Forecasting the success of the WEAVE Wide-Field Cluster Survey on the extraction of the cosmic web filaments around galaxy clusters

Daniel J. Cornwell [1]★ Ulrike Kuchner [1] Alfonso Aragón-Salamanca [1] Meghan E. Gray,[1]
Frazer R. Pearce,[1] J. Alfonso L. Aguerri,[2,3] Weiguang Cui [4,5] J. Méndez-Abreu [2,3]
Luis Peralta de Arriba [6,7] and Scott C. Trager[8]

[1] *School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK*
[2] *Instituto de Astrofísica de Canarias, Calle Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain*
[3] *Departamento de Astrofísica, Universidad de La Laguna, E-38200 La Laguna, Tenerife, Spain*
[4] *Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, United Kingdom*
[5] *Departamento de Física Teórica, Módulo 15, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain*
[6] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[7] *Centro de Astrobiología (CAB, CSIC-INTA), ESAC Campus, E-28692 Villanueva de la Cañada, Madrid, Spain*
[8] *Kapteyn Astronomical Institute, University of Groningen, Postbus 800, NL-9700 AV Groningen, the Netherlands*

ABSTRACT

Next-generation wide-field spectroscopic surveys will observe the infall regions around large numbers of galaxy clusters with high sampling rates for the first time. Here, we assess the feasibility of extracting the large-scale cosmic web around clusters using forthcoming observations, given realistic observational constraints. We use a sample of 324 hydrodynamic zoom-in simulations of massive galaxy clusters from TheThreeHundred project to create a mock-observational catalogue spanning $5R_{200}$ around 160 analogue clusters. These analogues are matched in mass to the 16 clusters targetted by the forthcoming WEAVE Wide-Field Cluster Survey (WWFCS). We consider the effects of the fibre allocation algorithm on our sampling completeness and find that we successfully allocate targets to 81.7 per cent $\pm$ 1.3 of the members in the cluster outskirts. We next test the robustness of the filament extraction algorithm by using a metric, $D_{\rm skel}$, which quantifies the distance to the filament spine. We find that the median positional offset between reference and recovered filament networks is $D_{\rm skel} = 0.13 \pm 0.02$ Mpc, much smaller than the typical filament radius of ~ 1 Mpc. Cluster connectivity of the recovered network is not substantially affected. Our findings give confidence that the WWFCS will be able to reliably trace cosmic web filaments in the vicinity around massive clusters, forming the basis of environmental studies into the effects of pre-processing on galaxy evolution.

Key words: methods: data analysis – methods: numerical – techniques: spectroscopic – galaxies: clusters: general – galaxies: haloes – large-scale structure of Universe.

1 INTRODUCTION

Galaxies spend a significant amount of time in cosmic web filaments. They are the backbone of a complex network of geometrical structures that indicate how matter is distributed in the Universe: as filaments, sheets, walls, knots and voids (Aragón-Calvo, van de Weygaert & Jones 2010), described by the Zel'dovich formalism (Zel'dovich 1970). The present-day cosmic web has evolved from small over- and underdensities already existent in the early Universe, and still grows and changes today as a consequence of gravitational collapse. Since $z \approx 2$, about half of the mass budget of the Universe is located within filaments, while only contributing to 6 per cent of its volume. By contrast, only about 10 per cent of the mass is located in galaxy clusters (Cautun et al. 2014).

However, clusters stand out as recognizable, bright peaks in the density field and have been the focus of observational galaxy evolution studies for decades (Gray et al. 2009; Balogh et al. 2017). They have uncovered a now well-supported relation that finds higher fractions of quiescent and early type galaxies in clusters compared to outside of clusters (Morphology–Density relation; Dressler et al. 1997). Typically, this is explained through astrophysical effects that quench and transform galaxies as they encounter the extremely dense intracluster medium of cluster cores during their infall. Ram pressure stripping is one of several possible mechanisms quenching galaxies infalling on to a cluster (Zinger et al. 2018; Arthur et al. 2019). However, the majority of the gas lies outside the boundaries where the clusters are virialized, and in the intergalactic medium within filaments (Galárraga-Espinosa, Langer & Aghanim 2022; Gouin, Gallo & Aghanim 2022; Walker & Lau 2022).

Galaxy clusters are therefore not isolated islands, but assemble, replenish, and grow via ongoing mergers with smaller clusters,

★ E-mail: daniel.cornwell@nottingham.ac.uk

groups, and clumps of gas, as well as through a constant flow of gas and galaxies from filaments. The most prominent of these filaments have hot gas temperatures and dense cores that have the possibility to strip the gas from galaxies, but also to replenish galaxies with pre-enriched filamentary gas (Vulcani et al. 2019), impacting their mass assembly and star formation histories in very different ways (Laigle et al. 2017; Song et al. 2020). It is clear that the challenge of understanding galaxy evolution must include the impact of the large-scale geometry and flows of the cosmic web, and that galaxy transformation begins well before the galaxies fall into the cluster (pre-processing; Zabludoff & Mulchaey 1998). Physical processes in the outskirts of galaxy clusters are therefore fundamentally different from cluster cores, and thus important areas for the study of cluster assembly and their connection to the filaments of the cosmic web (Sarron et al. 2019; Gouin et al. 2020; Salerno et al. 2020; Malavasi, Nicola et al. 2022). However, they are challenging to capture.

While filaments can be identified by mapping the gas distribution of galaxy clusters in simulations (Kuchner et al. 2020; Vallés-Pérez, Planelles & Quilis 2020; Gouin et al. 2022), galaxies tend to trace these features of the cosmic web and can therefore be used to detect filaments observationally (Einasto et al. 2020; Malavasi et al. 2020). To correctly identify filaments that feed clusters, we require a large area, high sampling density and depth to cover a sufficient number of galaxies over a broad range of masses. Large-area surveys such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Galaxy and Mass Assembly survey (GAMA; Liske et al. 2015), the Two-Degree Field Galaxy Redshift Survey (2dFGRS; Colless et al. 2001a), the VIMOS Public Extragalactic Redshift Survey (VIPERS; Colless et al. 2001b), and the Dark Energy Spectroscopic Instrument (DESI; Secco et al. 2022) all probe the distribution of galaxies over large redshift ranges, providing strong observational evidence for the presence of a cosmic web. However, they either lack statistically significant samples of galaxy clusters, or the necessary sampling or detail required for an investigation on pre-processing by filaments feeding clusters. Targeted spectroscopic studies that focus on clusters may provide the required sampling, but they are either only available as case studies of stand-out targets such as Virgo (Castignani et al. 2021, 2022), or do not extend far enough to bridge cluster infall regions to the large-scale cosmic web filaments (e.g. GOGREEN Balogh et al. 2017 or OmegaWINGS Gullieuszik et al. 2015; Moretti et al. 2017).

To address the need for observing programmes that combine high sampling and statistical power, we look towards next-generation wide-field, multi-object spectroscopic (MOS) surveys of galaxy clusters as they will enable detailed study into the far-reaching lower density cluster outskirts. They are designed to reveal the complex interplay between the properties of galaxies and their position in the cosmic web filaments that feed the clusters. Examples of next generation MOS surveys are the upcoming WEAVE Wide Field Cluster Survey (WWFCS; Kuchner et al. in preparation) and the 4MOST CHileAN Cluster galaxy Evolution Survey (CHANCES; Haines et al. in preparation). We motivate our studies with the WWFCS, which will cover 20 low redshift ($z \sim 0.05$) galaxy clusters in a mass range of $\log(M_{cl}/M_{\odot}) = 13.8-15.5$ out to and beyond $5R_{200}$. For each cluster, thousands of new spectra will be obtained with a galaxy stellar mass limit of $10^9 M_{\odot}$, extending our current understanding of these systems to include the infall regions and low-mass galaxies.

Given the challenging task of accurately mapping filaments in the vicinity of massive clusters (both in terms of extreme contrasts of interlaced high and low-density regions, 2D projections, and complications due to the Finger of God effect Kuchner et al.

2021a), careful preparation is required. We have planned, tested, and fine-tuned our steps to map and characterize the infall regions of these clusters with a large statistical sample of simulated clusters from The Three Hundred survey, and investigated strategies for doing so in redshift space (Kuchner et al. 2020, 2021a, b; Rost et al. 2021). To confidently carry out the observational programme with the 1000-fibre fed MOS WEAVE at the Wiliam Herschel Telescope (Dalton et al. 2014), we now take the final step from simulations to fully-informative mock observations to understand the success and limitations of identifying filaments around clusters.

Our goal for this paper is to quantify what effect the physical constraints of assigning fibres to targets – a necessary and important step in the design of a MOS survey – has on filament finding. Our previous investigations assumed that all theoretical cluster structure members are targeted and return spectra, thus featuring in the mapping and subsequent analysis. However, in reality, instruments only have a finite capacity to place fibers on targets, and physical restrictions imposed by the geometry and size of the instrumental components require us to make decisions that will ultimately influence the success of finding filaments. In addition, limitations of a realistic target selection may lead to losing valuable fibres to background galaxies. Fibre collisions in dense regions like groups and substructures in the outskirts, as well as decisions on which galaxies should receive higher priority than others, directly link to the input for filament identification algorithms and thus could impact the analysis of pre-processing in infall regions.

In this paper, we therefore close the circle of comparing simulations to observations – from a theoretical 3D volume to a fully configured 2D projection. We design a framework for determining the feasibility of reliably characterizing the large-scale structure from galaxies that can be observed in current wide-field cluster surveys, using concrete constraints that are matched to the WEAVE instrument and the WWFCS.

The paper is structured in the following way: Section 2 describes the data we have used. This includes the spectroscopic survey inspiring this paper and the numerical simulations used to create mock observations. Section 3 reports our generation of mock observations and a summary on how we extract the cosmic web. Section 4 displays the results and discussion, we describe the accuracy in which we trace the large-scale structure surrounding our simulated clusters and explain the importance of our results in probing the success of next-generation spectroscopic surveys. We present our conclusions on the likely success of filament retrieval from WWFCS in Section 5.

2 GALAXY CLUSTER INFORMATION

The framework presented in this paper is designed with the WWFCS in mind, but should also work for similar surveys making the adjustments required by the specific observational strategy and constraints. We will use simulation outputs in tandem with algorithms underpinning the observational processes of next-generation wide-field cluster surveys. This section describes the planned observations and simulated data relevant to this work.

