

# Adaptive friend-of-friends algorithm for identifying gravitationally bound cosmological structures

Prateek Gupta<sup>1,2,\*</sup> and Surajit Paul<sup>3,1,4,†</sup>

<sup>1</sup>*Department of Physics, Savitribai Phule Pune University, Pune-411007, India*

<sup>2</sup>*Indian Institute of Astrophysics, Bengaluru, India*

<sup>3</sup>*Manipal Centre for Natural Sciences, Centre of Excellence, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India*

<sup>4</sup>*Raman Research Institute, Bengaluru, India*



(Received 9 March 2023; accepted 14 August 2023; published 8 November 2023)

The Universe at the present epoch is found to be a network of matter overdense and underdense regions. Usually, the overdense regions are dominated by the dark-matter (DM) filaments where massive gravitationally bound structures such as groups and clusters of galaxies form at the nodes. At the cosmological timescales, the baryonic matter follows the flow of DM only, and together they form the cosmic web. To date, this picture of the Universe is best revealed through cosmological large-volume simulations and large-scale galaxy redshift surveys, in which, the most important step is the appropriate identification of structures. So far, these structures are identified using various group finding codes, mostly based on the friend of friends (FoF) or spherical overdensity (SO) algorithms. Although, the main purpose is to identify gravitationally bound structures, surprisingly, the mass information has hardly been used effectively by these codes. Moreover, while it is an established fact that the bound structures can best be formed at some particular mass overdense regions and practically the large-scale structures are barely spherical in shape, the methods used so far either constrain the overdensity or use the real unstructured geometry only. Even though these are key factors in the accurate determination of structures-mass information that can precisely constrain the cosmological models of the Universe, hardly any attempt has been made as yet to consider these important parameters together while formulating the grouping algorithms. In this paper, we present our proposed algorithm called the measure of increased tie with gravity order which takes care of all the above-mentioned relevant features and ensures the bound structures by means of physical quantities, mainly mass and the total energy information. Unlike the usual FoF method where a statistically chosen single linking length is used for all grouping elements, we introduced a novel concept of physically relevant arm length for each element depending on their individual gravity leading to a distinct linking length for each unique pair of elements. This proposed algorithm is thus fundamentally new such that, not only able to catch the gravitationally bound, real unstructured geometry very well, it does identify it roughly within a predefined physically motivated density threshold. Such a thing could not be simultaneously achieved before by any of the usual FoF or SO-based methods. We also demonstrate the unique ability of the code in the appropriate identification of structures, both from large volume cosmological simulations as well as from galaxy redshift surveys, highlighting the fact that it mitigates a few shortcomings of the basic FoF and SO algorithms and strengthens the foundation of clustering or halo-finding methods in general.

DOI: [10.1103/PhysRevD.108.103509](https://doi.org/10.1103/PhysRevD.108.103509)

## I. INTRODUCTION

The abundance of large-volume cosmological simulations (e.g., Millennium [1], Millennium-II [2], Bolshoi [3], Cosmo-OWLS [4], IllustrisTNG-300 [5], etc.) as well as the accessibility of increasingly large observational surveys (e.g., SDSS [6,7], KiDS-1000 [8], Euclid [9,10],

LSST [11,12], etc.) in recent times have enabled the extensive study of dynamics and evolution of structures at large scales ( $\gtrsim 100$  Mpc) of the Universe. Concurrently, they provide a unique opportunity for quantitative tests for the structure formation theories along with unraveling the fabric of the Universe. The key structural elements that truly reveal the evolutionary scenario of the Universe are the dark-matter (DM) halos in cosmological simulations or the clusters of galaxies in large galaxy surveys. With the appropriate structure-mass information, such as the halo

\*prateekgidolia@gmail.com

†surajit.paul@manipal.edu

mass function [13–15], one can precisely constrain the standard cosmological model parameters [16–18] or even may come up with an alternative evolutionary model for the Universe, e.g., [19,20], etc. The cluster counts as a function of their velocity dispersion and redshift, significantly constraining the dark energy models [21]. Moreover, the extraction of more accurate information about the location, mainly the position, the extent, and the bulk velocity of these structures leads to a fair understanding of their dynamics and the energy budget. Thus an obvious and very fundamental question that springs to mind is how to appositely identify these cardinal cosmic structures.

Historically, efforts have been made in this direction primarily through grouping or halo finding [13,22–25]. In the last few decades, many halo finders have been developed by several authors (see [26–28] and the references therein) to identify the structures and substructures from the copious simulated cosmological datasets. Likewise, various algorithms were developed to extract galaxy groups and clusters from the galaxy catalog or large galaxy surveys [29–34]. Ideally, halo finders search locally overdense, gravitationally bound systems in the matter density field (dark matter or dark + baryonic matter), either generated in the simulated realizations (N-body/hydrodynamic or hybrid-codes) or within any large galaxy-redshift surveys. Although numerous halo finders have been developed so far, nearly all of them have their foundation either in the spherical overdensity (SO) algorithm [13] or in the friends-of-friends (FoF) algorithm [22]; otherwise, they are the derivatives or combinations of these two basic methods.

The classical SO algorithm locates the density peaks in the matter density field; the spherical shells are then stacked up around these peaks until the enclosed average density drops below a certain predefined threshold value. The threshold value is assumed either as some multiples of the mean density of the simulated volume or as the critical density of the Universe at that redshift. Consequently, this enforces a spherical geometry in SO halos, whereas, the FoF halos are unstructured, as the FoF algorithm collects the particles that are close to each other with respect to a linking length, which is usually a fractional multiple of the mean separation of the particles in the searched volume. Certainly, each of these basic algorithms has its advantages and limitations.

While the SO-based algorithms, by enforcing a spherical symmetry, fail to capture the real shape of the halos, undoubtedly they pick the halo centers as the density peaks, fairly well, whereas, though the FoF algorithm is successful in discerning the real geometry of the halos to a great extent, it may not center the halos at the density peak of the systems so accurately. This is largely because it locates centers at the center of mass or simply at the mean position vector of the final structure. Furthermore, the FoF algorithm is fundamentally based on a free parameter called the “linking length” ( $l_f$ ), which can be fine-tuned depending on

the scientific problems to be addressed. All trial values that are assumed by various authors in this scheme are as yet chosen statistically, e.g., the fraction of mean interparticle separation in the volume of interest [ $l_f = b\bar{l}$ , where  $b \in (0, 1)$  and  $\bar{l}$  is the mean separation]. The major limitation of this parameter “ $b$ ” is that the choice mostly has no physical relevance and may connect two nearby gravitationally unbound structures into one, if the chosen linking length is too large. Even, the widely used value,  $b = 0.2$ , which some of the authors have claimed to roughly correspond to the halo overdensity of  $\sim 200$ , the value often assumed for the virialized systems, was found to be not so consistent when tested using simulated data with a different simulation setup. The difference may be as large as 100–200 of the mean density as reported by [35,36]. In general, no absolute relation has so far been established between  $b$  and the overdensity of the halos. This necessitates a physically motivated linking parameter that would catch the real gravitationally bound systems. Ironically, these methods ignore the mass information of individual elements of the systems, while connecting them, overlooking the fact that at large scales, gravitation is the only and strongest entity that effectively holds the elements together.

In this paper, we present our efforts to design an advanced halo-finding method comprised of a few novel features along with addressing aforesaid limitations of previously developed algorithms, invoking physically motivated grouping parameters. The paper is laid out as follows. After introducing the paper in Sec. I, we begin with the development of the proposed algorithm and illustrate the methods to estimate the associated basic physical properties of the identified halos in Sec. II. In the first part of Sec. III, we elaborate on the application of the proposed halo finder to the simulated data sample. In the later part of the section, we present the applicability of the algorithm (with a few essential modifications) to the observed galaxy catalog of the Sloan Digital Sky Survey (SDSS) in the region of the Saraswati supercluster. Finally, we summarize our findings and conclude the paper in Sec. IV.

## II. METHODS

Our primary motivation behind developing a new halo-finding algorithm is to connect it to more physical parameters pertinent to both the simulated as well as observational data. Moreover, the algorithm should have powerful features from the basic FoF and SO methods, and intrinsically it should be able to overcome some of the crucial shortcomings in them. For instance, it should avoid forcing a spherical symmetry in describing the halos; rather it should keep the unstructured geometry as the DM halos are not necessarily following the spherical symmetry during their formation through accretion or violent mergers inside the cosmic web. Furthermore, the code should follow

the physically more relevant overdensity parameter in describing the boundary of the halos, reducing the usual dependence on many nonphysical free parameters.

The proposed halo finder namely the measure of increased tie with gravity order (MITRO) has analogous basic features of the FoF algorithm. In the framework of FoF, the nearest neighbor of, say the  $i^{\text{th}}$  element (here, DM particle or galaxy) is searched within a predefined radius known as the linking length,  $l_f$ . The  $j^{\text{th}}$ 's elements in the examined volume, that are located within the distance  $l_f$  with respect to the  $i^{\text{th}}$  element, are considered as its friends (the direct friends). Likewise, a further search is done for the new friend(s) of these direct friends of the  $i^{\text{th}}$  element with the same linking length around them and are known to be the indirect friend(s) of the  $i^{\text{th}}$  element. The process is repeated until no new friend(s), direct or indirect, are found for the  $i^{\text{th}}$  element in the searched volume. The final configuration of these groups of friends (direct and indirect) will be considered as a DM halo (in the case of cosmological simulations) or a cluster of galaxies (in the case of galaxy surveys). In this setup, the FoF halos primarily depend on the choice of a free parameter, called the linking length. Choosing quite a small (or large) linking length will potentially lead to the formation of smaller (or bigger) size halos. Moreover, the enclosed overdensity of these halos with such an arbitrarily chosen linking length provides no absolute or even any general empirical relationship among them (see [35,37–39] and references therein).

### A. MITRO algorithm

The proposed algorithm MITRO introduces a new concept of adaptive and unique linking length between the pairs of particles in the search field. To visualize the basic concept of this novel algorithm, let us take a practical example. Imagine a flock ( $\mathbb{F}$ ) of people sitting randomly on the ground, and we need to search the groups of people in  $\mathbb{F}$  who can hold each other's hands. The simplest way to do this is to apply the basic FoF method in which, first a connecting length is set, which in FoF is usually a weighted fraction of the mean separation between the two people in the flock  $\mathbb{F}$ , and then the aforementioned steps of the FoF algorithm are applied. As we have already mentioned, the resulting FoF groups are very much sensitive to the chosen linking length and therefore need a careful determination of the parameter to produce meaningful results. For example, in this case, if we choose twice the average arm length as the connecting length for everyone in the flock, many people with shorter arm lengths will be counted in the group who in actual cases will hardly be able to hold themselves together to form a connecting group. In fact, arm length is an individual's physical property that depends on various parameters such as age, height, etc.; therefore, the assumption of a fixed linking length would make the group nonphysical.

The issue described above is not just a specific problem. It has been noticed several times that to achieve a meaningful grouping in galaxy survey data using FoF, the adopted linking length may be very large [31,40]. Since a single FoF linking length is used for the entire data, irrespective of the mass of individual galaxies in the sample, it is highly likely that many low-mass galaxies in the sample may not have gravitational influence to such a distance to make a bound system. In such a case, the groups that are chosen by the FoF algorithm hardly represent a gravitationally bound system in a physical sense.

To address this issue, one needs to rectify the friending method which is so far based on a fixed linking length. For simplicity, let us first assume that each person in the said flock has as many arms as s/he demands to connect with the neighbors. To get connected, they should stretch out their arms and try to hold the hands of their neighbors without changing their position on the ground. Here, two people are called their direct friends if they are able to hold each other's hands. And the people are called friends of friends (or indirect friends) if they are connected through intermediate friends. In this way, the resulting interconnected configuration of friendship circles is said to be the physically connected groups of people.

In an analogous situation as stated above, the MITRO algorithm addresses the same problem in a different and more efficient way. Let us assume that in a given cosmological dataset, there are two particles (either the dark-matter particle or the galaxy),  $A$  and  $B$ . Naturally, they may have different masses, say  $M_A$  and  $M_B$ , and therefore in general different arm lengths of  $L_A$  and  $L_B$ , respectively, depending on the effective field of attraction of these gravitating bodies as elaborated in Sec. II A 1. Accordingly, the linking length between them will be  $L_{AB} \equiv L_A + L_B$ . Now, if there is a third particle, say  $C$ , of mass  $M_C$  and arm length  $L_C$ , the possible linking length between  $A$  and  $C$  will be  $L_{AC} (\equiv L_A + L_C)$  and the same between  $B$  and  $C$  will be  $L_{BC} (\equiv L_B + L_C)$ . So, the significant characteristic of the MITRO algorithm is the introduction of the concept of arm length and the existence of as many linking lengths as the number of unique combinations of pairs of elements in the volume of interest. Overall, one may consider MITRO as a cousin of FoF with the major difference that the former adopts the concept of distinct linking length for each unique pair of particles instead of a common linking length for all the pairs, as is the basis of the latter.

Now, while searching for friends, any two elements are said to be connected or are direct friends, if and only if the Euclidean distance ( $D_{ij}$ ) between them is less than or equal to the sum of their individual arm lengths, i.e.,  $D_{ij} \leq L_i + L_j$ . In this way, one computes the direct friend(s) of  $i^{\text{th}}$  element. Subsequently, the friend(s) and further friend(s) of their friend(s) of  $i^{\text{th}}$  particle are altogether collected in a set. Finally, each disjoint set of

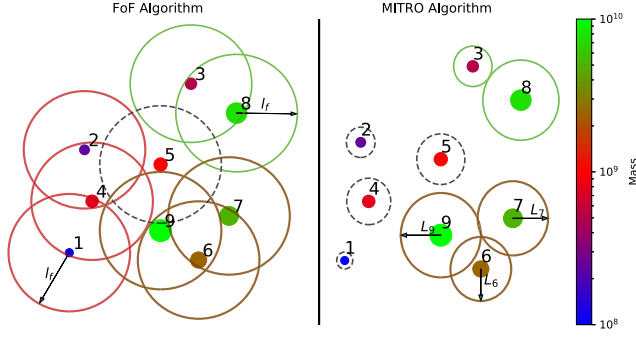


FIG. 1. The colored dots in the figure represent the particles in two-dimensional Euclidean space, where the color code and the size of the colored dots are indicative of the mass of the respective particle. Left panel: demonstrates the clustering with FoF algorithm. Big circles around the colored dots represent the linking length ( $l_f$ ) of each particle. Right panel: demonstrates the clustering with MITRO algorithm. Here, big circles around the colored dots represent the arm length ( $L_i$ ) of the individual particle based on its mass property. In both panels, the dashed circles represent the isolated particles while the solid circles are representing those who found their friend(s) or companion. The different colors for the solid circles are to differentiate the identified groups. FoF groups:  $\{P1, P2, P4\}$ ,  $\{P6, P7, P9\}$ , and  $\{P3, P8\}$ . MITRO-identified groups:  $\{P6, P7, P9\}$  and  $\{P3, P8\}$ .

friends will represent a group in the search volume. Moreover, this method also keeps in mind that the friendship between two individuals is not just the efforts of one; instead, the effort should be from both sides. A pictorial illustration of the above-discussed method can be seen in the right panel of Fig. 1.

In Fig. 1, we pictorially mimicked a specific clustering scenario with FoF (left panel) and with our proposed algorithm (right panel). Nine elements with different masses are randomly placed in a two-dimensional Euclidean space. For better representation, the particles of different masses are indicated by the filled dots of different sizes with dots getting larger for the higher masses. Colors are used for making them distinctly visible. In the left panel, we applied the FoF algorithm with a fixed linking length ( $l_f$ ), computed as the mean separation between the particles in the represented area. As per the usual FoF algorithm, as stated earlier, we obtained three different groups:  $G_1^F: \{P1, P2, P4\}$ ,  $G_2^F: \{P3, P8\}$ , and  $G_3^F: \{P6, P7, P9\}$ , and an isolated particle, P5, surrounded by the dashed-line circle. However, for the same set of particles, when we applied our proposed MITRO algorithm, we obtained only two different groups:  $G_1^M: \{P3, P8\}$ , and  $G_2^M: \{P6, P7, P9\}$  and the rest as the isolated particles, as demonstrated in the right panel of Fig. 1. Here, the arm lengths ( $L_i$ ) have been computed by scaling with the mass of individual particles, as proposed in the MITRO algorithm. Naturally, elements differing in mass have different arm lengths. Since, by definition of MITRO, an examined pair becomes friends only when the arm of one

overlaps or at least touches the other, thus accordingly the groups are formed. The results obtained in MITRO seem physically more relevant. Though FoF flags particles  $\{P1, P2, P4\}$  as groups, namely  $G_1^F$ , these are very small masses compared with the other elements in the set and in the actual case may not have enough potential to make a bound system, and therefore our proposed algorithm has truly flagged them out as isolated particles. We will further elaborate on this issue in various sections throughout the paper.

### 1. The concept of arm length

Since the major goal of running halo finder codes in large-scale matter distribution is to find the gravitationally bound systems, the first thing to ensure is that the elements of a group are under the gravitational field of influence. The basic theory of gravity says that particles with different masses will have different fields of influence. Therefore, in a system of particles with different masses, the heavier mass particles have a greater impact on the trajectories of other nearby smaller mass particles. This seeds the idea that in a gravitating system, the most effective physical parameter, which is definitely the mass of the elements (DM particles or galaxies), should determine the distance of influence. We call this as the effective radius or the arm length of the mass element. This in turn will decide the final possible linking length between any unique pair of particles in our proposed algorithm. Linking length is thus no more a free parameter, unlike in FoF, but rather an intrinsic physical property of the individual pair of particles in the system.

As a preliminary test step, we considered the gravitational potential as the measure of influence and determined the arm length. For any arbitrary mass  $M$ , the associated gravitational potential ( $V$ ) at any distance  $r$  is given by  $V = -\frac{GM}{r}$ , where  $G$  is the universal gravitational constant. The maximum arm length can therefore be taken as the distance from the particle where its influence almost vanishes, which is, by definition, at an infinite distance. However, a finite arm length is essential to get a meaningful clustering. For simplicity of calculation, one can attempt to get a distance at which the increment in the gravitational potential, i.e.,  $\frac{\partial V}{\partial r} \equiv \frac{GM}{r^2}$ , becomes considerably low. Assuming the radial rate of change of gravitational potential at a distance  $R$  (as  $F_R$ ) to be such a limit, we get

$$F_R = \left| \frac{\partial V}{\partial r} \right|_{r=R} \equiv \frac{GM}{R^2}. \quad (1)$$

This leads to an arm length of

$$R = \sqrt{\frac{GM}{F_R}} \approx 3.73 \times 10^{-9} \text{ kpc} \left( \frac{M}{M_\odot} \right)^{1/2} F_R^{-1/2}. \quad (2)$$



For the assumed cosmological setup and simulated dataset used in this study, we found that  $F_R \leq 10^{-7}$  produces a meaningful clustering.<sup>1</sup> However, the number (i.e.,  $\leq 10^{-7}$ ) obtained here is through trial and error and needs some robust method to compute such a limit, *which is out of the scope of this study*. Moreover, it should also be physically motivated. This led us to define the arm length in a generic way by relating it to the basic cosmological definition for the spherically collapsed objects.

