How robustly can clustering algorithms recover cosmic superstructures from X-ray galaxy cluster catalogues? A method comparison under input sparsity

AARYAN KUKAR
The British School
New Delhi
INDIA

Abstract: - The recent identification of Quipu — a 428 Mpc filament of 68 X-ray-luminous galaxy clusters and currently the largest known cosmic structure — relied on a friends-of-friends (FoF) analysis of the CLASSIX cluster catalogue at a fixed redshift-dependent linking length [1]. Whether such supercluster results hold under a different clustering algorithm, or with a sparser input catalogue, has not been systematically tested. We compared six clustering methods (FoF at fixed and redshift-scaled linking length, DBSCAN, HDBSCAN, single-linkage hierarchical clustering, and a Gaussian Mixture Model) applied to 317 X-ray clusters in the MCXC redshift slice $z = 0.03–0.07$ [2], scoring each against the 185 published superstructure members [1] by binary F1. Robustness was measured by random subsampling, with a second slice ($z = 0.04–0.08$) testing slice-dependence. At full sample, the fixed-scale single-linkage methods achieved the highest F1 (0.74) but collapsed to 0.00 by 60% subsampling, whereas adaptive-density HDBSCAN held $F1 \geq 0.31$ down to 20%. The results indicate a measurable trade-off between full-density accuracy and robustness under input sparsity, within the catalogue and hyperparameters reported here.

Key-Words: - galaxy clusters, superstructures, clustering algorithms, friends-of-friends, HDBSCAN, X-ray clusters, large-scale structure, computational physics

Received: March 11, 2026. Revised: April 16, 2026. Accepted: May 18, 2026. Published: November 19, 2026.

1 Introduction

Cosmic superstructures — coherent overdensities of galaxy clusters extending across hundreds of megaparsecs — provide direct probes of the matter distribution on the largest gravitationally-evolved scales of the observable Universe [1, 3]. The recent detection of the Quipu superstructure, a 428 Mpc filament of 68 X-ray-luminous galaxy clusters at $z = 0.03–0.06$, has been reported as the largest cosmic structure currently catalogued and is estimated to contain approximately 2.4×10^{17} solar masses [1]. The detection rests on a friends-of-friends (FoF) connectivity analysis of the CLASSIX X-ray cluster sample, with a redshift-scaled linking length tied to a uniform overdensity threshold.

The friends-of-friends algorithm itself was introduced by Huchra and Geller [4] to identify dynamically bound galaxy groups in redshift surveys, and has since become the dominant method for both galaxy-group identification and supercluster delineation [5]. X-ray-detected galaxy clusters became the preferred input for large-scale-structure work after the ROSAT All-Sky Survey [6] enabled a homogeneous sky-wide cluster sample, later compiled into the Meta-Catalogue of X-ray detected Clusters (MCXC) by Piffaretti et al. [2] and updated as MCXC-II by Sadibekova et al. [7]. Galaxy-density-based superstructure analyses on optical redshift surveys, notably the SDSS DR7 supercluster catalogue of Liivamägi et al. [5], have proceeded in parallel using the same FoF methodology applied at fixed linking length to luminosity-density fields.

The published cosmic-web literature includes several alternatives to FoF that are designed to recover filamentary or hierarchical structure rather than connected components alone. The DisPerSE method of Sousbie [8] uses persistent homology to extract topologically prominent filaments from a discrete tracer distribution. The Bisous model of Tempel et al. [9] applies a marked point process to identify filaments as random configurations of small interacting cylinders. The NEXUS family of methods, developed by Cautun et al. [10], performs multi-scale morphological classification of voids, walls, filaments, and clusters from density and tidal-field information. More recently, the Blooming Tree method of Yu and Diaferio [11] has proposed binding-energy-based hierarchical decomposition of superclusters into bound substructures. None of these alternatives has been systematically benchmarked

against FoF on the specific case of X-ray cluster catalogues, where the input density and selection function differ substantially from the optical redshift surveys on which the alternatives were developed.

A practically important property that has not been characterised in the published literature is the dependence of clustering-method recovery on input catalogue density. Upcoming X-ray cluster surveys, including the SRG/eROSITA all-sky survey [12], are expected to produce cluster catalogues with substantial heterogeneity in depth across the sky. The behaviour of FoF and competing clustering algorithms on partially complete catalogues is a question with direct implications for how surveys at varying flux limits should be analysed.

In this work, we compared six clustering algorithms applied to a single X-ray cluster catalogue (MCXC) within a fixed redshift slice, with the published superstructure member list of [1] used as a ground-truth reference for evaluation. We measured both full-sample recovery and degradation under random subsampling, and we tested whether the observed pattern reproduces in a second redshift slice. The reported results are descriptive comparisons of method behaviour at fixed catalogue and fixed hyperparameters; we do not claim that any method is causally preferable, nor that any new superstructure has been discovered.

2 Methods

2.1 Catalogue and redshift slice

The input cluster catalogue was the MCXC of Piffaretti et al. [2], retrieved from the VizieR archive (catalogue identifier J/A+A/534/A109, accessed 2026 May 8). The MCXC homogenises 11 ROSAT-based X-ray cluster catalogues (NORAS, REFLEX, BCS, SGP, NEP, MACS, CIZA, 160SD, 400SD, SHARC, WARPS, EMSS) to a common overdensity definition at R_{500}, providing 1743 clusters with right ascension, declination, redshift, X-ray luminosity L_{500}, and source-catalogue identifier. The published Quipu paper used the CLASSIX catalogue (REFLEX II [13] and NORAS II [14] combined), which extends to lower flux limits than the NORAS+REFLEX subset propagated into MCXC. Direct access to REFLEX II and NORAS II as machine-readable cluster catalogues was attempted via the Strasbourg astronomical Data Centre (cdsarc.cds.unistra.fr); the deposited entries were either unreachable or contained only spectroscopic supplement tables, so the analysis proceeded on the MCXC subset.

The primary redshift slice was z = 0.03–0.07, chosen to match the slice in which the Quipu superstructure is reported [1]. After this cut the working sample contained 317 clusters. A sensitivity slice z = 0.04–0.08 was constructed identically (324 clusters) for the slice-dependence robustness check.

2.2 Coordinate system and ground truth

Cluster positions were converted from equatorial (RA, Dec, z) to comoving Cartesian (x, y, z_c) coordinates assuming a spatially flat Lambda-cold-dark-matter cosmology with Hubble parameter H_0 = 70 km s^{-1} Mpc^{-1} and matter density Ω_m = 0.3, matching the assumption in [1] and consistent with the latest Planck constraints [15]. Comoving distances were computed using the FlatLambdaCDM class of the astropy v6.0 cosmology subpackage [16], which evaluates

$$D_C(z) = \frac{c}{H_0}\int_0^z \frac{dz'}{\sqrt{\Omega_m(1+z')^3+1-\Omega_m}} \quad (1)$$

for a cluster at redshift z, with c the speed of light. A cluster at right ascension α and declination δ is then placed at

$$(x, y, z_c) = D_C\,(\cos\delta\cos\alpha,\ \cos\delta\sin\alpha,\ \sin\delta) \quad (2)$$

Equation (2) is exact for a given redshift and cosmology. Separations derived from it inherit two approximations: the assumed cosmological parameters, and the treatment of the observed redshift as purely cosmological, which ignores the line-of-sight peculiar velocity of each cluster. Both are quantified in Section 3.5.

The ground-truth superstructure assignments were taken from Table A.1 of Böhringer et al. [1], retrieved from VizieR (identifier J/A+A/695/A59, accessed 2026 May 8). This table lists 185 cluster members across the five superstructures named in [1]: Quipu (SSID = 1; 68 members), Shapley (SSID = 2), Serpens-Corona Borealis (SSID = 3), Hercules (SSID = 4), and Sculptor-Pegasus (SSID = 5). Each cluster in the MCXC working sample was matched to the published table by sky position (angular separation < 5 arcmin) and redshift ($|\Delta z| < 0.005$); on a successful match, the cluster inherited the published SSID label. Of the 317 clusters in the primary slice, 102 carried a non-zero ground-truth SSID (35 Quipu, 19 Shapley, 11 Serpens-Corona Borealis, 11 Hercules, 26 Sculptor-Pegasus, with the Quipu count subset of the published 68 reduced to 35 because the deeper REFLEX II and NORAS II catalogues add 33 northern-hemisphere members that do not propagate into MCXC).

2.3 Clustering algorithms

Six clustering algorithms were applied to the same input. **Friends-of-Friends at fixed linking length** links clusters i and j whenever

$$d_{ij} = \| \mathbf{r}_i - \mathbf{r}_j \| \leq L \quad (3)$$

with L = 50 Mpc and r_i the position of cluster i from Eq. (2); groups are the connected components of the resulting graph. Links were found with a path-compressed union–find data structure built from the pair list returned by scipy.spatial.cKDTree.query_pairs of SciPy v1.13 [17]; this implementation is mathematically equivalent to the percolation procedure of [4]. **Friends-of-Friends with redshift-scaled linking length** applied the prescription of Böhringer et al. [1], in which the linking length is set per cluster as

$$l(z) = [2N_{\mathrm{Cl}}(z)]^{-1/3} \quad (4)$$

with the local cluster number density $N_{Cl}(z)$ estimated by counting clusters within a redshift shell $|\Delta z| < 0.01$ around each input and dividing by the comoving shell volume, scaled by the 86% sky coverage of CLASSIX [13, 14]; pair (i, j) was linked if

$$d_{ij} \leq \frac{l_i + l_j}{2} \quad (5)$$

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [18] was applied with ε = 50 Mpc and min_samples = m = 3, using the implementation in scikit-learn v1.5 [19]. A cluster is a core point if

$$\left|\{j : d_{ij} \leq \varepsilon\}\right| \geq m \quad (6)$$

where the count includes i itself; DBSCAN groups are the connected components of the core points under Eq. (3) with L = ε, each non-core cluster within ε of a core point is attached to one of them, and the remaining clusters are labelled noise. **Hierarchical Density-Based Spatial Clustering (HDBSCAN)** [20] was applied with parameters min_cluster_size = 10 and min_samples = 3, also through the scikit-learn implementation, which follows the accelerated algorithm of McInnes et al. [21]. **Single-linkage hierarchical clustering** was applied via scipy.cluster.hierarchy.linkage with method='single' and a flat-cluster cut at distance 50 Mpc; this is mathematically identical to friends-of-friends at L = 50 Mpc on the same input (Proposition 1, Section 2.4). **Gaussian Mixture Model (GMM)** with K = 10 components and full covariance matrices was applied through scikit-learn [19] with random_state = 42. The choice of K = 10 was set above the published count of five superstructures to give the model headroom to assign smaller and field components.