2.1 The WEAVE Wide-Field Cluster Survey

WEAVE (William Herschel Telescope Enhanced Area Velocity Explorer) is a next-generation MOS. The spectrograph makes use of $\sim$1000 individual fibres deployable over a 2-deg-diameter field of view. The instrument also includes 20 small deployable
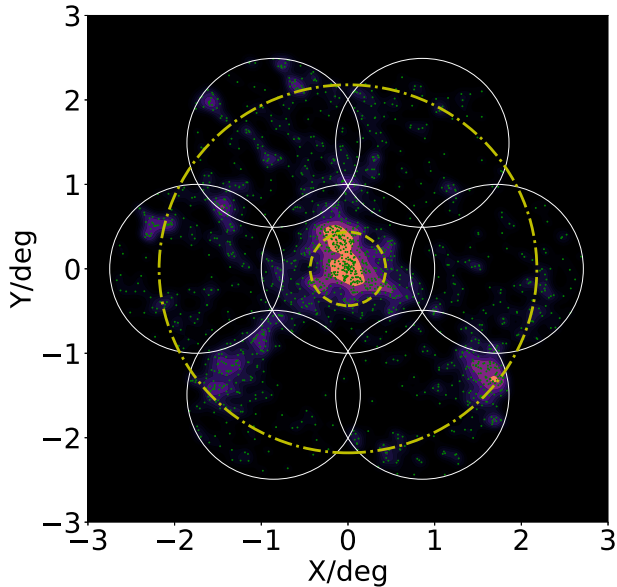
Figure 1. Example simulated cluster from THETHREEHUNDRED with a similar mass and redshift to the cluster Abell 602, one of the WWFCS targets. The projected dark matter density distribution is shown, derived using a Kernel Density Estimation (KDE) with a 500 kpc smoothing scale. This box has a depth of 10 Mpc. The green dots indicate the positions of galaxy–mass haloes. Each white circle encloses a WEAVE field with a 2-deg diameter. The central yellow dashed circle corresponds to R_{200} and the larger dot–dashed yellow circle to $5R_{200}$.

integral field units (mini-IFUs), as well as one large IFU. In this paper, we are only concerned with the (MOS) observing mode. Low- and high-spectral resolution observing modes are available. Further details on the instrument can be found in Balcells et al. (2010), Dalton et al. (2014), Dalton (2016), and Jin et al. (in preparation).

A number of surveys will be carried out with WEAVE in the next few years, one of which is the WEAVE Wide-field Cluster Survey (WWFCS; Kuchner et al. in preparation). The WWFCS will utilize the 1000 fibre-fed MOS to study the infall regions of galaxy clusters in unprecedented detail. The WWFCS will observe up to 20 clusters at low redshift ($0.04 < z < 0.07$) and will return spectra for thousands of cluster members for each cluster, out to several virial radii. The sample consists of galaxy clusters previously observed in the WINGS (Fasano et al. 2006) and OmegaWINGS (Moretti et al. 2017) surveys. The WINGS sample covers a wide range of cluster masses ($\sigma = 500-1200$ km s^{-1}; $\log L_X = 43.3-45$ erg s^{-1}; virial masses $\log (M_{cl}/M_{\odot}) = 13.8-15.5$). The WWFCS-selected clusters have velocity dispersions and X-ray luminosities that are statistically indistinguishable from the parent sample and are, therefore, unbiased in terms of their mass distribution (Kuchner et al. in preparation). The WWFCS will use the low-spectral resolution ($R \sim 5000$) mode and obtain optical spectra in the 366 nm $< \lambda <$ 959 nm range. These spectra will yield accurate redshifts, velocity dispersions as well as quantitative information on the star formation histories of the different galaxy populations.

The WWFCS observing strategy is illustrated by Fig. 1. We show a simulated galaxy cluster (cf. Section 2.2) overlaid with WEAVE 2-degree diameter MOS fields (white circles). The inner yellow dashed circle corresponds to the cluster's R_{200}, the radius at which the density is equal to two hundred times the critical density of the Universe. The outer yellow dot-dashed circle corresponds to

$5R_{200}$. Note the large over- and underdense regions reaching far out from the very dense cluster core. The large field of view we will be able to cover with WEAVE will allow us to explore and map these environments – including filaments – in great detail, reaching much larger cluster-centric distances than hitherto possible (beyond $5R_{200}$), and also study the properties of the galaxies that inhabit them.

2.2 THETHREEHUNDRED galaxy cluster simulations

THETHREEHUNDRED[1] project (Cui et al. 2018) is a set of zoom-in resimulations of the Multidark Dark Matter only (MDPL2) cosmological simulations (Klypin et al. 2016). MDPL2 is a periodic cube of comoving length $1\,h^{-1}$ Gpc containing 3840^3 dark matter particles, each with mass $1.5 \times 10^9\,h^{-1}M_{\odot}$. MDPL2 uses *Planck* cosmology ($\Omega_M = 0.307$, $\Omega_B = 0.048$, $\Omega_{\Lambda} = 0.693$, $h = 0.678$, $\sigma_8 = 0.823$, $n_s = 0.96$).

This simulation suite extracts the 324 most massive individual regions ($M_{vir} > 8 \times 10^{14}h^{-1}M_{\odot}$), follows them back to their initial conditions and resimulates the hydrodynamics of the volume surrounding a 15 h^{-1} Mpc radius sphere enclosing the cluster and its environment at a higher resolution. Outside of this high-resolution region are a set of consecutive shells, hosting lower mass resolution particles that reproduce the tidal fields of the large-scale structure at a reduced computational cost. The highest resolution dark matter particles are divided into dark matter and gas, following the cosmological baryonic fraction using the Planck 2015 cosmology: $\Omega_b/\Omega_M \approx 0.16$. This gives a combined mass resolution of $m_{DM} + m_{gas} = 1.5 \times 10^9 h^{-1}M_{\odot}$. There are 128 individual time snapshots for all 324 zoomed-in Lagrangian regions, ranging from $z = 17$ to $z = 0$. Throughout this work, we use the last snapshot at $z = 0$ given the low redshift of the WWFCS sample.

These zoom-in re-simulations have been run with the smoothed-particle hydrodynamics codes (SPH): GADGET-MUSIC (Sembolini et al. 2012), GADGET-X (Beck et al. 2015; Rasia et al. 2015), and mesh-less code GIZMO-SIMBA (Davé et al. 2019; Cui et al. 2022). We only focus on GADGET-X that incorporates full-physics galaxy formation, star formation, and feedback from both SNe and AGN. The work in this paper utilizes the AMIGA Halo Finder (Gill, Knebe & Gibson 2004; Knebe et al. 2011) to determine the halo properties.

THETHREEHUNDRED simulations provide a useful testbed to develop the observational strategy and forecast the performance of the WWFCS. First, the large volume of the parent dark-matter simulation (MDPL2) ensures a high number of massive clusters are available for statistical purposes. Secondly, the high-resolution re-simulations reach out as far as $15h^{-1}$ Mpc from each cluster centre, comparable to the area that the WEAVE observations will cover. The extensive information available fromg the cluster centre all the way to beyond $5R_{200}$ allows us to study all the environments present – from individual galaxy haloes to filaments, groups, and the cluster core.

This data set has already been used previously to generate theoretical expectations with the WWFCS in mind. For instance, Kuchner et al. (2021a) quantified the impact of redshift space distortions ('Fingers of God') on the identification of cosmic filaments. They found that trying to correct for this effect statistically in the virialized regions of clusters and groups does not lead to a more reliable extraction of the 'true' filamentary networks. For this reason, Kuchner et al. (2021a) forecast that the identification of the cosmic

---

[1]

web in the regions surrounding massive clusters using spectroscopic surveys should rely primarily on the 2D positions of the galaxies on the sky. However, they also point out that accurate spectroscopic redshifts are crucial in defining and isolating the cluster volume from which these galaxies should be selected.

3 GENERATION OF MOCK OBSERVATIONS

This section describes the framework we have developed to create mock observations and their optimization using the simulated clusters. We also discuss how we use WEAVE's fibre allocation algorithm to generate realistic WWFCS-like 'simul-observed' galaxy samples from the simulations, and the method for identifying the cosmic web using these samples. In other words, we describe the steps we take to go from simulations to observations.

3.1 Optimizing the WWFCS field positions

It is important to optimize the observational strategy of upcoming wide-field spectroscopic surveys to improve the reliability of the filament extraction process while maximizing the observational efficiency. The WWFCS can place MOS fibres on targets over a 2-deg-diameter field (Fig. 1). In order to map the filamentary structures that feed clusters it is necessary to maximize the spatial coverage, reaching out to and beyond $5R_{200}$. Such radial coverage is a good compromise between the available observing time and the need to cover as far as possible into the infall regions around the clusters (Kuchner et al. in preparation). We therefore need to design a tiling strategy to cover a circular region around the clusters that reaches $5R_{200}$ using the minimum number of WEAVE fields (or pointings). The tiling strategy we have used to find the optimal position of the WEAVE pointings for each WWFCS cluster is described in detail in Appendix A (see examples in Figs 1, 3, and 4). The 'simul-observations' described below follow the same strategy. We note that by applying this optimization process we have been able to reduce the required number of pointings (and thus the required observing time) by ∼ 15 per cent from our initial estimate, allowing us to increase the number of clusters we will be able to observe in the available time from ∼16 to ∼17–19 without compromising the accuracy of our filament mapping.

3.2 Deriving WWFCS cluster properties

In order to develop mock-observations from the simulations, we need to determine the properties of the clusters selected for the WWFCS. We firstly calculate R_{200} and M_{200} (the radius and mass where the density is 200 times the critical density of the Universe) of the clusters using their spectroscopic redshift z and velocity dispersion σ from the WINGS survey (Moretti et al. 2017) using the following equation (Finn et al. 2005; Poggianti et al. 2010):

$$R_{200} = \frac{1.73\sigma}{(\Omega_\Lambda + \Omega_M(1+z)^3)^{1/2}}. \tag{1}$$

The cluster mass inside R_{200} (M_{200}) can then be estimated using this value and the critical density of the universe. The complete list of the WWFCS targets and their properties can be found in Kuchner et al. (in preparation). The bottom panel of Fig. 2 shows the mass distributions of the WWFCS cluster sample (blue) and those from THETHREEHUNDRED simulations (orange). The mass distribution of the simulated clusters is skewed towards higher masses than those of the clusters selected for the WWFCS. This is to be expected since THETHREEHUNDRED resimulates the most massive haloes in

a large cosmological volume, making it possible to find the rarest objects. By contrast, the clusters selected for the WWFCS are more representative of clusters at low redshift and deliberately span a large range in X-ray luminosity (Section 2.1). We address this mismatch in next section.

3.3 Generating the simulated cluster and galaxy samples

We discuss now the generation of the clusters and galaxies that will be included in our mock observations using the halo data from THETHREEHUNDRED. The main aim is to create mock sample analogues to the ones we expect from the WWFCS.

First, we impose some quality constraints on the cluster haloes in the simulations so that we only select the highest quality data. We confine our study to the high resolution region of the cluster re-simulation, a spherical region with a radius of $15\,h^{-1}$ Mpc centred on the cluster centre. We then require that the mass fraction in high-resolution particles for the zoom simulation needs to be greater than 0.99 ($fM_{hires} > 0.99$). This criterion rejects low-resolution dark matter particles that may have travelled inwards into the high-resolution region during the simulation of the clusters evolution. Finally, we only accept haloes with a mass greater than the simulation resolution ($3 \times 10^{10} M_\odot$), corresponding to 20 dark matter particles, as explained in Kuchner et al. (2020).

3.3.1 Creating mass-matched cluster samples

In order to make a meaningful comparison between the clusters from THETHREEHUNDRED simulations and the WWFCS clusters, we need to create a sample of simulated clusters whose masses match those of the observational sample. Past studies have shown that for a flat Universe, on scales large enough to neglect baryonic physics, dark matter haloes evolve self-similarly (Kaiser 1986; Mostoghiu et al. 2018). Self-similarity implies that the dark matter distribution (and hence the location of dark-matter haloes) in less massive galaxy clusters is well represented by that of more massive clusters that have been scaled down appropriately taking into account their mass ratio.

