Case study

A Fast and Simple Algorithm for Detecting Large Scale Structures

Giovanni C. Baiesi Pillastrini[1*]

[1] *Sezione di Ricerca Spettroscopia - U.A.I. C/O I.A.S.F. Via del Fosso del Cavaliere 100, 00133 Roma, Italy*

ABSTRACT

Aims: We propose a gravitational potential method (GPM) as a galaxy structure finder based on the analysis of the local gravitational potential distribution derived from a fast and simple algorithm applied to the spatial density distribution of galaxy systems.
Methodology: the GPM is performed in two steps exploratory data analysis: first, we measure the comoving local gravitational potential generated by neighboring mass tracers at the position of a test point-like mass tracer. The computation extended to each mass tracer of the complete sample provides a detailed map of the negative potential fluctuations. The negative gravitational potential is directly dependent from mass density i.e., deeper are the potential fluctuations in a certain region of space and denser are the mass tracers in that region. Therefore, from a smoothed potential distribution, the deepest potential well detects unambiguously an overdensity in the mass tracer distribution. Second, using a density contrast criterion we extrapolate from the discovered overdensity, if any, the “bound core”.
Results: applying the GPM to a complete volume-limited sample of galaxy clusters, a huge concentration of galaxy clusters has been identified, but only 35 of them seem to form a massive and bound core enclosed in a spherical volume of 51 *Mpc* radius, centered at Galactic coordinates *l* ~ 63°.7, *b* ~ 63°.7 and at redshift ~ .36. It has a velocity dispersion of 1,183 *Km/s* with an estimated virial mass of $2.67 \pm .80 \times 10^{16}$ $M_{\odot}$.
Conclusions: the good agreement of our findings compared with those obtained using a different methodology, confirms that the GPM proposed as a cluster finder offers a straightforward and powerful as well as fast way to identify clustered structures from large datasets.

Keywords: methods: data analysis - galaxies: clusters - large scale structures of the Universe

* *permanent address: via Pizzardi, 13 - 40138 Bologna - Italy -*

1. INTRODUCTION

Galaxy redshift surveys show that galaxies are spread out in a fairly complicated way, over the so-called Cosmic Web. This network consists of the largest non–linear superstructures in the Universe which are interconnected through filaments and sheets as expected from the properties of the ΛCDM concordance model. However, is quite common to find giant superstructures at the nodes of the Cosmic Web as, for example, the well-known Shapley supercluster. In the past several decades, to search for such a superstructures, numerous clustering algorithms based on various theories have been proposed. Here, we present a brief survey on the widespread clustering algorithms applied to large data set provided by redshift surveys [1,2,3]. Traditionally galaxy systems of various scale have been selected from the cosmic web using quantitative methods, such as the Friends-of-Friends method (FoF) or the Density Field method. The FoF method is very common and suitable in searching systems of particles in numerical simulations where all particles have identical masses in volume-limited samples. Neighboring galaxies or clusters are searched using a fixed or variable linking lengths. Because galaxy systems contain galaxies of very different luminosity, the FoF method has the disadvantage that objects of different luminosity or mass are treated identically making difficult a clear distinction between poor and rich galaxy systems if their number density of galaxies is similar. The density field method overcome these difficulties because luminosities of galaxies are taken into account, both in the search of galaxy systems, as well as in the determination of their properties. A variant of the density field method is obtained by using cell sizes equal to the defined smoothing radii [4]. Another variant of the density smoothing is the use of the Wiener Filtering technique where data are covered by a grid whose cells grow in size with increasing distance from the observer as recently applied to the 2dFGRS to identify superclusters and voids by [5]. Adding important refinements to that variant as the use constant cell size and constant smoothing radius over the whole sample and an appropriate kernel, cell size and smoothing length, [6] identified superclusters in the 2dFGRS. The accomplishment of wide–area surveys of galaxies with spectroscopic follow up, such as Sloan Digital Sky Survey (SDSS) [7], allowed for the identification of superclusters directly from the large-scale galaxy distribution. Recently, from the SDSS–DR7 main and LRG samples [8], the largest catalogue of superclusters (SCLCAT hereafter) has been constructed by [9] using the density field method. In such catalogs, superclusters are operationally defined as objects within a region of positive galaxy density contrast and thus are subject to a certain degree of arbitrariness in the parameter selection [10]. For example, the method recently developed by [9] detects superstructures within the luminosity density fields of the samples identifying overdensities either above local adaptive or fixed thresholds to separate superclusters from poorer galaxy systems. It has the same meaning as the linking length in the FoF method. This is the key parameter which makes a clear distinction between rich and poor galaxy systems. Our goal is to find superstructures, at all distances from the observer until a certain limiting distance. To achieve this goal the selection procedure must be the same for all distances from the observer. In this paper, we consider a method capable of detecting large structures in a single data set. We observe that galaxy systems aggregate by following basic laws of gravity no matter how different they are. Inspired by this simple physics of clustering phenomena, we propose a new clustering detector to make use of universality of gravitational clustering behaviors. The basic idea is to regard data objects as particles with well-known mass and position at the present time. We follow an approach recently used to study star-forming gas cores in an SPH simulation of giant molecular clouds [11,12]. They define core-finding methods using the *deepest potential wells* to identify core boundaries. Using the prescriptions of the exploratory data analysis [13], this idea can be easily adapted to detect over–dense regions studying the distribution of local gravitational potentials defined by the spatial distribution of galaxy systems at intermediate redshifts. The close connection of the gravitational potential with the matter density field was studied by [14] which described the evolution of the large scale density perturbations using the characteristics of the potential field. They confirmed that the relation of the density perturbations to the potential established by the theory of gravitational instability is related to the formation of huge scale structures seen in the galaxy distribution that is, over–dense regions arise due to a slow matter flow into the negative potential regions. In other words, the detection of large scale structures can be carried out simply observing the regions where very deep gravitational potentials originate. Following this idea we address our study toward the clustering of galaxy clusters using a complete volume-limited cluster sample extracted from the GMBCG cluster catalog [15]. A two-step analysis is performed as follows: first, by an explorative investigation of the distribution of the local gravitational potentials measured at the position of each sampled cluster taken one at a time as a test-particle. Then, assuming that the deepest potential represents the *center* of a cluster concentration, we apply a clustering algorithm based on the “density contrast criterion” [16] in order to identify its cluster memberships. The paper is organized as follows: in Sect.2 we introduce the gravitational potential-based method. In Sect.3 we demonstrate how the method can be easily applied to a complete cluster sample studying in detail the region where the deepest gravitational potential is measured and describe the assumed criterion to identify the bound core of the cluster concentration which is compared with the findings of the SCLCAT. In Sect.4 we discuss various issues addressed by our study and the conclusions are drawn.

