	90	0.45	0.29	0.67	15	8	17	33	1.08
15	92	0.51	0.35	0.68	11	3	11	25	1.00
16	113	0.40	0.23	0.61	19	4	13	31	1.06
17	97	0.42	0.18	0.60	21	8	27	53	1.09
18	83	0.45	0.31	0.59	15	8	14	32	1.20
19	86	0.26	0.19	0.57	30	68	18	77	1.52
20	84	0.15	0.11	0.50	60	122	54	117	1.65
21	89	0.26	0.12	0.53	23	8	47	146	1.24
22	70	0.42	0.30	0.49	14	7	10	23	1.06
23	68	0.25	0.17	0.47	30	25	26	98	1.03
24	66	0.33	0.26	0.44	14	14	11	42	1.17
25	79	0.23	0.15	0.43	23	25	11	22	1.25
26	61	0.26	0.18	0.45	28	40	11	44	1.30
28	69	0.30	0.16	0.41	22	12	26	42	1.01
29	62	0.30	0.20	0.42	16	14	9	22	1.20
30	59	0.18	0.14	0.40	42	78	17	63	1.35
31	57	0.29	0.21	0.40	14	15	10	24	1.06
32	56	0.18	0.13	0.38	35	23	43	83	1.54
33	69	0.19	0.10	0.38	49	54	60	108	2.11
34	63	0.21	0.14	0.39	23	20	22	33	1.07
35	69	0.29	0.22	0.41	12	3	11	28	1.01
36	72	0.24	0.16	0.36	21	22	16	39	1.18
37	63	0.21	0.16	0.36	25	23	51	142	1.11
39	55	0.27	0.18	0.36	11	3	12	29	1.00
40	54	0.18	0.10	0.33	44	69	81	151	1.39
46	52	0.08	0.06	0.30	57	73	59	157	1.65
48	53	0.12	0.08	0.30	40	104	13	44	1.82
51	56	0.19	0.13	0.29	12	5	17	41	1.00
58	58	0.14	0.09	0.23	29	10	21	66	1.00

Figure C1. A box plot showing the distribution of the total normalized peak density for each cluster based on 768 projections. The red line shows the median of the projections, the box encompasses the 25% and 75% percentile of the distribution while the whiskers mark the 5% and the 95% percentile. The other black crosses are data points with extreme values beyond the 5% and 95% percentile.

Table C2. Summary statistic characterizing the offset distributions between the most bound particle and various summary statistics of the member galaxy population.

offset (kpc)	location	lower 68%	lower 95%	lower 99%	upper 68%	upper 95%	upper 99%
$\Delta y_{\rm BCG}$	0	-2	-2	-252	2	528	1107
$\Delta y'_{\rm centroid}$	0	-134	-491	-1176	134	491	1176
$\Delta y'_{\rm KDE}$	0	-19	-82	-1182	19	82	1182
$\Delta y'_{\rm num.dens}$	0	-83	-302	-1114	83	302	1114
$\Delta y'_{\rm shrink}$	0	-50	-288	-1025	50	288	1025

The offsets represented with the prime ′ symbols are estimated using the luminosity weighted galaxy data.

Table C3. Summary statistic characterizing the offset distributions for between the DM peak and the estimated galaxy location. All 43 clusters and all 768 projections are used in this table. The highest density values were used for the computation when there were more than one peak value estimated from the KDE.

kpc	mean	std	min	25%	50%	75%	max
$\|\Delta s_{\mathrm{BCG}}\|$	69	294	0	2	3	7	2335
Δx_{BCG}	-14	226	-2331	-2	-0	1	2327
Δy_{BCG}	23	197	-1980	-2	0	2	2332
$\|\Delta s'_{\mathrm{centroid}}\|$	261	209	2	114	202	317	1103
$\Delta x'_{\mathrm{centroid}}$	-42	224	-1022	-164	-37	66	1101
$\Delta y'_{\mathrm{centroid}}$	0	244	-1102	-111	-0	111	1100
$\|\Delta s'_{\mathrm{shrink}}\|$	118	156	0	21	60	165	1454
$\Delta x'_{\mathrm{shrink}}$	-7	131	-1089	-39	-3	23	969
$\Delta y'_{\mathrm{shrink}}$	0	145	-1091	-32	0	32	1109
$\|\Delta s'_{\mathrm{KDE}}\|$	37	35	0	14	26	49	498
$\Delta x'_{\mathrm{KDE}}$	-2	35	-330	-17	-2	12	386
$\Delta y'_{\mathrm{KDE}}$	-0	37	-439	-15	0	15	440
$\|\Delta s_{\mathrm{num.KDE}}\|$	136	161	1	56	92	147	2126
$\Delta x_{\mathrm{num.KDE}}$	-12	142	-1967	-55	-4	53	993
$\Delta y_{\mathrm{num.KDE}}$	-0	155	-1415	-54	-0	54	1417

The offsets represented with the prime $'$ symbols are estimated using the luminosity weighted galaxy data.

This paper has been typeset from a TEX/ LATEX file prepared by the author.