We define the arm length as the radius,  $R_\Delta$ , of a sphere around the particle of mass  $M$  within which we uniformly distribute the mass such that the mean density of the sphere becomes  $\Delta$  times the critical density ( $\rho_{cz}$ ):

$$\frac{M}{\frac{4}{3}\pi R_\Delta^3} = \Delta \times \rho_{cz}. \quad (3)$$

Here, the critical density,  $\rho_{cz}$  corresponds to the redshift of the particle and the assumed cosmology. The arm length is sensitive to the choice of  $\Delta$ ; however, its value can be limited to any physically motivated situation. For instance, if one needs to search for the structures at large scales ( $\gtrsim$  a few 10's of Mpc),  $\Delta \in [2, 2000]$ . The arm length for each particle can be determined as

$$R_\Delta = \left[ \frac{M}{\frac{4}{3}\pi\Delta\rho_{cz}} \right]^{1/3}. \quad (4)$$

An important advantage of overdensity mediated arm length is that it also enables us to fairly constrain the enclosed overdensity ratio ( $\bar{\rho}/\rho_c$ ) of the identified halos using the MITRO halo finder (discussed later in Sec. II B 4).

## 2. Unbinding process

Locating halos using a spatial clustering algorithm usually ensures the positional proximity of the halo elements. However, such clustering does not guarantee that all the elements in the halos are energetically bound to the system. Rather, the position-space-based halo finder will always include a few high-velocity particles which could be dynamically unrelated to the system [28]. Identification and removal of such unbound particles in a halo are known as the unbinding procedure. Implementation of the unbinding procedure allows us to make the halo catalogs free from significant contamination of spurious small objects and therefore an essential practice for filtering subhalos out of the main halo as well as computing more accurate halo centers and its bulk

velocities [41]. In our approach to this process, a particle is said to be unbound to the system (i.e., a halo identified by MITRO), if its rest-frame velocity is greater than or equal to the escape velocity computed at the particle position. So, in the formed cluster  $\mathbb{C}$ , with particles,  $P_j, j \in \mathbb{C}$ , the particle  $P_j$  is flagged as an unbound particle if the kinetic energy of  $P_j$  is greater than or equal to the total gravitational potential of the system at the particle position due to the rest of the particles in  $\mathbb{C}$ . Such particles will then be deflagged from the member list of  $\mathbb{C}$ . Since in the large-scale structures of the Universe, material flows along the filaments connecting the galaxy clusters and the whole network are in general in motion, to compute the kinetic energy of the particles, we have first corrected the velocity components ( $v_i$ 's) by subtracting the bulk flow of the cluster. The kinetic energy of the “ $j^{\text{th}}$ ” particle in the rest frame of cluster ( $\mathbb{C}$ ) is therefore given by

$$KE_j = \frac{1}{2}M_j \sum_{i \in \{x,y,z\}} (v_{ij} - V_i^{\text{bulk}})^2 \quad \forall j \in \mathbb{C}, \quad (5)$$

where  $M_j$  is the mass of the  $j^{\text{th}}$  particle in  $\mathbb{C}$ , and  $V_i^{\text{bulk}}$  denotes the  $i^{\text{th}}$  velocity component of the bulk flow of the cluster  $\mathbb{C}$  which is evaluated as

$$V_i^{\text{bulk}} = \sum_{j \in \mathbb{C}} v_{ij} M_j \quad \forall i \in \{x, y, z\}. \quad (6)$$

Although for a faster computation of gravitational potential one may use the tree code, the particle number being not so large ( $\leq 10^6$ ) in our specific problem, as well as, for a more precise evaluation of the same, we compute it as

$$PE_j = \sum_{k \neq j} G \frac{M_j M_k}{|\vec{r}_{jk}|} \quad \forall k \in \mathbb{C}. \quad (7)$$

Now, for each particle  $P_j$  in cluster  $\mathbb{C}$ , wherever  $KE_j \geq PE_j$ , it will be untagged from the membership of  $\mathbb{C}$ . We iterate the process of unbinding till all such unbound particles are removed from the system  $\mathbb{C}$ .

Nevertheless, switching on the unbinding process is not an absolutely necessary step for a halo finder. It mostly depends on the scientific problem that one deals with. The halo properties that are based on gravitating matters only, such as gravitational lensing, x-ray properties, Sunyaev-Zeldovich, etc., actually require all the particles that are positioned in the configuration, even if they are not bound to the system [41]. Moreover, if one is interested in studying the diffuse streaming or turbulence properties inside the group medium, the running unbinding process may even turn out to be detrimental as such calculations rather than have an absolute requirement for those high-velocity particles that are present in the system.

<sup>1</sup> $F_R$  in Eq. (2) does not express any direct physical quantity; rather it signifies the slope of the gravitational potential at some arbitrary distance  $R$  due to the particle of mass  $M$ . In other words,  $F_R$  may be realized as force (dynes, in cgs. units) applied on a unit mass of test particle placed at distance  $R$  from mass  $M$ .

Consequently, we introduced the unbinding process as an optional step in our code.

## B. Computed halo properties in MITRO-algorithm

### 1. Halo position

The successful determination of precise representative halo centers is an important task of a halo finder. This is mainly because of the fact that almost all basic halo properties such as radius, mass, bulk velocity, etc., are primarily computed on the basis of the chosen halo centers. In our proposed algorithm, we do it by implementing two different concepts: (i) the commonly used mass-weighted average position of all member particles of a halo, i.e.,  $r_{\text{com}}$  (the center of mass, c.m. position, hereafter), and (ii) the position of the halo member at the lowest gravitational potential, i.e.,  $r_{\text{lgp}}$  (the lowest gravity point, LGP position, hereafter), as the halo center. The latter one will be our main method to compute the halo positions, since this method was found to provide the halo centers closer to the actual matter density peak in the halo, in comparison with the former one as determined by the center offset study (see more detail in Sec. III A 3).

### 2. Halo overdensity ratio

In large-scale structures, the overdensity ratio is a key parameter that helps in defining roughly the regions that are successfully able to withstand the Hubble expansion and may eventually be collapsed to bound objects when their overdensity reaches a certain critical value [13]. In our proposed algorithm, we compute the overdensity ratio of a halo as the ratio of the matter density in the volume enclosed by the halo to the critical density of the Universe at the halo redshift. In our method, no spherical geometry for the halos has been enforced; rather it keeps the intrinsic shape of the surface that encloses all the halo-fellow particles.

We construct the envelope which encloses halo members using the convex hull algorithm. For any given set of points  $\mathcal{X}$  in 3D-Euclidean space, the convex hull algorithm selects the smallest number of convex set of points ( $\mathcal{Q}_{\text{hull}} \subset \mathcal{X}$ ) which encloses the given set  $\mathcal{X}$  in a closed convex surface, the minimum possible volume enclosing  $\mathcal{X}$ . For our case, we make use of the SciPy quick-hull algorithm [42] to construct the convex hull and subsequently compute the corresponding volume,  $\mathcal{V}$ , within the convex surface planes. However, in the MITRO algorithm, the set of points  $\mathcal{X}$  is not simply the set of positions of the halo members as used in the usual convex hull setup. In the proposed algorithm, we assign arm length to the individual particles, and therefore they assume a spherical shape of radius equal to their respective arm lengths (Sec. II A 1). Since these spheres will have non-negligible sizes, to meticulously compute the volume  $\mathcal{V}$ , one needs to apply the convex hull for a set of spheres instead of a set of points.

Furthermore, these spheres being of different sizes may even overlap with each other, demanding a very complex and modified convex hull method. To avoid such complexity, as it is out of the scope of this work, we have rather used a heuristic approach. We made a new set of points  $\mathcal{X}'$  taking the point on the spherical surface of each halo member that is farthest from the center (LGP center) of the host halo. In this way, instead of applying convex hull on the set of halo-member positions, we apply it on this new set of farthest points, i.e.,  $\mathcal{X}'$  and compute the volume  $\mathcal{V}$  of the halo.

With this, the mean density ( $\bar{\rho}$ ) of the halos is then computed as the ratio of the total enclosed mass ( $\Sigma M_i$ ) to its volume,  $\mathcal{V}$ . The overdensity ratio ( $\wp_{\text{halo}}$ ) is therefore defined as

$$\wp_{\text{halo}} = \frac{\bar{\rho}}{\rho_c} \quad \text{where} \quad \bar{\rho} = \frac{1}{\mathcal{V}} \sum_{i \in \mathcal{X}} M_i. \quad (8)$$

### 3. Halo radius and mass

The halo radius in our code is calculated in a very straightforward way. We report the distance between the LGP center and the farthest member of the halo as the radius or maximum radius,  $R_{\text{max}}$ , more precisely, the farthest point in the convex hull computed enclosing surface with respect to the LGP center. However, in highly mass-resolved simulations, the mass of DM particles being small, the arm length is negligible compared with  $R_{\text{max}}$ , and therefore searching the positions of the farthest members will be enough in such cases. Contrary to this, the presence of massive galaxies as the cluster member in galaxy surveys makes the arm length non-negligible, and thus in such cases, the farthest convex hull surface would be considered for measuring  $R_{\text{max}}$ . We further compute the mass of MITRO-identified halos by taking the sum of masses of each individual halo-member element and denoting them as  $M_{\text{tot}}$ .

### 4. Compactness parameter

Introduction of the concept of arm length allowed us to understand the stiffness of the configured halos by defining a new parameter called ‘‘compactness,’’ the ratio of the measured halo overdensity using the convex hull method,  $\wp_{\text{halo}}$ , (see Sec. II B 2) to the reference overdensity ( $\wp_{\text{ref}} = \Delta_{\text{ref}}/2$ ), i.e.,  $C_p = \wp_{\text{halo}}/\wp_{\text{ref}}$ . The two factors in the denominator of  $\Delta_{\text{ref}}/2$  will be discussed in the further text. This ‘‘ $C_p$ ’’ parameter figures out how closely or densely the member elements are packed in a halo compared to the search field element density. While defining the arm length, given in Eq. (4), the size of the sphere around the particle position has been computed assuming that the mass of the particle is uniformly distributed in the sphere of radius  $R_{\Delta}$ , such that the enclosed density is  $\Delta$  times the critical density ( $\rho_c$ ). For simplicity, let us assume a gravitationally bound system

with  $N$  number of elements of equal masses. Equal mass implies all spheres will have the same radii. If we consider them as solid spheres, while forming a gravitationally bound structure, they may simply touch each other if placed closely, similar to the atoms in a closely packed lattice structure. Much like the lattice, we may therefore define a term packing fraction of the halo

$$P_f = \frac{N \times \text{Volume of each sphere}}{\text{Total volume enclosed by the structure}}. \quad (9)$$

With a simplistic analogy of a cubic lattice filled with equal-sized spherical atoms, one gets a packing efficiency of  $\pi/6 \approx 0.5$ . Even if we consider spherical elements of arbitrary sizes or if we allow overlapping of spheres, the packing efficiency would only increase. This sets the lower limit to the enclosed overdensity of any MITRO-identified halo to roughly  $\Delta/2$ . Therefore, if the DM halos are searched with  $\Delta \equiv \Delta_{\text{ref}}$ , the enclosed overdensity of the identified halos must be  $\geq \Delta_{\text{ref}}/2$ . We may therefore take  $\Delta_{\text{ref}}/2$  as the density threshold in which the halos are searched, and accordingly, we should set the  $\Delta_{\text{ref}}$  as the specific problem may demand.

### III. APPLYING MITRO ALGORITHM ON DIFFERENT COSMOLOGICAL DATASETS

The proposed MITRO algorithm is so designed that in its basic framework, it universally works on the datasets from N-body/particle simulations as well as galaxy-redshift surveys. In further sections, we tested our algorithm on cosmological simulation data by analyzing the halo properties defined in Sec. II B. Subsequently, we implemented our algorithm on observational survey data to identify large-scale structures in the Universe, specifically the data from a previously known supercluster region.

#### A. Analysis of simulation data using the MITRO algorithm

The MITRO algorithm works on any cosmological N-body simulation that has recorded at least a few basic parameters. The minimum required parameters are the particle's (1) ID, (2) mass, (3) position (in Cartesian coordinates, in its current framework), and (4) Cartesian velocity components. The particle's ID helps in identifying individual particles. Primarily, the particle's position and mass play a significant role in the halo finding process, as discussed in Sec. II A. Whereas, particle velocity is the key information that is used to verify particle's association with the identified halo (see Sec. II A 2) by examining its gravitational boundedness to the system.

The simulated cosmological datasets used in this work are taken from Paul *et al.* [43]. Simulations were performed using the ENZO 2.1 code [44]. ENZO is a hybrid (N-body + Hydrodynamical), Eulerian adaptive mesh refinement (AMR) code dealing with the dark matter

as particle only and baryonic matter as the fluid that realizes the evolution of hydrodynamical parameters, i.e.  $\rho, P, T$ , in each grid or cell of the simulated volume.

For the cosmological model, to create the simulated realizations, a flat  $\Lambda - \text{CDM}$  cosmology was assumed with  $(\Omega_\Lambda, \Omega_m, \Omega_b) = (0.7257, 0.2743, 0.0458)$  and  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.702$ ; a spectral index for the primordial spectrum of initial matter fluctuations,  $n_s = 1.0$ ; and the rms amplitude of linear fluctuations in the spheres of radius  $8 \text{ Mpc} h^{-1}$ ,  $\sigma_8 = 0.812$  [45].

Regarding the simulation setup, in these simulations, a part of the Universe has been mimicked within the co-moving cosmological cubic volume,  $V$ , of sides,  $L_{\text{box}} = 128 \text{ Mpc} h^{-1}$ , filled with  $64^3$  number of dark-matter particles and  $64^3$  cells in the parent grid. Furthermore, two static/nested child grids ( $64^3$  cells) of volume  $V/2^3$  and  $V/2^6$  respectively, were inserted such that the center of both the child grids coincided with the center of the parent grid. With the insertion of two child grids, the dark-matter particle mass resolution of  $7.9 \times 10^9 M_\odot h^{-1}$  has been achieved in the central most region of the comoving volume  $(32 \text{ Mpc} h^{-1})^3$ . Another four levels of AMR on the matter density parameter have been initialized in the second child grid, of volume  $V/2^6$ , to obtain a peak (comoving) grid resolution of  $31.25 \text{ kpc} h^{-1}$ . This allowed one to produce enough resolution in mass and space to study both the DM and baryonic properties of the large-scale structures of the Universe, specifically the gravitationally bound galaxy clusters, in this work. For the hydrodynamic setup, an ideal gas equation of state, with  $\gamma = 5/3$ , has been used for the baryons. Radiative cooling and star formation feedback physics were also taken into account during the simulations (for more details, see [43]).

Considering the simulated data sample, with the above-said model, two realizations of parts of our Universe were produced, each of comoving volume  $(128 \text{ Mpc} h^{-1})^3$ . Each simulation has been initialized at redshift,  $z = 60$  using the transfer function of [46] and evolved till the present epoch, i.e.  $z = 0$ , with 48 snapshots taken at different redshift intervals. In this work, we have used only 11 snapshots taken at equal intervals of  $\Delta z = 0.02$ , in between redshift 0.2 to 0. To proceed further with the proposed algorithm, for each snapshot, the particles and the above-mentioned associated properties of the particles are recorded in a separate file. During the analysis, we worked only within the highly resolved central  $(32 \text{ Mpc} h^{-1})^3$  comoving volume of two realizations.

#### 1. MITRO algorithm for simulated data

While finding DM halos from DM particles of a cosmological simulation, such as ours, using the MITRO code, it follows the below-mentioned basic algorithm:



Step 1: Define the arm length,  $L_i$ , to each particle following Eq. (4) and appropriate parameters, i.e.,  $\Delta$ , in general  $\Delta = 200$ .

Step 2: After defining the arm length, it searches for friends of the particles in the volume of interest. Two particles,  $i$  and  $j$  are flagged as friends if the Euclidean separation between them ( $D_{ij}$ ) is smaller than the sum of their arm lengths, i.e.,  $D_{ij} \leq L_i + L_j$ . Likewise, the search for the friends of friends is done to achieve the final configuration of the halos.

Step 3: The unbinding process (described in Sec. II A 2) is then applied to each MITRO-identified halo when required. It is an optional step which can be switched on or off depending on the science problem that we address.

Step 4: Finally, it computes the halo properties defined in Sec. II B.

We put an additional restriction on particle numbers to flag an identified halo as the real halo (in both cases, i.e., before and after running the unbinding algorithm). We list only those halos having a minimum number of particles,  $N_{\min} = 100$ , decided based upon the highest particle mass resolution of  $\sim 10^{10} M_{\odot}$  in our simulation so that we obtain the halos with mass,  $M \geq 10^{12} M_{\odot}$ .

## 2. Overdensity ratio of MITRO halos and analysis of the unbinding process

As we see in Sec. II B 4, our proposed algorithm is capable of constraining the overdensities to a great extent, especially for the halos after running the unbinding algorithm. From the discussion in Sec. II A 1, it appears that to achieve a theoretical lower bound to the overdensity for the searched halos,  $\varphi_{\text{halo}}$ , one needs to keep in mind that an arm length,  $L_{\text{arm}}$  corresponds to an assumed overdensity of  $\Delta_{\text{arm}}$ , which would result in an approximate overdensity lower bound of halos,  $\varphi_{\text{halo}} \geq \Delta_{\text{arm}}/2$ . To statistically check the veracity of the above-stated feature of the MITRO algorithm, we take a practical example with  $\Delta_{\text{arm}} = 200$ , applied to the above-discussed simulated dataset. Here, we should mention that for further analysis in this paper, we created a DM halo list following the algorithm illustrated in Sec. III A 1 and taking all the identified halos with mass  $\geq 10^{12} M_{\odot}$  (MITRO-200 hereafter).