To have reasonable statistics, our goal is to create a mass-matched sample of simulated WWFCS cluster analogues containing 10 simulated clusters for each WWFCS cluster. Because dark matter haloes evolve self similarly on large scales, we are able to do so using the large sample of simulated clusters from THETHREEHUNDRED project.

We describe in detail the methodology behind the mass-scaling of the simulated clusters in Appendix B. In short, when the mass of a WWFCS cluster is too small to be able to find similar mass clusters in THETHREEHUNDRED simulations, we use a mass scaling factor M_F to reduce the mass of all the simulated clusters to approximately match the mass of the WWFCS cluster. The mass of each individual dark matter halo in the relevant cluster simulation is also scaled down by the same factor, and included in the halo sample if its mass is above a mass threshold. These mass thresholds are chosen to ensure that the number of galaxy-mass haloes in each simulated cluster approximately matches that expected in the observed cluster. For each WWFCS cluster, ten analogue clusters with similar masses are randomly selected from the scaled clusters.

Fig. 2 demonstrates that three values of the mass-scaling factor M_F (1, 2, and 5) are sufficient to provide enough mass-matched simulated clusters. Of course, $M_F = 1$ implies no scaling is applied. The bottom panel shows the mass distribution of the WWFCS clusters (black)
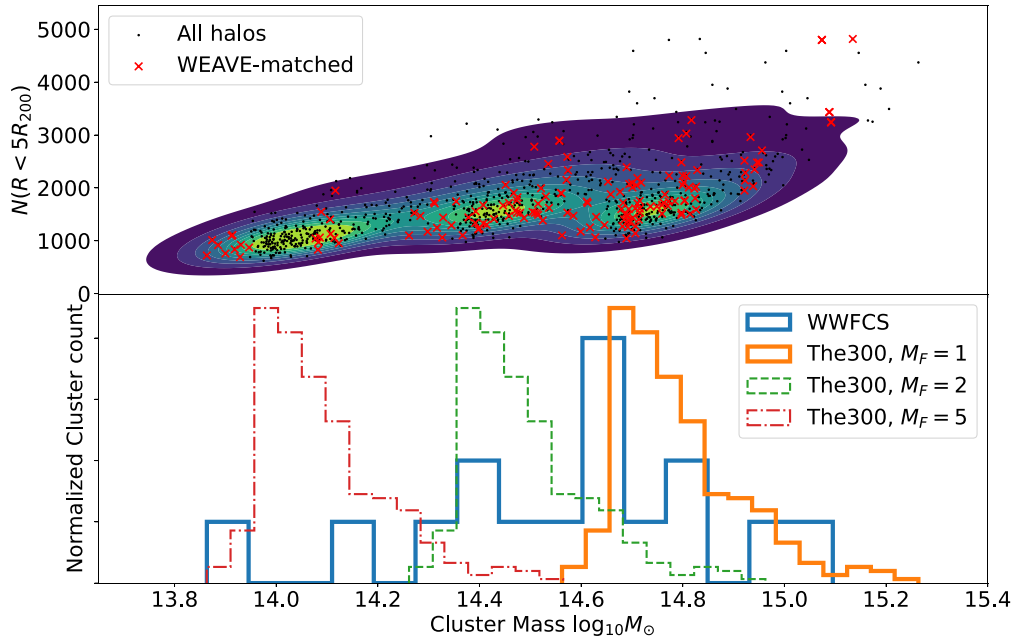
Figure 2. Bottom panel: Normalized cluster count as a function of cluster mass showing the mass distribution of the WWFCS clusters (solid blue line) and THETHREEHUNDRED un-scaled clusters ($M_F = 1$, solid orange line), and scaled by $M_F = 2$ (green dashed line) and $M_F = 5$ (red dot–dashed line). Three scaling values ($M_F = 1$, 2, and 5) are sufficient to fully cover the mass range of the WWFCS cluster sample. Top panel: The number of galaxy haloes contained within $5R_{200, \text{scaled}}$ for each simulated cluster as a function of cluster mass is displayed by the small dots and the coloured density distribution. Red crosses correspond to the mass-matched simulated clusters (ten per WWFCS cluster). This gives an indication of the approximate number of galaxies that can be 'observed' in each cluster simul-observation.

together with that of the un-scaled and scaled THETHREEHUNDRED clusters. The orange, green, and red histograms represent the mass distributions of all 324 simulated clusters with $M_F = 1$, 2, and 5, respectively.

3.3.2 Scaling cluster properties

Having built a mass-matched sample of simulated clusters, we now describe how the spatial properties of the clusters and subhaloes (linear and angular size, halo positions) are likewise scaled.

The radius of the mass-scaled clusters $R_{200, \text{scaled}}$ can be derived from the relationship between R_{200} and M_{200}, adapted from Poggianti et al. (2010),

$$R_{200}^3 = \frac{M_{200}}{K \times h^2(\Omega_\Lambda + (1+z)^3 \Omega_0)}, \quad (2)$$

where $K = 2.32 \times 10^{14} \, M_\odot \text{Mpc}^{-3}$, making $M_{200} = M_{200, \text{scaled}}$. The spatial coordinates x, y, and z of all the haloes in the cluster (and thus their clustercentric distances) are therefore multiplied by a factor $R_{200, \text{scaled}}/R_{200, \text{unscaled}}$.

For each WWFCS cluster the angular diameter distance is calculated using their redshifts (Moretti et al. 2017) with the adopted cosmology. Next, for each of the corresponding 10 analogue simulated clusters, we convert 3D positions to angular distances between the haloes and the centre of their clusters, and thus their relative positions on the simulated sky.

The top panel of Fig. 2 displays the number of galaxy haloes contained within $5R_{200, \text{scaled}}$ from their cluster centres as a function of cluster mass (small dots and density distribution). The red crosses correspond to the clusters in the mass-matched sample (ten per WWFCS cluster). Note that the planned WWFCS observations will generally cover well beyond $\sim 5R_{200}$ (cf. Kuchner et al. in prepration;

see also Fig. 1 and Appendix A), and therefore, the number of potential targets for each cluster shown in the figure is a conservative lower limit.

3.3.3 Defining galaxy sample and properties

To bring our mock 'observational' sample closer to the real observations, each of the WWFCS clusters' mass-matched simulated analogues are placed at the appropriate redshift and sky location. We then allocate WEAVE pointings using the field positions determined in Appendix A. Only haloes covered by these pointings will be considered as possible spectroscopic targets.

Galaxy-size dark-matter haloes in each simulated cluster are then given in-fibre magnitudes in the SDSS r-band (similar to the ones that will be used the observational target selection) following a simple procedure that ensures the target galaxies have comparable numbers and magnitude ranges to the planned observations. The actual galaxy magnitudes have no impact on the findings of this paper, but the fibre allocation programme `Configure` (Terrett et al. 2014) requires them as input. Explicitly, the total r-band magnitude of a galaxy is estimated from the mass of the simulated halo using the equation

$$r_{\text{total}} = W - 2.5 \log_{10}(M_{\text{halo}}/M_\odot), \quad (3)$$

where W is a constant that is calculated by mapping the least massive haloes in each simulated cluster (Appendix B) to the planned r-band limit of the WWFCS spectroscopic observation ($r_{\text{total}} < 19.75$, corresponding to an approximate galaxy stellar mass limit of $\sim 10^9 \, M_\odot$, Kuchner et al. in preparation). An average offset between total and in-fibre magnitudes of 1.75 mag, estimated through a least-square fit to the appropriate SDSS magnitudes, is then applied. The in-fibre magnitude limit of the WWFCS galaxy sample is therefore $r_{\text{fibre}} < 21.50$, which sets the planned exposure times of ~ 1 hour.

Table 1. Target priorities used in `Configure`.

Object type	Target priority p
Cluster galaxy	10
Background galaxy	1
Sky position	1
Cluster galaxy with one fibre already allocated	1
Background galaxy with one fibre already allocated	0
Sky position with one fibre already allocated	0

This exposure time is expected to yield reasonable signal-to-noise ratio (S/N > 5 per Å, for all the spectra up to this magnitude limit), and we therefore expect close to 100 per cent redshift completeness for the observed (and thus mock) galaxies (see Kuchner et al. in preparation). We use a simple procedure to allocate magnitudes to the galaxy-sized dark matter haloes here, as the results in this paper only require accurate spatial distributions of mock galaxies and their expected number densities. Our simple approach ensures this without relying on uncertain model galaxy properties. As Cui et al. (2018) show (see e.g. their fig. 8), large uncertainties still remain in the model observed magnitudes and colours, and the results depend strongly on the specific baryonic model used, particularly at low galaxy masses. While the simulations have appropriate resolution to yield reliable masses and locations for the dark matter haloes, the additional step of predicting observable properties through the available hydrodynamic or semi-analytic models would require making uncertain choices that are not necessary for our purposes.

At this point, we have created a set of 160 simulated galaxy clusters (10 per WWFCS target cluster), populated them with mock galaxies, placed them at the appropriate redshift and sky position, and covered them with WEAVE pointings exactly as those planned in the observations.

3.4 Allocating spectroscopic fibres to mock galaxies using `Configure`

An integral part of creating mock observations is to carry out a realistic allocation of spectroscopic fibres to the mock galaxies, since this process can potentially distort and limit the spatial information that can be derived from the real observations. Geometric and mechanical constraints (such as fibre collisions and overlap) mean that it is not possible to assign fibres to all the galaxies on a pointing.

Optimizing fibre allocation is not a trivial task, and sophisticated software is generally used to reduce costly human intervention. `Configure` is the programme that WEAVE will use to find an optimal set of assignments of fibres to positions on the sky (Terrett et al. 2014). Each field (or pointing) will contain not only science targets, but also a set of calibration objects and guide stars. `Configure` uses a probabilistic technique named 'simulated annealing' (Kirkpatrick & Gelatt 1983) to simulate the thermal motion of a system to be optimized. The 'energy' of each fibre with a target assigned to it is given by $(1.0 + s)/p$, where s is a measure of how straight the fibre is and p, the target priority, is an integer value between 0 and 10 that is used to prioritize objects on the fields. In our case, we assign a maximum priority of 10 to all of the cluster mock galaxy members, and lower values to other targets (Table 1). The algorithm then optimizes the fibre allocation by finding the configuration with the lowest 'energy' by swapping the position of fibres until the minimum is found.

This process determines the galaxies that will be allocated a fibre and therefore decides which galaxies will have spectroscopic information. It may play a crucial role in determining the accuracy in extracting cosmic-web information from the WWFCS observations, and its effect will be thoroughly tested in this paper.