For all six methods, output groups containing fewer than 20 clusters were re-classified as field (label = noise) before scoring, applying the minimum-membership cut N_{min} = 20 used by [1]. The hyperparameter values listed above are deliberately unoptimised, representing the comparison a researcher running each method "out of the box" would obtain; the effect of tuning them is quantified in Section 3.6.

2.4 Equivalence of friends-of-friends and single-linkage clustering

The identity between friends-of-friends and single-linkage clustering used in Section 2.3 follows from a standard property of single linkage [22], which we state and prove for the cut used here.

Proposition 1. Let $r_1, \ldots, r_N$ be points with Euclidean separations d_{ij}, and let G_L be the graph that joins i and j whenever Eq. (3) holds. Cut the single-linkage dendrogram of the points at height L, so that i and j share a flat cluster exactly when their cophenetic distance c_{ij} (the height at which they first share a cluster) satisfies $c_{ij} \leq L$. Then the flat clusters are the connected components of G_L, that is, the friends-of-friends groups at linking length L.

Proof. Single linkage starts from N singletons and repeatedly merges the two clusters A and B with the smallest separation $D(A, B) = \min d_{ab}$ over $a \in A$ and $b \in B$, recording D(A, B) as the merge height. Merge heights never decrease: after A and B merge, $D(A \cup B, C) = \min[D(A, C), D(B, C)] \geq D(A, B)$ for every other cluster C, because A and B were the closest pair. (i) Every cluster formed at a height of at most L is connected in G_L: a merge of A and B at height $D(A, B) \leq L$ is realised by a pair $a \in A$, $b \in B$ with $d_{ab} \leq L$, which is an edge of G_L, so if A and B are each connected in G_L so is $A \cup B$, and induction over the merges gives the claim. Hence $c_{ij} \leq L$ implies that i and j are connected in G_L. (ii) Conversely, let $i = k_0, k_1, \ldots, k_n = j$ be a path in G_L, and stop the procedure after the last merge of height at most L. By monotonicity every remaining pair of clusters is then separated by more than L. If two consecutive points k_t and k_{t+1} were in different clusters, those clusters would be separated by at most the length of the link between them, which is at most L; this is a contradiction, so all points of the path, and in particular i and j, share a cluster, and $c_{ij} \leq L$. Together, (i) and (ii) show that $c_{ij} \leq L$ exactly when i and j are connected in G_L. ■

Equivalently, the cophenetic distance is the minimax path distance [22]

$$c_{ij} = \min_{P \in \mathcal{P}_{ij}} \max_{(k,l) \in P} d_{kl} \quad (7)$$

where $\mathcal{P}_{ij}$ is the set of paths from i to j through the point set; two sets of clusters have members in a common friends-of-friends group exactly when the smallest c_{ij} between them is at most L. This form is used in Section 3.1 to explain merged superstructures. Proposition 1 also fixes the relation to DBSCAN: with m = 1 every cluster is a core point and DBSCAN reduces to friends-of-friends with L = ε, while with m > 1 its labels can differ only at non-core clusters (noise

or border assignment) and where a friends-of-friends group is held together by non-core clusters. We verified the proposition numerically on both redshift slices and on all 1000 subsamples of Section 2.6: friends-of-friends and single-linkage labellings were identical (ARI = 1) in every case, as were friends-of-friends and DBSCAN with m = 1. With m = 3, 222 of the 317 clusters in the primary slice are core points and 70 are labelled noise, which gives the ARI of 0.707 between the DBSCAN and friends-of-friends labellings reported in Section 3.1.

2.5 Evaluation metrics

Three complementary metrics scored each method's output against the ground truth. The **adjusted Rand index (ARI)** [23], with its known correction for chance and label-permutation invariance [24], measured overall partition agreement on the full label vector (including the field label). **Binary F1** measured agreement on the auxiliary "in-any-superstructure" classification, treating any output cluster whose group contained 20 or more clusters as a positive prediction and any ground-truth member as a positive label. With TP the number of clusters that are both, FP the number of positive predictions without a published label and FN the number of published members not predicted,

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad F_1 = \frac{2PR}{P+R} \qquad (8)$$

Mean per-superstructure Jaccard found, for each of the five published superstructures, the output group with maximum overlap. With T_s the set of clusters in the working sample that carry label s and G ranging over the output groups of 20 or more clusters,

$$J_s = \max_G \frac{|T_s \cap G|}{|T_s \cup G|}, \quad \langle J \rangle = \frac{1}{5}\sum_{s=1}^{5} J_s \qquad (9)$$

J_s = 1 only if a single output group contains every member of superstructure s present in the sample and nothing else. Because T_s is restricted to the working sample, members of a published superstructure that are absent from MCXC (Section 2.2) do not lower J_s. A group can be the best match for more than one superstructure, which is how merged structures appear in J_s (Section 3.1). Together these measures separate three failure modes: ARI captures over- and under-merging at the full-graph level; F1 captures false-positive contamination of any superstructure; Jaccard captures whether an individual superstructure is recovered as a unit or split across multiple groups.

2.6 Sparsity sweep and slice sensitivity

The sparsity sweep emulates catalogues of lower completeness by random thinning. At each of five sample fractions f ∈ {1.0, 0.8, 0.6, 0.4, 0.2}, 100 subsamples of

$$n_f = [\, fN \,] \qquad (10)$$

clusters, where [·] denotes rounding to the nearest integer, were drawn without replacement from the N clusters of the working set (NumPy default_rng, seed 42). The same subsamples were given to every method, so differences between methods within a subsample are paired. Each method was rerun on every subsample with hyperparameters recomputed where appropriate (l(z) is data-dependent and was recomputed each time; all other hyperparameters were fixed). Because the matching of Section 2.2 is done cluster by cluster, the ground-truth labels of a subsample are those of its members, so on average a fraction f of each superstructure's members is present. The draws are subsamples without replacement rather than bootstrap resamples in the sense of Efron [25], which draw with replacement: a duplicated cluster sits at zero separation from its copy and is always linked to it, which would bias the connectivity the sweep is designed to probe. At f = 1.0 every subsample is a permutation of the full slice. Friends-of-friends, DBSCAN, single-linkage clustering and the l(z) prescription do not depend on input order and give zero spread there; HDBSCAN breaks ties in its mutual-reachability graph in input order (F1 = 0.462 or 0.468 across permutations of the primary slice), and the GMM initialisation depends on input order (16th–84th percentile range of F1 = 0.47–0.53). For each method and fraction we report the median F1 over the 100 subsamples, the 16th–84th percentile range, the inter-quartile range, the standard deviation and the detection fraction, defined as the fraction of subsamples in which the method returned at least one group of 20 or more clusters. A second sweep on the finer grid f = 1.00, 0.95, …, 0.15 (100 subsamples per fraction, seed 2026; 2000 per fraction on a grid of step 0.025 for friends-of-friends) resolves the transition. The slice-dependence check repeated the full sample analysis (and the sparsity sweep) on the second redshift slice z = 0.04–0.08, providing a check that the observed pattern is not specific to the particular density structure of the primary slice. The benchmarking design follows the prior halo-finder comparison methodology of Knebe et al. [26].

All analysis was performed on a standard laptop CPU (no GPU); the complete pipeline runs in under an hour. Code is openly available at https://doi.org/10.5281/zenodo.20580102 with explicit version pins (Python 3.11.4, NumPy 1.26.4, SciPy 1.13.0, astropy 6.0.0, scikit-learn 1.5.0); all input and derived data products regenerate from the two public VizieR catalogues (J/A+A/534/A109 and J/A+A/695/A59) via the included pipeline.

2.7 Percolation model of the completeness dependence

To describe the sparsity dependence of the fixed-scale methods analytically, we model a catalogue of completeness f as one in which each cluster is retained independently with probability f; the fixed-size subsamples of Eq. (10) are the same process conditioned on the number retained. Friends-of-friends at linking length L is a percolation process on the graph G_L of Eq. (3) [27]. If k_i is the number of clusters within L of cluster i (its degree in G_L), the mean number of neighbours within L is

$$B(f) = f\, n_0\, \frac{4\pi}{3} L^3 (1 + \xi_L) = f\, B(1) \qquad (11)$$

where n_0 is the mean number density of clusters and ξ_L the two-point correlation function of clusters averaged over a sphere of radius L. Thinning removes neighbours in proportion to f but leaves ξ unchanged, so the mean degree falls linearly with completeness. For points without correlations, a connected component of unbounded size exists only above the continuum percolation threshold B_c; with the critical density $n_c = 0.652960$ for unit linking length [28],

$$f_{\mathrm{cont}} = \frac{B_c}{\langle k \rangle}, \qquad B_c = \frac{4\pi}{3} n_c = 2.735 \qquad (12)$$

so a region whose clusters have mean degree $\langle k \rangle$ at full completeness stays above threshold only for $f > f_{\mathrm{cont}}$.

Superstructures are finite and their linking graphs have heterogeneous degrees, so we also treat each friends-of-friends group G of the full sample as a graph with the degree distribution p_k measured in G_L. In the configuration-model (generating-function) treatment of random site removal [29, 30, 31], a component spanning a finite fraction of the group survives only if $f > f_{\mathrm{MR}}$, where

$$f_{\mathrm{MR}} = \frac{1}{\kappa - 1}, \qquad \kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} \qquad (13)$$

and the expected fraction of the group's clusters in its largest surviving component is

$$S(f) = f[1 - G_0(u)], \quad u = 1 - f + f\, G_1(u) \qquad (14)$$

with $G_0(x) = \Sigma_k\, p_k\, x^k$ and $G_1(x) = G_0'(x)/G_0'(1)$ [31]. A method that applies the 20-member cut detects the structure only if this component keeps at least $N_{\mathrm{min}} = 20$ clusters, which in expectation requires the first condition below; since on average only f|G| clusters of the group survive, the expected component can reach N_{min} only if the second condition, which does not depend on connectivity, also holds:

$$|G|\, S(f) \geq N_{\mathrm{min}}, \qquad f \geq f_{\mathrm{size}} = \frac{N_{\mathrm{min}}}{|G|} \qquad (15)$$

For a Poisson degree distribution with mean $\langle k \rangle$, Eq. (14) has the closed form

$$s(\lambda) = 1 + \frac{W_0(-\lambda e^{-\lambda})}{\lambda}, \qquad S(f) = f\, s(\langle k \rangle f) \qquad (16)$$

where W_0 is the principal branch of the Lambert W function. Equations (11)–(16) give the completeness at which fixed-scale recovery of a structure fails in terms of its size |G|, its mean degree and degree spread at the chosen L, and the membership cut N_{min}. Because Eqs. (13)–(16) treat the group as a locally tree-like random graph, we also thinned the actual linking graph of each group 4000 times at each f on a grid of step 0.025 and recorded $P_{\mathrm{det}}(f)$, the probability that a component of at least 20 clusters survives, and f_{50}, the completeness at which $P_{\mathrm{det}} = 0.5$. Thinning cannot join clusters that were not already linked, so friends-of-friends returns a group of 20 or more clusters only if one of the full-sample groups keeps such a component; the probability that it returns any group is then $1 - \Pi_G[1 - P_{\mathrm{det},G}(f)]$, which can be compared directly with the detection fraction of Section 2.6.