Figure 2 shows the distribution of overdensity of the MITRO halos with the arm length corresponding to  $\Delta = 200$  ( $\varphi_{\text{halo}} - \text{dist}$ , hereafter). Figure 2(a) depicts MITRO-identified halos with  $M_{\text{halo}} \geq 10^{12} M_{\odot}$ , whereas, the lower panel [Fig. 2(b)] represents the halos with  $M_{\text{halo}} \geq 10^{13} M_{\odot}$ . In both the panels, dark-cyan color has been assigned to halos identified before running the unbinding process, and orange depicts the halos after the removal of unbound particles from the initially identified halos. The  $\varphi_{\text{halo}} - \text{dist}$  in Figs. 2(a) and 2(b) shows that the halos after going through the unbinding process strictly

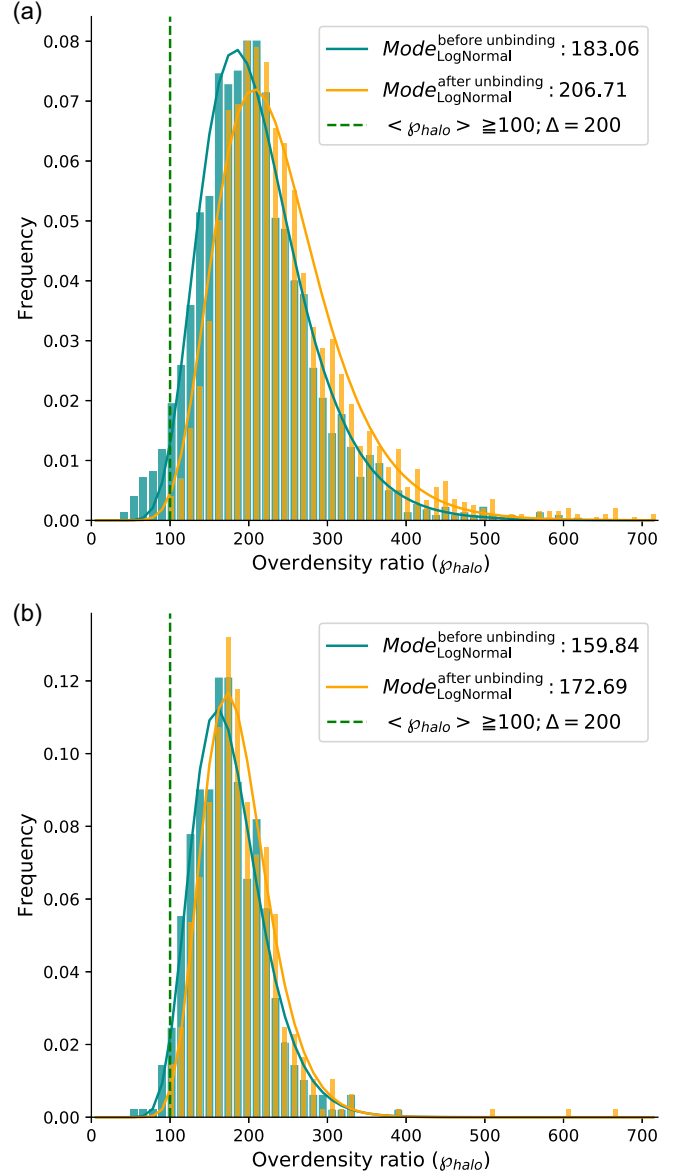


FIG. 2. Overdensity ratio of identified halos for a search overdensity of  $\Delta_{\text{arm}} = 200$ . (a) Histogram plot of  $\varphi_{\text{halo}}$  for all MITRO clusters having Mass  $\geq 10^{12} M_{\odot}$  (b) The same for all MITRO clusters having Mass  $\geq 10^{13} M_{\odot}$ .

follow the expected lower bound, i.e.,  $\varphi_{\text{halo}} \geq 100$ , with 99.6% of halos falling above the overdensity lower limit, whereas, before the removal of unbound particles, slightly more than 96% of the identified halos fall above this determined lower bound of overdensity. We further notice that the maximum probable overdensity ratio in  $\varphi_{\text{halo}} - \text{dist}$  is roughly greater than the average of the lower bound and the actual assumed overdensity threshold used for defining the arm length, i.e.,  $\gtrsim 150$ .

While, the  $\varphi_{\text{halo}} - \text{dist}$  for  $M_{\text{halo}} \geq 10^{12} M_{\odot}$  [Fig. 2(a)], shows a significant peak shift as well as a reduction in peak value between the distribution before and after the removal of unbound particles, the same is not so prominent in the



distribution with  $M_{\text{halo}} \geq 10^{13} M_{\odot}$ , as shown in Fig. 2(b).<sup>2</sup> This actually suggests that many low-mass objects,  $10^{12} M_{\odot} \leq M_{\text{halo}} \leq 10^{13} M_{\odot}$ , are largely affected by the unbinding process,<sup>3</sup> indicating that they are very unstable and are possibly nonvirialized, as also suggested by [43]. This may also indicate an environment with shallower gravitational potential, such as filaments, easily allowing them to disturb their spherical geometry by any minor dynamical events. Moreover, the large difference in maxima values of  $\varphi_{\text{halo}} - \text{dist}$  seen in Fig. 2(a) for low- and high-mass halos makes it apparent that a significant fraction of the stable and gravitationally bound low-mass systems are compact in nature as further discussed in Sec. III A 4.

### 3. Position offsets in MITRO halos

For studying halo position offsets, the local peak matter density of the configured halos,  $\mathbf{r}_{SO}$ , have been taken as the reference position for the halos. It has been noticed that, under the smooth radial-density profile fits, the halo finders that work with the center at the highest local density are more accurate than the one that computes the centers by taking the position average (simple or mass weighted) of all the halo-fellow particles, such as in FoF or any position-space-based halo finders (see Knebe *et al.* [26], Gao and White [47]). Here, we demonstrate the accuracy of the MITRO-halo centers by defining the position offsets as

$$\delta r_{\text{com}} = |\mathbf{r}_{\text{com}} - \mathbf{r}_{SO}|; \quad \delta r_{\text{main}} = |\mathbf{r}_{\text{main}} - \mathbf{r}_{SO}|. \quad (10)$$

To search the local-density peak in the configured halos, we first create a particle density field with a uniform-grid structure in the halo-defined region, having the grid positions the same as given by ENZO snapshot outputs. The particle mass density field is produced using *Triangular Shaped Cloud* interpolation method ([48]), in which the mass of the particle will be distributed in 27 cells based on the spatial location of the particle in the grid cells. Finally, the maximum density location is searched within the halo-defined region and is referred to as  $\mathbf{r}_{SO}$ . Although we implemented AMR for our simulations, a uniform grid cell structure is created at the highest resolution (i.e., 31.25 kpc/h) of our simulation during the reconstruction of the matter density field. However, it should be noted that such a crude process may affect the density field if any

<sup>2</sup>A slight increment in the peak of distribution after the unbinding process is due to the increment in the compactness of low-mass objects ( $10^{13} M_{\odot} \leq M_{\text{halo}} \leq 5 \times 10^{13} M_{\odot}$ ). The unbinding was found to bring the overdensity ratio close to the searched overdensity threshold ( $\Delta_{\text{arm}} = 200$ ).

<sup>3</sup>A few halos in the mass range  $10^{12} M_{\odot} \leq M_{\text{halo}} \leq 5 \times 10^{12} M_{\odot}$  are removed as they fail to qualify the minimum particle condition for a halo, i.e.,  $N_{\text{min}} = 100$  after losing particles due to the unbinding process.

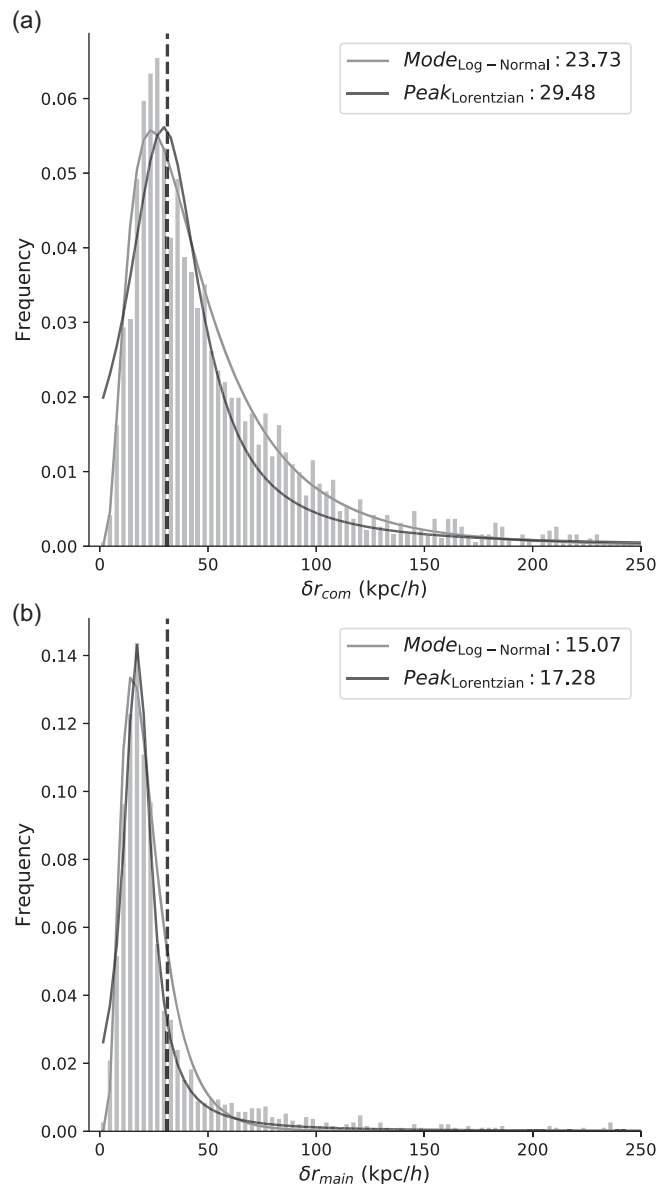


FIG. 3. Halo position offsets between the mass-weighted average of positions of all halo particles,  $\mathbf{r}_{\text{com}}$ , and the highest local density in the configured halo,  $\mathbf{r}_{SO}$ , in (a), and between the highest gravitational potential energy particle position,  $\mathbf{r}_{\text{main}}$  and  $\mathbf{r}_{SO}$  [panel (b)]. The dashed black line represents the peak grid resolution of the simulations.

relatively high-mass particle(s) resides at the periphery ( $\gtrsim r_{200}$ ) of the configured halo.

The distribution of center offsets computed following Eq. (10) is plotted in Figs. 3(a) and 3(b). It is evident from the plots that MITRO computed centers (both  $r_{\text{com}}$  and  $r_{\text{lgp}}$ ) are precisely closed to the highest local density position, as the peak of center-offset distributions is found to be well within the size of the highest resolution cell (i.e., 31.25 kpc/h) in our simulations. The halo positions ( $r_{\text{com}}$ ) calculated based on the mass-weighted position average have a broader offset distribution than the one calculated

based on the lowest gravity point ( $r_{\text{lgp}}$ ). The peak of the distribution of the former ( $\delta r_{\text{com}}$ ) is also found considerably away compared to the latter ( $\delta r_{\text{lgp}}$ ). Furthermore, while 73% of the position offset with respect to LGP centers (i.e.,  $\delta r_{\text{lgp}}$ ) are found to be within the highest grid resolution of the simulation, i.e., 31.25 kpc/h, only 37% of position offsets for c.m. centers (i.e.,  $\delta r_{\text{com}}$ ) come within this resolution limit. All these show  $r_{\text{lgp}}$  centers are more accurate and therefore used as the preferred definition for halo position in this work.

#### 4. Compactness of MITRO-halos

“Compactness” of a halo is defined in Sec. III B 4 as  $C_p = \varrho_{\text{halo}}/\varrho_{\text{ref}}$ , and it is a measure of how closely the halo-member elements are packed in a halo. By definition, unit compactness indicates a simply packed system, whereas the higher values point toward a tighter packing or overlapping of the arms of the halo elements. Values lower than unity are indicative of a loosely bound system. Figures 4(a) and 4(b) exhibit a scatter plot of  $C_p$  values of halos for both, without and with the enacted unbinding process, respectively. We mark  $C_p \geq 2$  (red dashed line) as the limit for the bound to tightly bound systems, while,  $C_p \leq 1$  (green dashed line) is the limit for marginally bound to loosely bound systems. It can be seen that with the chosen arm length corresponding to  $\Delta = 200$ , the high-mass systems ( $M_{\text{halo}} \geq 5 \times 10^{13} M_{\odot}$ ) have compactness within the range  $C_p \sim 1$  to 2, whereas, the compactness of low-mass systems are spread over a wide range of values,  $C_p \approx 1$  to 6. It is also evident from the plots that the unbinding process filters out the larger number of loosely bound low-mass systems compared with that of high-mass systems, as also discussed in Sec. III A 2. We should mention here that a few low-mass halos exit the MITRO-halo list while the unbinding process was enacted as they could not meet the minimum particle condition ( $N_{\text{min}} \geq 100$ ; see Sec. III A 1) any more, as during the unbinding it may have lost a few member particles. By explicit inspection, it was also found that some of the low-mass halos lost their peripheral or satellite particles that resulted in a marginal decrement in their masses as well as in their extents, leading to enhancement of the mean density of the halos and consequently the increment of  $C_p$  values. This can be easily observed in halos with mass  $M_{\text{halo}} < 5 \times 10^{13} M_{\odot}$  in Fig. 4(b).

Furthermore, if we were to understand the situation in which a halo become highly compact, we need to understand how the halos are formed. The hierarchical clustering model says, these structures continuously accrete material from the surroundings. Therefore, if the matter is available in the surroundings, the member particle distribution will fall gradually toward the outskirts, and cuspsiness in the halo will be less, while lack of surrounding materials may lead to a cuspy halo because of the presence of a dominant core,

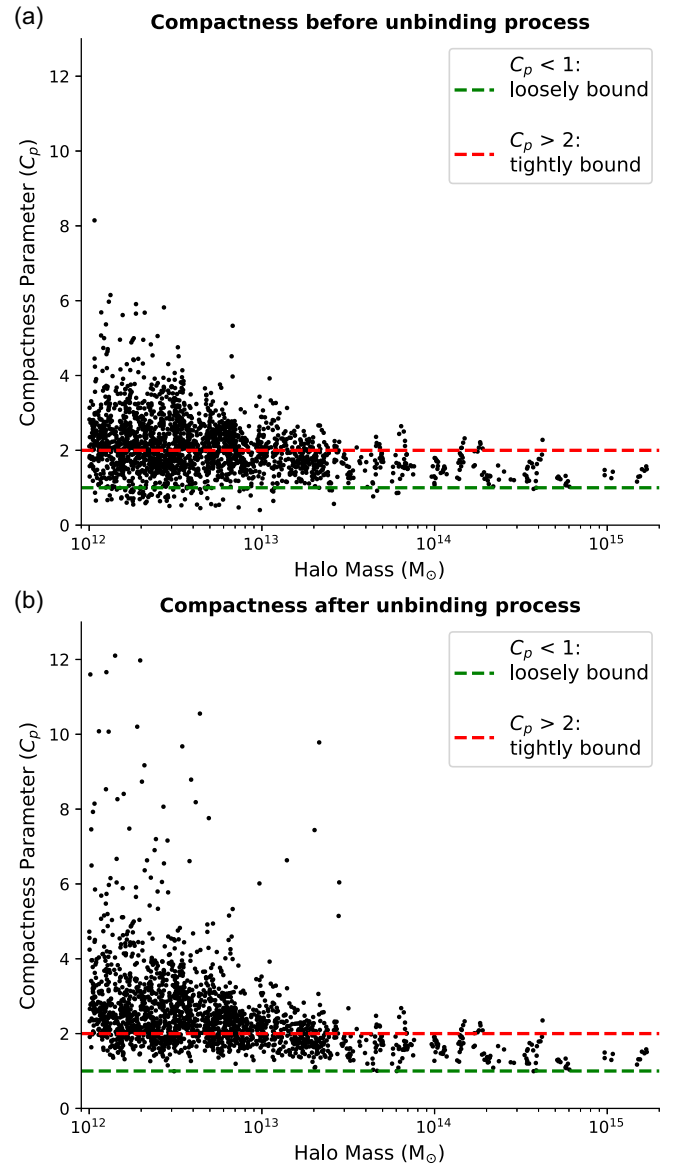


FIG. 4. The scatter plot of compactness parameter, i.e.,  $C_p$ , for formed halos with  $\Delta_{\text{arm}} = 200$  as a decisive factor to assign the arm length to the particles in search of member particles, for all MITRO halos having mass  $\geq 10^{12} M_{\odot}$ : without initializing the unbinding process (a) and after the unbinding process (b). The green dashed line corresponds to the unit  $C_p$  below which halos are considered to be the loosely bound structures whereas the red dashed line corresponds to  $C_p = 2$  above which halos are tightly bound systems with respective overdensity threshold in which the halos were searched.

leading to high  $C_p$  values. In cosmic web structures, a lack of material can be foreseen as the void or a low-density region such as the cosmic filaments. Therefore,  $C_p$  values can be a good indicator of the environment of the halos in which they are evolving. Here, we should clarify that it is not feasible to draw a general conclusion on the specific surrounding environment based on  $C_p$  values only, mainly because of its intrinsic dependence on  $\varrho_{\text{ref}}$  which is controlled by the

chosen value of  $\Delta_{\text{arm}}$  for assigning the arm length to each particle. Nevertheless, in general, a higher value of compactness parameter  $C_p$  would mean the lower surrounding matter density for the halo with respect to the searched overdensity threshold, as well as revealing the cusps of the identified halos. This roughly allows us to derive the type of environment in which the halo is evolving, and when the search value of overdensity is known.

## B. Finding supercluster scale structures in large redshift surveys using the MITRO algorithm

Mapping the Universe at large scales is best done with large galaxy surveys (i.e., 2dFGRS, SDSS, 2MASS, WISE etc.). Various observatories have done or are currently doing sky surveys with different science goals. However, in general, they collect the basic information, i.e., the photon counts of the sky patches in different optical color bands. This is collectively processed in a database after filtering the objects and their associated parameters: basically, object type, color bands magnitude, its position on the celestial sphere, etc. The most crucial parameter that helps unfold the 3D structure of the Universe is the redshift information. A few of the surveys provide either photometric or spectroscopic redshift to infer the line of sight distance—such surveys are called the galaxy redshift surveys [49,50]. Though galaxy surveys are done at various frequency bands (e.g., x ray, radio, optical etc.), among them the largest and deepest (in redshift) coverage is provided by the optical surveys, e.g., SDSS [6,7]. Using the 2.5-meter Sloan Telescope [51] operated at the Apache Point Observatory, the SDSS has provided by far the largest map of the Universe ( $\sim 14,555 \text{ deg}^2$  sky coverage, and star-forming galaxy's redshift depth  $\sim 0.8$ ) by constantly surveying the sky for over the last two decades [52]. This large galaxy map of the Universe provides critical cosmological information on the key physical processes that govern the evolution of the Universe through its specific clustering patterns. To produce the clustering pattern from the raw galaxy survey data, one needs to identify the grouping of galaxies as precisely as possible as we have elaborated in Sec. I.

Here, we discuss the application of the MITRO algorithm to real observational data. We chose the Sloan Digital Sky Survey (SDSS-III DR12) galaxy data [6,53] in the region of the most massive supercluster of galaxies, the Saraswati supercluster [54] for our study, primarily to compare our results with the ones given in the literature and understand how robust our algorithm is in the context of identifying most massive superclusters and its properties.