To make a more realistic analogue to the planned observations, before running `Configure` we pollute our target catalogues with background objects. We randomly place 1400 background objects on each WEAVE field, corresponding to a number density ~ 450 deg^{-2}. This is somewhat larger than the galaxy number density corresponding to an in-fibre magnitude limit $r_{\rm fibre} < 21.5$, the planned WWFCS limit, but overpopulating the background is not a problem because these objects will be assigned a low priority of 1 (see Table 1). We note that the background objects are not designed to be representative of the larger-scale cosmic web, but are implemented to test the usage of free fibres (i.e. fibres that have not been assigned to cluster members according to their photometric redshifts). This is especially relevant further away from the cluster centre as fibres are then free to be placed on background targets if they are brighter than the magnitude limit. In this paper, we stress-test this assumption by putting a slightly exaggerated number of background targets in the catalogue to compete with higher priority cluster targets. If their spectroscopic redshifts from the WWFCS reject them from cluster membership, they will not feature in the filament finding, as described in Section 3.5.1. For this paper, we assume that the vast majority of our WEAVE target selection – based on magnitude, colour, and precise photometric redshifts – correctly rejects galaxies that lie outside the volume in space that corresponds to `TheThreeHundred` volume. This is supported by our tests using available observational data of the cluster centres. While we will only know the exact number after analysing early WEAVE observations, we can expect that not every background galaxy will be identified correctly. In practice, we will therefore use a range of priorities for background objects. However, given the high quality of the J-PLUS photometric redshifts (Cenarro et al. 2019), in combination with conservative colour and magnitude cuts, we expect the percentage of interlopers to be small and that the majority of background objects can be accurately de-prioritized. We keep them in the target catalogue for the sole reason that unallocated fibres can be used and they do not feature in our filament finding, as described in Section 3.5.1.

Finally, the input catalogues fed to `Configure` contains also blank sky positions for sky-subtraction purposes. These correspond to real celestial positions devoid of objects visible in SDSS images of the target clusters. They are given a priority of 1. Although in all cases we have more suitable sky positions, `Configure` is set to allocate a maximum of 50 sky fibres per field, as per the observational strategy described in Kuchner et al.

Therefore, each WEAVE field target list consists of $N_{\rm gal}$ cluster galaxies, determined by the simulated cluster galaxy sample (see Section 3.3.3 and Fig. 2), $N_{\rm sky}$ sky positions, and $N_{\rm back} = 1400 - N_{\rm gal} - N_{\rm sky}$ background targets. As mentioned above, the exact number of background objects does not matter, and we limit $N_{\rm back}$ in this way to keep the size of the target catalogues small enough to keep the `Configure` running time manageable.

To fully 'configure' a cluster's mock-observation we need to apply the `Configure` software to each individual WEAVE field (or pointing) sequentially, taking into account that these pointings overlap (Figs 3 and 4) and that the central pointing will be observed twice in order to deal with the high density of targets in the cluster core (Kuchner et al. in preparation). The aim of the process is to maximize the number of target galaxies with at least one fibre allocated.
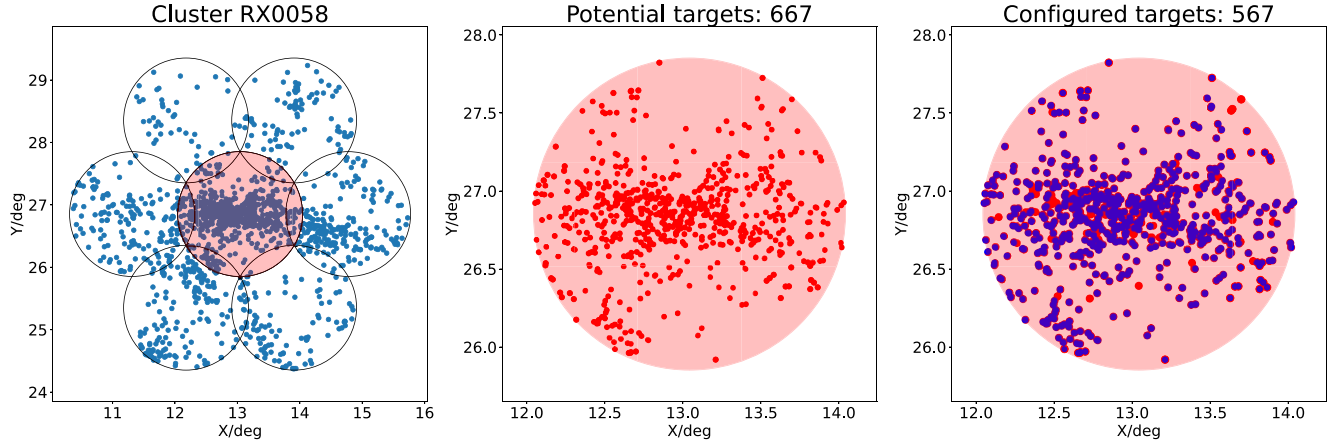
Figure 3. A demonstration of the process of 'configuration', the allocation of spectroscopic fibres to targets, on a simulated cluster mass-matched to WWFCS cluster RX0058 ($M \sim 4.3 \times 10^{14} M_\odot$, $R_{200} \sim 1.54$ Mpc), an average mass WWFCS cluster. Left-hand panel: Mock 'observation' of the simulated cluster containing eight 2-deg diameter WEAVE fields. Blue points show the positions on the sky of the simulated galaxies. Middle panel: Zoom-in on the central field from the plot on the left, with the simulated target galaxies before fibre configuration shown as red points. There are 667 simulated cluster members that are potentially observable in this field. Right-hand panel: The same central field after configuration, where 567 cluster members have been assigned a fibre (blue dots), while galaxies without a fibre are shown in red.
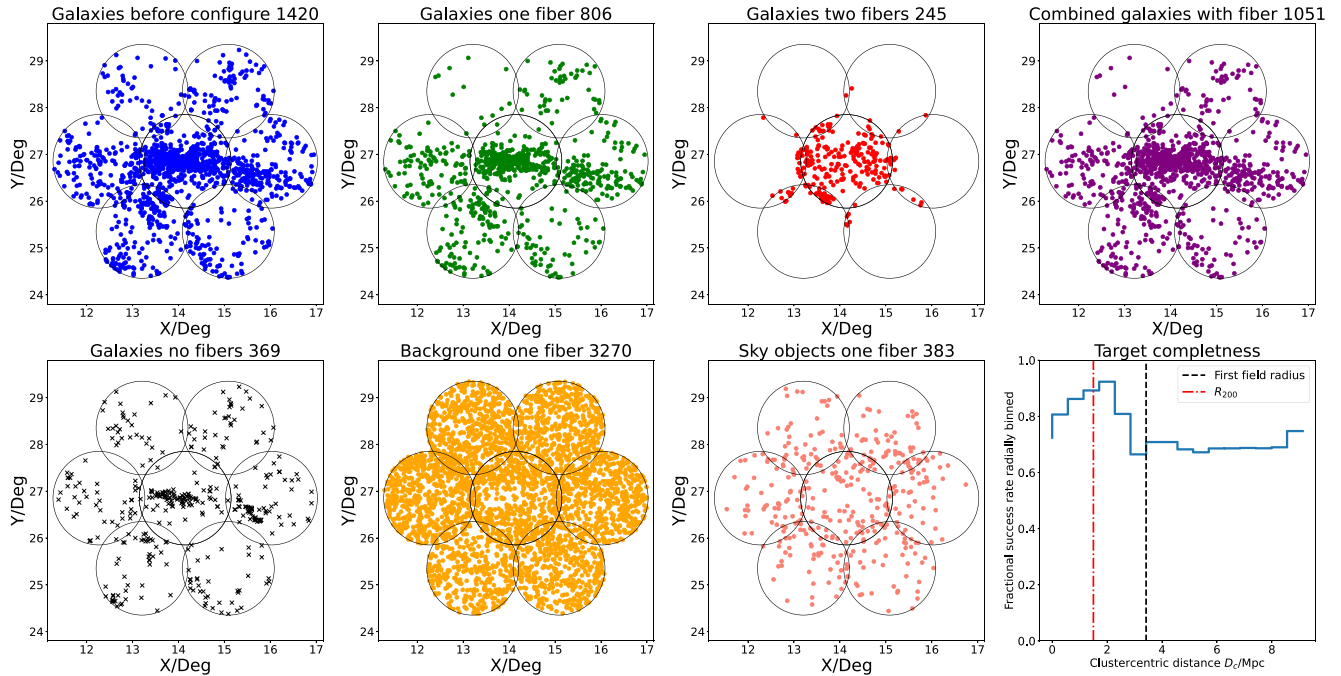
Figure 4. Results for the full configuration process of one simulated analogue to cluster RX0058. We plan to observe this cluster using eight WEAVE individual pointings (black circles). The central pointing will be observed twice given the high density of targets in the cluster core. For this cluster, 74 per cent of its simulated galaxies have at least one fibre allocated and 69 per cent background objects also have allocated fibres. Top left: All cluster members (blue dots) that could be assigned a fibre. Top middle-left: Galaxies with one fibre assigned (green dots). Top middle-right: Galaxies with two fibres assigned (red dots). Top right: combined sample of target galaxies with one or two fibres (purple dots). Bottom left: Galaxies with no fibre allocated (grey crosses). Bottom middle-left: background objects allocated one fibre (orange dots) and blank sky positions assigned one fibre (salmon dots). Bottom right: Fractional target completeness (i.e. fraction of cluster galaxy targets with one or two fibres; see Section 4.1) in clustercentric distance bins. R_{200} and the radius of the central WEAVE field are plotted for reference. The bin width is one-sixth of the radius of one WEAVE pointing (i.e. one-sixth of one degree).

Maximizing the number of galaxies with measured spectroscopic redshifts, particularly in the cluster outskirts and infall regions, is a key goal of the observational strategy that will enable a more accurate mapping of the cosmic web.

Each field intersects a minimum of three other fields, meaning that target objects in the overlap region have multiple chances of having a fibre allocated. To obtain information on data quality and repeatability, it is desirable to have some repeated observations, but we do not want these to have a significant impact on the final sample of galaxies with redshifts. We therefore aim at no more than ~10–20 per cent of the cluster galaxies to be observed twice and we chose to not artificially select an upper limit on the number of galaxies

observed twice as we want to minimize the number of empty fibers. We thus allow target galaxies to have at most two fibres allocated (in separate pointings), but sky positions and background galaxies are only allocated one fibre at most. This process is controlled by the `Configure` targeting priorities (Table 1).[2]

For all simulated clusters, we sequentially apply `Configure` to each WEAVE field (see Fig. 4). We start with the central one, which we configure twice, and then continue with the outer fields. After each step, we update the priorities for all objects in the target list taking into account whether an object (cluster galaxy, background galaxy, or sky position) has been allocated a fibre in a previous iteration. If a cluster galaxy has already been allocated one fibre, its priority is reduced to 1. If it has already been allocated 2 fibres, its priority goes to 0. Background galaxies and sky positions with fibres allocated previously get also a priority of 0 (Table 1). The process is illustrated with one example for the cluster RX0058 in Figs 3 and 4. Obviously, we are not able to allocate a fibre to each target cluster galaxy. In the typical cluster shown in these figures, 74 per cent of the cluster galaxies have at least one fibre assigned, with little radial variation beyond the radius of the inner WEAVE field. The success rate there is higher despite the higher density because this field is observed twice. Beyond that, no strong spatial biases are apparent, but we will analyse quantitatively the effect the configuration process has on our ability to map the large-scale structure and filaments around the clusters next.