2. THE GRAVITATIONAL POTENTIAL-BASED METHOD (GPM)

The very slow evolution of the spatial distribution of large scale structures allows to investigate the distribution of potential fields to explain properties of the matter distribution at the present time. Here, we perform an elementary numerical simulation to demonstrate the close links of the spatial distribution of matter density with the potential distribution. To simplify the demonstration, we adopt a simple toy model assuming: *i*) the mass distribution is represented by point like masses placed in a Bravais lattice with periodic boundary conditions where the unit cell is a cube (all sides of the same length and all face perpendicular to each other) with a point like mass at each corner. The unit cell completely describes the structure of the space, which can be regarded as a finite repetition of the unit cell. We restrict the present simulation to a cubic lattice with a side length of 10*l* and a side length of the unit cell equal *l* (corresponding to 1,000 corners/points); *ii*) each point lies at positions (*x, y, z*) in the Cartesian three-space (where *x, y* and *z* are integer multiples of *l*); *iii*) all points have the same mass *m* so that, within the cubic lattice the mass distribution is perfectly uniform. Now, it is well-known that gravity is a *superposable* force which implies that the gravitational force exerted on some test mass by a collection of point masses is simply the sum of the forces exerted on the test mass by each point mass taken in isolation. It follows that the gravitational potential generated by a collection of point masses at a certain location in space is the sum of the potentials generated at that location by each point mass taken in isolation. Hence, if within a spherical volume V_j, centered on a generic test-particle *j* at position vector d_j from the observer, with a fixed radius R_V, there are N_{V_j} point mass m_i $(i = 1,....,N_{V_j})$, located at position vectors d_i (from the observer), then the gravitational potential generated at position vector d_j is given by the well-known equation $\Phi_j = -G \sum_{i=1, i\neq j, i\in V_j}^{N_{V_j}} m_i \left(d_i - d_j\right)^{-1}$ where *G* is the gravitational constant. Repeating the calculation for each point mass at *jth* position of the cubic lattice, we provide the whole Φ_j distribution. The value of potential is always negative, denoting that the force between particles is attractive. According to equation of the potential, the attractive interaction between objects decreases with distance and becomes 0 when the distance is greater than R_V assumed large enough so that masses outside V_j should have little influence on the potential determination (we will discuss this assumption later). If we assume for simplicity $R_V = l$, for the geometric properties of our Cubic lattice system, each point mass has $N_{V_j} = 6$ nearest neighbors at a distance *l* but external points i.e., the points lying on the external faces, edges and corners of the cubic lattice (wherein they have 5, 4 and 3 nearest neighbors, respectively). Note that we will not measure Φ_j at the positions of such external points because of evident incompleteness effect. Under these assumptions, the calculation of Φ_j will turn out constant at each position *j* i.e., the Φ_j distribution is perfectly flat as expected from a uniform mass distribution and no potential wells will appear. We consider now a perturbation on the perfect lattice. If we create an artificial clump of point masses modifying the number density within the spherical volume of radius *l* centered on the central point P(5*l*, 5*l*, 5*l*), for example adding 6 new point masses at *l*/2 around P and remaking the calculation, we will show that this clump influences locally the potential distribution. As expected, the contour plot of the Φ_j distribution of Fig. 1 shows the potential well centered on P.

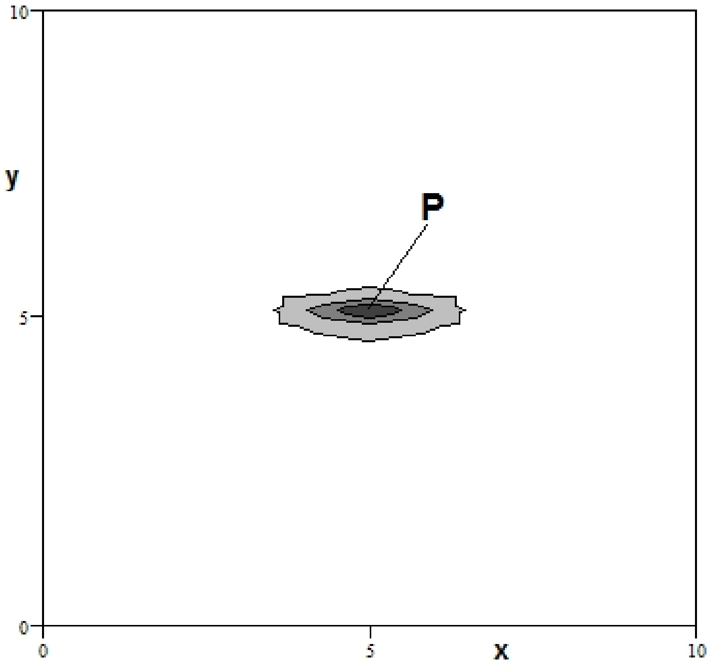


Fig.1 – Contour plot of the potential distribution in the (x,y)-projection. The potential well centered on P (5*l*,5*l*,5*l*) is shown. It shows the most negative Φ_j provided by the artificial clump obtained adding 6 new point mass at *l*/2 around P.