### 1. Data Sample

We obtained the galaxy data from the SDSS data archive using the same constraint as Bagchi *et al.* [54] (hereafter JB17) in the field of the Saraswati supercluster to maintain

consistency in data selection criteria. This would allow us to compare the results reported in JB17 with ours, which uses the proposed MITRO algorithm in a consistent manner. We queried the spectroscopic data (using CAS-Job submission portal) from three different programs: “LEGACY,” “BOSS,” and “SOUTHERN”; of the SDSS-III DR12 database for galaxies having redshift errors  $< 1\%$ , and clean photometry within the wedge of Right Ascension (R.A.) and Declination (Dec.) range  $336^\circ \leq \text{RA} \leq 16^\circ$ ,  $-1.25^\circ \leq \text{Dec.} \leq +1.25^\circ$ , and spectroscopic redshift range  $0.23 \leq z \leq 0.33$ . The data from all three programs were combined together to give out a total of 3136 galaxies. Furthermore, we also obtained the SDSS apparent magnitude for each of the galaxies as an additional physical parameter required by the MITRO algorithm (see Sec. III B 2 for more details). We adopt the extinction corrected *model* apparent magnitude calculated from the best of two fits (a de-Vaucouleurs and an exponential luminosity profile) in the r band. Here, to mention, we preferred the *model* apparent magnitude in order to measure the unbiased colors of the galaxies, since their flux is measured in the equivalent aperture for all the SDSS bands, which makes the galaxy color an aperture-independent parameter. The absolute magnitudes of galaxies are  $k$  corrected using *K-CORRECT* (version 4.2 Python) software [hereafter [55] BR07] for the rest frame at  $z = 0$ , and evolution corrected  $E_b(z) = E_{0,b} z$  using  $E_{0,b} = \{2.3, 1.6, 1.3, 1.1, 1.0\}$  in  $b = \{u, g, r, i, z\}$  [hereafter [56] B03].

For further calculations, required to run our group finding algorithm, we produced the comoving Cartesian co-ordinates of each galaxy in the above-discussed (flux-limited) sample using the relations

$$\begin{aligned} x_i &= R_i \cos(\delta_i) \sin(\theta_i) \\ y_i &= R_i \cos(\delta_i) \cos(\theta_i) \\ z_i &= R_i \sin(\delta_i), \end{aligned} \quad (11)$$

where  $R_i$  is the comoving distance of the  $i^{\text{th}}$  galaxy, and  $\theta_i$  &  $\delta_i$  are their respective right ascension and declination. For computing various quantities, we adopted the cosmological parameters based on five-year WMAP results [45];  $\Omega_M = 0.279$ ,  $\Omega_\Lambda = 0.721$ ,  $\Omega_R = 8.493 \times 10^{-5}$  and Hubble parameter,  $H_0 = 70.1 \text{ km s}^{-1} \text{ Mpc}^{-1}$  (same as followed by JB17, *however, we verified that the adopted value of  $\Omega_R$  does not affect the final results*). The galaxy distribution within the selected wedge in the translated comoving Cartesian coordinate space has been shown in Fig. 5.

### 2. Applying the MITRO algorithm on the survey data

Before we started finding the overdense regions in the survey data, we first made the data compatible with the structure of the proposed algorithm that is originally written on the basis of the simulated data. The principal basis of the proposed algorithm is to assign the physically motivated arm length to each and every element (here the galaxies) in



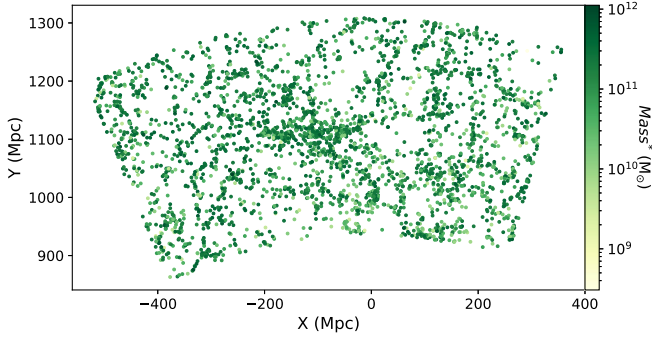


FIG. 5. Selected wedge of 3136 galaxies plotted in comoving coordinates, projected on the two-dimensional  $X$ - $Y$  plane. The declination along the  $Z$  axis has been suppressed.

the search volume, as discussed in Sec. II A, which we achieved by correlating the arm length with the mass of each galaxy. The choice of mass as the parameter for assigning the arm length is robust as gravitational force is the dominant of all forces acting at large scales, essential for building the bound structures in the Universe. However, unlike the particles in simulated data, no direct mass information is available for the galaxies in the SDSS data sample, and therefore we need to estimate the masses of galaxies in the selected SDSS wedge.

It is well known that the mass composition of the galaxies is largely dominated by dark matter along with a non-negligible visible baryonic mass. It is the total mass (i.e., baryon + DM) of individual galaxies that determines the gravitational field of influence, and is essential information required by the MITRO algorithm for computing the arm-length parameter. Although, as a first approximation, the stellar-to-halo mass relation [57–61] would provide a better proxy to compute the total mass of galaxies, this relation does not always provide the mass of local halo of individual galaxies; rather, it usually provides the mass of cluster DM halo in which the galaxy and their companion galaxies are evolving, making it unfit for our case. Therefore, as an alternative, we first estimate the color-derived stellar mass,  $M_*$  of each galaxy in the selected wedge by adopting the correlation, discussed in Bell *et al.* [56] (hereafter B03), between the stellar mass-to-light ratio and color of the galaxy. Bell *et al.* [56] report that at redshift,  $z = 0$ ,  $\log(M_*/L_r) = 1.097^0(g-r) - 0.306 - \text{offset}$ , where the  $M_*/L_r$  ratio is in solar units. We took the effective offset of 0.10 corresponding to the Kroupa [62] initial mass function [63]. Stellar masses  $M_*$  were then obtained by multiplying  $M_*/L_r$  to the rest-frame r-band luminosity  $L_r$  (which is not evolution corrected). The final relation that we obtained for the stellar masses of the galaxies is

$$\log \left[ \frac{M_*}{M_\odot} \right] = 1.097^0(g-r) - 0.406 + 0.4[M_{\odot,r} - {}^0M_r], \quad (12)$$

where  $M_{\odot,r} = 4.64$  is the r-band absolute magnitude of the Sun in the AB system (BR07), and  ${}^0(g-r)$  and  ${}^0M_r$  are the  $(g-r)$  color and r-band k-corrected absolute magnitude to redshift  $z = 0$ , respectively. We computed the  $K + E$  corrected absolute magnitude in band-pass  $b$  for each galaxy using

$${}^0M_b = m_b + \Delta b_{AB} - DM(z) - K_b - E_b(z), \quad (13)$$

where  $\Delta b_{AB} = \{-0.036, 0.012, 0.010, 0.028, 0.040\}$  for  $b = \{u, g, r, i, z\}$  [64] is the latest zero-point correction for converting the apparent magnitudes ( $m_b$ ) of SDSS to the AB system.  $DM(z) = 5 \log[D_L/\text{Mpc}] + 25$  is the bolometric distance modulus corresponding to the luminosity distance  $D_L$  in WMAP cosmology (given in Sec. III B 1).  $K_b$  and  $E_b(z)$  are the k correction and evolution correction to  $z = 0$ . Out of 3136 galaxies in the list, a  ${}^0(g-r)$  color-derived stellar mass of two galaxies was found to be unrealistically high, i.e.,  $> 10^{14}M_\odot$ , plausibly due to incorrect color magnitude in the specific band. For them, we obtained the stellar mass from  ${}^0(r-i)$  color using the relation (refer to Table 7 of B03)

$$\log \left[ \frac{M_*}{M_\odot} \right] = 1.431^0(r-i) - 0.122 + 0.4[M_{\odot,r} - {}^0M_r]. \quad (14)$$

Finally, we obtained stellar mass for all the galaxies that came out to be in the range  $10^9 - 10^{12}M_\odot$ , and plotted them as a color map in Fig. 5. Once we computed the stellar masses, the next task was to find the dark-matter mass associated with each of the galaxies which is not a very straightforward task to do. Furthermore, in the spectroscopic galaxy survey (SDSS-DR12) that we took for this study (same as the JB17), the number density of the galaxies is quite low in the examined volume and therefore would introduce a significant missing galaxy problem. Since the proposed algorithm defines its arm length on the basis of the overdensity ratio (see Sec. II A 1), the corresponding arm length associated with the stellar mass would be unphysically shorter in the sample wedge where the galaxy number density is lower because of missing galaxies. In the following section, we come up with an approximate solution from the basic principles of cosmology that would significantly mitigate this problem.

### 3. Accounting for the missing matter (DM and the missing galaxies)

Since the MITRO algorithm defines its arm length on the basis of the overdensity ratio parameter (see Sec. II A 1) while searching the structures, having information on all materials present in the volume of interest is of paramount importance. However, unlike the simulations, where the complete information about matter (DM and baryons) and energy densities is readily available, optical galaxy surveys can only provide the information of stars and galaxies,



i.e., merely a fraction of baryonic matter density of the Universe. Information on the intergalactic medium or the DM halos is completely missing. Although the indirect evidence of DM can be obtained from the gravitational lensing effects, it needs dedicated observations and analysis of the fields and is therefore out of the scope of this work. Furthermore, galaxy surveys are intrinsically flux limited depending on the depth of the observations. Owing to the dimming effect [65], objects with the same luminosity but placed at a higher redshift may remain undetected, while at a lower redshift, they may get detected if they are above the sensitivity limit of the instrument(s). So, many faint galaxies present at higher redshifts will remain undetected if they are below the sensitivity of the instrument of a galaxy survey, causing missing matter problems in the volume of interest.

To take account of this missing matter (due to undetected galaxies and DM content as well as intergalactic medium), we provide a very simple solution from the basic principles of cosmology. We began by dividing the 3D redshift-space of the sample in  $n$  parts ( $n = 10$  for our analysis) along the line of sight keeping the comoving volume enclosed by each part to be equal and greater than  $(100 \text{ Mpc})^3$ . This particular volume is chosen to retain the homogeneity scale of the Universe, the scale at which the statistical properties of the Universe have transnational invariance. Bharadwaj *et al.* [66] and Ntelis *et al.* [67] have reported this scale to be  $\sim 63h^{-1} \text{ Mpc} \approx 90 \text{ Mpc}$ , at which the observables, i.e., the mean counts of neighboring points in the  $D$ -dimensional space and the fractal correlation dimension, approaches its homogeneous value. The cosmological mean-matter density would therefore account for all the material present in each of these sections (say,  $\rho_{m_c}^i$  in  $i^{\text{th}}$  section), whereas, the matter density measured using the color-derived stellar mass of the galaxies (see Sec. III B 2) in each section (say,  $\rho_{m_{\text{SDSS}}}^i$  in  $i^{\text{th}}$  section) will be much lower as it considers only some fraction of mass (i.e., baryonic) information luminous only in optical energy band. Therefore, the difference between these two densities may tentatively account for the missing matter in the sampled volume. To compensate for this mass, we simply distributed the unaccounted mass to each galaxy in the respective comoving volumetric section. For a rationalized distribution, a weight factor ( $w_i$ ) is computed for each  $i^{\text{th}}$  section by taking the ratio of cosmological mean matter density at the mean redshift of the comoving volumetric section to the corresponding total stellar mass density,  $w_i = \rho_{m_c}^i / \rho_{m_{\text{SDSS}}}^i$ , and multiplied it by their individual stellar mass, i.e.,  $w_i \times M_{*ki}$ , where  $M_{*ki}$  is the stellar mass of the  $k^{\text{th}}$  galaxy in the  $i^{\text{th}}$  volume sector. The missing matter from the surroundings thus has been distributed to each of the galaxies in the  $i^{\text{th}}$  section in proportion to their stellar masses. The best part of this approach is, even if the number of galaxies in a field increases due to the higher sensitivity of any future surveys or if we include the SDSS

photometric data along with our current spectroscopic sample (where available), the results would not get affected considerably (provided the increased number density of galaxies is uniformly distributed over the examined volume). Since the change in galaxy number density due to the availability of more galaxies in the field would modify the stellar mass density ( $\rho_{m_{\text{SDSS}}}^i$ ) of the field, the corresponding weight factor will get adjusted accordingly. For the same reason, the issue of increasingly high missing galaxies in the deeper redshift fields due to the dimming effect and flux limitations, which usually needs extra attention while computing various clustering parameters such as the linking length (in FoF), would automatically get adjusted in our proposed MITRO algorithm (more details in Sec. III B 4).

#### 4. Arm length in RA-DEC plane and along the line of sight

A great deal of discussion has been done in Sec. II A 1 regarding the concept of arm length in our proposed algorithm. To assign arm length for individual galaxies in a field, the most crucial information is the total mass ( $M$ ) of each galaxy, which, in the case of observational surveys, we attempted to obtain by the method described in Sec. III B 3. The arm length of each galaxy has been determined by Eq. (4), where  $M$  is the effective mass of the galaxy, given as,  $M_k = M_{*ki} w_i$ , and  $\rho_c$  is the critical density of the Universe at the redshift of the galaxy. The other important parameter is the choice of  $\Delta$  which decides the arm length for the individual galaxies. As we mentioned earlier, while finding the local overdense regions,  $\Delta$  is the measure of the matter overdensity with respect to which the structures are identified. Since we are interested in identifying the very large structures, specifically the galaxy superclusters, appropriately the overdensity parameter has been chosen. Based on the adopted cosmology, to ensure sustainable gravitationally bound structures at the supercluster scale, the required bare minimum density contrast is suggested to be  $\rho/\rho_c > 2.36$  [68]. Accordingly, we chose  $\Delta = 2.36$  as a reasonable approximation for our further analysis. In the entire volume of interest (Sec. III B 1), with  $\Delta = 2.36$  and the other above-discussed parameters, we obtained the individual arm length for the galaxies in the range of 680 kpc to 10.8 Mpc according to their weighted masses.

While this fixes the arm length in R.A.-Dec. plane, finding an arm length along the line of sight needs careful consideration. The redshift (cosmological redshift) of the extra-galactic sources (here, galaxies) has been used to estimate their distances along the line of sight (LoS) which provides the de-projected position information from the inner surface of the celestial sphere. However, this de projection of galaxy positions on the LoS is largely affected by redshift distortions (RSD). It is generally assumed that the observed redshift is the result of cosmological expansion only; however, in practice, the recession velocity of the

objects also manifests the peculiar velocity induced by the gravity of neighboring objects. This perturbs the cosmological redshifts of the objects and hence distorts the clustering pattern of galaxies in redshift space. On the small scales, this elongates the galaxy cluster along the LoS, known as the Finger of God effect [69], whereas, it squashes the clusters at large scales, called as Kaiser effect [70]. Although there are additional effects induced by general relativity, the gravitational redshift distortion [71] and Sachs-Wolfe effect [72], the effects are quantitatively insignificant. Apart from these dynamical distortions to the redshift space, the difference between the adopted cosmology (while computing the line of sight distances to the galaxies, using their redshift information) and the true underlying cosmology of the Universe, artificially induces the geometrical distortions in the clustering of galaxies, known as Alcock-Paczynski effect [73]. There is no direct method to correct all these possible distortions in redshift space, at least to the observational data [74–77].

Here, we attempted to address this issue by borrowing an idea from the basic approach used for defining the linking length in traditional FoF. The traditional FoF-group finders while searching clusters in galaxy surveys, adopted two linking lengths approach. One is along the LoS (in the radial direction) and the other one is transverse of it (on the 2D projected sky plane). Therefore, instead of searching for friends within the spherical region around the individual galaxy, one needs to do the same within an ellipsoidal region. However, in practice, it is done in a cylindrical region, as a cylindrical kernel is found to recover groups better than the ellipsoidal one [29]. In the cylindrical kernel, the transverse linking length ( $l_{\perp}$ ) is the radius of the cylinder, and the length of the cylinder is taken as the linking length ( $l_{\parallel}$ ) in the radial direction. Observationally, due to RSD effects, the comoving LoS separations between two galaxies within the group appear larger than the projected separations. However, these linking lengths cannot be estimated directly from the galaxy survey data. Instead, they are determined from the analysis of mock galaxy catalogs from N-body simulations, where the true clusters are known. In literature, the typical ratio for the radial to transverse linking length (i.e.,  $\gamma_{gr} \equiv l_{\parallel}/l_{\perp}$ ), is found to be in the range between 5 and 20 [31,40], and is chosen as a fixed number for all galaxies in a particular study, depending on their science goals. Such an approach of fitting all galaxies with a single LoS arm length (in MITRO) would go against our basic idea of finding linking parameters from the physical quantities reflected from the contribution of individual elements, as done for the transverse linking length. Also, an exercise to find the most suitable combination of  $\{l_{\parallel}, l_{\perp}\}$  from any simulated mock galaxy catalog to test the recovery of true simulated galaxy clusters is out of the scope of this study. For a straightforward application of the MITRO algorithm on galaxy redshift survey data, keeping the main essence of

our proposed algorithm intact, we adopted a different but theoretically motivated way to compute the radial-to-transverse linking length ratio,  $\gamma_{gr}$ .