3 Results

3.1 Full-sample recovery

At full input density (317 clusters in the primary redshift slice), three of the six methods produced identical binary F1 scores: friends-of-friends at fixed L = 50 Mpc, DBSCAN, and single-linkage hierarchical clustering each reached binary F1 = 0.74 (precision 0.62, recall 0.91), with pairwise differences of exactly zero across all 100 subsamples at full sample. Pairwise adjusted-Rand-index comparison of these three methods' output labellings confirmed that single-linkage hierarchical clustering and friends-of-friends at L = 50 Mpc are mathematically identical (ARI = 1.000), as required by Proposition 1. DBSCAN's labellings agreed with both at ARI = 0.707, with the disagreement concentrated at peripheral clusters that DBSCAN classifies as noise (because they fail the min_samples = 3 neighbour-count requirement) but that single-linkage chains into the nearest connected component. Because both methods' disagreements occurred almost entirely on clusters absent from the published ground-truth list, the binary F1 score collapsed the difference.

The redshift-scaled friends-of-friends prescription scored F1 = 0.67 at full sample (precision 0.54, recall 0.89). The reduction relative to fixed-L FoF reflects that the prescription $l(z) = (2N_{\mathrm{Cl}})^{-1/3}$ returns a per-cluster mean of approximately 53 Mpc on our half-density-relative-to-CLASSIX sample, slightly enlarging the linking length and admitting a small number of additional false-positive clusters from the southern-sky periphery of the recovered Quipu group. The Gaussian Mixture Model scored F1 = 0.50

(precision 0.34, recall 0.96), with the high recall and low precision indicating systematic over-grouping; HDBSCAN scored F1 = 0.47 (precision 0.37, recall 0.64), with the more balanced recall–precision ratio reflecting HDBSCAN's adaptive density threshold producing more conservative groupings than the fixed-scale methods. The full-sample precision–recall positions of all six methods are shown in Fig.1, and Table 1 lists the full-sample scores of all six methods in both slices.

Table 1. Full-sample scores of the six methods in the primary (z = 0.03–0.07, 317 clusters) and sensitivity (z = 0.04–0.08, 324 clusters) slices. P, R and F_1: Eq. (8); ⟨ J⟩ : Eq. (9); N_G: number of output groups with at least 20 clusters.

	Primary slice						Sensitivity slice					
Method	**P**	**R**	**F_1**	**ARI**	**⟨ J⟩**	**N_G**	**P**	**R**	**F_1**	**ARI**	**⟨ J⟩**	**N_G**
FoF, L = 50 Mpc	0.62	0.91	0.74	0.48	0.50	4	0.62	0.69	0.65	0.50	0.36	3
FoF, l(z)	0.54	0.89	0.67	0.33	0.43	4	0.47	0.69	0.56	0.35	0.29	3
DBSCAN	0.62	0.91	0.74	0.48	0.50	4	0.62	0.69	0.65	0.50	0.36	3
HDBSCAN	0.37	0.64	0.47	0.17	0.33	4	0.25	0.96	0.39	0.03	0.24	6
Single linkage	0.62	0.91	0.74	0.48	0.50	4	0.62	0.69	0.65	0.50	0.36	3
GMM, K = 10	0.34	0.96	0.50	0.08	0.34	8	0.20	0.87	0.33	0.02	0.26	8

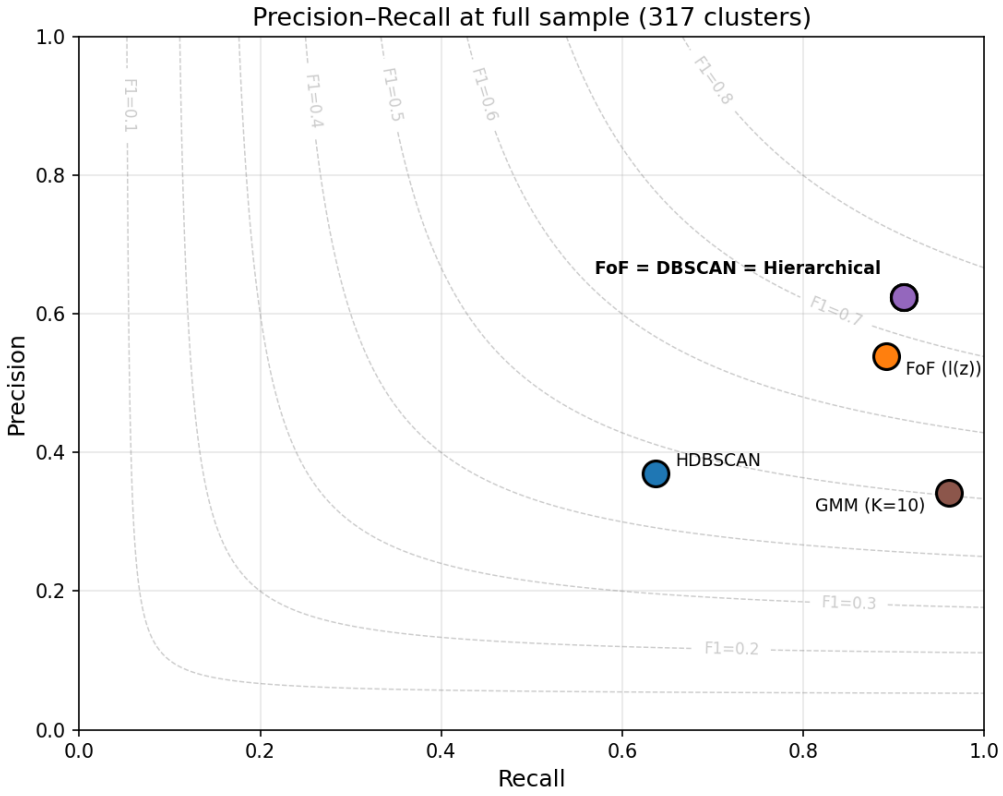


Fig.1. Precision against recall at the full input sample of 317 clusters in the primary redshift slice z = 0.03–0.07, with isocontours of constant binary F1 shown as dashed grey curves. The three fixed-scale single-linkage methods coincide at the same point near (recall = 0.91, precision = 0.62). The redshift-scaled friends-of-friends prescription sits at (0.89, 0.54). HDBSCAN occupies a more conservative position at (0.64, 0.37). The Gaussian Mixture Model occupies a permissive position at (0.96, 0.34) characteristic of over-grouping.

Fig.2 and Table 2 break the best-match Jaccard index of the highest-F1 method (friends-of-friends at L = 50 Mpc) down into the composition of each matched group. Shapley is matched by a group of 23 clusters that contains all 19 Shapley members in the sample and 4 clusters without a published label, so J = 19/23 = 0.83. Sculptor-Pegasus is matched by a group of 34: its 26 members, 2 Quipu members and 6 unlabelled clusters (J = 26/34 = 0.76). Quipu is matched by a group of 50 that contains 30 of the 35 Quipu members in the sample and 20 unlabelled clusters (J = 30/55 = 0.55); of the other 5 Quipu members, 2 sit in the Sculptor-Pegasus group and 3 in groups below the 20-member cut. Serpens-Corona Borealis and Hercules share one group of 42 clusters: 10 of the 11 Hercules members, 6 of the 11 Serpens-Corona Borealis members and 26 unlabelled clusters, giving J = 10/43 = 0.23 for Hercules and 6/47 = 0.13 for Serpens-Corona Borealis. The decomposition separates three failure modes contained in J_s: contamination of a

matched group by unlabelled clusters (Quipu, and the shared group), merging of two published superstructures into one group (Serpens-Corona Borealis and Hercules), and loss of members to other groups or to fragments below the cut (5 of 35 Quipu members, 5 of 11 Serpens-Corona Borealis members). Table 2 also lists J_s for the other methods: DBSCAN and single-linkage clustering reproduce the friends-of-friends values exactly, and no method reaches J_s = 0.3 for either Serpens-Corona Borealis or Hercules.

The merger follows from Eq. (7). The closest pair of Hercules and Serpens-Corona Borealis members (RA 240.6°, Dec +16.0°, z = 0.0353 and RA 230.8°, Dec +8.6°, z = 0.0355) is 31.7 Mpc apart, so the two structures share a group for every L ≥ 31.7 Mpc. Joining all Hercules members to one another requires L ≥ 54.2 Mpc, and all Serpens-Corona Borealis members L ≥ 69.5 Mpc. Because the gap between the two structures is smaller than the gaps within either, no single linking length separates them while keeping either one whole: a scan over L = 30–60 Mpc in steps of 2 Mpc placed members of both in one group for every L ≥ 32 Mpc. The two structures overlap on the sky (RA 223°–258° for Serpens-Corona Borealis and 235°–263° for Hercules) and in redshift (z = 0.035–0.067 and 0.030–0.041), so the merger is a property of how their MCXC clusters are distributed under any fixed-scale single-linkage criterion, not of the particular choice L = 50 Mpc.

Table 2. Composition of the best-matching friends-of-friends (L = 50 Mpc) group G for each published superstructure in the primary slice, and the best-match Jaccard index J_s (Eq. 9) for each method. DBSCAN and single-linkage clustering give the friends-of-friends values. $|T_s|$: members in the working sample; other: members of another published superstructure in G; unlabelled: clusters in G without a published label.