The astrophysical meaning of such elementary simulation can be summarized in a straightforward concept: denser is a clump of celestial bodies (assumed as mass points) embedded in homogeneous and isotropic background and deeper is the corresponding gravitational potential well. *The deepest of the potential wells identifies unambiguously the densest clump of a mass density distribution.* Obviously, to apply this potential method to real data space, each data object should be regarded as a particle with the proper mass. The proposed method provides relevant advantages: *i*) it enables the identification of clustered structures based on dynamical principles by knowing positions in space and individual masses of a complete volume-limited sample of celestial bodies; *ii*) the gravitational potential distribution is smoother than the density distribution since the contribution to local potential fields due to small density fluctuations is irrelevant and *iii*) it does not require any threshold to be set overcoming the problem of defining arbitrary density thresholds in the clustering analysis (see Sect. 5). However, some misunderstandings may arise from very close objects which can measure very deep potential wells due to the inverse dependency of Φ_j on the spatial separation. Since we are not interested in detecting small and isolated structures (cluster pairs or triplets, to overcome this problem we calculate the *mean potential field* $\langle\Phi\rangle_j$ at the position *j* and obtained averaging all Φ located within V_j. Undesirable outliers are flattened providing a smoothed evaluation of the potential field and a more reliable signal in detecting huge structures.

3. APPLICATION

3.1. The data

In gravitational studies, the use of cluster samples overcome some of the problems faced by galaxy samples since clusters are luminous enough for samples to be volume-limited out to large distances and trace the peaks of the density fluctuation field although the spatial distribution only sparsely sample the underlying density field. Besides, cluster mass can be quickly estimated offering a fundamental advantage which have a direct effect on the gravitational The astrophysical meaning of this elementary simulation can be summarized in a straightforward concept: denser is a clump of celestial bodies (assumed as mass points) embedded in homogeneous and isotropic background and deeper is the corresponding gravitational potential well. The deepest of the potential wells identifies the densest clump of the mass density distribution. Herefore, a complete volume-limited galaxy cluster sample is extracted from the GMBCG optical cluster catalog [15] derived from the SDSS DR7 survey data [8] in redshift space, using coordinates of the brightest cluster galaxy (BCG) as the origin. As mentioned in their work, the Authors have developed an efficient cluster finding algorithm named GMBCG method to identify the BCG plus red sequence galaxies with a spatial smoothing kernel to measure the clustering strength of galaxies around BCGs. This provided a catalog of over 55,424 rich galaxy clusters in the redshift range .1 < *z* < .55. The catalog is approximately volume limited up to redshift *z* = .4 and shows high purity and completeness when tested against a mock catalog, or compared to the well-established cluster catalog derived from the SDSS DR6 [17 WHL hereafter]. From the GMBCG cluster catalog, we select a cluster sample belonging to a complete volume-limited spherical shell constrained by the galactic coordinates 0° < *l* < 360° and 60°< *b* < 82° and, a radial thickness of .1 < *z* < .4. Using information on *z* and the Galactic coordinates *l* and *b*, for each cluster we determine the comoving radial distance *d* where the metric is defined by the ΛCDM cosmological parameters: H_0 = 70 *Km s*$^{-1}$*Mpc*$^{-1}$, Ω_m = .28 and Ω_Λ = .72.

3.2. Simplifying assumptions:

i) the GPM measures the local gravitational potential generated by neighboring masses at the position of a point-mass taken as a test-particle on the assumption that the gravitational potential is *time-independent;*
ii) an assumption concerning the relation between luminous and dark matter should be made. The most simplest one is that galaxy clusters trace the peaks of the underlying matter density field, i.e., the galaxy cluster density is linearly biased with respect to the dark matter density. The exact relationship between the cluster power spectrum and the dark matter power spectrum is well understood theoretically [18], and this relationship or biasing is a function of cluster mass. Then, we may reasonably assume that fluctuations of the gravitational potential generated by the galaxy cluster distribution also reflect those in the full matter distribution;
iii) there is the well-known problem to find a finite solution of Φ_j for infinite number of gravitating masses. To overcome this problem in our case, we need to assume the form of the spatial distribution of these masses. By considering that at the position of each cluster, the gravitational potential is mainly influenced by its nearest neighbors and much less by other distant masses, we assume a simplified version where a superstructure (clusters of clusters) are approximated by a system of point-like masses (clusters) forming a gravitationally bound system. In this case, we consider such a system as a point-like mass concentrated in its center of mass, which do not interact gravitationally

with each other. Further, we assume that this system is surrounded by an empty sphere of fixed radius R_V embedded in a uniform background. Such supposed segregation provides the finiteness of the gravitational potential at any test point inside the sphere but outside where the potential vanishes. Then, we assume R_V = 80 *Mpc* which is capable of incorporate the characteristic scale of superclusters i.e. ~ 100-150 h^{-1}*Mpc* [9] and the major share of the gravitational influence exerted on a test cluster by the nearest neighbors (a massive cluster placed beyond R_V has a tiny gravitational influence equivalent to that of a close single galaxy) and, large enough to avoid the shoot noise error;

iv) we do not take into account the bias affecting high photometric redshift measurements due to relativistic effect since it has been found negligible on the scales of interest herein [19]. Therefore, in the error analysis, we consider only the redshift error given in the GMBCG cluster catalog which does not exceed the 10% [15].

v) if a spherical volume V_j centered at the cluster position j overlaps the boundaries of our volume-limited cluster sample, the measured Φ_j is removed to avoid edge effects in the computation;

vi) Cluster mass is not directly available from the catalog, so that we must overcome this major problem with a quite questionable assumption: contrary to the cluster mass, the cluster richness may be reliably predicted allowing its use as a proxy for mass (e.g., [20]). However, even if it could be defined precisely for the observational sample under consideration, its use should be taken with care since richness vary depending upon survey characteristics and cluster identification methods. Therefore, we need a richness-mass relation which best fits the GMBCG dataset. From the WHL cluster catalog, [21] determined a richness-mass relation for galaxy clusters calibrated using accurate X-ray and weak-lensing mass determinations of a complete sample of clusters defined within .17 < *z* < .26. The substantial agreement of the richness classification provided by the GMBCG catalog with that of the WHL catalog for objects in common allows us to adopt that richness-mass relation to determine the cluster mass of our sample.