As we already discussed, the usual elongation of the linking parameter, along the LoS, depends on the motions of the galaxies inside the halo. Thus the general notion of the peculiar velocity of the galaxies inside the halos may provide a clue to overcoming the RSD effect. In a simplistic approach, we assumed each galaxy of the sample as a part of a halo, with halo mass  $M_h$  and radius  $r_h$ , where, the virialization process would ensure the one-dimensional velocity dispersion [78] to be

$$\sigma_{v,z} = \left(\frac{\Delta}{16}\right)^{\frac{1}{6}} (GM_h H_z)^{\frac{1}{3}}, \quad (15)$$

where  $\Delta$  is the mean density contrast relative to the critical density at the halo redshift, and  $H_z$  is the redshift-dependent Hubble parameter. Using  $\rho_c = \frac{3H_z^2}{8\pi G}$ ,  $M_h = \frac{4\pi}{3} \Delta \rho_c r_h^3$ , and with some elementary mathematics, Eq. (15) reduces to

$$\frac{\sigma_{v,z}}{r_h} = \sqrt{\frac{2}{3}} \pi G \Delta \rho_c. \quad (16)$$

Equation (16) expresses a velocity equivalent to the peculiar velocity of the constituent galaxies inside the virial radius of the halo. At the present redshift, for the adopted cosmology,  $\Omega_M = 0.279$ ,  $\Omega_{\Lambda} = 0.721$ , the spherical collapse model provides the value of  $\Delta \approx 100$  for barely virialized clusters. This gives rise to

$$\frac{\sigma_v}{r} \approx 350 \text{ km s}^{-1} \text{ Mpc}^{-1}. \quad (17)$$

Consequently, in a virialized structure of radius  $r$ , the peculiar motions of the constituent galaxies can stretch the structure to  $1\sigma \sim \frac{3.5}{h} r$  along the line of sight. This makes the ratio of LoS to the transverse size of the object independent of  $r$  and is  $\frac{3.5}{h} \sim 5$  ( $h = 0.701$ ). Intriguingly, such a theoretically motivated number turns out to be quite consistent with the elongation factors  $\sim 5$ – $15$  (i.e., well overlapping within  $1\sigma$ – $3\sigma$  range) adopted by various authors [e.g., [31,40]] while finding groups using the traditional FoF method. Furthermore, we speculate that each galaxy in  $\mathcal{S}_{\text{SDSS}}$  as a part of some pseudogalaxy cluster; therefore the maximum distortion it can introduce in the cosmological redshift is  $\sim (1\sigma$ – $3\sigma)$  with respect to the searched overdensity threshold. However, while clustering, instead of correcting the galaxy redshift space due to RSD, we compensated for it by introducing an elongation in the ‘‘arm length’’ along the line of sight. We assumed the  $1\sigma$  elongation factor while assigning the arm length along the line of sight ( $l_{\parallel}$ ) in our algorithm. Therefore, it boils down to the elongation ratio of  $\gamma_{gr} = \frac{\sigma_v}{r H_0}$ , where  $H_0$  is the Hubble parameter. Since the critical density  $\rho_c$  is redshift

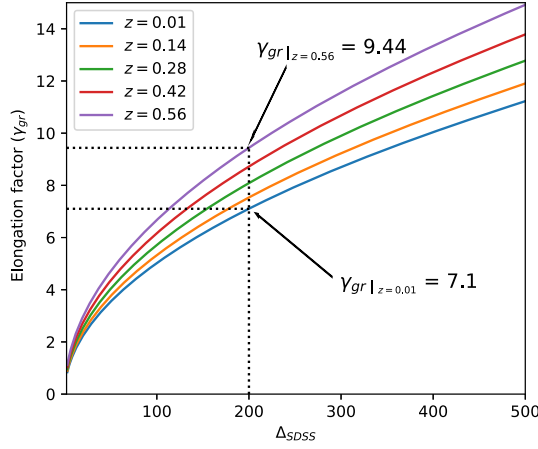


FIG. 6. Variation of the elongation ratio with the chosen density threshold  $\Delta$  while defining the arm length to each galaxy. At higher redshift, the elongation factor increases for the definite choice of density threshold  $\Delta$ .

dependent, the overdensity parameter,  $\Delta$  (here,  $\Delta_{\text{SDSS}}$ ) and concomitantly the elongation in arm length for individual galaxies would be different depending on their redshift. In the selected  $\Delta_{\text{SDSS}}$  (Sec. III B 4) and adopted cosmology with  $H_0 = 70.1 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , the values of  $\gamma_{gr}$  ranges between 0.86 and 0.9. For a better illustration of the variation of the elongation factor with respect to the overdensity parameter and redshift of the examined volume, we produced a plot shown in Fig. 6. The plot highlights that for the usual virialized structures (with  $\Delta = 200$ ) formed within redshift  $z = 0.01$ – $0.60$ , the elongation factor ranges between 5 and 10. However, the structures at supercluster scales, whose overdensity factor is usually much lower at about  $\Delta = 2$ – $3$ , result in an elongation factor of 0.77–1.05 indicating wall-like structures. *This situation is quite similar to the Kaiser squashing effect, however, may not represent this as a general case.*

### 5. MITRO algorithm applied to galaxy survey data

The basic MITRO algorithm for finding large-scale bound structures is as described below:

- Step 1: Compute the color-derived stellar mass,  $M_{*i}$ , of each galaxy in the sample [Eq. (12)].  $M_{*i} \forall i \in \mathcal{S}_{\text{SDSS}}$ .
- Step 2: Estimate the weight factor,  $w_i$ , for each galaxy, to account for missing galaxies, Intergalactic medium (IGM), and DM in the survey sample (Sec. III B 3).
- Step 3: Define the arm length,  $L_i$ , to each galaxy using Eq. (4) and relevant parameters as discussed in Sec. III B 4.
- Step 4: Estimate the radial-to-transverse link-length ratio, i.e., the elongation ratio  $\gamma_{gr}$  (Sec. III B 4).
- Step 5: With the arm length as the radius (transverse length) in the R.A.-Dec. plane, and line of sight (radial length, determine from  $\gamma_{gr}$ ) as the depth, create cylindrical regions for each galaxy such that the

galaxy is placed at the middle of the axis of the cylinder.

Step 6: Now search for friends in the selected wedge—whenever two cylinders correspond to the pair of galaxies intersect each other, those pairs of galaxies are considered as friends. And finally, search for the friends of friends to get the final configuration of the bound structures in examined sky volume, with a minimum of ten galaxy constraints.

To define friend, two galaxies,  $i$  and  $j$ , are friends if they satisfy the following conditions:

- (1) the angular separation ( $\theta_{ij}$ ) between two galaxies at comoving distances  $R_i$  and  $R_j$ , and arm length  $L_i$  and  $L_j$  respectively,

$$\theta_{ij} \leq \frac{L_i}{R_i} + \frac{L_j}{R_j}, \quad (18)$$

- (2) and the LoS difference

$$|R_i - R_j| \leq \gamma_{gr}^i L_i + \gamma_{gr}^j L_j. \quad (19)$$

As a preliminary test, we implemented this modified MITRO algorithm on a simulated mock galaxy sample that roughly mimics the real observational data. Furthermore, before applying the proposed algorithm to a real observational dataset, we produced the halo-mass function of MITRO-identified structures and compared them with the standard theoretical halo-mass functions (see Appendix A for more details).

### 6. MITRO-identified supercluster scale structures in SDSS-DR12 sample

Finally, we implemented our algorithm on galaxy survey data to identify supercluster scale structures. In order to search the overdense regions in observational data, we introduced various additions and modifications to the data as well as to the algorithm, as discussed in the above sections. We choose the spectroscopic galaxy survey sky wedge from SDSS-DR12 in the field of the Saraswati supercluster by running relevant SQL queries (Sec. III B 1) as the dataset. The choice of the overdensity value is the most crucial parameter for defining an identified structure as A supercluster [see [79]], specifically to ensure that the structure remains gravitationally bound at present or will eventually collapse in the future. Since our principal aim is to identify supercluster scale structures, as discussed in Sec. III B 4, as a reasonable approximation derived from the spherical collapse model, we choose the overdensity parameter to be  $\Delta \equiv \Delta_{\text{SDSS}} = 2.36$  [68,79] for running our algorithm.

The selected SDSS data wedge consists of 3036 galaxies. It is intrinsically a flux-limited sample, and also not uniformly sampled in the redshift space. Furthermore, as only a fraction of baryonic matter is observed in the form of the galaxy in these surveys, most of the matter density is



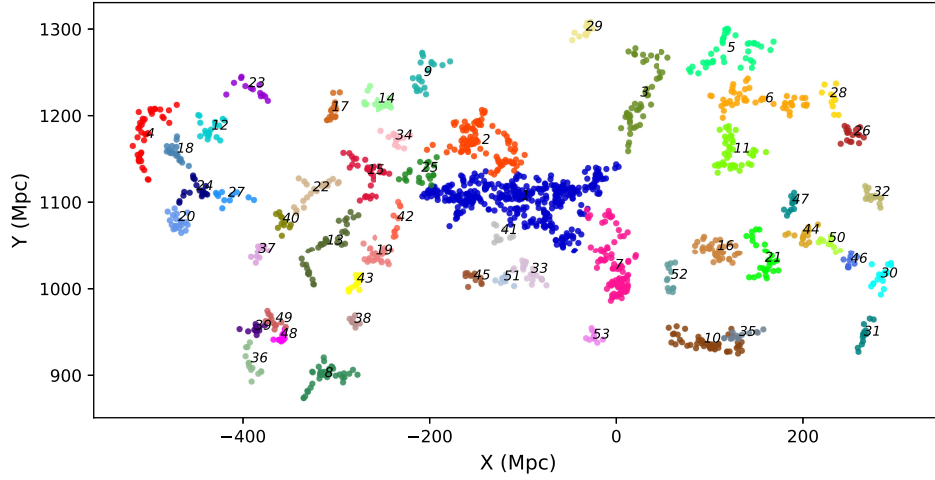


FIG. 7. Structures in  $\mathcal{S}_{\text{SDSS}}$  using the MITRO algorithm having a minimum of ten galaxies. Different colors indicate different structures with their unique structure IDs as plotted in comoving coordinates, projected on the two-dimensional  $X$ - $Y$  plane. Declination along the  $Z$  axis has been suppressed.

missing throughout. To compensate for this surrounding missing matter, we assigned a weight ( $w_i$ ) factor to each galaxy in the sample following the method described in Sec. III B 3. Eventually, the arm length to each galaxy is defined using the weighted mass of each galaxy, i.e.,  $w_i M_i^*$ , at a specified overdensity and identified the overdense structures in  $\mathcal{S}_{\text{SDSS}}$  by running the MITRO clustering algorithm (Sec. III B 5).

Our algorithm spotted a total of 53 structures within the specified analysis region of the sky. The MITRO-identified structures are shown in Fig. 7, each with a different color and a unique cluster ID (CID). A unique ID has been assigned in ascending order of the total weighted mass of the structures. As expected, the largest and most massive, consequently the first structure (CID-1) in the list is the Saraswati supercluster. Our algorithm identified this supercluster with an extension of  $\sim 210$  Mpc (comoving) located at R.A.  $354.41^\circ$ , and Dec.  $0.267^\circ$  (LGP center) and at the mean redshift of  $z = 0.277$ . We found 384 member galaxies in the MITRO-identified Saraswati supercluster (Fig. 8) with an estimated average comoving linking length of  $\approx 9.4$  Mpc. The entire structure encloses  $\approx 6.47 \times 10^5 \text{ Mpc}^3$  comoving volume or  $\approx 2.23 \times 10^5 \text{ Mpc}^3 h^{-3}$  comoving volume in scale of the Hubble parameter having an effective mass of  $1.86 \times 10^{17} M_\odot$  and houses 42 known WHL<sup>4</sup>-galaxy clusters (see Table IV). The enclosed comoving volume is computed with the SciPy

<sup>4</sup>WHL is the shorthand of the initials of the three authors, named Wen Z. L., Han J. L., and Liu F. S. who identified 132,684 galaxy clusters using the photometric redshifts of galaxies from the SDSS in the redshift range of  $0.05 < z < 0.8$  using the FoF algorithm and constraint on cluster richness (ratio between the total  $r$ -band luminosity within the radius  $R_{200}$  to the evolved characteristic luminosity of galaxies in the  $r$  band). The detailed cluster detection algorithm can be found in Wen *et al.* [80,81].

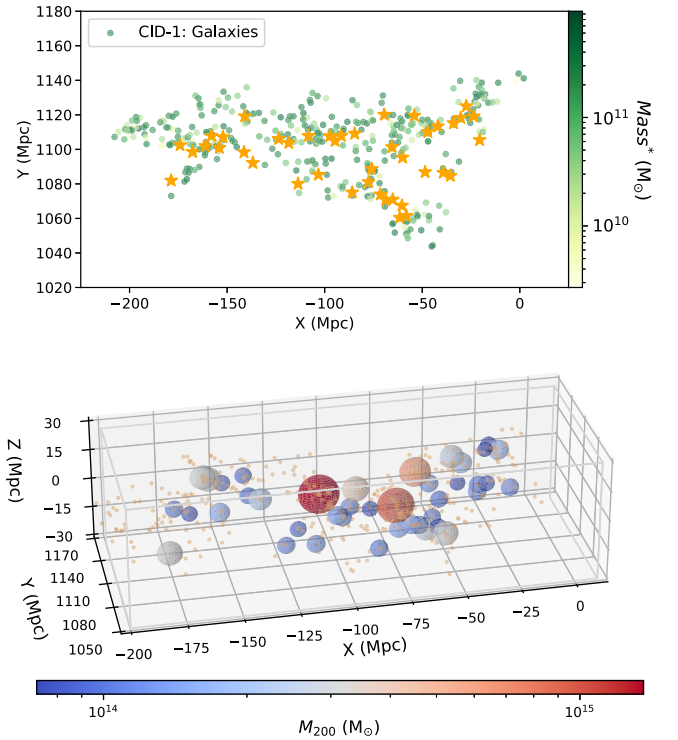


FIG. 8. Upper panel: Saraswati supercluster plotted in comoving coordinates, projected on the two-dimensional  $X$ - $Y$  plane, spanning  $\sim 200$  Mpc with 384 galaxies in it. Colors represent the stellar mass of fellow galaxies, and the color scale is shown with the color bar. The orange star marker symbolizes the known 42 WHL clusters of galaxies in the identified Saraswati structure. Lower panel: 3D distribution of galaxies and 42 WHL galaxy clusters comprised in the MITRO-identified Saraswati structure. Spheres represent the known WHL-galaxy cluster, and the corresponding color represents the mass of the cluster in log scale as shown with the color bar. The radius of each sphere is proportional to its radius  $r_{200}$  corresponding to the mass  $M_{200}$  defined using Eq. (2) in [81].



quick-hull algorithm [42] for the convex hulls as discussed in Sec. II B 2.

Previously, Bagchi *et al.* [54] (JB17) in their discovery paper reported the Saraswati supercluster to be a structure spanning  $\sim 200$  Mpc roughly centering at Abell 2631 and at a mean redshift of  $z = 0.28$ . They implemented the FoF algorithm to identify the structure, and the linking length of  $\ell_0 = 12$  Mpc was chosen by them so that they could identify the maximum number of FoF clusters in the searched volume with at least ten galaxies in each. Within the filamentary structure of Saraswati, they found 43 WHL clusters, and the computed volume of the convex hull of these 43 cluster points is reported to be  $V \approx 2.05 \times 10^5$  Mpc<sup>3</sup>. The authors estimated the mass of Saraswati using a very crude method of summing the so-called bound halo mass,  $M_{5,6}$  of 43 WHL clusters associated with the structure which they reported as  $M_{ss} \sim 2 \times 10^{16} M_\odot$ . The authors further mentioned that their estimation is strictly a lower limit as a lot of mass may still remain uncounted.

It is noteworthy to mention that our findings are fairly consistent with the JB17, especially the extent, the enclosed comoving volume in the Hubble scale, the average linking length, and the number of identified WHL clusters within the structure. Interestingly, the LGP center and the average redshift for this supercluster, as identified by MITRO code, are found to be very close to the most massive galaxy cluster inside the structure, i.e., Abell 2631 ( $M_{500} \approx 10^{15} M_\odot$ ). However, the estimated total mass and the computed overdensity for the structure do not compare well. The difference may be attributed to the method of estimation of mass as well as enclosed volume. We considered the comoving volume and estimated the weighted mass of the member galaxies that well accounts

for the uncounted mass, whereas JB17 computed the overdensity contrast by considering the sum of the turnaround mass of each WHL-clusters and the Hubble parameter scaled comoving volume of the enclosed structure. The overdensity factor (over critical density) reported by JB17 is  $\rho/\rho_c = 1.17$ , while our estimation is slightly higher  $\rho/\rho_c = 1.62$ . However, both the numbers are far lesser than the required critical value of  $\rho/\rho_c = 2.36$  that is necessary for the supercluster to remain gravitationally bound. Further trials revealed that, at the place of Saraswati structure, theoretically bound halos can only be found when searched with  $\Delta_{\text{SDSS}} = 3.78$ . However, the structure does not remain intact; rather, it splits into three parts with this choice of  $\Delta_{\text{SDSS}}$  (more on this in Appendix B 1). It is very important to mention here that the density contrast of  $\rho/\rho_c = 2.36$  actually comes from the spherical collapse model which may not be fully applicable to our case of a real unstructured supercluster. Moreover, the finding could also be very much dependent on the limitations of the current observations and therefore opens up the possibility of a further detailed study on this matter.

In addition to the Saraswati supercluster, we identify another 52 structures by applying the same overdensity threshold inside the specified analysis region. The associated physical properties, such as the total effective mass, location, richness, overdensity ratio, maximum radius, and largest linear size (LLS) of all the identified structures are listed in Table I (see Table III). The location mentioned in the table is the LGP center of each of the structures. It has been found that the final overdensity values of 51 out of 52 of these structures are above the threshold value of a bound supercluster. However, many of them are not massive or large enough in extent to call them superclusters. We put a reasonable constraint on the physical parameters to identify the most

TABLE I. List of 53 MITRO-identified structures in JB17-Saraswati wedge region (“LEGACY,” “BOSS,” and “SOUTHERN” program of SDSS-III DR12). “\*” indicating the Saraswati supercluster.

CID	Mass ( $10^{15} M_\odot$ )	$z$	R.A. (deg.)	Dec. (deg.)	Richness	$\frac{\bar{\rho}}{\rho_c}$	$R_{\text{max}}$ (Mpc)	LLS (Mpc)	$\bar{l}$ (Mpc)
1*	185.602	0.277	354.409	0.267	384	1.619	124.976	226.306	9.397
2	75.259	0.3	352.942	0.069	92	3.074	71.179	112.115	11.204
3	59.827	0.302	1.018	-0.009	53	4.111	83.19	138.652	12.276
4	54.982	0.33	336.031	-1.002	40	7.168	96.034	96.034	13.122
5	50.929	0.318	4.709	-0.094	41	3.546	74.782	111.204	12.518
6	49.397	0.307	6.554	0.084	48	2.756	78.104	116.96	11.992
7	43.796	0.253	359.801	0.251	99	1.951	77.811	116.988	9.256
8	34.449	0.235	340.97	-0.123	42	5.085	45.148	75.611	11.802
9	30.623	0.321	351.079	-0.029	20	8.501	44.268	66.634	13.792
10	30.237	0.232	6.156	-0.026	45	3.134	53.775	95.147	10.786

*Note.* Column 1: Structure unique ID. Column 2: Total effective mass of the structure (Stellar mass + surrounding missing matter (baryonic as well as dark matter)). Column 3: Structure redshift. Column 4: R.A. (J2000) of structure (degree). Column 5: Declination (J2000) of structure (degree). Column 6: Richness number of galaxies (according to the SDSS-III DR12). Column 7: Structure overdensity with respect to the critical density at the structure’s redshift. Column 8: Maximum radius of the structure (comoving). Column 9: comoving largest linear size of the structure. Column 10: Average linking lengths in structure. (This table is available in its entirety in a machine-readable form.)

TABLE II. List of “candidate-superclusters” of galaxies in JB17-Saraswati wedge region (“LEGACY,” “BOSS,” and “SOUTHERN” program of SDSS-III DR12).