### 3.5 Cosmic web extraction method

The rationale of this paper is to assess the ability of upcoming spectroscopic surveys such as the WWFCS to accurately map and characterize the cosmic web. We describe in this section the techniques we use for that purpose.

3.5.1 DisPerSE

To map the large-scale structure around clusters and extract filaments, we utilize the topological structures extractor `DisPerSE` (Sousbie 2011; Sousbie, Pichon & Kawahara 2011), which uses the concept of Morse theory (Stein et al. 1963) applied to matter distributions. It identifies structures, such as nodes, walls, knots, and voids as components of the Morse-Smale complex – the set of all ascending manifolds of the input function – and is able to classify regions using critical points and integral lines (nulls of the density field and tangents to the critical points). `DisPerSE` identifies topologically significant regions in the Delaunay-2D/3D tessellation density field by taking a set of discrete 2D or 3D positions (positions of haloes, mock galaxies, gas particles, etc.). It computes a set of segments representing the skeleton of the filamentary network.

In what follows, `DisPerSE` is run in 2D on the sky positions of the simulated cluster galaxies to mimic the observations since, uncertainties in the radial position of the galaxies due to peculiar velocities mean than filament extraction in 2D is preferable when redshifts (and not true distances) are available (Kuchner et al. 2020, 2021a).

`DisPerSE` analyses the topology of the pre-defined region and finds the saddle points that join together the main nodes set by a pre-defined 'persistence' threshold. The theory of persistence allows a user to account for uncertainty and Poisson noise in data

---
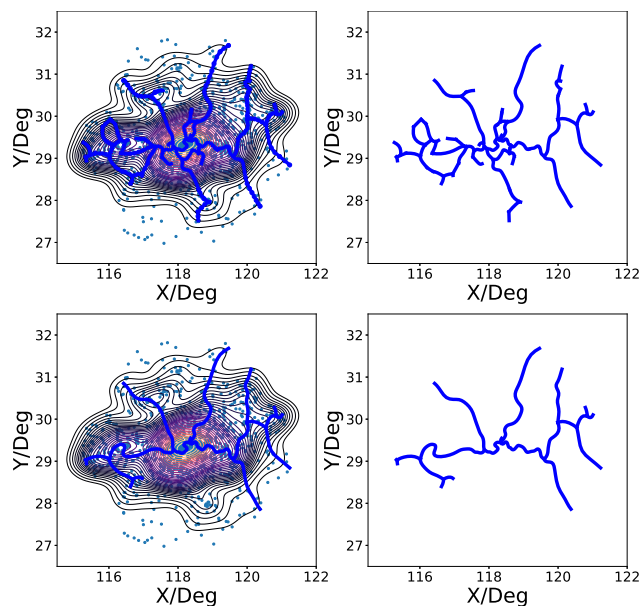
Figure 5. Example filament network extracted with different persistence thresholds. Top row: Persistence = 2σ. Bottom row: Persistence = 3σ. All plots correspond to the same cluster. Left column: KDE smoothed matter density distribution with the extracted filament network overlaid. Right column: Filament network only.

sets (in terms of σ) and is analogous to signal-to-noise ratio for observations. When extracting filaments with `DisPerSE`, setting a higher persistence threshold returns only the most robust, topologically significant, large-scale structure. Lower persistence values enable the detection of smaller tendrils. Therefore, there is a trade-off between the number of filaments that are extracted and their astrophysical significance. Fig. 5 demonstrates the implementation of different persistence thresholds. Both filamentary networks shown are derived by `DisPerSE` with the same input galaxy positions but using two different persistence values, 2σ and 3σ. It is evident that the least robust filament segments are not present in the network associated with higher persistence. Therefore, persistence strongly dictates the level of structure that is extracted, and it is important the its value is optimized for the scientific application intended. In the analysis that follows we will use a persistence of 2.5σ for the simulated reference network and 2.1σ for the network obtained from the analogue observations (after `Configure` is applied). We use different persistence values as the underlying density field will change upon target selection for each cluster. These choices are justified in Appendix C. There is also a smoothing parameter that is input into the Disperse runs that influences the rigidity of the identified filament paths. We chose a smoothing parameter of 5, as used in Kuchner et al. (2020), although using values between 1 and 5 has virtually no impact our results.

3.5.2 Dealing with boundary conditions

The complex 2D geometrical shape defined by the positions of the WEAVE pointings that will tile each cluster and its environment (Appendix A) influences the features that are detected by `DisPerSE`, particularly near the boundaries. If the shape of the field tiling is not properly accounted for, artificial nodes are detected that trace the outer boundary of the sky region covered by the WEAVE fields. To avoid that we populate the region outside the boundaries of the area
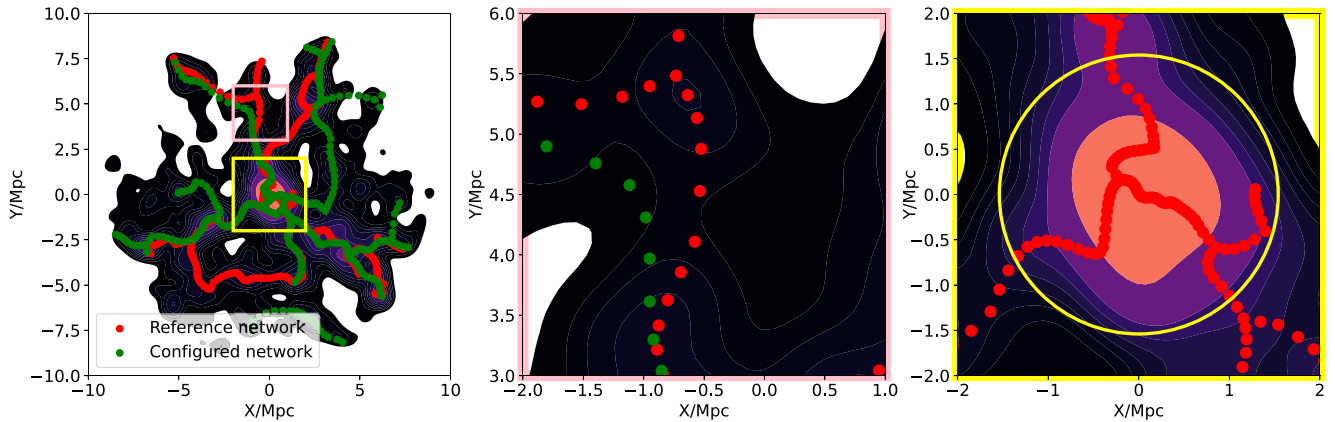
Figure 6. Illustration of the methods used to calculate D_{skel} and cosmic connectivity. Left-hand panel: The reference filament network (red) and 'configured' network (green) are plotted on top of the KDE-smoothed halo density distribution of a simulated cluster, an analogue to WWFCS cluster RX0058. A filament is the amalgamation of many discrete segments, as clearly seen in the middle panel. Middle panel: Zoom-in on the region shown by the pink box on the left-hand panel, where the two networks show large differences to demonstrate how D_{skel} is calculated (see Section 4.2.1). Right-hand panel: Zoom-in on the cluster core (yellow box in the left-hand panel), only plotting the reference network for illustrative purposes. The circle corresponds to R_{200} and is used to calculate the connectivity as the number of filaments that stem from the main node and cross the R_{200} circle (see Section 4.2.2). For this cluster, the connectivity has a value of three.
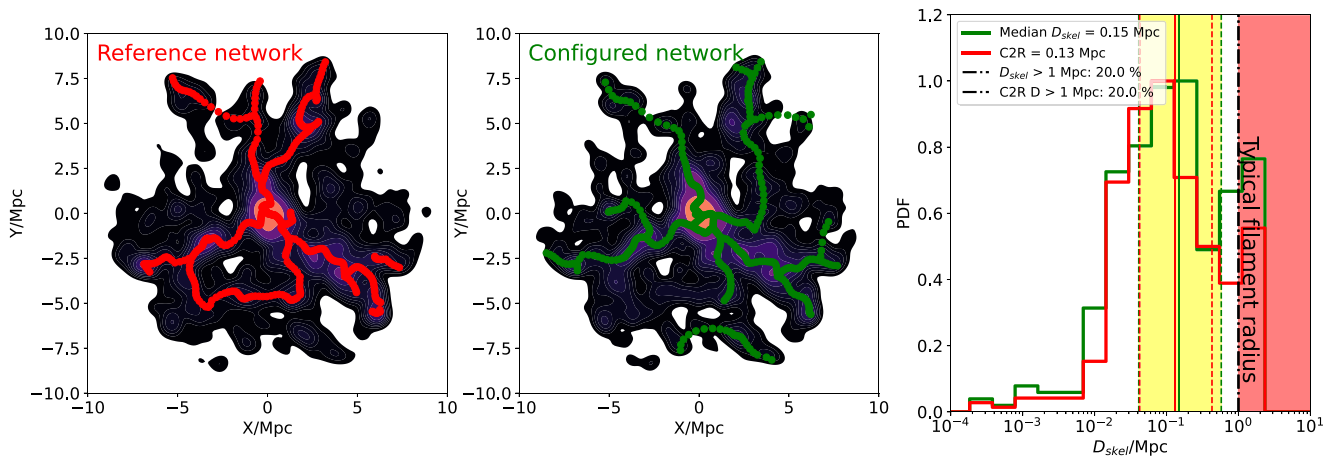
Figure 7. Illustration of the recovery of the filament network around a simulated analogue of the RX0058 cluster, the same cluster as Fig. 6. Left-hand panel: KDE-smoothed density distribution of the simulated cluster galaxies with reference filament network in red. Middle panel: As the left-hand panel, but showing the density distribution and filament network (green line) recovered from the 'configured' (mock-observed) galaxy sample. Right-hand panel: D_{skel} distribution function obtained going from the reference network to the 'configured' network (R2C) and vice-versa (C2R), as described in the text. The thick vertical line represents the medians of each distribution, while the shaded yellow region corresponds to the 25th and 75th percentiles. The dot–dashed black line represents the typical radius of a filament (~ 1 Mpc). The values of the medians and the percentage of segments with $D_{skel} > 1$ Mpc are shown in the legend. We normalized the PDFs to have a common peak value.

covered by the planned WEAVE pontings with a random uniform distribution of artificial galaxies that will act as 'guard particles' (cf. Sousbie 2011) to prevent the appearance of these artificial nodes and their associated filaments. The surface density of the artificial galaxies is chosen to be similar to that of cluster galaxies in the outer regions. In practice, the number density of galaxies that lie beyond $2R_{200}$ is computed for each cluster, and a random uniform distribution of 'guard particles' with this number density is added outside the outer boundary of the 'observed' fields, reaching $7.5R_{200}$. After testing different values for this radius, we find that the recovered networks are very similar when one increases the guard particle boundary beyond $7.5R_{200}$. This is sufficiently far away from the cluster centre to prevent the true filamentary network being distorted by the irregular boundaries.