3.3 Errors on Φ_j

Much of the uncertainty concerning the evaluation of Φ_j comes from the determination of the cluster mass (which is not observable) as a function of the abundance of the cluster members (richness). The assumed 30% error estimation of m_i given by [21 their Eq.4] becomes somewhat arbitrary and poorly determined due to the larger redshift range (*z* ~ .4) of our sample than that used by [21]. Of course, this assumption is quite questionable because it is well known that the slope of the mass function varies with scales. However, it is worth mentioning that we are dealing with an "exploratory" analysis where a precise mass determination is desirable but not mandatory. In any case, we quantify the error affecting the determination of Φ using a Monte-Carlo simulation based on the resampling technique [22]. For each data point entering in the calculation of Φ we assume a Gaussian error distribution on distances with σ as established by [15] and masses by [21]. From these distributions we can now randomly sample new data points to estimate the simulated Φ. Repeating this resampling task 10,000 times, we get a distribution of the simulated data from which we can then infer the uncertainty given by the standard deviation. We find that the estimated standard error for Φ is ~ 32% while the error for $\langle\Phi\rangle_j$ is reduced by a factor $1/\sqrt{N}$ (N =number of Φ measured within V_j).

3.4. Finding cluster concentrations

As stated in Sect.2, when an idealized sample of mass tracers has a perfect uniform distribution, the map of the gravitational potential fields is expected to appear uniform, without wells. In a real sample however, even if deviations from that uniformity should be tiny at large scales as predicted by the Cosmological principle, at intermediate redshift, deviations are expected to be relevant and the influence of such anisotropies should create extended and deep gravitational potential wells. It follows that an extreme minimum in the $\langle\Phi\rangle_j$ distribution is the key measurement to detect a huge overdensity in the spatial distribution of our cluster sample. This can be clearly seen in the most negative region of the potential distribution of Fig. 2. The isolines of the potential distribution are roughly elliptical for very deep wells, and they become more and more complicated near the zero level as expected for random fields. The point P shows the deepest $\langle\Phi\rangle_j$ = -1.715 x 10^6 (Km/s)^2 located at Galactic coordinates *l* ~ 62°.7, *b* ~ 63°.1 and *z* ~ .367. However, from the visual inspection of the Fig. 2, close to the deepest well at P appears a secondary very deep well indicating another cluster concentration. Curiously, they form a binary-like system lying in the same redshift range of 0.34 < *z* < 0.37 where they centers are separated by more than 180 *Mpc*, prefiguring two distinct cluster concentrations as part of a huge overdensity already detected in the SCLCAT (see Sect.5). However, for the setting limits of this study we take into account only the cluster concentration identified by the deepest well (measured at the position P of Fig.2).

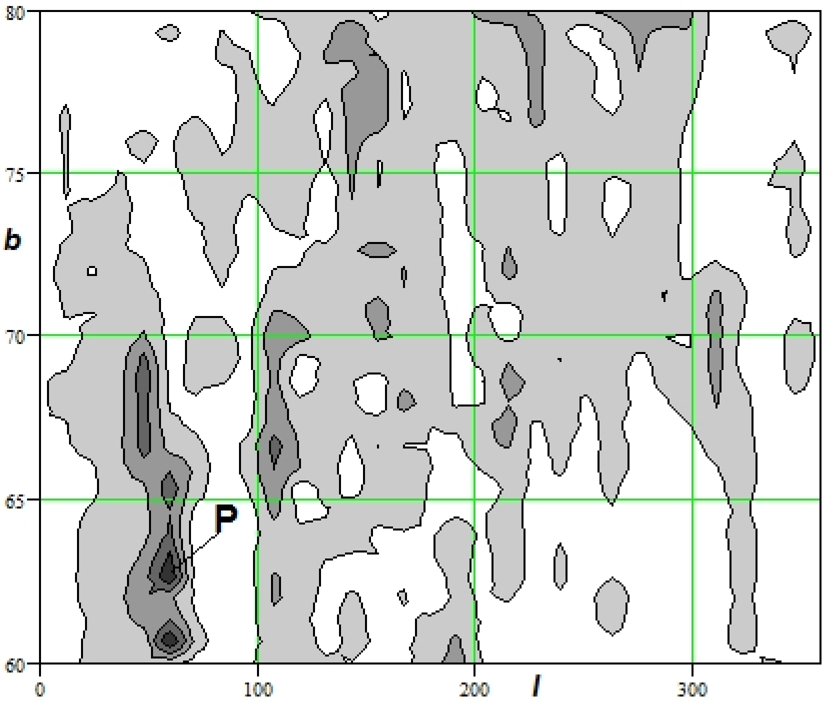


Fig.2 – Contour plot of the $\langle \Phi \rangle_j$ distribution in the (*l*,*b*)-projection. At the point P (*l=62°.7*, *b=63°.1, z=.367*), the position of deepest $\langle \Phi \rangle_j$ = -1.715 x 10^6 (Km/s)^2 is shown. Note also the large chain of overdensities around P (see Sect. 3.7.).