CID	Mass ( $10^{15}M_{\odot}$ )	$z$	R.A. (deg.)	Dec. (deg.)	Richness	$\frac{\bar{\rho}}{\rho_c}$	$R_{\max}$ (Mpc)	LLS (Mpc)	$\bar{l}$ (Mpc)	Number of	
										WHL clusters	Shape
2	75.259	0.3	352.942	0.069	92	3.074	71.179	112.115	11.204	6	Circular
3	59.827	0.302	1.018	-0.009	53	4.111	83.19	138.652	12.276	5	Elongated
4	54.982	0.33	336.031	-1.002	40	7.168	96.034	96.034	13.122	4	Elongated
5	50.929	0.318	4.709	-0.094	41	3.546	74.782	111.204	12.518	4	Circular
6	49.397	0.307	6.554	0.084	48	2.756	78.104	116.96	11.992	8	Wall-like
10	30.237	0.232	6.156	-0.026	45	3.134	53.775	95.147	10.786	4	Wall-like
13	23.894	0.274	343.581	0.688	36	4.242	62.194	106.461	10.583	3	Elongated

*Note.* Column 1: Structure unique ID. Column 2: Total effective mass of the structure (stellar mass + surrounding missing matter (baryonic as well as dark matter)). Column 3: Structure redshift. Column 4: R.A. (J2000) of structure (degree). Column 5: Declination (J2000) of structure (degree). Column 6: Richness number of galaxies (according to the SDSS-III DR12). Column 7: Structure overdensity with respect to the critical density at the structure’s redshift. Column 8: Maximum radius of the structure (comoving). Column 9: comoving largest linear size of the structure. Column 10: Average linking lengths in structure. Column 11: Number of known WHL-galaxy clusters in the structure (Table IV). Column 12: Morphology/shape of structure in declination projected/suppressed 2-*dim* plane.

Elongated: stretched (or LLS) along the LoS; wall-like: stretched (or LLS) along the angular plane and squeezed along the LoS (along the redshift axis); and circular: stretched almost equally in both directions.

probable candidate superclusters in the list. We classified the structures as candidate superclusters if the object has a mass greater than  $10^{16}M_{\odot}$ , i.e., roughly  $\sim 10$  times  $10^{15}M_{\odot}$ , the observed massive galaxy clusters and LLS of  $\gtrsim 90$  Mpc, the homogeneity scale at which the statistical properties of the Universe have translational invariance [66,67]. By applying these constraints, in total, seven MITRO-identified structures were found to qualify as the “candidate superclusters,” and are given in Table II. Based on their morphology on the declination-suppressed 2D plane, we also defined the shape of these structures which are given in column 12 of Table II. Interestingly, all the candidate superclusters have an overdensity factor of bound structures and encompass at least three known WHL-galaxy clusters (respectively, columns 7 and 11 of Table II).

#### IV. SUMMARY AND CONCLUSIONS

Large-scale structures in the Universe are known to be comprised of groups and clusters of galaxies embedded in a filamentary structure of dark matter forming the cosmic web. Baryonic matter flows and drains through these DM channels and gets trapped at the large gravitationally bound structures, such as clusters and groups of galaxies that form at the nodes or the crossroads of these filaments. Studies that revealed these structures are majorly based on cosmological large-volume simulations and large-scale galaxy redshift surveys. The most important step in these studies is the appropriate identification of structures in the Universe. Whether the cosmological simulations or the galaxy surveys, the unit element in the dataset is a point source, either a DM particle or a galaxy. The known information about them is basically the position and the velocity. While, mass

information is readily available in the case of simulations, in galaxy surveys, mass can just be a derived quantity.

The bound structures in these studies are so far identified using various group finding codes, mostly based on the FoF or SO algorithms. Although, the main goal of these codes is to identify gravitationally bound structures, unfortunately, the mass information (i.e., the origin of the strongest binding force at the large scales), has hardly been effectively used by these codes. Moreover, while both from theory and observations, it is well known that the bound structures can best be formed at some particular mass overdense condition and real structures are ideally never spherical, the methods used so far constrain either the overdensity or the real unstructured geometry. Although, these are very essential parameters to extract accurate structures-mass information and precisely constrain the cosmological models of the Universe, almost no attempt has been made so far to retain these important parameters together while formulating the grouping algorithms.

In this paper, we present our algorithm called MITRO that takes care of all the above-mentioned relevant features as well as ensures the bound structures using the physical quantities, mainly mass and the total energy information. The proposed MITRO algorithm adopts a noble approach of clustering to mitigate a few shortcomings of the basic FoF and SO algorithms strengthening the foundation of clustering or halo-finding methods. Here we briefly summarize the salient features of this proposed algorithm below.

- (i) Our major goal was to retain the real geometry, as well as, to identify physically bound cosmological structures in galaxy surveys or simulated cosmological volume. To achieve this, for the first time, we used the

dominant gravitational force to determine the linking length between any two elements of a group. In the proposed algorithm, the arm length has been computed at a user-defined mass overdensity (see Sec. II A 1) for each element depending on their individual mass. This leads to having a distinct linking length (addition of two arm lengths) for each unique pair of elements. This is notably different from the usual FoF method, where a single linking length is used for each and every element, usually determined statistically as the scaled mean particle distance in the examined volume and has no dependence on the mass (i.e., gravitational force) of the elements.

- (ii) We demonstrate that the proposed algorithm has been able to successfully identify the unstructured DM halos roughly within the predefined density threshold which cannot be achieved simultaneously by usual FoF or SO-based methods. This method is therefore a significant improvement as it identifies appropriate overdense regions without imposing the spherical geometry; moreover, it uses physically relevant linking parameters, as demonstrated through various tests (see Secs. III A 2 and III A 4) presented in this paper. At the same time, the algorithm adopted the important features from both SO and FoF methods such as overdensity and percolation.
- (iv) We could demonstrate that the overdensity contrast of the MITRO-identified halos have a lower bound, at least the half of the chosen reference overdensity ( $\Delta_{ref}$ ) used to define the arm-length. This enabled us to define a new parameter called compactness. This parameter is the indicator of how closely the elements of MITRO halos are packed or how stiff the distribution is. Eventually, it reveals the cusps of the density profile divulging information about the surrounding environment in which the particular halo is evolving. No similar quantity can be defined in usual SO halos, as all halos have the same predefined overdensity and consequently, compactness is  $C_p = 1$  for all. On the other hand, since FoF halos are not bound by any overdensity contrast linked to the predefined linking length, computing compactness is not possible. Our analysis shows compactness of MITRO-identified halos varies as  $C_p \geq 1$ . Here, the higher  $C_p$  values hint at the lack of matter in the vicinity of evolving halos for the searched  $\Delta_{arm}$ .
- (v) The critical test that these clustering codes face is while analyzing galaxy redshift survey data, mainly because, such surveys do not directly provide any mass information, as well as, because of missing matter issues due to detection limitations. The usual SO method, which should primarily find the peak of

matter density in the search volume, would fail due to the lack of mass information. On the other hand, the application of FoF may result in incorrect grouping because of the use of an unphysical single linking length for all elements in the search volume. In the line of sight, dealing with the redshift distortion effect in the usual way is rather complicated as it requires dedicated cosmological simulations to estimate the equivalent elongation factor. Instead, we overcome the issue considerably using a generalized theoretical approach. Moreover, such an approach helped us to compute the required elongation for all kinds of density contrast, making life easier.

For better mass estimation of each galaxy, the color-derived stellar mass has been used in our code. Additionally, the remaining unaccounted matter due to undetected galaxies, IGM and DM content, has been estimated by considering the cosmological mean-matter density at each redshift bin within which it obeys the predicted theoretical mean-matter density. Such an assumption automatically adjusts the computed arm length of each element and eventually compensates for the missing mass information. Our theoretical approach helped us to estimate an appropriate elongation factor along the line of sight to overcome the RSD effect. The elongation factor has been determined by assuming the virialization of the halos while computing the peculiar velocities of the galaxies in our setup. Interestingly, the estimated LoS elongation factor of  $\sim 5-15$  (for overdensity density  $\Delta \sim 180$ ) in our study is found to be fairly consistent with many earlier studies.

- (vi) Finally, we examined the sky region centered at the recently discovered Saraswati supercluster, by running the MITRO algorithm on SDSS-DR12 data. Our code has successfully identified the supercluster extending  $\sim 210$  Mpc (comoving) at a mean redshift of  $z = 0.277$  having 384 member galaxies and housing 42 WHL galaxy clusters. The findings are largely consistent with the previous study by JB17. However, with a more robust and different mass estimation formalism that better accounts for the missing matter, we found the structure to be of  $M_{SCL} = 1.86 \times 10^{17} M_{\odot}$ . This is an order higher than the previously reported mass and makes it so far the most massive identified supercluster. Furthermore, by setting reasonable criteria, we report seven more MITRO-identified ‘‘candidate superclusters: in the field of Saraswati, indicating that gravitationally bound structures larger than the usual galaxy clusters are not extremely rare. Thus, in the era of increasingly large galaxy redshift surveys (SDSS, Euclid, LSST, etc.), our proposed algorithm MITRO may become a



very useful tool in unraveling the structure formation, constraining the cosmological models, and subsequently understanding the nature of dark energy in the evolutionary history of the Universe.

## ACKNOWLEDGMENTS

P. G. acknowledges the financial support from Council of Scientific and Industrial Research, Govt. of India for Senior research fellowship (File No. 09/137/0574/2018 EMR-I). P. G. wants to thank *Shishir Sankhyayan* and *Shahdab Alam* for helpful discussions on SDSS data. P. G. also wants to thank *Surhud More* for the useful conversation on the FoF algorithm. S. P. wants to thank Department of Science and Technology, Government of India for funding his research group through INSPIRE Faculty scheme (IF12/PH-44). The authors want to thank the referee(s) for their useful comments and suggestions. The authors are also thankful to the Inter-University Centre for Astronomy and Astrophysics (IUCAA) for providing the HPC facility. Computations described in this work were performed using the Enzo code developed by the Laboratory for Computational Astrophysics at the University of California in San Diego [86]. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III website is [87]. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, the Brookhaven National Laboratory, Carnegie Mellon University, the University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University. The CosmoSim database used in this paper is a service by the Leibniz Institute for Astrophysics Potsdam (AIP). The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. [88] and the Partnership for Advanced Supercomputing in Europe (PRACE, [89]) for funding the MultiDark simulation project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, [90]). Finally, the authors want to acknowledge the developers of the following Python packages used extensively in this work: SciPy [82], NumPy [83] and Matplotlib [84].

## APPENDIX A: MITRO IMPLEMENTATION ON SIMULATED MOCK GALAXY SAMPLE

### 1. Mock galaxy sample

To investigate and compare the halo mass function (HMF) of MITRO and standard FoF-identified galaxy clusters with respect to a theoretical HMF, we utilized a large-volume mock galaxy sample produced by Cora *et al.* [85]. This Semi-Analytical Galaxy (SAG) catalog uses the DM halos' information as extracted from a cosmological DM simulation (MultiDark simulation MDPL2), which is a part of the CosmoSim database (see [85] for more details). MDPL2 simulation follows the evolution of  $3840^3$  particles within a box of 1 Gpc/h each side, with a flat  $\Lambda$ CDM cosmology characterized by Planck cosmological parameters:  $(\Omega_\Lambda, \Omega_m, \Omega_b, n_s) = (0.693, 0.307, 0.048, 0.96)$ ,  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.678$ , and DM particle mass resolution  $\sim 1.5 \times 10^9 M_\odot/h$ . A few realizations of MDPL2 simulation is publicly available<sup>5</sup> on the CosmoSim database.<sup>6</sup>

We prepared a subcatalog of semianalytical galaxies from the SAG catalog<sup>7</sup> by extracting data from the CosmoSim database within the central volume of  $(400 \text{ Mpc}/h)^3$  at redshift  $z = 0.0$ . The following information was recorded in a data file: halo mass, x position, y position, z position, the absolute magnitude of galaxies with respect to SDSS Camera *r*-band response function (assuming airmass = 1.3, in rest frame), mass of stars in the spheroid/bulge, mass of stars in the disk, mass of gas in the spheroid/bulge, mass of gas in the disk, hot gas mass, and mass of central black hole. Here, the total mass of each galaxy is the sum of all extracted masses. Finally, we put a constraint on the absolute magnitude of galaxies in order to mimic the observations, i.e., *r*-band absolute magnitude  $< -17$  (as obtained in our SDSS sample). The constraint on absolute magnitude, removes many faint galaxies from the examined volume. This exercise introducing the missing matter situation in the final mock galaxy sample. In the final list, a total of 4243667 galaxies were retained in the selected volume.

### 2. MITRO galaxy clusters identified from the SAG catalog

We implemented our proposed MITRO algorithm following the steps mentioned in Sec. III B 3 for accounting for the missing matter within the selected volume, and accordingly, the weight factor ( $w$ ) was computed. Finally, the overdense regions, in this case, the galaxy clusters, are identified with the MITRO algorithm, assuming arm length corresponding to  $\Delta_{\text{arm}} = 200$  and weighted mass of each galaxy as  $w \times M_i^{\text{gal}}$ . In the given sample with the above-stated parameters and Eq. (4), the {max, min, mean, median} arm length

<sup>5</sup><https://www.cosmosim.org/metadata/mdpl2/>.

<sup>6</sup><https://www.cosmosim.org/>.

<sup>7</sup><https://www.cosmosim.org/metadata/mdpl2/sag/>.



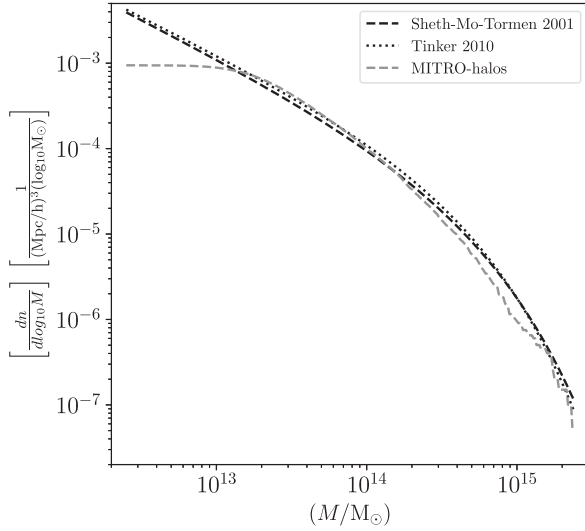


FIG. 9. Halo mass function of MITRO-galaxy clusters identified from SAG catalog.

would be  $\{2645.27, 25.14, 122.14, 101.41\}$  kpc. Since the given sample is produced from cosmological simulations and is not affected by the redshift distortion, the elongation factor need not be considered. The search for the overdense regions is therefore done in spherical kernels around the galaxies instead of using a cylindrical kernel. Furthermore, with the constraints on minimum DM particles,  $N_{\min} = 10$ , and on the total mass of galaxy cluster  $M_{\text{halo}} \geq 10^{12} M_{\odot}$ , the MITRO algorithm identified 18821 galaxy clusters with the massive most having mass  $2.4 \times 10^{15} M_{\odot}$ . The corresponding HMF has been shown in Fig. 9, which is in good agreement with the standard Sheth-Tormen 2001 and Tinker 2010 HMFs  $\geq 10^{13} M_{\odot}$ .

### 3. FoF galaxy clusters identified from SAG catalog

Since there is no direct (co)relation between the linking length and searched overdensity threshold, the linking length ( $l_f$ ) in the FoF algorithm is still a quest. Thus, for a meaningful comparison between FoF and MITRO, we chose two statistically important pointers, i.e., mean and median of the arm lengths of all objects in the MITRO algorithm when searched at  $\Delta_{\text{arm}} = 200$ . Since with this approximation, the final linking length should be equivalent to twice the chosen arm length; the equivalent linking lengths would be  $l_{f_{\text{mean}}} \sim 250$  kpc and  $l_{f_{\text{median}}} \sim 200$  kpc, respectively. The produced FoF halo-mass function (Fig. 10) shows they do not match well with the expected halo-mass function of Sheth-Tormen 2001 or Tinker *et al.* 2010. The number of clusters identified by FoF was also found to be considerably less (2347 and 6119, respectively)

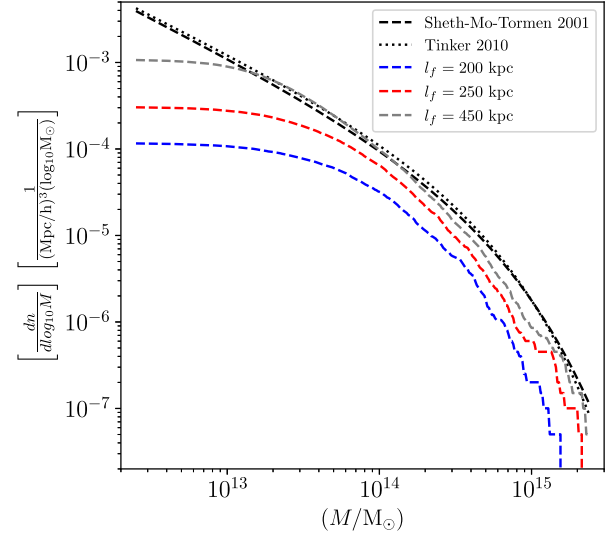


FIG. 10. Halo mass function of FoF galaxy clusters identified from the SAG catalog. Different colored dashed lines correspond to different linking lengths as mentioned in the legend.

in comparison to MITRO (i.e., 18821). This led us to do a further iterative test that resulted in a linking length of  $l_f = 450$  kpc, where FoF could roughly be able to produce a match with the halo-mass function of Tinker 2010 and Sheth-Tormen 2001. This certainly indicates the real strength of the MITRO algorithm as MITRO produces desired results without the need for an iterative trial. In addition, our algorithm also provides a tentative overdensity threshold for the identified structure, i.e.,  $\gtrsim \Delta_{\text{arm}}/2 \sim 100$ .

## APPENDIX B: STRUCTURES IN SDSS SELECTED WEDGE NEAR SARASWATI REGION

### 1. Choice of overdensity search in SDSS sample

We carried out an iterative test (each step of 10% increment) to understand the sensitivity of the  $\Delta_{\text{SDSS}}$  parameter. We found, even if we increase the value just by 10% (i.e.,  $\Delta_{\text{SDSS}} = 2.60$ ) of the chosen value (i.e.,  $\Delta_{\text{SDSS}} = 2.36$ ), the Saraswati supercluster splits into two parts. The winglike structure on the right side of the supercluster gets detached [see Fig. 11(a)]. Likewise, when we increase the value by 60% of 2.36 ( $\Delta_{\text{SDSS}} = 3.78$ ), the structure splits into three parts [left wing, center part + lower tail, right wing; see Fig. 11(b)]. However, similar to  $\Delta_{\text{SDSS}} = 2.36$ , the density threshold of identified two split parts at  $\Delta_{\text{SDSS}} = 2.60$  is also not above the theoretical limit of gravitationally bound structures. It is only at  $\Delta_{\text{SDSS}} = 3.78$  that all three split structures become gravitationally bound.