	Members		Matched group G				J_s			
Superstructure	**Published**	**$\|T_s\|$**	**$\|G\|$**	**$\|T_s \cap G\|$**	**Other**	**Unlabelled**	**FoF**	**FoF, l(z)**	**HDBSCAN**	**GMM**
Quipu	68	35	50	30	0	20	0.55	0.51	0.03	0.38
Shapley	23	19	23	19	0	4	0.83	0.73	0.66	0.51
Serpens-Corona Borealis	34	11	42	6	10	26	0.13	0.14	0.08	0.23
Hercules	28	11	42	10	6	26	0.23	0.25	0.12	0.29
Sculptor-Pegasus	32	26	34	26	2	6	0.76	0.52	0.76	0.30

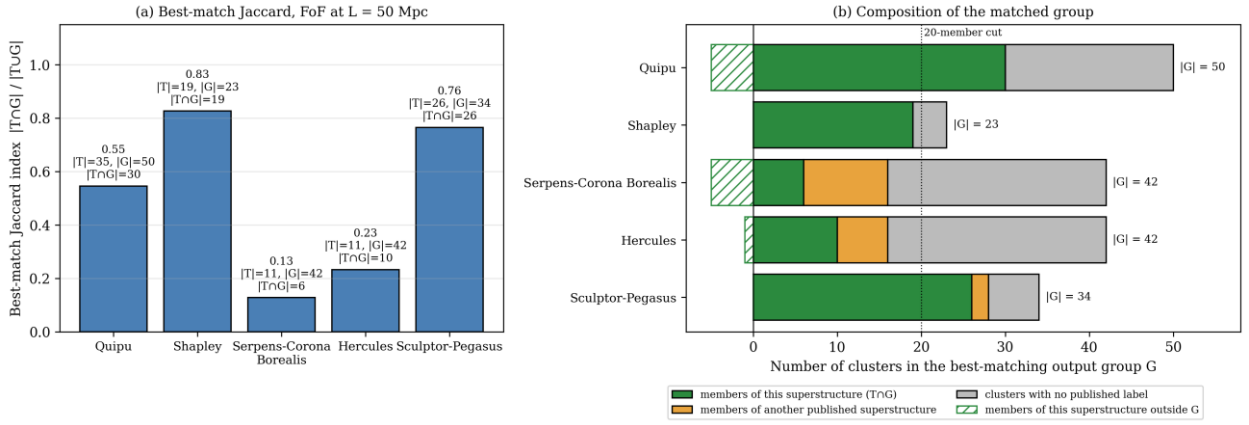


Fig.2. Per-superstructure recovery by friends-of-friends at L = 50 Mpc in the primary slice. (a) Best-match Jaccard index J_s (Eq. 9) for each published superstructure, annotated with the number of its members in the sample |T|, the size of the matched output group |G| and their overlap |T∩G|. (b) Composition of the matched group: members of the superstructure (green), members of another published superstructure (orange) and clusters without a published label (grey); the hatched bar to the left of zero counts members of the superstructure that fall outside the matched group. Serpens-Corona Borealis and Hercules are matched by the same 42-cluster group (Section 3.1).

The recovered Quipu candidate from this method is shown in Fig.3, with true positives, false positives, and the 38 published Quipu members not recovered (33 of which have no counterpart in the MCXC slice; Section 2.2) highlighted by colour.

Eighteen of the 20 false positives in the recovered Quipu group lie in one compact region at RA 48°–68° and Dec −27° to −61°, behind the recovered part of Quipu in redshift (median z = 0.062, against 0.040 for the 30 true positives; Fig.3b). Fifteen of the 20 have recession velocities in the range cz = 17 000–22 500 km s^{-1}, and 11 lie within 10° of RA 03^h19^m, Dec −50°02′; these are the velocity range and centre of the Horologium–Reticulum supercluster [32]. None of the 185 published members of [1] lies within 10° of that centre, and the supercluster lies largely beyond the 130–250 Mpc range in which [1] report their five superstructures, so these clusters belong to a known supercluster that is not part of the reference list. They are false positives with respect to that list rather than

spurious groupings produced by the method, and they form the concentration in the southern sky near RA 50° visible in Fig.3a.

Two tests distinguish a physical extension from chance alignments along the line of sight. First, the block is internally denser than the recovered part of Quipu (median nearest-neighbour separation 15.6 Mpc, against 21.0 Mpc for the true positives). It is linked to Quipu only through a small outlying fragment that holds 3 Quipu members, which it reaches through three of its own clusters with links of 9.3, 26.1 and 40.5 Mpc; that fragment joins the main body of Quipu only at L = 48.6 Mpc. The block therefore enters the recovered Quipu group only for L ≥ 48.6 Mpc (all 18 of its clusters by L = 49.2 Mpc), and for L = 44–48 Mpc the matched group excludes it (J = 0.73; Section 3.5). It is a separate overdensity attached to Quipu by a narrow margin at the adopted L = 50 Mpc, not a scatter of isolated clusters. Second, chance alignments produced by redshift errors or peculiar velocities would appear as links oriented preferentially along the line of sight. For every link in the groups of 20 or more clusters we measured the angle θ between the link and the line of sight to its midpoint (Fig.3c). The mean |cos θ| is 0.53 for the 153 links that involve an unlabelled cluster and 0.49 for the 302 links between labelled members; isotropic orientations give 0.5, and a Kolmogorov–Smirnov test against the isotropic (uniform) distribution of |cos θ| gives p = 0.36 in both cases. The false positives therefore show no line-of-sight elongation. The remaining two false positives (RA 46.6°, Dec −9.7°, z = 0.034 and RA 52.0°, Dec +2.6°, z = 0.030) lie 26–29 Mpc from Quipu members along the northern part of the recovered filament. The effect of the coordinate conversion on these boundaries is quantified in Section 3.5.

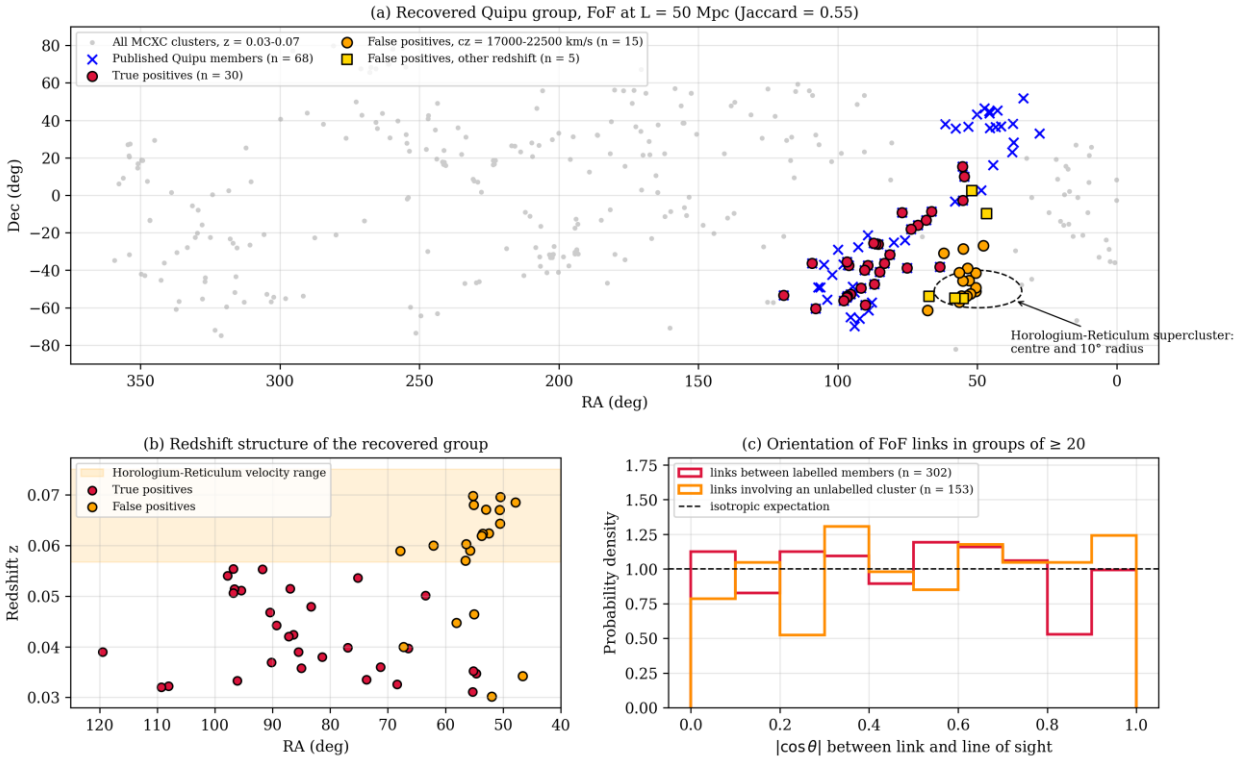


Fig.3. The recovered Quipu group from friends-of-friends at L = 50 Mpc in the primary slice. (a) Sky positions: red circles, 30 true positives; orange circles, the 15 false positives with recession velocities of 17 000–22 500 km s^{-1}; yellow squares, the other 5 false positives; blue crosses, all 68 published Quipu members, of which 38 are not recovered (33 have no counterpart in the MCXC slice and 5 fall outside the recovered group); grey points, all clusters in the slice. The dashed contour marks 10° around the centre of the Horologium–Reticulum supercluster [32]. (b) Redshift against right ascension for the group's members; the shaded band is the supercluster's velocity range. (c) Distribution of |cos θ|, where θ is the angle between a link and the line of sight, for all links in groups of 20 or more clusters; the dashed line is the isotropic expectation.

3.2 Sparsity-degradation curves

Random subsampling of the input catalogue at decreasing fractions revealed sharply different degradation patterns across the six methods. In the primary slice (Fig.4, upper left), friends-of-friends at fixed L = 50 Mpc, DBSCAN, and single-linkage hierarchical clustering all collapsed to F1 = 0.00 (median across 100 subsamples) at the 60% sample fraction (190 clusters), having retained F1 = 0.61 at the 80% fraction (254 clusters). The redshift-scaled friends-of-friends prescription degraded more gradually, retaining F1 = 0.49 at 60% and F1 = 0.33 at

40%, before reaching F1 = 0.00 at 20%. The Gaussian Mixture Model retained F1 = 0.47 at 60% before cliffing at 40% (F1 = 0.05) and 20% (F1 = 0.00). HDBSCAN was the only method to maintain F1 ≥ 0.30 across the entire sparsity range, returning F1 medians of 0.58, 0.50, 0.49, and 0.41 at the 80%, 60%, 40%, and 20% fractions respectively.

Table 3 lists the spread of F1 across subsamples. For the fixed-scale methods the spread is concentrated at the transition and comes from a bimodal outcome rather than a broad distribution. In the primary slice at f = 0.6, 41 of the 100 subsamples still contain a group of 20 or more clusters, with median F1 = 0.42 among them, and the other 59 contain none, so their F1 is 0; the inter-quartile range at this fraction is 0.38 and the standard deviation 0.20, against 0.15 and 0.12 at f = 0.8, where every subsample keeps a group. On the finer grid the friends-of-friends detection fraction falls from 0.91 at f = 0.725 to 0.09 at f = 0.50 and passes 0.5 near f = 0.62 (Fig.4, lower row). In the sensitivity slice the same transition is centred near f = 0.79, and 51 of 100 subsamples keep a group at f = 0.8. The other three methods show no comparable step: the detection fraction of the l(z) prescription stays above 0.9 down to f = 0.50 in both slices and that of the GMM down to f = 0.45, while HDBSCAN returns a group in at least 87% of subsamples down to f = 0.2. Their spread grows gradually as f decreases; for HDBSCAN in the primary slice the inter-quartile range rises from 0.09 at f = 0.8 to 0.24 at f = 0.2.

Table 3. Spread of binary F1 across the 100 subsamples at each sample fraction f: median, with the 16th–84th percentile range in brackets, followed by the detection fraction (fraction of subsamples with at least one output group of 20 or more clusters). Single linkage: friends-of-friends, DBSCAN and single-linkage clustering, whose F1 values are identical in every subsample.