3.5. The bound core of the detected cluster concentration

To identify unambiguously the memberships of a hypothetical *bound* core of the detected cluster concentration is very difficult since they are not yet fully formed, virialized and clearly separated from each other. Generally, these structures have been defined by quite arbitrary criteria, mostly on the basis of a statistical algorithms like percolation, Friend of Friends code and density threshold. Here we adopt the radial density contrast criterion proposed by [15] even if, one could replace this algorithm with entirely different density criteria. They assure an accurate process to constrain a massive overdensity with respect to background using simulations in ΛCDM cosmology and establish a criterion based on the application of the spherical collapse model to constrain regions enclosed by a spherical shell that eventually evolve into virialized systems working out a lower density limit for gravitationally bound structures. This limit is based on the density contrast that a spherical shell needs to enclose to remain bound to a spherically symmetric overdensity. If ρ_c is the cluster mass density enclosed by the critical shell and ρ_{bck} is the background density (given by $\rho_{crit} \cdot \Omega_m$ where ρ_{crit} is the critical density of the Universe), the mass density enclosed by the last bound shell of a structure must satisfy the density threshold $\delta_c = \rho_c / \rho_{bck} =$ 8.67 (note that in [15], δ_c = 7.88 due to Ω_Λ = .70 instead of .72 adopted in the present study). All density parameters are determined in unit of $M_\odot$ *Mpc*$^{-3}$. To apply the density criterion, we simply assume that the core of the cluster concentration is defined down to the deepest potential well which it shares with neighboring objects. In this scheme the test cluster where the deepest gravitational potential is measured forms the head of the structure and the center of mass of the densest parts of the cluster concentration. Then, we calculate the density contrast parameter $\delta_{sph,n}$ for *n* concentric spheres with increasing radius until the condition $\delta_{sph,n}$ < δ_c = 8.67 will be satisfied. Subsequently, we calculate the center of mass of this sphere and repeat the process iterating until the shift in the center between successive iterations is less than 1% of the radius. With the final center of mass, we identify the angular position, radius and cluster memberships of the bound spherical region corresponding to the core of the cluster concentration. In Fig.3 the radial density contrast profile is apparent. It shows a cusp constrained within 10 *Mpc* radius from the center while in the outer part it gradually fades up to 51 *Mpc* radius where $\delta_{sh,7}$ = 0, resulting in an increasingly uniform cluster distribution. This is the characteristic cusp expected from the collapse of a large structure where the continuous sharpening of the internal mass distribution is reflected in the steepening of its density profile.

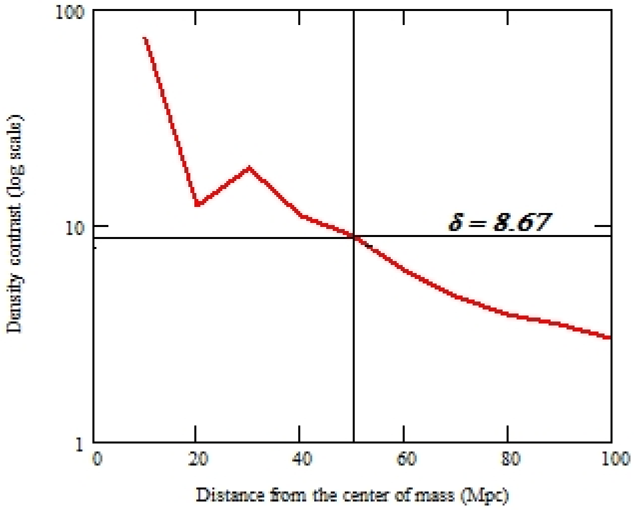


Fig.3 - Plot of the radial density contrast profiles obtained from the application of the clustering algorithm to identify bound core of the cluster concentration. The intersection with the horizontal line showing the limit δ = 8.67 of the density contrast criterion identifies the radius of the critical shell.

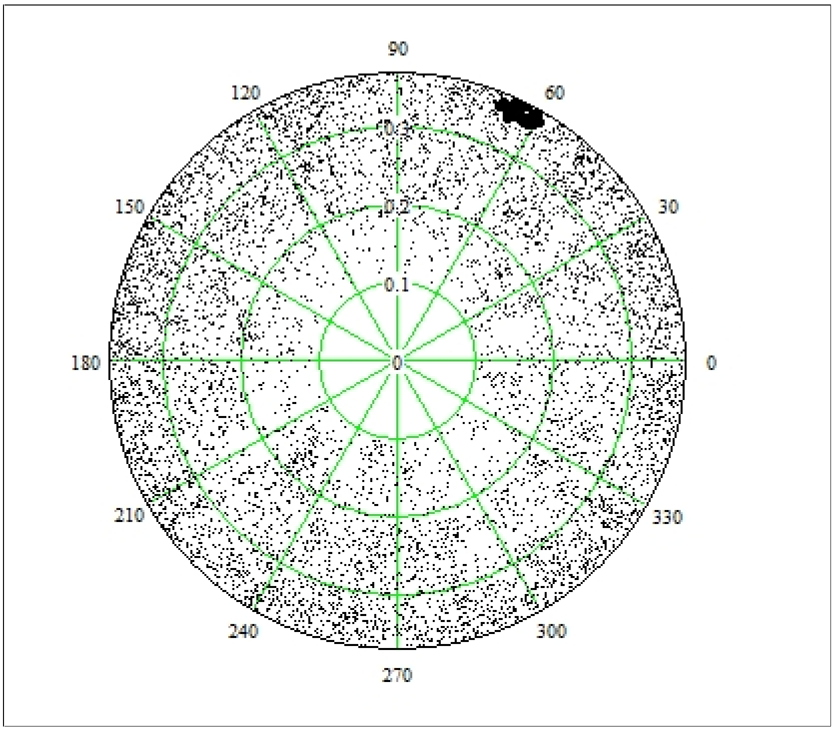
The identified bound core lies at Galactic coordinate l=63°.71 and b=63°.72 or J2000 coordinate RA = 14^h 46^m 18^s and Dec = 37° 37' 40" (J221°.54+37°.64 in decimal degrees) and z ~ .36. It is assembled by 35 clusters enclosed in a sphere of 51 *Mpc* radius. The main properties of its cluster members are summarized in Table 1 as follows: in Col.1, the GMBCG-ID J2000 coordinates in decimal degrees; Col.2 and 3, the photometric redshift and richness class, respectively (these columns are taken from the GMBCG cluster catalog); Col.4, the cluster mass estimation obtained from the richness-mass relation of [21]; Col.5, the gravitational potential $\langle\Phi\rangle_j$.