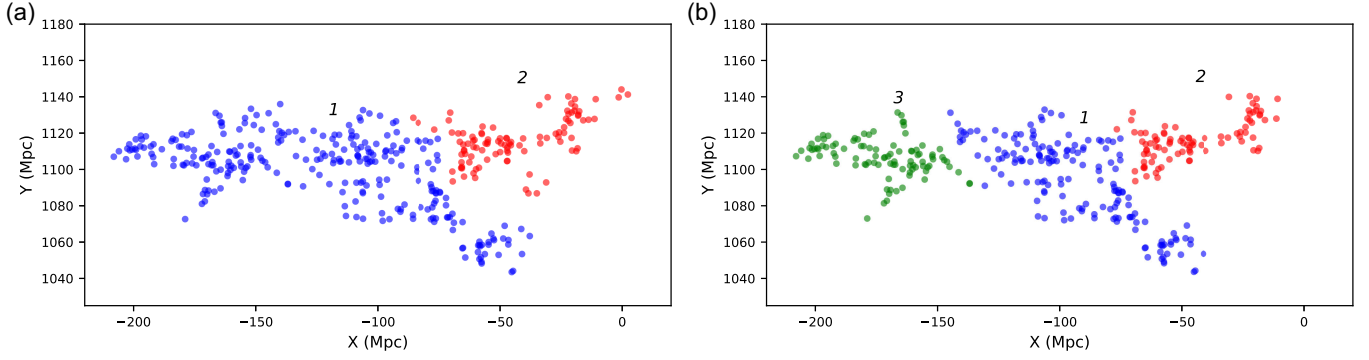


FIG. 11. (a) MITRO-identified structure at the position of Saraswati supercluster splits into two parts with the choice of  $\Delta_{\text{SDSS}} = 2.60$ . Part 2 (red) parts away from the main structure. (b) The same structure when searched at  $\Delta_{\text{SDSS}} = 3.78$ , splits into three distinct parts.

TABLE III. List of 53 MITRO identified structures from the JB17-Saraswati region (“LEGACY,” “BOSS,” and “SOUTHERN” program of SDSS-III DR12).

CID	Mass ( $10^{15} M_{\odot}$ )	$z$	R.A. (deg.)	Dec. (deg.)	Richness	$\frac{\bar{r}}{\rho_c}$	$R_{\text{max}}$ (Mpc)	LLS (Mpc)	$\bar{l}$ (Mpc)
1*	185.602	0.277	354.409	0.267	384	1.619	124.976	226.306	9.397
2	75.259	0.3	352.942	0.069	92	3.074	71.179	112.115	11.204
3	59.827	0.302	1.018	-0.009	53	4.111	83.19	138.652	12.276
4	54.982	0.33	336.031	-1.002	40	7.168	96.034	96.034	13.122
5	50.929	0.318	4.709	-0.094	41	3.546	74.782	111.204	12.518
6	49.397	0.307	6.554	0.084	48	2.756	78.104	116.96	11.992
7	43.796	0.253	359.801	0.251	99	1.951	77.811	116.988	9.256
8	34.449	0.235	340.97	-0.123	42	5.085	45.148	75.611	11.802
9	30.623	0.321	351.079	-0.029	20	8.501	44.268	66.634	13.792
10	30.237	0.232	6.156	-0.026	45	3.134	53.775	95.147	10.786
11	28.789	0.293	5.959	0.568	43	4.131	51.605	75.443	10.554
12	24.32	0.317	339.708	1.078	19	10.283	52.201	56.732	12.956
13	23.894	0.274	343.581	0.688	36	4.242	62.194	106.461	10.583
14	23.052	0.312	347.422	1.104	18	10.979	50.394	57.309	13.238
15	22.924	0.292	347.077	-0.187	32	3.93	47.131	78.523	10.811
16	22.121	0.259	5.826	0.192	46	3.777	36.005	65.306	9.781
17	20.143	0.31	345.581	0.603	17	13.993	54.98	71.364	12.427
18	19.328	0.315	337.573	0.68	18	10.365	35.539	48.453	12.229
19	18.484	0.268	345.786	-0.92	26	7.053	39.845	50.391	10.955
20	17.365	0.294	336.605	0.893	21	11.901	24.156	42.673	11.278
21	16.918	0.259	9.098	0.533	30	4.835	48.112	69.719	10.01
22	16.135	0.29	344.517	0.357	20	8.2	51.161	66.819	11.507
23	15.632	0.323	342.799	0.704	13	7.809	45.396	53.84	12.134
24	15.374	0.3	338.308	0.131	18	12.646	38.697	57.745	11.413
25	15.336	0.287	350.167	-1.014	20	7.672	42.938	56.825	11.433
26	14.433	0.302	11.856	0.679	17	7.564	32.612	49.729	11.026
27	12.616	0.298	338.9	-1.077	11	14.114	47.898	47.898	12.767
28	12.294	0.313	10.766	0.049	10	8.875	33.85	49.348	12.517
29	12.009	0.326	358.504	-0.58	12	11.029	26.54	46.212	11.748
30	11.338	0.259	15.307	-0.719	18	11.907	35.758	47.51	11.163
31	10.43	0.243	15.626	-0.848	15	21.887	30.041	54.335	10.623
32	10.419	0.285	14.34	0.849	18	16.294	27.533	40.096	10.094
33	10.063	0.256	354.491	0.776	22	9.457	41.29	51.335	9.616
34	9.454	0.301	348.575	0.191	13	4.96	39.153	54.615	11.118
35	8.573	0.235	7.463	0.754	14	10.082	42.411	55.644	10.124
36	7.844	0.242	337.205	-0.95	13	19.201	38.987	51.832	10.01

(Table continued)

TABLE III. (Continued)

CID	Mass ( $10^{15}M_{\odot}$ )	$z$	R.A. (deg.)	Dec. (deg.)	Richness	$\frac{\bar{r}}{\rho_c}$	$R_{\max}$ (Mpc)	LLS (Mpc)	$\bar{l}$ (Mpc)
37	7.786	0.275	339.645	0.743	11	15.349	21.771	35.626	10.918
38	7.099	0.247	343.733	-1.042	10	22.859	21.615	27.324	11.731
39	6.968	0.254	337.905	0.259	11	17.649	23.835	40.85	10.882
40	6.926	0.283	341.549	0.494	11	10.472	25.987	39.252	9.881
41	6.734	0.264	353.209	-1.041	10	15.8	24.33	34.619	10.722
42	6.578	0.272	347.447	0.055	10	16.784	39.977	51.357	10.245
43	6.554	0.26	344.632	-0.111	13	6.266	29.5	47.385	9.828
44	6.223	0.268	10.69	0.473	14	6.655	27.749	47.466	9.581
45	6.075	0.251	351.795	-1.139	11	14.511	30.705	32.196	10.283
46	5.217	0.262	13.614	0.461	10	22.064	21.594	24.636	9.712
47	4.789	0.276	9.9	-0.494	12	10.204	24.451	34.629	8.939
48	4.751	0.252	339.093	-0.514	10	24.829	21.683	33.723	9.555
49	4.278	0.254	339.177	-0.275	12	7.481	24.232	37.156	8.338
50	4.274	0.268	12.31	-0.158	10	7.884	22.534	37.062	8.97
51	4.271	0.252	353.213	1.174	10	7.958	35.918	39.626	9.821
52	3.823	0.248	3.508	0.794	12	15.294	34.126	41.634	8.599
53	3.239	0.232	358.556	-0.321	11	7.416	25.007	36.919	8.273

TABLE IV. WHL clusters in MITRO-identified galaxy clusters and superclusters.

CID	WHL clusters	$z$	R.A. (deg.)	Dec. (deg.)	CID	WHL clusters	$z$	R.A. (deg.)	Dec. (deg.)
1*	WHL J232229.7 - 002223	0.2728	350.62363	-0.37294		WHL J002648.8 - 001610	0.2311	6.70325	-0.26937
	WHL J232406.1 + 002147	0.278	351.02548	+0.36311		WHL J003130.2 - 010303	0.232	7.87587	-1.05083
	WHL J232517.4 + 001639	0.2767	351.32236	+0.27741					
	WHL J232643.3 + 010803	0.2774	351.68036	+1.13423	11	WHL J002400.3 + 003007	0.2923	6.00129	+0.50202
	WHL J232731.9 + 005634	0.2788	351.88278	+0.94281		WHL J002813.8 - 002154	0.2912	7.05762	-0.36497
	WHL J232809.2 + 001109	0.2768	352.03851	+0.18593					
	WHL J232842.6 + 004939	0.2784	352.17755	+0.82753	12	WHL J223818.0 + 005559	0.3173	339.57483	+0.9331
	WHL J233040.2 + 003634	0.2757	352.66733	+0.6094					
	WHL J233116.7 + 004336	0.2812	352.81964	+0.72678	13	WHL J224807.9 + 002040	0.2657	342.03287	+0.34439
	WHL J233126.4 + 003655	0.2739	352.8602	+0.61525		WHL J225424.4 + 004101	0.2738	343.60159	+0.68368
	WHL J233430.3 - 005906	0.2772	353.62631	-0.98501		WHL J225729.0 + 003102	0.273	344.37103	+0.51717
	WHL J233530.9 - 003152	0.2764	353.87866	-0.53116					
	WHL J233558.5 - 002904	0.27	353.99393	-0.48434	16	WHL J002314.8 + 001158	0.2597	5.81154	+0.19944
	WHL J233739.7 + 001617	0.2772	354.41553	+0.27137		WHL J002358.9 - 003039	0.2575	5.99551	-0.51081
	WHL J233816.0 + 000700	0.2711	354.56683	+0.1168		WHL J002427.4 + 002450	0.2598	6.11418	+0.41389
	WHL J233955.0 - 002558	0.2769	354.97913	-0.43282					
	WHL J234024.5 - 000535	0.2761	355.10202	-0.09299	17	WHL J230408.4 - 010642	0.3147	346.03494	-1.11163
	WHL J234106.9 + 001833	0.2768	355.27872	+0.30925					
	WHL J234144.1 - 004031	0.268	355.43375	-0.67539	19	WHL J230227.7 - 002516	0.2647	345.61533	-0.42125
	WHL J234233.1 - 001719	0.277	355.63776	-0.28874		WHL J230316.7 - 005114	0.2685	345.81976	-0.85396
	WHL J234335.7 + 001951	0.2694	355.89862	+0.33093		WHL J230348.8 - 004024	0.2662	345.95328	-0.67331
	WHL J234403.1 + 001335	0.2714	356.01273	+0.22644					
	WHL J234446.7 - 000520	0.2674	356.19455	-0.0888	20	WHL J222426.4 + 002737	0.2938	336.1102	+0.46019
	WHL J234519.2 - 000312	0.2665	356.3298	-0.05331		WHL J222542.8 + 010339	0.2952	336.42816	+1.06081
	WHL J234548.2 - 010740	0.2797	356.45068	-1.12771		WHL J222627.3 + 005329	0.294	336.61365	+0.89143
	WHL J234604.7 - 001109	0.2665	356.5195	-0.18573					
	WHL J234623.9 + 004458	0.2746	356.59955	+0.74942	21	WHL J003200.8 + 001353	0.2532	8.00329	+0.23131
	WHL J234643.2 + 005003	0.2637	356.67987	+0.83429		WHL J003511.8 + 004348	0.262	8.79937	+0.73002
	WHL J234703.0 - 000508	0.2655	356.76236	-0.08544		WHL J003554.5 + 000924	0.2597	8.97729	+0.15671
	WHL J234727.0 + 003140	0.2729	356.86264	+0.52773		WHL J003555.1 + 002123	0.2592	8.97972	+0.35646

(Table continued)



TABLE IV. (*Continued*)

CID	WHL clusters	$z$	R.A. (deg.)	Dec. (deg.)	CID	WHL clusters	$z$	R.A. (deg.)	Dec. (deg.)
	WHL J234727.6 – 000914	0.2639	356.86499	–0.15381		WHL J003614.8 + 000940	0.2569	9.06173	+0.16113
	WHL J234856.4 – 005328	0.2793	357.2352	–0.8912		WHL J003619.8 + 002418	0.2576	9.08261	+0.40494
	WHL J234946.6 + 003846	0.2705	357.44406	+0.64598		WHL J003629.8 + 003520	0.2607	9.1241	+0.58899
	WHL J235016.5 + 005301	0.2767	357.56891	+0.88373		WHL J003711.8 + 000546	0.2588	9.29924	+0.09603
	WHL J235121.3 + 003701	0.2775	357.83884	+0.61691					
	WHL J235143.9 + 003339	0.2703	357.93292	+0.56078	22	WHL J225211.0 + 003531	0.2877	343.04565	+0.59185
	WHL J235229.5 + 003622	0.2698	358.12292	+0.60623					
	WHL J235301.3 – 001351	0.2779	358.25555	–0.23089	24	WHL J223304.8 + 000818	0.2989	338.2699	+0.13823
	WHL J235347.5 + 004859	0.2787	358.448	+0.81641					
	WHL J235424.8 + 004838	0.2805	358.60349	+0.81066	25	WHL J231948.3 – 010932	0.2909	349.95108	–1.15889
	WHL J235508.9 + 004512	0.2789	358.7869	+0.75341					
	WHL J235543.8 – 000014	0.2752	358.93271	–0.00393	28	WHL J004051.3 + 000858	0.3132	10.2137	+0.14932
2	WHL J232420.2 – 003229	0.293	351.08423	–0.54147	29	WHL J235424.3 – 001451	0.3277	358.60135	–0.24737
	WHL J232902.7 – 003847	0.2988	352.26129	–0.6463					
	WHL J232923.7 – 004856	0.3007	352.34872	–0.81546	30	WHL J010154.7 – 005216	0.2616	15.47781	–0.87118
	WHL J233317.2 + 000741	0.3025	353.32187	+0.12797		WHL J010214.5 – 010851	0.2611	15.56023	–1.14737
	WHL J233531.2 – 005320	0.2974	353.87994	–0.88882		WHL J010331.1 – 005932	0.2565	15.87962	–0.99216
						WHL J010339.2 – 004508	0.2646	15.91314	–0.75235
3	WHL J000201.8 + 002049	0.3007	0.50762	+0.34701					
	WHL J000224.7 – 003253	0.2895	0.60304	–0.54798	31	WHL J010229.8 – 004848	0.2375	15.62412	–0.81343
	WHL J000242.3 – 001320	0.2989	0.67632	–0.22217					
	WHL J000455.1 + 003415	0.3197	1.22949	+0.57073	32	WHL J005338.1 + 005515	0.2851	13.40886	+0.92088
	WHL J000825.6 + 004152	0.3124	2.10683	+0.69786		WHL J005434.9 + 004819	0.2847	13.6454	+0.80528
						WHL J005526.0 + 004731	0.2846	13.85833	+0.79205
4	WHL J222456.1 – 002115	0.3154	336.23373	–0.35415					
	WHL J222523.4 – 005410	0.3204	336.3476	–0.90289	33	WHL J234145.8 + 010728	0.2507	355.44077	+1.12444
	WHL J222731.7 – 003945	0.3226	336.88223	–0.66258					
	WHL J222810.8 – 004848	0.327	337.04517	–0.8134	34	WHL J231308.1 – 010632	0.3007	348.28375	–1.10895
5	WHL J001848.6 – 000430	0.3184	4.70257	–0.07488	35	WHL J003407.0 – 001147	0.2365	8.52936	–0.19632
	WHL J002020.7 – 001925	0.3251	5.08638	–0.32355					
	WHL J002025.3 – 001216	0.3267	5.10561	–0.20444	36	WHL J222652.2 – 005607	0.2483	336.71762	–0.93536
	WHL J002105.0 + 001406	0.3291	5.27082	+0.23505		WHL J222720.8 – 005610	0.2468	336.83655	–0.93623
6	WHL J002200.7 – 001839	0.304	5.50288	–0.31075	39	WHL J223126.2 + 001743	0.2547	337.85898	+0.29532
	WHL J002334.8 + 001907	0.3084	5.89512	+0.31861					
	WHL J002557.8 + 000734	0.3076	6.49081	+0.12619	40	WHL J224641.8 + 002837	0.2846	341.67407	+0.47684
	WHL J002920.4 – 004032	0.3085	7.3349	–0.67557					
	WHL J003401.5 – 003314	0.3076	8.50633	–0.55392	41	WHL J233608.8 – 011100	0.265	354.03653	–1.18347
	WHL J003508.9 – 010803	0.3055	8.78714	–1.13429					
	WHL J003542.0 + 000158	0.3102	8.92503	+0.03278	43	WHL J225830.5 – 000608	0.2587	344.62704	–0.1021
	WHL J003744.4 – 011424	0.3106	9.43509	–1.23996					
					44	WHL J004249.9 + 001255	0.2691	10.70773	+0.21535
7	WHL J235751.3 + 002133	0.2596	359.46381	+0.3593		WHL J004252.7 + 004306	0.2699	10.71962	+0.71844
	WHL J235914.3 + 001405	0.2533	359.8096	+0.23484					
	WHL J235952.8 + 004155	0.2674	359.96988	+0.69856	45	WHL J232708.4 – 011250	0.2517	351.7851	–1.21393
	WHL J000024.3 + 003804	0.2522	0.1013	+0.63434					
	WHL J000050.5 + 004705	0.2634	0.21052	+0.78478	48	WHL J223556.1 – 001411	0.2488	338.98395	–0.2364
	WHL J000117.8 – 005612	0.2473	0.32417	–0.93668		WHL J223622.8 – 003049	0.2488	339.09488	–0.51352
	WHL J000126.3 – 000143	0.2479	0.35969	–0.02867					
8	WHL J224249.1 – 001648	0.2363	340.70477	–0.2801	49	WHL J223629.1 + 000118	0.2539	339.12143	+0.02159
10	WHL J001637.4 – 002739	0.2325	4.15592	–0.46084	52	WHL J001229.8 + 003911	0.2505	3.12436	+0.65318
	WHL J002435.2 – 001307	0.2308	6.14679	–0.21867		WHL J001357.8 + 004706	0.2465	3.49083	+0.78508