Slice: method	f = 1.0	f = 0.8	f = 0.6	f = 0.4	f = 0.2
Primary: Single linkage	0.74 [0.74–0.74] 1.00	0.61 [0.46–0.68] 1.00	0.00 [0.00–0.43] 0.41	0.00 [0.00–0.00] 0.00	0.00 [0.00–0.00] 0.00
Primary: FoF, l(z)	0.67 [0.67–0.67] 1.00	0.56 [0.47–0.64] 1.00	0.49 [0.37–0.64] 0.99	0.33 [0.00–0.53] 0.79	0.00 [0.00–0.00] 0.07
Primary: HDBSCAN	0.47 [0.47–0.47] 1.00	0.58 [0.50–0.64] 1.00	0.50 [0.43–0.59] 1.00	0.49 [0.39–0.59] 0.99	0.41 [0.17–0.53] 0.87
Primary: GMM, K = 10	0.51 [0.47–0.53] 1.00	0.52 [0.46–0.56] 1.00	0.47 [0.34–0.57] 1.00	0.05 [0.00–0.39] 0.68	0.00 [0.00–0.00] 0.00
Sensitivity: Single linkage	0.65 [0.65–0.65] 1.00	0.00 [0.00–0.31] 0.51	0.00 [0.00–0.00] 0.08	0.00 [0.00–0.00] 0.00	0.00 [0.00–0.00] 0.00
Sensitivity: FoF, l(z)	0.56 [0.56–0.56] 1.00	0.35 [0.24–0.43] 1.00	0.26 [0.03–0.40] 0.96	0.24 [0.00–0.38] 0.81	0.00 [0.00–0.12] 0.17
Sensitivity: HDBSCAN	0.39 [0.39–0.39] 1.00	0.39 [0.34–0.44] 1.00	0.37 [0.29–0.42] 1.00	0.33 [0.24–0.41] 1.00	0.31 [0.16–0.39] 0.88
Sensitivity: GMM, K = 10	0.36 [0.33–0.37] 1.00	0.37 [0.31–0.41] 1.00	0.37 [0.29–0.47] 1.00	0.12 [0.00–0.33] 0.66	0.00 [0.00–0.00] 0.00

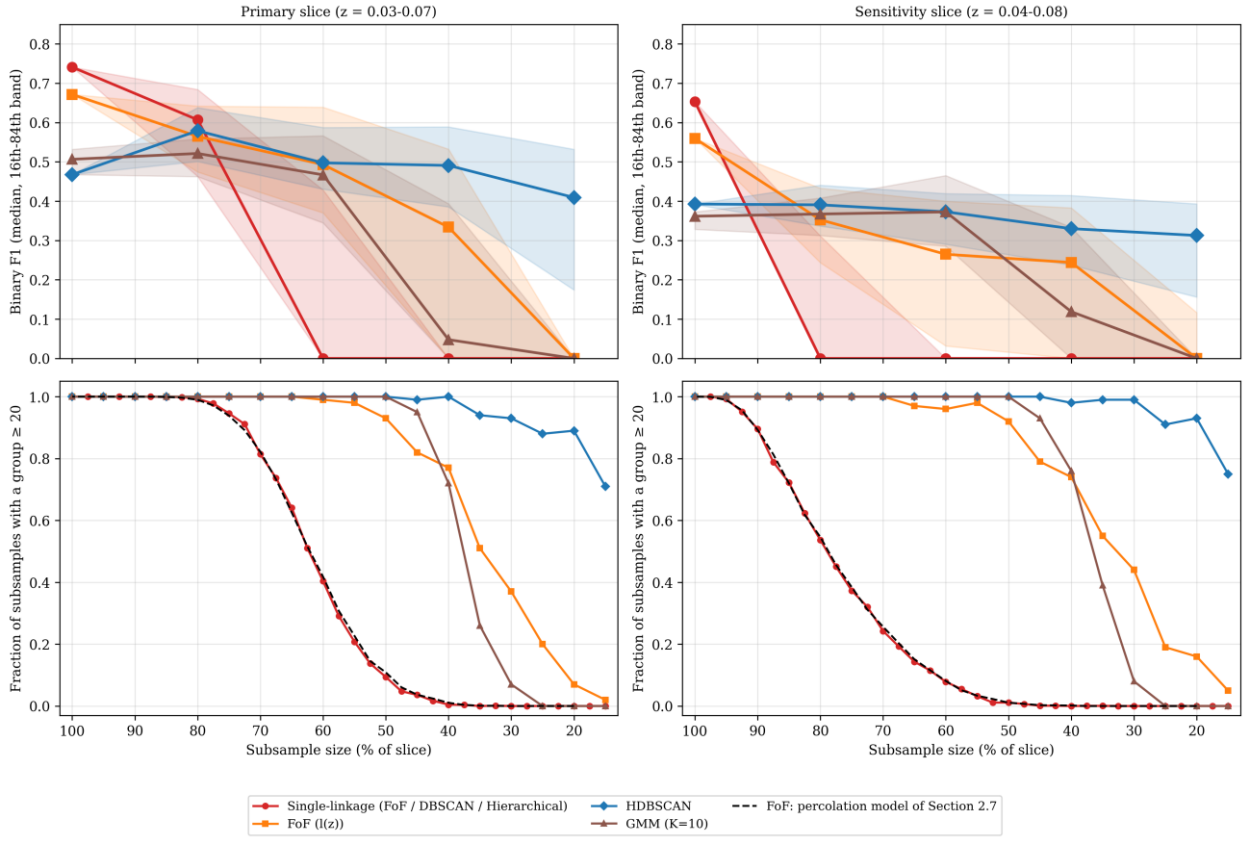


Fig.4. Upper row: median binary F1 against the 185 published superstructure members of [1] as a function of subsample fraction, for the four distinct method outputs in the primary slice z = 0.03–0.07 (left; 317 clusters at full sample) and the sensitivity slice z = 0.04–0.08 (right; 324 clusters). Lines show the median over 100 subsamples at each fraction and shaded bands the 16th–84th

percentile range. The single-linkage curve consolidates friends-of-friends at fixed L = 50 Mpc, DBSCAN and single-linkage hierarchical clustering, whose F1 values are identical in every subsample (Section 2.4). Lower row: fraction of subsamples in which each method returned at least one group of 20 or more clusters, on the fine grid of Section 2.6 (100 subsamples per fraction; 2000 for the single-linkage curve). The dashed curve is the percolation model of Section 2.7 for friends-of-friends.

3.3 Slice-dependence check

The pattern reproduced on the sensitivity slice z = 0.04–0.08, with one quantitative shift (Fig.4, upper right). At full density (324 clusters), the three fixed-scale single-linkage methods again tied at the top (F1 = 0.65), with the redshift-scaled friends-of-friends following at F1 = 0.56, and HDBSCAN and GMM at the lower end (F1 = 0.39 and 0.33 respectively). Under subsampling, however, the fixed-scale methods cliffed earlier in the sensitivity slice: F1 collapsed to 0.00 already at the 80% fraction (median across 100 subsamples), rather than the 60% fraction observed in the primary slice. The redshift-scaled friends-of-friends and the Gaussian Mixture Model retained intermediate recovery to 60% before reaching 0.00 at 20%. HDBSCAN maintained F1 in the range 0.31–0.39 across all five sample fractions in the sensitivity slice. Across the two slices, HDBSCAN was the only method whose median F1 stayed above 0.30 at the 20% subsample size; all other methods scored F1 = 0.00 at this fraction in at least one slice.

The cross-slice comparison (Fig.4, upper row) supports two descriptive findings. The pattern of fixed-scale methods cliffing under sparsity reproduced in both slices tested, with the cliff occurring at a higher sample fraction in the sensitivity slice (suggesting the cliff threshold is sample-dependent within the same algorithm). HDBSCAN's relative robustness reproduced in both slices, although its full-sample F1 was lower in both slices than the fixed-scale alternatives. The 16th–84th percentile bands in Fig.4 widen most strongly around the cliff transitions; Section 3.4 shows that this is a percolation transition of the friends-of-friends linking graph and explains why it occurs at a higher completeness in the sensitivity slice.

3.4 Percolation analysis of the fixed-scale collapse

At full completeness the primary slice has $n_0 = 3.21 \times 10^{-6}$ Mpc^{-3} (317 clusters in 9.88×10^7 Mpc^3; mean separation $n_0^{-1/3} = 67.8$ Mpc). Without clustering this density would give B(1) = 1.68 neighbours within 50 Mpc, below the continuum threshold $B_c = 2.735$ of Eq. (12): uniformly distributed clusters at this density would not percolate at L = 50 Mpc even with a complete catalogue. The measured mean degree is 3.91, a factor $1 + \xi_L \approx 2.3$ higher, and groups of 20 or more clusters exist only in the overdense regions that make up the superstructures. In the sensitivity slice the corresponding values are a mean separation of 75.4 Mpc, B(1) = 1.22 and a mean degree of 3.01.

Table 4 applies Eqs. (12)–(16) to each friends-of-friends group of the full samples and compares them with thinning of the group's own linking graph (Fig.5). Three results follow. (i) The configuration-model threshold f_{MR} (0.07–0.34) lies far below the observed transitions, so the loss of a giant component in the random-graph sense is not what limits recovery. (ii) For the compact Shapley group (23 clusters, mean degree 13.2, diameter 3 links) Eq. (15) gives $f^* = 0.87$, against $f_{50} = 0.85$ from thinning: the threshold is set by the membership cut, since 20 of the 23 clusters must survive. (iii) For the three elongated groups (Quipu, Sculptor-Pegasus, and Hercules with Serpens-Corona Borealis; diameters of 11–14 links), f_{50} = 0.67–0.77 lies 0.11–0.23 above the generating-function estimate f^* = 0.47–0.64. In these groups 14–26% of the clusters are cut vertices, whose removal splits the group, and the largest surviving component shrinks faster than Eq. (14) predicts (Fig.5a). The generating-function estimate describes a random graph with the same degree distribution, in which distant parts of a group are joined by many independent paths; in an elongated structure they are joined by few, so such a structure needs a higher completeness than its size and mean degree alone would suggest.

Combining the groups, the probability that friends-of-friends returns any group of 20 or more clusters, $1 - \Pi_G[1 - P_{det,G}(f)]$, with $P_{det,G}$ taken from thinning of each group's linking graph, reproduces the detection fraction measured with 2000 subsamples per fraction to within 0.025 at every f in both slices (Fig.4, lower row). The fixed-scale collapse is therefore accounted for entirely by the fragmentation of the four (primary) or three (sensitivity) full-sample groups. The transition is abrupt because these few groups have similar f_{50} (0.67–0.85 in the primary slice) and each group's own P_{det} falls from 0.9 to 0.1 within $\Delta f \approx 0.19$–0.33 (Fig.5b). It occurs at a higher completeness in the sensitivity slice because the groups there are small relative to the cut: they have 20, 35 and 21 clusters, so f_{size} = 1.00, 0.57 and 0.95, and the Sculptor-Pegasus and Shapley groups are lost when one or two of their clusters are removed (f_{50} = 0.96 and 0.92, against 0.80 for the Quipu group).