Table 1. Properties of the cluster members

GMBCG ID	z	R	M $10^{14}M_{sun}$	‹Φ› $10^6(Km\ s^{-1})^2$
GMBCG J219.58196+37.61495	0.358	13	1.43	-1.443
GMBCG J219.58832+38.09427	0.352	8	0.67	-1.444
GMBCG J219.64068+36.67152	0.350	12	1.27	-1.408
GMBCG J219.84475+38.54932	0.357	13	1.43	-1.428
GMBCG J220.29901+37.53316	0.345	15	1.79	-1.43
GMBCG J220.34301+36.96806	0.362	12	1.27	-1.52
GMBCG J220.85463+36.65632	0.364	22	3.24	-1.602
GMBCG J220.90808+36.73450	0.365	8	0.67	-1.611
GMBCG J220.91481+35.24239	0.353	9	0.81	-1.412
GMBCG J221.03525+35.79696	0.345	17	2.17	-1.391
GMBCG J221.06604+35.95363	0.358	62	16.15	-1.537
GMBCG J221.11459+35.85098	0.358	19	2.58	-1.524
GMBCG J221.18593+35.36778	0.348	8	0.67	-1.426
GMBCG J221.19438+36.16468	0.352	20	2.80	-1.484
GMBCG J221.31667+36.44627	0.343	50	11.57	-1.423
GMBCG J221.47047+35.61850	0.362	18	2.37	-1.554
GMBCG J221.60125+37.98198	0.350	28	4.71	-1.527
GMBCG J221.62931+38.02956	0.356	27	4.45	-1.595
GMBCG J221.65575+38.10294	0.353	49	11.22	-1.596
GMBCG J221.68131+37.99997	0.341	18	2.37	-1.418
GMBCG J221.73575+37.21649	0.352	38	7.56	-1.5
GMBCG J221.88634+39.25153	0.351	14	1.61	-1.459
GMBCG J222.05890+35.19966	0.353	8	6.76	-1.463
GMBCG J222.13249+35.47027	0.355	11	1.11	-1.501
GMBCG J222.15362+37.98939	0.359	56	13.79	-1.633
GMBCG J222.25885+38.00825	0.362	11	1.11	-1.653
GMBCG J222.44332+37.31708	0.367	34	6.36	-1.715 deepest
GMBCG J222.46480+37.40898	0.347	13	1.43	-1.555
GMBCG J222.46816+37.58336	0.354	19	2.58	-1.603
GMBCG J222.49434+38.09762	0.356	23	3.47	-1.616
GMBCG J222.70127+35.63375	0.359	9	0.81	-1.566

GMBCG J222.77598+38.55925	0.347	8	0.67	-1.566
GMBCG J222.92370+38.06561	0.356	14	1.61	-1.644
GMBCG J223.50442+35.82607	0.354	26	4.20	-1.61
GMBCG J223.64751+37.60505	0.358	17	2.17	-1.681

In Fig. 4, we show the (*l,z*)-polar projection of our volume-limited cluster sample, displaying only the 8348 galaxy clusters where at their *j* position $\langle \Phi \rangle_j$ has been measured (small black dots). The 35 cluster members of the bound core of the cluster concentration identified by the deepest potential well are also shown (large black dots).



.

Fig.4 - The (*l,z*)-polar projection of our volume-limited cluster sample. The core membership listed in Table 1 is outlined with large black dots.

3.6. Quantifying the virial mass of the bound core

As a first approximation, a mass estimation is obtained by summing up the individual cluster masses which yields a value of 1.23 x 10^{16} $M_{\odot}$. Of course, the total mass is expected to be considerably larger than this, because lower mass as field galaxies are expected to contribute significantly to the total. The lack of lensing and X-ray data prevents the use of accurate mass estimators forcing us to use less accurate mass estimator based on kinematical data as, for example, the virial mass estimator based on the virial theorem. This estimator is applied under the assumption of dynamical equilibrium of the system, an assumption quite questionable for large scale structures because many effects like the halo asphericity, the secondary infall or the lack of the virial equilibrium may affect heavily the result. However, according to [23], this estimator has the advantage of providing a "conservative" evaluation as demonstrated by simulations where the cluster virial mass estimations, on average, turn out 20% underestimated. Besides, this result was also confirmed by [24] which, on the basis of 10,000 Monte Carlo simulations, demonstrated that at least 87% of the virial mass estimations are below the true mass. According to [25] which estimated the mass of the Corona Borealis supercluster we use the equation $M_{vir} = 3\sigma_v R_{vir} G^{-1}$ (see also [26]) where R_{vir} is estimated as in [27] and the line-of-sight velocity dispersion in the center of mass frame is computed using the prescriptions of [28]. Then, we find σ_v=1,183 *Km/s* and M_{vir} = 2.67±0.80 x 10^{16} $M_{\odot}$ which is a factor of ~ 2 more massive than the individual mass summation. The estimation of the 1-σ error of M_{vir} has been calculated according to the resampling technique described in Sec. 2.3(*vii*). To better appreciate the properties of the supercluster, we make a comparison with the properties of one of the most massive structures found in the local Universe: the Shapley supercluster (SSC). A detailed study of the SSC was performed by [29] establishing that SSC is composed of 21 clusters within a sphere of ~ 50 *Mpc* radius and a total mass of 4.4±0.44 x 10^{16} $M_{\odot}$. In comparison, our supercluster shows almost the same extension but is less massive than the SSC in spite of having a more numerous cluster population (actually, looking over the richness class of each object listed in Table 1, one can easily recognize that many of them look like galaxy *groups* rather than clusters).