The positions of cluster members and 'guard particles' are fed into `DisPerSE` and, once the filament network is computed by the filament finder, we truncate the network outside of the 'observed' region, keeping only the filament segments inside. This procedure works remarkably well, and visual inspection indicates that spurious nodes and filaments associated with the boundaries are eliminated without affecting the filament network inside the observed fields.

4 RESULTS AND DISCUSSION

With all the necessary elements in place, in what follows we will compare the filamentary networks that are extracted using `DisPerSE` before and after applying the MOS 'fibre configuration' process. In other words, we will quantify the difference between
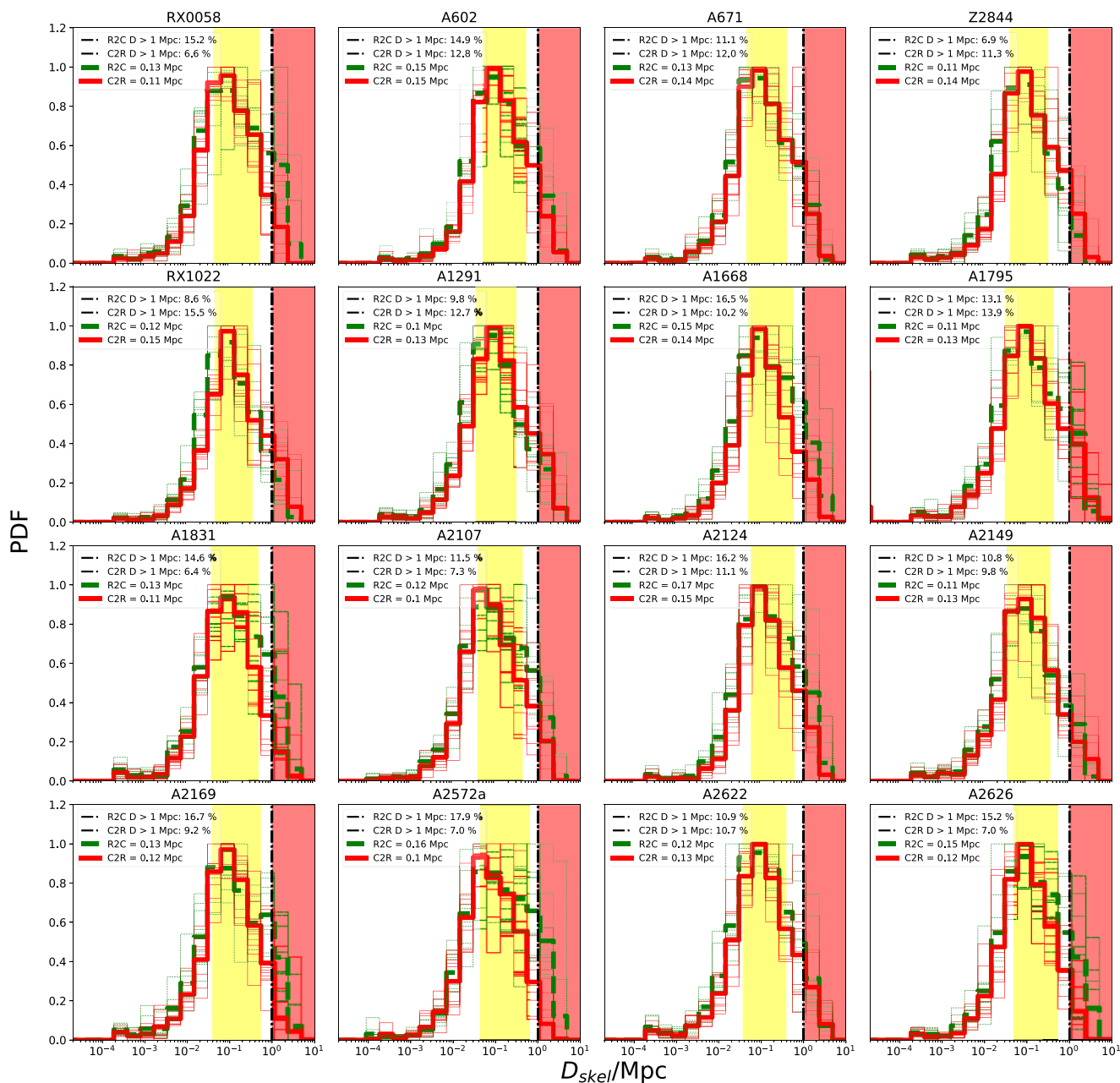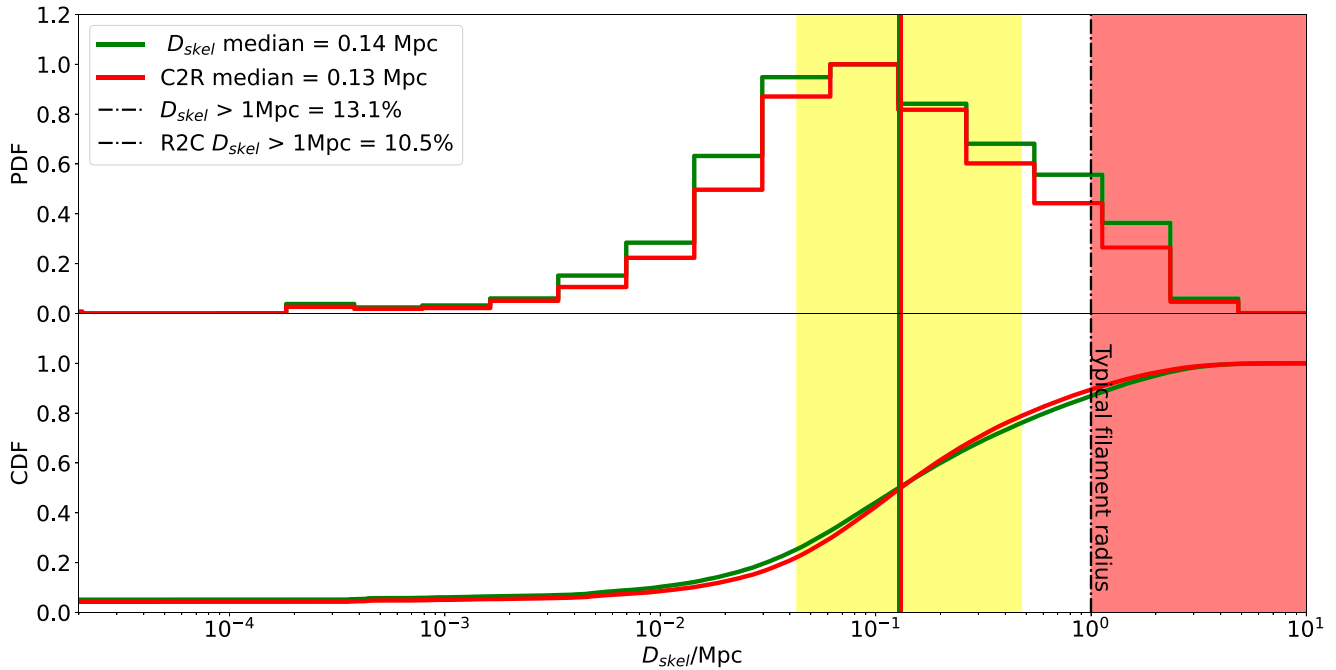
Figure 8. D_{skel} distributions for all the simulated WWFCS cluster analogues. Each panel shows the individual cluster comparison (thin lines) and the average for the 10 simulated cluster mass-matched to each WWFCS cluster (thick lines). The format of each panel follows that of the right-hand panel of Fig. 7. There is little variation in D_{skel} over different WWFCS analogue clusters and the D_{skel} median is always much less than a typical filament radius of 1 Mpc.

the filaments extracted when all simulated galaxies are fed into `DisPerSE` with those we obtained if we only use the 'mock-observed' galaxy sample, where some galaxies are lost due to fibre-positioning constraints. This will allow us to forecast the impact that realistic observational constraint will have on the information we can derive about the filamentary networks around clusters from spectroscopic survey like the WWFCS.

4.1 Recovery of cluster galaxies after `Configure`

Physical constraints from the fibre positioner imply that we will never have a 100 per cent completeness of cluster galaxies. Some galaxies won't be targeted as they will appear too close together as

well as the constraint of fibre overlap. Therefore, the first test to quantify the success of the WEAVE-like MOS fibre configuration is to estimate the fraction of simulated cluster galaxies with at least one fibre assigned. A high fraction – particularly outside the cluster core – will help us achieve our science goals. The overall average fraction of galaxies covered by the WEAVE pointings with at least one fibre allocated (overall target completeness) is 72.7 per cent ± 1.7 per cent, where the errors denote the scatter of the values for the 160 simulated clusters. If we restrict our calculation to the regions outside the central pointing, which is dominated by the cluster core, the corresponding fraction (outer target completeness) is 81.7 per cent ± 1.3 per cent. We argue that this value is more relevant than the overall one when dealing with the characterization of the filament network since the

Figure 9. Stacked probability density function (top panel) and cumulative density function (lower panel) of the D_{skel} distributions corresponding to all the simulated clusters shown in Fig. 8. The positional difference in the networks is minimal – our mock observations of filaments around galaxy clusters successfully recreate the 'true' simulated filament network.

whole cluster core will behave just as a single node (Kuchner et al. 2020, 2021a).

We have checked whether the fraction of galaxies selected by configure depends on galaxy mass. If we divide the galaxy sample at the median mass into two equal subsamples, we find that the fraction of high-mass galaxies that are 'configured' is ∼77 per cent, while the corresponding fraction of low-mass galaxies is ∼69 per cent. This is due to the fact that the central regions of the cluster, which contain a higher fraction of massive galaxies – high-mass galaxies cluster more strongly than low-mass ones – are observed twice. Beyond ∼$2R_{200}$ the fractions are approximately equal.

We find that, while the overall target completeness stays relatively constant as a function of cluster mass, the outer target completeness decreases slightly for higher cluster masses. This not surprising since the more massive clusters will have larger cores and a higher surface density of galaxies at all radii. In any case, the sample size reduction induced by the observational constraints seems moderate at all radii.

Note also that the very high number of WEAVE fibres will allow us to observe thousands of background objects per cluster (Fig. 4), providing a thorough test of the accuracy of our photometric redshifts and the quantification of any the possible biases their inclusion in the target selection may introduce (see Kuchner et al. in preparation).

4.2 Filament network comparison metrics

We have found that the completeness rates we find are encouragingly high, suggesting that the sample size statistics will not be very severely impacted by the observational constraints. We now need to check whether this sample reduction introduces any biases or changes in the properties of the recovered filamentary networks.