HDBSCAN has no length scale. Its hierarchy is built from mutual-reachability distances, which scale with the coordinates, and its cluster selection depends only on their ordering, so its labels are invariant under a uniform rescaling of all positions; we confirmed this

for scale factors from 0.5 to 2.0, and Section 3.5 shows that its output does not change when H_0 is varied. Random thinning at completeness f increases the distance to the m-th nearest neighbour, and hence every core distance, by a factor of about $f^{-1/3}$ throughout the volume, so the density ordering on which HDBSCAN builds its hierarchy is approximately preserved. The quantities that do respond to f are counts: min_samples, min_cluster_size and the 20-member cut. As f decreases HDBSCAN therefore returns fewer and relatively larger groups (median number of groups of 20 or more: 4 at f = 1.0, 3 at f = 0.6 and 1 at f = 0.2 in the primary slice) while flagging a roughly constant share of the subsample (median 0.55–0.69 of the clusters in the primary slice and 0.58–0.81 in the sensitivity slice over f = 1.0–0.2), so its F1 changes gradually rather than through a detection threshold. Section 3.6 shows how this depends on the count parameters.

Table 4. Percolation analysis of the friends-of-friends (L = 50 Mpc) groups with at least 20 clusters in the full samples. $\langle k \rangle$ and κ: mean degree and $\langle k^2 \rangle / \langle k \rangle$ in the linking graph; f_{MR}, f_{cont} and f_{size}: Eqs. (13), (12) and (15); f*: completeness at which Eq. (15) holds with S(f) from Eq. (14); f_{50}: completeness at which thinning of the group's own linking graph leaves a component of at least 20 clusters in half of 4000 trials; C: transitivity; D: diameter in links; CV: fraction of clusters that are cut vertices. A dash marks a group for which Eq. (15) is not satisfied for any f ≤ 1.

Slice: group	|G|	$\langle k \rangle$	κ	f_{MR}	f_{cont}	f_{size}	f*	f_{50}	C	D	CV
Primary: Sculptor-Pegasus	34	4.5	5.8	0.21	0.61	0.59	0.64	0.75	0.68	12	0.26
Primary: Quipu	50	4.8	5.7	0.21	0.57	0.40	0.47	0.67	0.62	11	0.14
Primary: Hercules + Serpens-Corona Borealis	42	5.1	7.1	0.16	0.53	0.48	0.54	0.77	0.67	14	0.26
Primary: Shapley	23	13.2	15.3	0.07	0.21	0.87	0.87	0.85	0.80	3	0.04
Sensitivity: Sculptor-Pegasus	20	3.3	4.0	0.34	0.83	1.00	–	0.96	0.52	10	0.35
Sensitivity: Quipu	35	4.9	6.1	0.20	0.56	0.57	0.62	0.80	0.65	10	0.29
Sensitivity: Shapley	21	13.2	14.9	0.07	0.21	0.95	0.95	0.92	0.84	3	0.05

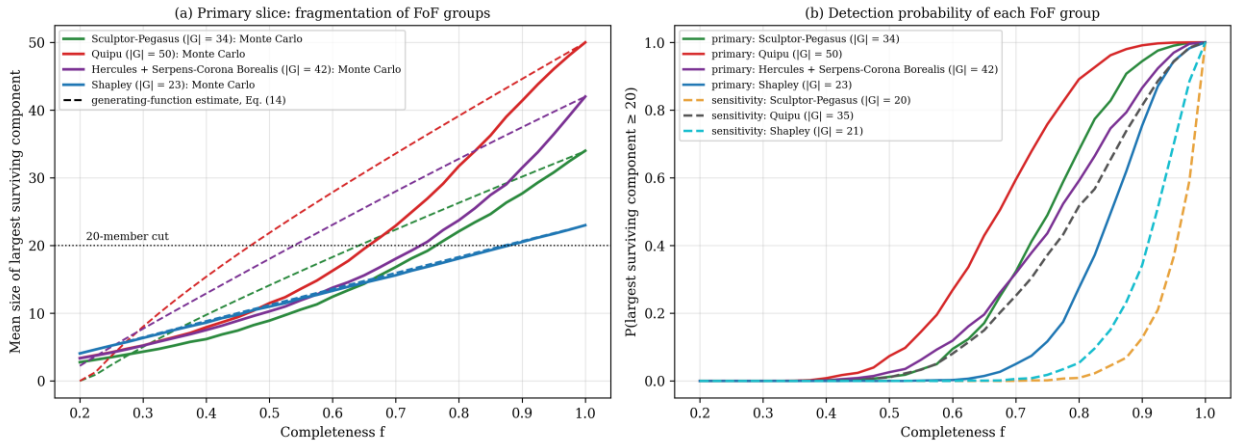


Fig.5. Percolation analysis of the friends-of-friends groups. (a) Mean size of the largest surviving component of each primary-slice group under random thinning to completeness f, from 4000 trials per f (solid), and the generating-function estimate |G|S(f) of Eq. (14) (dashed); the dotted line is the 20-member cut. (b) Probability that a component of at least 20 clusters survives, for the groups of both slices (solid: primary; dashed: sensitivity).

3.5 Sensitivity to the coordinate conversion and to peculiar velocities

Equation (2) affects the boundaries of recovered structures through the cosmological parameters and through the treatment of redshift as distance. At fixed Ω_m, Eq. (1) gives $D_C \propto 1/H_0$, so changing H_0 rescales every separation by the same factor and is equivalent to changing the linking length of the fixed-scale methods by the factor $H_0/70$. With the Planck 2018 parameters ($H_0 = 67.4$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.315$ [15]) the fixed-scale F1 rises from 0.74 to 0.78 and the Quipu Jaccard index from 0.55 to 0.73; with $H_0 = 73$ km s^{-1} Mpc^{-1} the fixed-scale F1 falls to 0.67 (Table 5). HDBSCAN's output is unchanged in both cases: the change of H_0 is a uniform rescaling, to which it is invariant, and the small change of Ω_m in the Planck case does not alter its labels. The Serpens-Corona Borealis and Hercules members share a group in every case.

The observed redshift also contains the line-of-sight peculiar velocity $v_\parallel$ of each cluster, which displaces it radially in Eq. (2) by

$$\Delta d \simeq \frac{(1+z)\, v_\parallel}{H(z)} \approx \frac{v_\parallel}{H_0} \qquad (17)$$

Because the MCXC provides no peculiar-velocity estimates, these displacements cannot be removed, and the analysis is carried out in redshift space. To

measure their effect we added Gaussian velocities with dispersion σ_v = 150, 300, 600 and 1000 km s^{-1} to the redshifts (rms displacements of 2.1–14.3 Mpc), kept the published labels fixed, and reran all six methods on 200 realisations for each σ_v. The fixed-scale F1 changes from 0.74 to medians of 0.73, 0.73, 0.70 and 0.68 (Table 5); the median Quipu Jaccard index stays at 0.55, and Serpens-Corona Borealis and Hercules remain in one group in every realisation. The fixed-scale methods keep the highest median F1 at every σ_v. Redshift-space displacements of this size therefore shift the boundaries of the recovered structures by a few clusters but do not change the merger, the false-positive block or the ranking of the methods at full sample. HDBSCAN's full-sample F1 is the exception: it changes from 0.47 to medians of 0.59–0.63 under all four dispersions, which shows that its full-sample labelling of this catalogue is sensitive to displacements of a few megaparsecs. The sparsity cliff of the fixed-scale methods arises from removing clusters (Section 3.4); the perturbation test moves clusters without removing any and leaves the fixed-scale F1 within 0.07 of its unperturbed value.

Table 5. Sensitivity of the full-sample results in the primary slice to the cosmological parameters (H_0 in km s^{-1} Mpc^{-1}) and to line-of-sight peculiar velocities with dispersion σ_v (medians over 200 realisations; $\Delta d_{rms} = \sigma_v/H_0$). Single linkage: friends-of-friends, DBSCAN and hierarchical clustering. J_{Quipu}: best-match Jaccard index of Quipu for friends-of-friends; H+S merged: fraction of cases in which Hercules and Serpens-Corona Borealis members share a group.

Conversion	Δd_{rms} (Mpc)	F_1 single linkage	F_1 FoF, l(z)	F_1 HDBSCAN	F_1 GMM	J_{Quipu}	H+S merged
H_0 = 70, Ω_m = 0.3 (adopted)	0	0.74	0.67	0.47	0.50	0.55	1.00
Planck 2018: H_0 = 67.4, Ω_m = 0.315	0	0.78	0.75	0.47	0.50	0.73	1.00
H_0 = 73, Ω_m = 0.3	0	0.67	0.60	0.47	0.50	0.55	1.00
σ_v = 150 km s^{-1}	2.1	0.73	0.68	0.59	0.50	0.55	1.00
σ_v = 300 km s^{-1}	4.3	0.73	0.64	0.62	0.51	0.55	1.00
σ_v = 600 km s^{-1}	8.6	0.70	0.62	0.63	0.51	0.55	1.00
σ_v = 1000 km s^{-1}	14.3	0.68	0.60	0.62	0.51	0.55	1.00

3.6 Hyperparameter sensitivity of the sparsity cliff

Table 6 repeats the sparsity sweep, on the same subsamples, for DBSCAN with ε = 30–70 Mpc and m = 2, 3 and 5, for DBSCAN with a linking length rescaled to the subsample density, $\varepsilon(f) = 50\ f^{-1/3}$ Mpc, and for HDBSCAN with min_cluster_size = 5–30 and min_samples = 1–10 (36 configurations in all; Table 6 lists a representative subset). For DBSCAN, m = 2 and m = 3 give identical F1 at every fraction, and m = 5 raises the full-sample F1 at ε = 50 Mpc slightly (0.77) but lowers it under thinning (0.45 at f = 0.8). The linking length sets where the cliff falls rather than whether it occurs. With ε = 30–40 Mpc and m = 2 or 3 only the most compact structures are recovered at full sample (F1 = 0.31–0.57 in the primary slice and 0 in the sensitivity slice), and their median F1 is zero from f = 0.8 downwards. With ε = 50, 60 and 70 Mpc the median F1 reaches zero at f = 0.6, 0.4 and 0.2 respectively in the primary slice, as expected from Eq. (11), in which the mean number of neighbours grows as L^3; the price is full-sample F1, which falls from 0.74 to 0.55 and 0.52 as the larger linking lengths merge more unlabelled clusters into the groups. The density-matched ε(f) keeps the mean degree expected for uniformly distributed clusters constant and moves the primary-slice cliff to between f = 0.4 (median F1 = 0.37) and f = 0.2, but in the sensitivity slice it still reaches zero by f = 0.6, because there the 20-member cut (f_{size}, Table 4) rather than connectivity is the binding condition. No tested DBSCAN configuration keeps a non-zero median F1 at f = 0.2 in either slice. The cliff therefore remains a feature of fixed-scale connectivity combined with an absolute membership cut: tuning moves it, in the direction that Eq. (11) implies, but does not remove it.