3.7. Comparison with the SCLCAT

The volume occupied by our cluster sample has been studied by [9] and their results are reported within the SCLCAT. A first glance in the SCLCAT we note a giant overdensity detected at the lowest density limit of 2.20 and identified as ID=226+034+0359 (RA+Dec+z) composed of 6,962 galaxies with a box-diagonal of 2,162 $Mpc\ h^{-1}$ which corresponds to the large and most negative region appearing around the point P in Fig. 2 confirming the fair agreement between the two clustering methods. In the SCLCAT, at higher density limits, this huge overdensity fragments in several denser structures. We find a tight correspondence of our cluster concentration with the supercluster ID=222+037+0357 identified at the density limit of 5.40 and composed of 91 galaxies with a box_diagonal of 178 $Mpc\ h^{-1}$. The comparison between of the angular positions of the two centers of the structures gives a substantial agreement even if our cluster concentration is segregated in smaller and denser volumes. Taking into account that either the SCLCAT and the GMBCG catalogs, are both derived from the SDSS DR7 survey, the observed discrepancies may be attributed to the different methods used in the process of identification: first, our analysis is based on the hierarchical chain: clusters→superclusters i.e. the GMBCG cluster catalog is used as a collection of point-mass tracers, while the SCLCAT is based on the chain: galaxies→superclusters i.e. the galaxy sample is used to search for galaxy overdensities, identified *directly* as "superclusters" within regions of positive density contrast from low to high density thresholds. Second, a general problem of the modern hierarchical data clustering algorithms is that clustering quality highly depends on how certain parameters are set. What makes the situation even more complicated is that optimal parameter setting is data dependent. As a result, it may happen that different parts of a given data set require different parameter settings for optimizing clustering quality that, on the contrary, applying a global parameter setting to the entire data set may compromise the final result. Thus, if a selection effect affects clustering algorithms, it may depend on a certain degree of arbitrariness in the parameter selection. In any case, the substantial agreement between our findings and the SCLCAT counterpart confirms the reliability of our GPM in the clustering analysis.

3.8 Some remarks

Fig. 1 shows two extended minima in the potential distribution of the cluster sample segregated in tight redshift range between $.34 < z < .37$. The sources of that potential wells are two close but separated massive cluster concentrations. The reason of this mass segregation is presently unclear, but it may carry important cosmological implications which requires a deeper analysis since it detects an alignment of high density regions in the cluster distribution situated in the same redshift range. Such a coherent cluster segregation seems to be in tension with the theoretical expectations of the Cosmological principle which predicts an ever increasing matter homogeneity toward larger scale i.e., in a perfect homogeneous background the gravitational potential fields smooth toward uniformity as well as the local gravitational potentials should tend to a common energy. However, we cannot exclude the hypothesis that the observed mass segregation may be an artifact due to an unknown bias in the data. It is thus necessary to be cautious in interpreting the consequences of our finding in terms of a full 3D cluster distribution since the GMBCG catalog was compiled using photometric redshift and there are not convincing proofs that allows to overcome the suspect that large uncertainties in the measurements may affect our results. In fact, relevant discrepancies were found by [30] (see their Fig.5) superimposing a window of the HectoMAP (based on spectroscopic data) on the corresponding part of the GMBCG cluster catalog where many GMBCG clusters do not match the spectroscopic counterpart positions. Besides, [9] report that the weighting factors of their clustering algorithm used to derive the SCLCAT are too high for the highest distances, which cause densities that are too high at the farthest edge of the field. Evidently, if the GMBCG catalog suffers of similar bias and large uncertainties in the photometric data, our result may be incorrect. On the other hand, the data used in the present study were derived by two teams [14,17] applying different clustering algorithms to the SDSS data. However, the derived cluster catalogs show a fair concordance among angular positions, redshifts and richness classification for 22,000 objects in common [14]. Besides, both purity and completeness of the above catalogs were compared by [31] with their new catalog of 55,121 groups and clusters (also derived from the SDSS DR7) obtaining a substantial agreement and comparable quality. Then, if we can reasonably assume that the SDSS database itself is unaffected by large selection effects, unlikely systematic errors due to data processing may affect the GMBCG catalog concluding that the observed mass segregation hardly could be interpreted as an artifact. However, the discrepancies claimed by [30] are robust enough to preclude any conclusion as long as accurate spectroscopic data will confirm our findings. If so, the discovered binary system would turn out one of the most massive concentrations of galaxy clusters detected at intermediate redshift and would have a direct cosmological implication since their estimated masses seem to be in tension with the allowable locations predicted in the mass-redshift plane by the ΛCDM model [32].

4. CONCLUSIONS

In this work, we explore the use of a gravitational potential-based method as a clustering finder, focusing its application on a volume-limited cluster sample extracted from the recent GMBCG cluster catalog. We adopt the three-dimensional framework which enables us to investigate the relation between the local potential distribution and the volume overdensities. To identify large cluster concentrations, the analysis is performed in two steps: 1) we measure the comoving local gravitational potential generated by neighboring cluster masses inside fixed spherical volume centered at the position of each sampled cluster taken one at a time as a test particle. The computation extended to each cluster of the selected sample provides a detailed map of the negative potential fluctuations. The deepest potential well identified in the potential map detects unambiguously the cluster overdensity in the cluster distribution. 2) a density contrast criterion has been applied to constrain the bound core of such overdensity. Using the gravitational potential to identify clustered structures is advantageous because it enables a cluster finder based on gravity theory where the local gravitational potential is computed from volume density and identified from the contours of the projected surface of the potential distribution. Being gravity a long range force, the distribution of the potentials is smoother than the density distribution enabling us to constrain overdensity boundaries with a clear physical meaning i.e. clustered structures are identified by very deep fluctuations in the global potential distribution. Besides, it shows much less complexity in comparison with conventional clustering algorithms that require parameter tuning. It allows refinements or modifications, for example, if one needs to study the clustering properties of cluster pairs, triplets or small groups, a contour plot of the local potential distribution Φ_j is more appropriate rather than the smoothed potential fields $\langle \Phi \rangle_j$ used here to detect large structures. Therefore, we may conclude that the proposed GPM offers a promising cluster finder suitable for application to large datasets. As an example, we have applied our method to a complete sample of galaxy clusters as mass tracers. Mapping the gravitational potential distribution, we have found that the deepest potential well is generated by a huge concentration of galaxy clusters. It has a bound core of 35 galaxy clusters enclosed in a sphere of 51*Mpc* radius is located at *l* ~ 63°.7, *b* ~ 63°.7, and redshift *z* ~ .36 with velocity dispersion of 1,183 *Km/s* and an estimated virial mass of $2.67 \pm .80 \times 10^{16}$ $M_{\odot}$. The good agreement of our findings compared with those obtained using a different methodology, confirms that our GPM offers a straightforward, powerful as well as fast way to identify clustered structures from large datasets. The uncertainty affecting our result is mainly due to the richness-mass relation adopted here. Therefore, the major refinement expected to improve the basic GPM outlined here is to reduce the scatter between the observable (richness) and the predicted quantity (mass). This can be achieved checking that richness-mass relation hold for an optically selected cluster sample compared with X-ray or lensing selected counterparts and, measuring how the relation scales towards high redshift.