- [1] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly *et al.*, Simulations of the formation, evolution and clustering of galaxies and quasars, *Nature (London)* **435**, 629 (2005).
- [2] M. Boylan-Kolchin, V. Springel, S. D. M. White, A. Jenkins, and G. Lemson, Resolving cosmic structure formation with the millennium-II simulation, *Mon. Not. R. Astron. Soc.* **398**, 1150 (2009).
- [3] A. A. Klypin, S. Trujillo-Gomez, and J. Primack, Dark matter halos in the standard cosmological model: Results from the Bolshoi simulation, *Astrophys. J.* **740**, 102 (2011).
- [4] A. M. C. Le Brun, I. G. McCarthy, J. Schaye, and T. J. Ponman, Towards a realistic population of simulated galaxy groups and clusters, *Mon. Not. R. Astron. Soc.* **441**, 1270 (2014).
- [5] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist, M. Vogelsberger, S. Genel, P. Torrey, F. Marinacci, and J. Naiman, First results from the IllustrisTNG simulations: Matter and galaxy clustering, *Mon. Not. R. Astron. Soc.* **475**, 676 (2018).
- [6] C. Stoughton, R. H. Lupton, M. Bernardi, M. R. Blanton, S. Burles, F. J. Castander, A. J. Connolly, D. J. Eisenstein, J. A. Frieman, G. S. Hennessy *et al.*, Sloan digital sky survey: Early data release, *Astron. J.* **123**, 485 (2002).
- [7] R. Ahumada, C. Allende Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, R. Arcodia, E. Armengaud, M. Aubert *et al.*, The 16th data release of the Sloan digital sky surveys: First release from the APOGEE-2 southern survey and full release of eBOSS spectra, *Astrophys. J. Suppl. Ser.* **249**, 3 (2020).
- [8] K. Kuijken, C. Heymans, A. Dvornik, H. Hildebrandt, J. T. A. de Jong, A. H. Wright, T. Erben, M. Bilicki, B. Giblin, H. Y. Shan *et al.*, The fourth data release of the Kilo-Degree Survey: UGRI imaging and nine-band optical-IR photometry over 1000 square degrees, *Astron. Astrophys.* **625**, A2 (2019).
- [9] R. Laureijs, J. Amiaux, S. Arduini, J. L. Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet *et al.*, Euclid definition study report, [arXiv:1110.3193](https://arxiv.org/abs/1110.3193).
- [10] Euclid Collaboration, A. Blanchard, S. Camera, C. Carbone, V. F. Cardone, S. Casas, S. Clesse, S. Ilić, M. Kilbinger, T. Kitching, M. Kunz *et al.*, Euclid preparation. VII. Forecast validation for Euclid cosmological probes, *Astron. Astrophys.* **642**, A191 (2020).
- [11] Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew *et al.*, LSST: From science drivers to reference design and anticipated data products, *Astrophys. J.* **873**, 111 (2019).
- [12] LSST Dark Energy Science Collaboration, Large synoptic survey telescope: Dark Energy Science Collaboration, [arXiv:1211.0310](https://arxiv.org/abs/1211.0310).
- [13] W. H. Press and P. Schechter, Formation of galaxies and clusters of galaxies by self-similar gravitational condensation, *Astrophys. J.* **187**, 425 (1974).
- [14] A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole, A. E. Evrard, H. M. P. Couchman, and N. Yoshida, The mass function of dark matter haloes, *Mon. Not. R. Astron. Soc.* **321**, 372 (2001).
- [15] J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz, Toward a halo mass function for precision cosmology: The limits of universality, *Astrophys. J.* **688**, 709 (2008).
- [16] H. Böhringer, P. Schuecker, L. Guzzo, and C. A. Collins, Constraining cosmological parameters with observations of galaxy clusters, in *Astronomy, Cosmology and Fundamental Physics*, edited by P. A. Shaver, L. Dilella, and A. Giménez (Springer-Verlag, 2003), p. 35.
- [17] S. W. Allen, A. E. Evrard, and A. B. Mantz, Cosmological parameters from observations of galaxy clusters, *Astron. Astrophys.* **49**, 409 (2011).
- [18] J. L. Tinker, E. S. Sheldon, R. H. Wechsler, M. R. Becker, E. Rozo, Y. Zu, D. H. Weinberg, I. Zehavi, M. R. Blanton, M. T. Busha, and B. P. Koester, Cosmological constraints from galaxy clustering and the mass-to-number ratio of galaxy clusters, *Astrophys. J.* **745**, 16 (2012).
- [19] L. Salvati, M. Douspis, and N. Aghanim, Impact of systematics on cosmological parameters from future galaxy cluster surveys, *Astron. Astrophys.* **643**, A20 (2020).
- [20] A. Klypin, V. Poulin, F. Prada, J. Primack, M. Kamionkowski, V. Avila-Reese, A. Rodriguez-Puebla, P. Behroozi, D. Hellinger, and T. L. Smith, Clustering and halo abundances in early dark energy cosmological models, *Mon. Not. R. Astron. Soc.* **504**, 769 (2021).
- [21] J. A. Newman, C. Marinoni, A. L. Coil, and M. Davis, Measuring the cosmic equation of state with galaxy clusters in the DEEP2 redshift survey, *Publ. Astron. Soc. Pac.* **114**, 29 (2002).
- [22] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, The evolution of large-scale structure in a universe dominated by cold dark matter, *Astrophys. J.* **292**, 371 (1985).
- [23] E. L. Turner and J. R. III. Gott, Groups of galaxies. I. A catalog., *Astrophys. J. Suppl. Ser.* **32**, 409 (1976).
- [24] J. P. Huchra and M. J. Geller, Groups of galaxies. I. Nearby groups, *Astrophys. J.* **257**, 423 (1982).
- [25] I. B. Zeldovich, J. Einasto, and S. F. Shandarin, Giant voids in the universe, *Nature (London)* **300**, 407 (1982).
- [26] A. Knebe, S. R. Knollmann, S. I. Muldrew, F. R. Pearce, M. A. Aragon-Calvo, Y. Ascasibar, P. S. Behroozi, D. Ceverino, S. Colombi, J. Diemand *et al.*, Haloes gone MAD: The halo-finder comparison project, *Mon. Not. R. Astron. Soc.* **415**, 2293 (2011).
- [27] J. Onions, A. Knebe, F. R. Pearce, S. I. Muldrew, H. Lux, S. R. Knollmann, Y. Ascasibar, P. Behroozi, P. Elahi, J. Han, M. Maciejewski, M. E. Merchán, M. Neyrinck, A. N. Ruiz, M. A. Sgró, V. Springel, and D. Tweed, Subhaloes going Notts: the subhalo-finder comparison project, *Mon. Not. R. Astron. Soc.* **423**, 1200 (2012).
- [28] A. Knebe, F. R. Pearce, H. Lux, Y. Ascasibar, P. Behroozi, J. Casado, C. C. Moran, J. Diemand, K. Dolag, R. Dominguez-Tenreiro *et al.*, Structure finding in cosmological simulations: The state of affairs, *Mon. Not. R. Astron. Soc.* **435**, 1618 (2013).

- [29] V. R. Eke, C. M. Baugh, S. Cole, C. S. Frenk, P. Norberg, J. A. Peacock, I. K. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon *et al.*, Galaxy groups in the 2dFGRS: The group-finding algorithm and the 2PIGG catalogue, *Mon. Not. R. Astron. Soc.* **348**, 866 (2004).
- [30] X. Yang, H. J. Mo, F. C. van den Bosch, and Y. P. Jing, A halo-based galaxy group finder: Calibration and application to the 2dFGRS, *Mon. Not. R. Astron. Soc.* **356**, 1293 (2005).
- [31] E. Tago, E. Saar, E. Tempel, J. Einasto, M. Einasto, P. Nurmi, and P. Heinämäki, Groups of galaxies in the SDSS data release 7. Flux- and volume-limited samples, *Astron. Astrophys.* **514**, A102 (2010).
- [32] B. F. Gerke, J. A. Newman, M. Davis, A. L. Coil, M. C. Cooper, A. A. Dutton, S. M. Faber, P. Guhathakurta, N. Konidaris, D. C. Koo *et al.*, The DEEP2 galaxy redshift survey: The Voronoi-Delaunay method catalog of galaxy groups, *Astrophys. J.* **751**, 50 (2012).
- [33] S. Pereira, L. E. Campusano, N. Hitschfeld-Kahler, D. Pizarro, C. P. Haines, R. G. Clowes, G. Marinello, and I. K. Söchting, A 3D Voronoi + Gapper galaxy cluster finder in redshift space to  $z \approx 0.2$  I: An algorithm optimized for the 2dFGRS, *Astrophys. J.* **838**, 109 (2017).
- [34] F. Rodriguez and M. Merchán, Combining friend-of-friend and halo-based algorithms for the identification of galaxy groups, *Astron. Astrophys.* **636**, A61 (2020).
- [35] M. S. Warren, K. Abazajian, D. E. Holz, and L. Teodoro, Precision determination of the mass function of dark matter halos, *Astrophys. J.* **646**, 881 (2006).
- [36] S. More, A. V. Kravtsov, N. Dalal, and S. Gottlöber, The overdensity and masses of the friends-of-friends halos and universality of halo mass function, *Astrophys. J. Suppl. Ser.* **195**, 4 (2011).
- [37] C. S. Frenk, S. D. M. White, M. Davis, and G. Efstathiou, The formation of dark halos in a universe dominated by cold dark matter, *Astrophys. J.* **327**, 507 (1988).
- [38] C. Lacey and S. Cole, Merger rates in hierarchical models of galaxy formation—part two—comparison with N-body simulations, *Mon. Not. R. Astron. Soc.* **271**, 676 (1994).
- [39] E. Audit, R. Teyssier, and J.-M. Alimi, Non-linear dynamics and mass function of cosmic structures. II. Numerical results, *Astron. Astrophys.* **333**, 779 (1998).
- [40] E. Tempel, A. Tamm, M. Gramann, T. Tuvikene, L. J. Liivamägi, I. Suhhonenko, R. Kipper, M. Einasto, and E. Saar, Flux- and volume-limited groups/clusters for the SDSS galaxies: Catalogues and mass estimation, *Astron. Astrophys.* **566**, A1 (2014).
- [41] J. Onions, Y. Ascasibar, P. Behroozi, J. Casado, P. Elahi, J. Han, A. Knebe, H. Lux, M. E. Merchán, S. I. Muldrew *et al.*, Subhaloes gone Notts: Spin across subhaloes and finders, *Mon. Not. R. Astron. Soc.* **429**, 2739 (2013).
- [42] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, The quick-hull algorithm for convex hulls, *ACM Trans. Math. Softw.* **22**, 469 (1996).
- [43] S. Paul, R. S. John, P. Gupta, and H. Kumar, Understanding “galaxy groups” as a unique structure in the universe, *Mon. Not. R. Astron. Soc.* **471**, 2 (2017).
- [44] G. L. Bryan, M. L. Norman, B. W. O’Shea, T. Abel, J. H. Wise, M. J. Turk, D. R. Reynolds, D. C. Collins, P. Wang, S. W. Skillman *et al.* (Enzo Collaboration), ENZO: An adaptive mesh refinement code for astrophysics, *Astrophys. J. Suppl. Ser.* **211**, 19 (2014).
- [45] E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page *et al.*, Five-year Wilkinson microwave anisotropy probe observations: Cosmological interpretation, *Astrophys. J. Suppl. Ser.* **180**, 330 (2009).
- [46] D. J. Eisenstein and W. Hu, Power spectra for cold dark matter and its variants, *Astrophys. J.* **511**, 5 (1999).
- [47] L. Gao and S. D. M. White, Asymmetries in the inner regions of  $\Lambda$ CDM haloes, *Mon. Not. R. Astron. Soc.* **373**, 65 (2006).
- [48] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles* (Hilger, Bristol, 1988).
- [49] M. Davis, J. Huchra, D. W. Latham, and J. Tonry, A survey of galaxy redshifts. II. The large scale space distribution., *Astrophys. J.* **253**, 423 (1982).
- [50] M. J. Geller and J. P. Huchra, Mapping the universe, *Science* **246**, 897 (1989).
- [51] J. E. Gunn, W. A. Siegmund, E. J. Mannery, R. E. Owen, C. L. Hull, R. F. Leger, L. N. Carey, G. R. Knapp, D. G. York, W. N. Boroski *et al.*, The 2.5 m telescope of the Sloan Digital Sky Survey, *Astron. J.* **131**, 2332 (2006).
- [52] Abdurro’uf, K. Accetta, C. Aerts, V. Silva Aguirre, R. Ahumada, N. Ajaonkar, N. Filiz Ak, S. Alam, C. Allende Prieto, A. Almeida *et al.*, The seventeenth data release of the Sloan Digital Sky Surveys: Complete release of MaNGA, MaStar, and APOGEE-2 data, *Astrophys. J. Suppl. Ser.* **259**, 35 (2022).
- [53] S. Alam, F. D. Albareti, C. Allende Prieto, F. Anders, S. F. Anderson, T. Anderton, B. H. Andrews, E. Armengaud, É. Aubourg, S. Bailey *et al.*, The eleventh and twelfth data releases of the Sloan Digital Sky Survey: Final data from SDSS-III, *Astrophys. J. Suppl. Ser.* **219**, 12 (2015).
- [54] J. Bagchi, S. Sankhyayan, P. Sarkar, S. Raychaudhury, J. Jacob, and P. Dabhade, Saraswati: An extremely massive  $\sim 200$  megaparsec scale supercluster, *Astrophys. J.* **844**, 25 (2017).
- [55] M. R. Blanton and S. Roweis, K-corrections and filter transformations in the ultraviolet, optical, and near-infrared, *Astron. J.* **133**, 734 (2007).
- [56] E. F. Bell, D. H. McIntosh, N. Katz, and M. D. Weinberg, The optical and near-infrared properties of galaxies. I. Luminosity and stellar mass functions, *Astrophys. J. Suppl. Ser.* **149**, 289 (2003).
- [57] A. Vale and J. P. Ostriker, Linking halo mass to galaxy luminosity, *Mon. Not. R. Astron. Soc.* **353**, 189 (2004).
- [58] Z. Zheng, A. L. Coil, and I. Zehavi, Galaxy evolution from halo occupation distribution modeling of DEEP2 and SDSS galaxy clustering, *Astrophys. J.* **667**, 760 (2007).
- [59] X. Yang, H. J. Mo, F. C. van den Bosch, Y. Zhang, and J. Han, Evolution of the galaxy-dark matter connection and the assembly of galaxies in dark matter halos, *Astrophys. J.* **752**, 41 (2012).
- [60] A. Rodríguez-Puebla, J. R. Primack, V. Avila-Reese, and S. M. Faber, Constraining the galaxy-halo connection over the last 13.3 Gyr: Star formation histories, galaxy mergers



- and structural properties, *Mon. Not. R. Astron. Soc.* **470**, 651 (2017).
- [61] G. Girelli, L. Pozzetti, M. Bolzonella, C. Giocoli, F. Marulli, and M. Baldi, The stellar-to-halo mass relation over the past 12 Gyr. I. Standard  $\Lambda$ CDM model, *Astron. Astrophys.* **634**, A135 (2020).
- [62] P. Kroupa, On the variation of the initial mass function, *Mon. Not. R. Astron. Soc.* **322**, 231 (2001).
- [63] A. Borch, K. Meisenheimer, E. F. Bell, H. W. Rix, C. Wolf, S. Dye, M. Kleinheinrich, Z. Kovacs, and L. Wisotzki, The stellar masses of 25 000 galaxies at  $0.2 \leq z \leq 1.0$  estimated by the COMBO-17 survey, *Astron. Astrophys.* **453**, 869 (2006).
- [64] X. Yang, H. J. Mo, F. C. van den Bosch, A. Pasquali, C. Li, and M. Barden, Galaxy groups in the SDSS DR4. I. The catalog and basic properties, *Astrophys. J.* **671**, 153 (2007).
- [65] R. C. Tolman, On the estimation of distances in a curved universe with a non-static line element, *Proc. Natl. Acad. Sci. U.S.A.* **16**, 511 (1930).
- [66] S. Bharadwaj, A. K. Gupta, and T. R. Seshadri, Nature of clustering in the Las Campanas redshift survey, *Astron. Astrophys.* **351**, 405 (1999).
- [67] P. Ntelis, J.-C. Hamilton, J.-M. Le Goff, E. Burtin, P. Laurent, J. Rich, N. Guillermo Busca, J. Tinker, E. Aubourg, H. du Mas des Bourboux *et al.*, Exploring cosmic homogeneity with the BOSS DR12 galaxy sample, *J. Cosmol. Astropart. Phys.* **06** (2019) 019.
- [68] R. Dünner, P. A. Araya, A. Meza, and A. Reisenegger, The limits of bound structures in the accelerating Universe, *Mon. Not. R. Astron. Soc.* **366**, 803 (2006).
- [69] J. C. Jackson, A critique of Rees's theory of primordial gravitational radiation, *Mon. Not. R. Astron. Soc.* **156**, 1P (1972).
- [70] N. Kaiser, Clustering in real space and in redshift space, *Mon. Not. R. Astron. Soc.* **227**, 1 (1987).
- [71] P. McDonald, Gravitational redshift and other redshift-space distortions of the imaginary part of the power spectrum, *J. Cosmol. Astropart. Phys.* **11** (2009) 026.
- [72] R. K. Sachs and A. M. Wolfe, Perturbations of a cosmological model and angular variations of the microwave background, *Astrophys. J.* **147**, 73 (1967).
- [73] C. Alcock and B. Paczynski, An evolution free test for non-zero cosmological constant, *Nature (London)* **281**, 358 (1979).
- [74] H.-M. Zhu, Y. Yu, U.-L. Pen, X. Chen, and H.-R. Yu, Nonlinear reconstruction, *Phys. Rev. D* **96**, 123502 (2017).
- [75] Y. Yu and H.-M. Zhu, Nonlinear reconstruction of the velocity field, *Astrophys. J.* **887**, 265 (2019).
- [76] Y. Wang, B. Li, and M. Cautun, Iterative removal of redshift-space distortions from galaxy clustering, *Mon. Not. R. Astron. Soc.* **497**, 3451 (2020).
- [77] E. Paillas, Y.-C. Cai, N. Padilla, and A. G. Sánchez, Redshift-space distortions with split densities, *Mon. Not. R. Astron. Soc.* **505**, 5731 (2021).
- [78] G. L. Bryan and M. L. Norman, Statistical properties of x-ray clusters: Analytic and numerical comparisons, *Astrophys. J.* **495**, 80 (1998).
- [79] G. Chon, H. Böhringer, and S. Zaroubi, On the definition of superclusters, *Astron. Astrophys.* **575**, L14 (2015).
- [80] Z. L. Wen, J. L. Han, and F. S. Liu, Galaxy clusters identified from the SDSS DR6 and their properties, *Astrophys. J. Suppl. Ser.* **183**, 197 (2009).
- [81] Z. L. Wen, J. L. Han, and F. S. Liu, A catalog of 132,684 clusters of galaxies identified from Sloan Digital Sky Survey III, *Astrophys. J. Suppl. Ser.* **199**, 34 (2012).
- [82] E. Jones, T. Oliphant, and P. Peterson, *SciPy: Open Source Scientific Tools for Python* (IEEE, 2001).
- [83] S. van der Walt, S. C. Colbert, and G. Varoquaux, The NumPy array: A structure for efficient numerical computation, *IEEE Pervasive Computing* **13**, 22 (2011).
- [84] J. D. Hunter, Matplotlib: A 2D Graphics Environment, *IEEE Pervasive Computing* **9**, 90 (2007).
- [85] S. A. Cora, C. A. Vega-Martínez, T. Hough, A. N. Ruiz, Á. A. Orsi, A. M. Muñoz Arancibia, I. D. Gargiulo, F. Collacchioni, N. D. Padilla, S. Gottlöber, and G. Yepes, Semi-analytic galaxies—I. Synthesis of environmental and star-forming regulation mechanisms, *Mon. Not. R. Astron. Soc.* **479**, 2 (2018).
- [86] <http://lca.ucsd.edu>.
- [87] <http://www.sdss3.org/>.
- [88] [www.gauss-centre.eu](http://www.gauss-centre.eu).
- [89] [www.prace-ri.eu](http://www.prace-ri.eu).
- [90] [www.lrz.de](http://www.lrz.de).