4.2.1 Skeleton distance

A useful metric designed to quantify the accuracy of filament extraction is the 'skeleton distance' D_{skel} (Malavasi et al. 2016; Laigle et al. 2017; Sarron et al. 2019). After running the cosmic web extractor software (DisPerSE in our case), we obtain a series of segments that delineate the cosmic filamentary structure. When comparing two different networks derived in the same region of space, D_{skel} measures the distance between the start of a segment in the reference network and the nearest one in the other network. The segments that are found are much smaller than the length of a typical filament, allowing us to use the position of the start of a segment as a proxy of a segments position. This is illustrated in Fig. 6. The left-hand panel shows two filamentary networks, the reference one derived from the full simulated cluster galaxy sample in red, and the 'configured' network recovered from the mock-observed galaxy sample in green. The middle panel shows an enlarged version of the pink-boxed region of the left-hand panel, where the differences between the red and green networks are largest, showing the individual segments. The right-hand panel illustrates how we calculate cluster connectivity and is discussed in Section 4.2.2. The D_{skel} values are calculated for each segment in the reference network by finding the distance to the nearest segment in the 'configured' network. Note that the calculation can also be done in the opposite direction, starting from the segments in the 'configured' network instead, and the distribution of D_{skel} values will not be necessarily the same (see below). In both cases, the distribution of D_{skel} values quantifies how well both filament networks match each other.

Fig. 7 shows the reference network (left-hand panel) and the 'configured' network (middle panel) for one of the simulated clusters mass-matched to one of the WWFCS target clusters, RX0058. The right-hand panel shows the normalized probability distribution function for D_{skel}, calculated going from the reference network to the

'configured' one (R2C, in green) and vice versa (C2R, in red). The median values are indicated. Both medians are much smaller than the typical radius of filaments (~ 1 Mpc; Kuchner et al. 2020). A large proportion of $D_{skel} > 1$ Mpc would indicate that a filament in this cluster has no counterpart in the corresponding mock-observational cluster. Note that, typically, the median $D_{skel, R2C}$ is smaller than the median $D_{skel, C2R}$ because there are generally more segments in the reference network than in the 'configured' one, and thus the likelihood of finding a nearer corresponding segment is higher in the R2C direction. If both networks are very similar, both D_{skel} median values will not only be very small but also very similar to each other. Therefore, the median values of D_{skel} and their ratio can be used to quantify the accuracy of filament network reconstruction and also to derive the optimal parameters used by DisPerSE (see Appendix C).

Another measure of the similarity between the reference and 'configured' filament networks is provided by the fraction of D_{skel} values that are larger than ~ 1 Mpc (the typical radius of filaments). The right-hand panel of Fig. 7 shows that this fraction is also reassuringly small (~ 10 per cent) in both cases.

4.2.2 Cluster connectivity

Another useful parameter to quantify the accuracy of the filament network derived from the mock observations is the cluster connectivity C. We define connectivity as the number of filaments that stem from the main node (cluster core) and terminate beyond R_{200} away from the cluster centre. This definition is slightly different from that of Laigle et al. (2017), where the authors use the cluster virial radius instead of R_{200}. The last panel in Fig. 6 gives an example of how C is calculated – there are three filaments stemming from the main node of the network (cluster core) that cross the circle with R_{200} radius, resulting in a cluster connectivity of three. A weak positive correlation between cluster connectivity and cluster mass has been reported in the literature (Darragh Ford et al. 2019; Kraljic et al. 2019; Sarron et al. 2019; Gouin, Bonnaire & Aghanim 2021), albeit with considerable scatter. Our simulated clusters show a similar correlation. If the recovered filament network is similar to the reference one, their connectivity should be the same. Therefore, comparing network connectivities will also allow us to assess the accuracy of the recovered filaments.

4.3 Quantifying the quality of the recovered filament networks

We are now in a position to use the D_{skel} and connectivity metrics to assess quantitatively the impact of the WWFCS observational strategy and constraints on the recovery of the filament networks surrounding galaxy clusters.

As described above, Fig. 7 illustrates the filament network comparison process for a single-simulated cluster, mass-matched to RX0058. Visually, there is remarkable similarity in the reference and 'configured' filament networks. For this particular cluster, 71.8 per cent of the simulated cluster galaxies have at least one MOS fibre allocated, which is close to the average for the whole sample. The median values of D_{skel} are 0.12 Mpc going in the R2C direction and 0.16 Mpc going in the C2R direction (cf. Section 4.2.1). These values are much smaller than ~ 1 Mpc, the typical radius of filaments. Moreover, only 8 and 10 per cent of the filamentary segments lie at a distances greater than ~ 1 Mpc.

The cluster connectivity of the reference network is 3, while one of the filaments in the 'configured' network bifurcates inside R_{200}, increasing the connectivity to 4. Changes in connectivity of ± 1 are
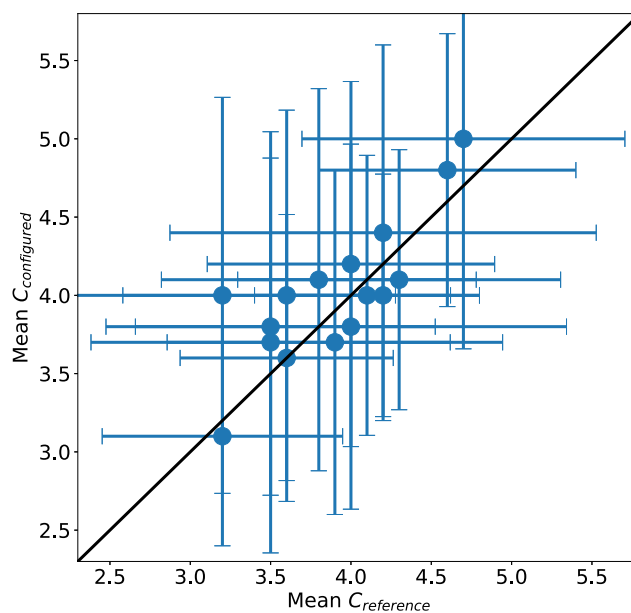
Figure 10. Comparison between the cluster connectivity C of the reference and 'configured' filament networks. The 1-to-1 line is shown in black. Each point corresponds to the mean connectivity of the 10 mass-matched analogues of each WWFCS cluster, and the error bars show the $\pm 1\sigma$ scatter.

not uncommon, indicating that the recovery is not perfect. However, larger changes in connectivity are rare (see below).

These results, if replicated for the whole cluster sample, are very encouraging, suggesting that the data provided by the WWFCS will allow the reliable recovery of the filamentary structures around clusters since the impact of the observational constraints will be moderate.

Fig. 8 confirms that the D_{skel} results shown for the RX0058 analogue are indeed typical of the whole sample. We can therefore stack the D_{skel} distributions for the 160 simulated clusters (Fig. 9) and derive representative average quantities for the whole sample. On average, the median values of $D_{skel, R2C}$ and $D_{skel, C2R}$ are 0.13 ± 0.02 Mpc. The values are not only reassuringly small but also almost exactly the same when going in both directions, strongly suggesting the compared filament networks are very similar. Furthermore, typically only 11–13 per cent of the corresponding filamentary segments are more that 1 Mpc away from each other.

Fig. 10 shows a comparison between the connectivity of the reference filament networks $C_{reference}$ and that of the 'configured' ones $C_{configured}$. The mean values corresponding to each WWFCS match very well, with no significant bias, and the small scatter indicates changes in C are generally kept within the ± 1 range.

In summary, the quantitative tests we have carried out for the whole cluster sample confirm our initial assessment that the impact of the observational constraints imposed by the WWFCS on the recovery of the filament networks around galaxy clusters will be very moderate.

5 SUMMARY AND CONCLUSIONS

The outskirts and infall regions of galaxy clusters act as the points of contact linking the large-scale structure of the Universe to the highly dense virialized cores of the clusters themselves, containing some of the key environments affecting galaxy evolution. Since next-

generation spectroscopic surveys such as the WWFCS[3] will explore and map in detail these complex regions, in this paper we forecast how successful such surveys will be at identifying the filaments that link together the 'nodes' in the large-scale structure – clusters and groups – and channel galaxies into them.

We aim at quantifying the impact the observational limitations will have on our ability to detect the filamentary structures that feed the clusters in the WWFCS. To achieve that aim, we have used a large sample of simulated massive galaxy clusters from `TheThreeHundred` project (Cui et al. 2018) and created a set of simulated cluster galaxy samples closely matching the selection and observational constraints imposed by the WWFCS (Kuchner et al. in preparation). For each one of the 16 WWFCS target clusters, we have extracted 10 mass-matched analogue clusters from the simulations and built mock-observed galaxy samples reaching beyond $\sim 5R_{200}$, where cosmic filaments trace and connect galaxy clusters to the cosmic web. We summarize our main results as follows:

(i) We have then followed closely the strategy, selection, and observational constraints of the WWFCS. Applying the same MOS fibre configuration tool that the WEAVE spectrograph will use, we find that, on average, we are able to allocate fibres to 72.7per cent $\pm$ 1.7percent of all the target galaxies. More importantly, outside the cluster core – in the outer regions that are crucial for filament identification – the success rate increases to $81.7\% \pm 1.3\%$. The number of cluster galaxies that are targeted ranges from 1284 to 4062. The high completeness that the WEAVE observations will allow, together with the large field coverage, are key to the success of the survey.

(ii) In each of the simulated cluster regions we have used the filament finder `DisPerse` (Sousbie 2011) to trace the cosmic-web filament skeleton before and after the observational constraints (including MOS fibre positioning) are imposed on the galaxy samples. We then compared quantitatively the resulting filament networks and find that we are able to recover the original network remarkably well. Specifically, we find that the median distance between corresponding filament segments D_{skel} in the reference and recovered networks is only 0.13 ± 0.02 Mpc on average, an much smaller than the typical filament radius of ~ 1 Mpc. Furthermore, only ~ 11–13 per cent of all recovered filament segments lie at a distance larger than 1 Mpc away from their corresponding reference segment.

(iii) As a further test on the integrity of the recovered filament networks, we computed their *connectivity*, the number of filaments that stem from the cluster core and terminate beyond R_{200} away from the cluster centre. We find that the connectivities of the reference and recovered networks match very well, without any significant bias, indicating that their global properties are also recovered well.

These findings make us confident that the WWFCS will be able to reliably trace cosmic-web filaments in the vicinity of massive galaxy clusters. The next step, when we start receiving data from WEAVE, will be to identify the galaxies that belong to these filaments, and compare their properties (e.g. mass, metallicity, star formation, and stellar populations) to those of galaxies inhabiting other environments such as groups, the clusters cores, and the general field. With the combination of a statistical sample of clusters together with high target sampling rate, the WWFCS will provide a detailed look at the influence of all environments in the cluster infall regions on galaxy evolution.

---

[3]Although in this paper we have focused on the WWFCS, our methodology could easily be adapted and applied to other wide-field spectroscopic surveys.

DATA AVAILABILITY

Data available on request to `TheThreeHundred` collaboration: https://www.the300-project.org.