REFERENCES

1. Martinez V, Saar E. Statistics of the Galaxy Distribution, Chapman and Hall, London, 2002
2. Jang W, Hendry M. Cluster analysis of massive datasets in astronomy. Stat. Comput. 2007, 17, 253
3. Fu B. Algorithms for Large-Scale Astronomical Problems. Doctor of Philosophy, Computer Science Department, Carnegie Mellon University, 2011.
4. Basilakos S, Plionis M, Rowan-Robinson M. PSCz superclusters: detection, shapes and cosmological implications. 2001, MNRAS, 323, 47
5. Erdogdu P, Lahav O, Zaroubi S, Efstathiou G, Moody S, Peacock J et al. The 2dF Galaxy Redshift Survey: Wiener reconstruction of the cosmic web. 2004, MNRAS,352, 939
6. Einasto J, Einasto M, Tago E, Saar E, Hütsi G, Jõeveer M et al. Superclusters of galaxies from the 2dF redshift survey. I. The catalogue. 2007, AA, 462, 811
7. Stoughton C, Lupton RH, Bernardi M, Blanton MR, Burles S, Castander FJ et al. Sloan Digital Sky Survey: Early Data Release. 2002, AJ, 123, 485
8. Abazajian KN, Adelman-McCarthy JK, Agüeros MA, Allam SS, Allende Prieto C, An D et al. The Seventh Data Release of the Sloan Digital Sky Survey. 2009, ApJS, 182, 543
9. Liivamagi LJ, Tempel E, Saar E. SDSS DR7 superclusters. The catalogues. 2012, AA, 539,80
10. Costa-Duarte MV, Sodre Jr L, Durret F. Morphological properties of superclusters of galaxies. 2011, MNRAS, 411,1716
11. Smith RJ, Clark PC, Bonnell IA. Fragmentation in molecular clouds and its connection to the IMF. 2009, MNRAS, 396, 830
12. Gong H, Ostriker EC. Dense Core Formation in Supersonic Turbulent Converging Flows. 2011, ApJ, 729, 120
13. Tukey JW. Exploratory Data Analysis, Addison-Wesley, Reading, 1977
14. Madsen S, Doroshkevich AG, Gottlober S, Muller V. The Cross Correlation between the Gravitational Potential and the Large Scale Matter Distribution. 1998, AA, 398, 1
15. Hao J, McKay TA, Koester BP, Rykoff ES, Rozo E, Annis J et al. A GMBCG Galaxy Cluster Catalog of 55,424 Rich Clusters from SDSS DR7. 2010, ApJS, 191, 254
16. Dunner R, Araya PA, Meza A, Reisenegger A. The limits of bound structures in the accelerating Universe. 2006, MNRAS, 366, 803
17. Wen ZL, Han JL, Liu FS. Galaxy Clusters Identified from the SDSS DR6 and Their Properties. 2009, ApJS, 183, 197
18. Mo HJ, White SDM. An analytic model for the spatial clustering of dark matter haloes. 1996, MNRAS, 282, 347
19. Yoo J. General relativistic description of the observed galaxy power spectrum: Do we understand what we measure? 2010, Phys. Rev.D, 82, 3508
20. Tinker J, Kravtsov AV, Klypin A, Abazajian K, Warren M, Yepes G et al. Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. 2008, ApJ, 688, 709
21. Wen ZL, Han JL, Liu FS. Mass function of rich galaxy clusters and its constraint on $\sigma_{8,}$ 2010, MNRAS, 407, 533
22. Andrae R. Error estimation in astronomy: A guide. 2010, arXiv10123754
23. Cen R. Toward Understanding Galaxy Clusters and Their Constituents: Projection Effects on Velocity Dispersion, X-Ray Emission, Mass Estimates, Gas Fraction, and Substructure. 1997, ApJ, 485, 39
24. Evans NW, Wilkinson MI, Perrett KM, Bridges TJ. New Mass Estimators For Tracer Populations. 2003, ApJ, 583, 752
25. Small TA, Ma CP, Sargent WLW, Hamilton D. Results from the Norris Survey of the Corona Borealis Supercluster. 1998, ApJ, 492, 45
26. Zapata T, Perez J, Padilla N, Tissera P. The influence of halo assembly on galaxies and galaxy groups. 2009, MNRAS, 394, 2229
27. Merchan ME, Zandivarez A. Galaxy Groups in the Third Data Release of the Sloan Digital Sky Survey. 2005, ApJ, 630, 759
28. Danese L, de Zotti G, di Tullio G. On velocity dispersions of galaxies in rich clusters. 1980, AA, 82, 322
29. Munoz JA, Loeb A. The density contrast of the Shapley supercluster. 2008, MNRAS, 391, 1341
30. Geller MJ, Diaferio A, Kurtz MJ. Mapping the Universe: The 2010 Russell Lecture. 2011, AJ, 142, 133
31. Budzynski JM, Koposov S, McCarthy IG, McGee SL, Belokurov V. The radial distribution of galaxies in groups and clusters. 2012, MNRAS, 423, 104
32. Harrison I, Coles P. Testing cosmology with extreme galaxy clusters. 2012, MNRAS, 421,19