None of the 20 HDBSCAN configurations shows a cliff of this kind between f = 1.0 and f = 0.4: in the primary slice their detection fraction at f = 0.4 is at least 0.59 and their median F1 lies between 0.25 and 0.56. At f = 0.2, where a subsample has 63 clusters (65 in the sensitivity slice), the outcome depends on the count parameters. The median F1 stays above zero for 7 of the 20 configurations in the primary slice, those with min_cluster_size = 10–15 and min_samples ≤ 5 and the combination min_cluster_size = 20, min_samples = 1, and for 12 of the 20 in the sensitivity slice; it falls to zero for min_cluster_size = 5, for min_samples = 10 and for the larger min_cluster_size values (20 with min_samples ≥ 3, and 30) in the primary slice. HDBSCAN's retention of recovery at f = 0.2 (Section 3.2) therefore depends on the default min_cluster_size = 10 being small compared with the subsample size, whereas its gradual behaviour between f = 1.0 and 0.4 holds for every configuration tested.

Table 6. Median binary F1 over the 100 subsamples at each sample fraction f for DBSCAN and HDBSCAN configurations, on the subsamples of the main sweep. ε in Mpc; m: DBSCAN min_samples; mcs and ms: HDBSCAN min_cluster_size and min_samples. The configurations used in the main comparison are in bold.

	Primary slice					Sensitivity slice				
Configuration	**f = 1.0**	**0.8**	**0.6**	**0.4**	**0.2**	**f = 1.0**	**0.8**	**0.6**	**0.4**	**0.2**
DBSCAN ε = 30, m = 3	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DBSCAN ε = 40, m = 3	0.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DBSCAN ε = 50, m = 3	**0.74**	**0.61**	**0.00**	**0.00**	**0.00**	**0.65**	**0.00**	**0.00**	**0.00**	**0.00**
DBSCAN ε = 50, m = 5	0.77	0.45	0.00	0.00	0.00	0.38	0.00	0.00	0.00	0.00
DBSCAN ε = 60, m = 3	0.55	0.59	0.56	0.00	0.00	0.41	0.42	0.23	0.00	0.00
DBSCAN ε = 70, m = 3	0.52	0.53	0.53	0.45	0.00	0.39	0.40	0.34	0.00	0.00
DBSCAN ε = 50 $f^{-1/3}$, m = 3	0.74	0.63	0.56	0.37	0.00	0.65	0.24	0.00	0.00	0.00
HDBSCAN mcs = 5, ms = 3	0.46	0.45	0.41	0.28	0.00	0.41	0.29	0.25	0.18	0.03
HDBSCAN mcs = 10, ms = 1	0.59	0.57	0.49	0.48	0.37	0.32	0.37	0.36	0.33	0.31
HDBSCAN mcs = 10, ms = 3	**0.47**	**0.58**	**0.49**	**0.49**	**0.41**	**0.39**	**0.39**	**0.37**	**0.33**	**0.31**
HDBSCAN mcs = 10, ms = 10	0.62	0.62	0.54	0.52	0.00	0.44	0.44	0.40	0.38	0.25
HDBSCAN mcs = 15, ms = 3	0.56	0.55	0.52	0.51	0.41	0.39	0.38	0.37	0.35	0.32
HDBSCAN mcs = 20, ms = 3	0.53	0.54	0.52	0.51	0.00	0.39	0.38	0.38	0.37	0.27
HDBSCAN mcs = 30, ms = 3	0.53	0.52	0.51	0.50	0.00	0.38	0.37	0.36	0.36	0.00

4 Discussion

The findings can be summarised as a three-axis trade-off: fixed-scale single-linkage methods deliver the highest binary F1 at full catalogue density but collapse under random subsampling; the adaptive-density HDBSCAN method gives lower F1 at full density but maintains useful recovery when the input is sparsified; the Gaussian Mixture Model and the redshift-scaled friends-of-friends prescription both occupy intermediate positions on this trade-off.

The fixed-scale cliff has a mechanical explanation (Section 3.4). At full density the mean inter-cluster separation in the primary slice (67.8 Mpc) already exceeds the 50 Mpc linking length, and uniformly distributed clusters at this density would not percolate; friends-of-friends groups exist only in the overdense superstructure regions. Thinning lowers the number of neighbours within L in proportion to the completeness (Eq. 11). The recovered superstructures, which are elongated chains held together by cut vertices, then fragment, and their largest surviving pieces fall below the 20-member threshold of [1] between f ≈ 0.8 and 0.6. The earlier cliff in the sensitivity slice (between 100% and 80% sample size, rather than between 80% and 60%) follows from the sizes of its full-sample groups: two of its three groups have only 20 and 21 clusters and are lost when one or two members are removed.

Adaptive-density methods, of which HDBSCAN is the canonical example [20, 21], do not impose a fixed length scale. HDBSCAN's labels are invariant under a uniform rescaling of the coordinates, and thinning acts on them approximately as such a rescaling, which leaves its count parameters as the only quantities that respond to completeness (Sections 3.4 and 3.6). Their cluster identification is governed instead by the relative density contrast within the input distribution, so the same algorithm operating on a sparser input still identifies the densest local cluster cores (which remain locally over-dense even when the overall sample is reduced). The price for this property is reduced full-sample F1, because the conservative density-threshold criterion excludes the periphery of each superstructure that single-linkage chains in. Whether this trade-off should be made in any specific application depends on the use case: full-sample comparison studies that have access to a complete input catalogue have a clear preference for fixed-scale methods, whereas studies operating on partially complete catalogues (heterogeneous-depth surveys, foreground-masked regions, sparse-redshift sub-samples) have a clear preference for adaptive-density methods.

The redshift-scaled prescription of [1] occupies an intermediate position because it adapts the linking length to local cluster density but otherwise retains single-linkage's connectivity structure. Its survival to lower sample fractions than fixed-L FoF (while still cliffing at 20%) is consistent with this hybrid character. The Gaussian Mixture Model's behaviour, in turn, reflects that GMM is in a strict sense the wrong tool for filamentary structure: it fits ellipsoidal Gaussians to a distribution that is not Gaussian, and its over-grouping on small subsamples reflects model misspecification rather than algorithmic robustness.

The per-superstructure Jaccard breakdown (Section 3.1) revealed that fixed-L single-linkage at 50 Mpc merges Serpens-Corona Borealis and Hercules into one connected component, and that in this catalogue no fixed linking length separates them, because the closest pair across the two structures (31.7 Mpc) is closer than the largest gaps within each. This is a

known limitation of single-linkage clustering when two real structures have inter-cluster separation comparable to or smaller than the linking length, and it is one of the stated reasons for the development of binding-energy-based hierarchical decomposition by Yu and Diaferio [11] and persistent-homology approaches by Sousbie [8].

Several limitations bound the present comparison. First, the input catalogue (MCXC NORAS+REFLEX subset) contains 317 clusters in the analysis slice, compared with the 345 clusters in the CLASSIX sample used by [1]. The published Quipu has 33 northern-hemisphere members that do not propagate into MCXC and are therefore unrecoverable in the present analysis at any algorithmic skill. Second, the main comparison uses off-the-shelf hyperparameters. The tuning sweep of Section 3.6 shows that the location of the fixed-scale cliff moves with ε and that HDBSCAN's recovery at the lowest completeness depends on min_cluster_size, but it did not include cross-validated optimisation, which would shift the absolute F1 values reported here. Third, the comparison was restricted to clustering of comoving Cartesian positions; methods that incorporate additional information (cluster mass or X-ray luminosity) were not benchmarked, and redshift-space distortions were assessed only by the perturbation test of Section 3.5, since no peculiar-velocity reconstruction is available for the MCXC clusters. Fourth, the reported uncertainties are based on random subsampling and do not account for systematic uncertainties in the published ground truth, which are not separately quantified in [1]. Fifth, the analysis is restricted to two redshift slices of a single catalogue and should not be extrapolated to data from other surveys (galaxy-redshift surveys, weak-lensing maps, or simulation catalogues) without a parallel comparison.

The findings have direct relevance for upcoming X-ray cluster catalogues from the SRG/eROSITA all-sky survey [12], where the depth varies substantially across the sky. The current results suggest that supercluster analyses on the heterogeneous-depth eROSITA catalogue would benefit from cross-checking results obtained with fixed-scale FoF against an adaptive-density alternative such as HDBSCAN, especially in regions where catalogue completeness is below approximately 80%.

5 Conclusion

This work compared six clustering algorithms (friends-of-friends at fixed and redshift-scaled linking length, DBSCAN, HDBSCAN, single-linkage hierarchical clustering, and a Gaussian Mixture Model) for recovering published cosmic superstructures from the MCXC X-ray cluster catalogue, using the 185 superstructure members of Böhringer et al. [1] as ground truth. At full input density of 317 clusters in the primary redshift slice $z = 0.03–0.07$, fixed-scale single-linkage methods (FoF, DBSCAN, hierarchical) achieved the highest binary F1 of 0.74. Under random subsampling at five density fractions and across two redshift slices, the same fixed-scale methods collapsed to $F1 = 0.00$ by the 60% subsample fraction in the primary slice and by 80% in the sensitivity slice. The adaptive-density HDBSCAN method maintained $F1 \geq 0.31$ across the entire 100%–20% sparsity range in both slices. Thinning the friends-of-friends linking graphs group by group reproduces the fixed-scale collapse to within 0.025 in detection fraction and traces it to the 20-member cut and to the fragmentation of elongated superstructures at cut vertices, which the analytical percolation model captures for compact groups but underestimates for elongated ones; tuning the linking scale moves the collapse but did not remove it in the range tested. Most false positives attached to the recovered Quipu belong to the Horologium–Reticulum supercluster, which is not among the reference superstructures. The reported associations between method choice, input density, and recovery accuracy are descriptive within the catalogue and hyperparameter choices reported here and should not be interpreted as a universal ranking. The findings are most directly relevant for cluster-based superstructure analyses on heterogeneous-depth X-ray catalogues such as those expected from the SRG/eROSITA all-sky survey.

Acknowledgments

This work used the MCXC catalogue (Piffaretti et al. 2011) and Böhringer et al. (2025) Table A.1, both retrieved from the VizieR archive operated by the Centre de Données astronomiques de Strasbourg. All analysis code is open source. A condensed version of the central result has been submitted concurrently to Research Notes of the AAS. That note reports the full-sample F1 of the six methods and the sparsity sweep of the four distinct method outputs in the two redshift slices, the latter as a single figure equivalent to the upper row of Fig.4, and shares with this paper the input catalogues, the analysis code and those F1 values. It contains none of the other analyses presented here (the precision, recall and ARI comparison, the per-superstructure decomposition, the false-positive analysis, the equivalence proof, the percolation model, and the coordinate-sensitivity and tuning tests), and its text was written separately.

References:

[1] H. Böhringer, G. Chon, J. Trümper, R. C. Kraan-Korteweg, N. Schartel. Unveiling the largest structures in the nearby universe: discovery of the quipu superstructure. Astronomy and Astrophysics. Vol. 695, pp. A59, 2025, https://doi.org/10.1051/0004-6361/202453582.

[2] R. Piffaretti, M. Arnaud, G. W. Pratt, E. Pointecouteau, J.-B. Melin. The mcxc: a meta-catalogue of x-ray detected clusters of galaxies. Astronomy and Astrophysics. Vol. 534, pp. A109, 2011, https://doi.org/10.1051/0004-6361/201015377.

[3] M. Einasto, L. J. Liivamägi, E. Tago, E. Saar, E. Tempel, J. Einasto, V. J. Martínez, P. Heinämäki. Sdss dr7 superclusters: morphology. Astronomy and Astrophysics. Vol. 532, pp. A5, 2011, https://doi.org/10.1051/0004-6361/201116564.

[4] J. P. Huchra, M. J. Geller. Groups of galaxies. i. nearby groups. The Astrophysical Journal. Vol. 257, pp. 423–437, 1982, https://doi.org/10.1086/160000.

[5] L. J. Liivamägi, E. Tempel, E. Saar. Sdss dr7 superclusters – the catalogues. Astronomy and Astrophysics. Vol. 539, pp. A80, 2012, https://doi.org/10.1051/0004-6361/201016288.

[6] W. Voges, B. Aschenbach, T. Boller, H. Bräuninger, U. Briel, W. Burkert, K. Dennerl, J. Englhauser, R. Gruber, F. Haberl, G. Hartner, G. Hasinger, M. Kürster, E. Pfeffermann, W. Pietsch, P. Predehl, C. Rosso, J. H. M. M. Schmitt, J. Trümper, H. U. Zimmermann. The rosat all-sky survey bright source catalogue. Astronomy and Astrophysics. Vol. 349, pp. 389–405, 1999.

[7] T. Sadibekova, M. Arnaud, G. W. Pratt, M. Maturi, M. Sereno, M. Radovich, M. Roncarelli, S. Bocquet, R. F. J. van der Burg, M. E. Ramos-Ceja, A. Streblyanska, R. Adam, P. Ade, P. André, M. Arnaud, G. Aussel, P. Ausset, A. Beelen, A. Benoît. Mcxc-ii: second release of the meta-catalogue of x-ray detected clusters of galaxies. Astronomy and Astrophysics. Vol. 688, pp. A187, 2024, https://doi.org/10.1051/0004-6361/202449427.

[8] T. Sousbie. The persistent cosmic web and its filamentary structure i. theory and implementation. Monthly Notices of the Royal Astronomical Society. Vol. 414, pp. 350–383, 2011, https://doi.org/10.1111/j.1365-2966.2011.18394.x.

[9] E. Tempel, R. S. Stoica, V. J. Martínez, L. J. Liivamägi, G. Castellan, E. Saar. Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the sdss. Monthly Notices of the Royal Astronomical Society. Vol. 438, pp. 3465–3482, 2014, https://doi.org/10.1093/mnras/stt2454.

[10] M. Cautun, R. van de Weygaert, B. J. T. Jones. Nexus: tracing the cosmic web connection. Monthly Notices of the Royal Astronomical Society. Vol. 429, pp. 1286–1308, 2013, https://doi.org/10.1093/mnras/sts416.

[11] H. Yu, A. Diaferio. The hierarchy of superclusters and the blooming tree algorithm. Monthly Notices of the Royal Astronomical Society. Vol. 481, pp. 4836–4852, 2018, https://doi.org/10.1093/mnras/sty2542.

[12] E. Bulbul, A. Liu, M. Kluge, X. Zhang, J. S. Sanders, Y. E. Bahar, V. Ghirardini, E. Artis, R. Seppi, C. Garrel, M. E. Ramos-Ceja, J. Comparat, F. Balzer, K. Böckmann, M. Brüggen, N. Clerc, K. Dennerl, K. Dolag, M. Freyberg, S. Grandis, D. Gruen, F. Kleinebreil, S. Krippendorf, G. Lamer, A. Merloni, K. Nandra, F. Pacaud, P. Predehl, T. H. Reiprich, T. Schrabback, B. Whelan, A. Zenteno, S. Zelmer. The srg/erosita all-sky survey: the first catalog of galaxy clusters and groups in the western galactic hemisphere. Astronomy and Astrophysics. Vol. 685, pp. A106, 2024, https://doi.org/10.1051/0004-6361/202348264.

[13] H. Böhringer, G. Chon, M. Bristow, C. A. Collins. The extended rosat-eso flux-limited x-ray galaxy cluster survey (reflex ii). iii. clusters and superclusters in the southern sky. Astronomy and Astrophysics. Vol. 555, pp. A30, 2013, https://doi.org/10.1051/0004-6361/201220026.

[14] H. Böhringer, G. Chon, J. Trümper. The extended northern rosat galaxy cluster survey (noras ii). i. survey construction and first results. The Astrophysical Journal Supplement Series. Vol. 229, pp. 20, 2017, https://doi.org/10.3847/1538-4365/229/2/20.

[15] Planck Collaboration: N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J.-M. Delouis, E. Di Valentino, J. M. Diego, O. Doré, M. Douspis, A. Ducout, X. Dupac, S. Dusini, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen. Planck 2018 results. vi. cosmological parameters. Astronomy and Astrophysics. Vol. 641, pp. A6, 2020, https://doi.org/10.1051/0004-6361/201833910.

[16] A. M. Price-Whelan, P. L. Lim, N. Earl, N. Starkman, L. Bradley, D. L. Shupe, A. A. Patil, L. Corrales, C. E. Brasseur, M. N'Diaye, J. Greco, R. Hosseinzadeh, D. M. Hercher, M. Cara, E. Tollerud, K. Conroy, B. M. Morris, T. Robitaille, J. T. VanderPlas, K. L. Cruz, A. Ginsburg, B. Sipőcz, R. White, A. Bostroem, T. Aldcroft, P. Velikov, J. Speagle, T. Dotter, A. Crawford, T. P. Robitaille. The astropy project: sustaining and growing a community-oriented open-source project and the latest major release (v5.0) of the core package. The Astrophysical Journal. Vol. 935, pp. 167, 2022, https://doi.org/10.3847/1538-4357/ac7c74.

[17] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A.

Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt. Scipy 1.0: fundamental algorithms for scientific computing in python. Nature Methods. Vol. 17, pp. 261–272, 2020, https://doi.org/10.1038/s41592-019-0686-2.

[18] M. Ester, H.-P. Kriegel, J. Sander, X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). pp. 226–231, 1996.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay. Scikit-learn: machine learning in python. Journal of Machine Learning Research. Vol. 12, pp. 2825–2830, 2011.

[20] R. J. G. B. Campello, D. Moulavi, J. Sander. Density-based clustering based on hierarchical density estimates. Advances in Knowledge Discovery and Data Mining (PAKDD 2013). Vol. 7819, pp. 160–172, 2013, https://doi.org/10.1007/978-3-642-37456-2_14.

[21] L. McInnes, J. Healy, S. Astels. hdbscan: hierarchical density based clustering. The Journal of Open Source Software. Vol. 2, pp. 205, 2017, https://doi.org/10.21105/joss.00205.

[22] J. C. Gower, G. J. S. Ross. Minimum spanning trees and single linkage cluster analysis. Journal of the Royal Statistical Society Series C (Applied Statistics). Vol. 18, pp. 54–64, 1969, https://doi.org/10.2307/2346439.

[23] L. Hubert, P. Arabie. Comparing partitions. Journal of Classification. Vol. 2, pp. 193–218, 1985, https://doi.org/10.1007/BF01908075.

[24] N. X. Vinh, J. Epps, J. Bailey. Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. Journal of Machine Learning Research. Vol. 11, pp. 2837–2854, 2010.

[25] B. Efron. Bootstrap methods: another look at the jackknife. The Annals of Statistics. Vol. 7, pp. 1–26, 1979, https://doi.org/10.1214/aos/1176344552.

[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, K. Dolag, B. L. Falck, P. Fasel, J. Gardner, S. Gottlöber, C.-H. Hsu, F. Iannuzzi, A. Klypin, Z. Lukić, M. Maciejewski, C. McBride, M. C. Neyrinck, S. Planelles, D. Potter, V. Quilis, Y. Rasera, J. I. Read, P. M. Ricker, F. Roy, V. Springel, J. Stadel, G. Stinson, P. M. Sutter, V. Turchaninov, D. Tweed, G. Yepes, M. Zemp. Haloes gone mad: the halo-finder comparison project. Monthly Notices of the Royal Astronomical Society. Vol. 415, pp. 2293–2318, 2011, https://doi.org/10.1111/j.1365-2966.2011.18858.x.

[27] S. More, A. V. Kravtsov, N. Dalal, S. Gottlöber. The overdensity and masses of the friends-of-friends halos and universality of the halo mass function. The Astrophysical Journal Supplement Series. Vol. 195, pp. 4, 2011, https://doi.org/10.1088/0067-0049/195/1/4.

[28] C. D. Lorenz, R. M. Ziff. Precise determination of the critical percolation threshold for the three-dimensional "swiss cheese" model using a growth algorithm. The Journal of Chemical Physics. Vol. 114, pp. 3659–3661, 2001, https://doi.org/10.1063/1.1338506.

[29] M. Molloy, B. Reed. A critical point for random graphs with a given degree sequence. Random Structures and Algorithms. Vol. 6, pp. 161–180, 1995, https://doi.org/10.1002/rsa.3240060204.

[30] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin. Resilience of the internet to random breakdowns. Physical Review Letters. Vol. 85, pp. 4626, 2000, https://doi.org/10.1103/PhysRevLett.85.4626.

[31] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, D. J. Watts. Network robustness and fragility: percolation on random graphs. Physical Review Letters. Vol. 85, pp. 5468, 2000, https://doi.org/10.1103/PhysRevLett.85.5468.

[32] M. C. Fleenor, J. A. Rose, W. A. Christiansen, R. W. Hunstead, M. Johnston-Hollitt, M. J. Drinkwater, W. Saunders. Large-scale velocity structures in the horologium-reticulum supercluster. The Astronomical Journal. Vol. 130, pp. 957–967, 2005, https://doi.org/10.1086/431972.