REFERENCES

Aragón-Calvo M. A., van de Weygaert R., Jones B. J. T., 2010, MNRAS, 408, 2163

Arthur J. et al., 2019, MNRAS, 484, 3968

Balcells M. et al., 2010, Proceedings of the SPIE, 7735, 77357G

Balogh M. L. et al., 2017, MNRAS, 470, 4168

Beck A. M. et al., 2015, MNRAS, 455, 2110

---

[4]https://ingconfluence.ing.iac.es/confluence//display/WEAV/The+WEAVE+Project

[5]https://ingconfluence.ing.iac.es/confluence/display/WEAV/WEAVE+Acknowledgements

Castignani G. et al., 2021, A&A, 657, A9
Castignani G. et al., 2022, ApJS, 259, 43
Cautun M., van de Weygaert R., Jones B. J. T., Frenk C. S., 2014, MNRAS, 441, 2923
Cenarro A. J. et al., 2019, A&A, 622, A176
Colless M. et al., 2001a, MNRAS, 328, 1039
Colless M. et al., 2001b, MNRAS, 328, 1039
Cui W. et al., 2018, MNRAS, 480, 2898
Cui W. et al., 2022, MNRAS, 514, 977
Dalton G., 2016, in Skillen I., Balcells M., Trager S., eds, ASP Conf. Ser. Vol. 507, Multi-Object Spectroscopy in the Next Decade: Big Questions, Large Surveys, and Wide Fields. Astron. Soc. Pac., San Francisco, p. 97
Dalton G. et al., 2014, Proceedings of the SPIE, 9147, 91470L
Darragh Ford E. et al., 2019, MNRAS, 489, 5695
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Dressler A. et al., 1997, ApJ, 490, 577
Einasto M. et al., 2020, A&A, 641, A172
Fasano G. et al., 2006, A&A, 445, 805
Finn R. A. et al., 2005, ApJ, 630, 206
Galárraga-Espinosa D., Langer M., Aghanim N., 2022, A&A, 661, A115
Gill s. P., Knebe A., Gibson B. K., 2004, MNRAS, 351, 399
Gouin C., Aghanim N., Bonjean V., Douspis M., 2020, A&A, 635, A195
Gouin C., Bonnaire T., Aghanim N., 2021, A&A, 651, A56
Gouin C., Gallo S., Aghanim N., 2022, A&A, 664, A198
Gray M. E. et al., 2009, MNRAS, 393, 1275
Gullieuszik M. et al., 2015, A&A, 581, A41
Kaiser N., 1986, MNRAS, 222, 323
Kirkpatrick S., Gelatt C. D. V. M. P., 1983, Science, 220, 671
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Knebe A. et al., 2011, MNRAS, 415, 2293
Kraljic K. et al., 2019, MNRAS, 491, 4294
Kuchner U. et al., 2020, MNRAS, 494, 5473
Kuchner U. et al., 2021a, MNRAS, 503, 2065
Kuchner U. et al., 2021b, MNRAS, 510, 581
Laigle C. et al., 2017, MNRAS, 474, 5437
Liske J. et al., 2015, MNRAS, 452, 2087
Malavasi N. et al., 2016, MNRAS, 465, 3817
Malavasi N., Aghanim N., Tanimura H. B. V., Douspis M., 2020, A&A, 634, A30
Malavasi N., Langer M., Aghanim N., Galárraga-Espinosa D., Gouin C., 2022, A&A, 658, A113
Moretti A. et al., 2017, A&A, 599, A81
Mostoghiu R., Knebe A., Cui W., Pearce F. R., Yepes G., Power C., Dave R., Arth A., 2018, MNRAS, 483, 3390
Poggianti B. M., De Lucia G., Varela J., Aragon-Salamanca A., Finn R., Desai V., von der Linden A., White S. D. M., 2010, MNRAS, 405, 995
Rasia E. et al., 2015, ApJ, 813, L17
Rost A. et al., 2021, MNRAS, 502, 714
Salerno J. M. et al., 2020, MNRAS, 493, 4950
Sarron F., Adami C., Durret F., Laigle C., 2019, A&A, 632, A49
Secco L. F. et al., 2022, Phys. Rev. D, 105, 023515
Sembolini F., Yepes G., De Petris M., Gottlöber S., Lamagna L., Comis B., 2012, MNRAS, 429, 323
Song H. et al., 2020, MNRAS, 501, 4635
Sousbie T., 2011, MNRAS, 414, 350
Sousbie T., Pichon C., Kawahara H., 2011, MNRAS, 414, 384
Stein E., Milnor J. W., Spivak M., Wells R., Wells R., Mather J. N., 1963, Morse Theory. Princeton Univ. Press, Princeton, NJ
Terrett D. L., Lewis I. J., Dalton G., Abrams D. C., Aguerri J. A. L., Bonifacio P., Middleton K., Trager S. C., 2014, SPIE, 9152, 216
Vallés-Pérez D., Planelles S., Quilis V., 2020, MNRAS, 499, 2303
Vulcani B. et al., 2019, MNRAS, 487, 2278
Walker S., Lau E., 2022, preprint (arXiv:2202.07056)
York D. G. et al., 2000, AJ, 120, 1579
Zabludoff A. I., Mulchaey J. S., 1998, ApJ, 496, 39
Zel'dovich Y. B., 1970, A&A, 5, 84
Zinger E., Dekel A., Kravtsov A. V., Nagai D., 2018, MNRAS, 475, 3654
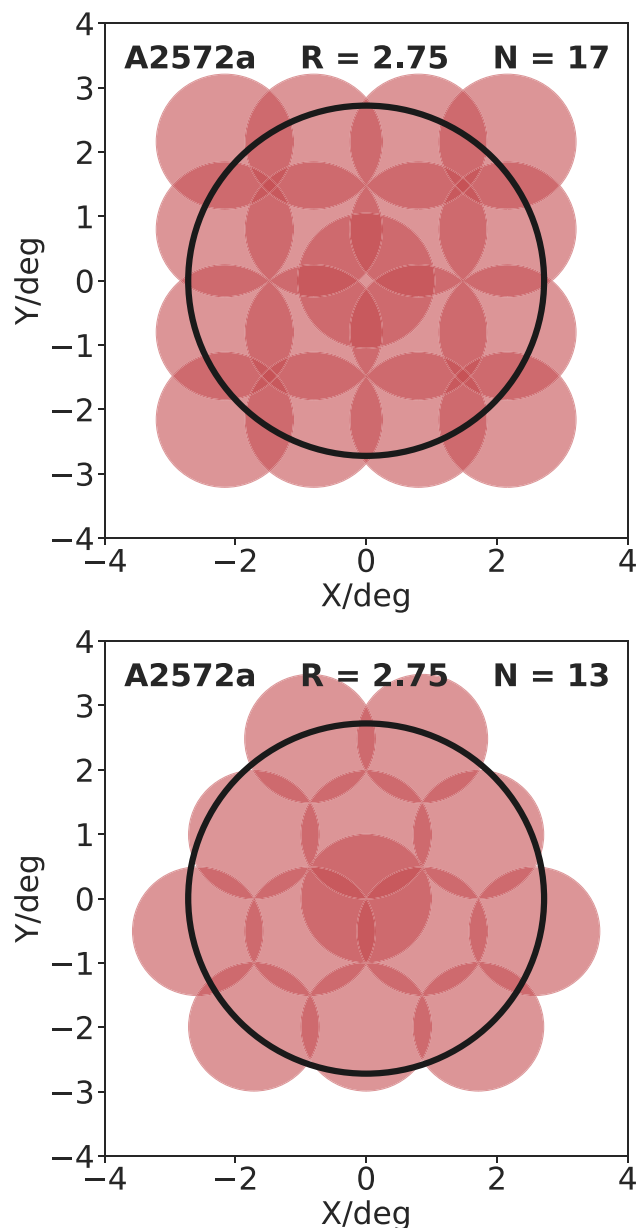
Figure A1. An example field layout of cluster A2572a before this work (top) and after this work (bottom). The red circles are individual 2-deg diameter WEAVE fields, while the black outer circle represents the angular diameter corresponding to $5R_{200}$ of this cluster. The numbers displayed are the total number of fields required to cover this cluster N, the cluster redshift and R_{200} taken from Moretti et al. (2017).

APPENDIX A: PROCESS OF OPTIMIZING THE WWFCS FIELD POSITIONS

As mentioned in Section 3.1, the WWFCS performs observations by arranging 2-deg fields into a mosaic pattern, covering the cluster core, infall region, and outskirts (Fig. 1). To optimize the observational strategy, we aim to design the field positions in a way that maximizes the cluster coverage. First, we place two fields at the core of each cluster, the region of the highest number density. This is so we can maximize the number of targeted cluster members over the total field of view. To optimize the tiling for each cluster, we adopt the following regime, such that if

$$\frac{\text{Area within } 5R_{200} \text{ for } N-1 \text{ fields}}{\text{Area within } 5R_{200} \text{ for } N \text{ fields}} > 97 \text{ per cent}, \qquad (A1)$$
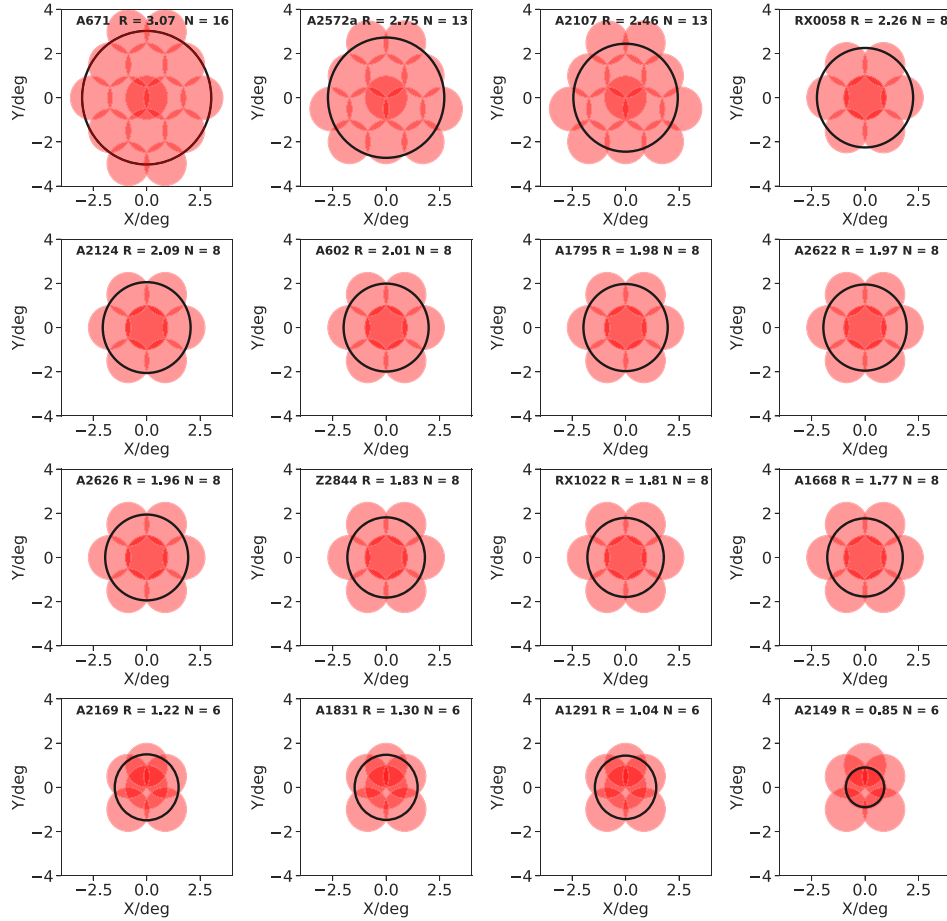
Figure A2. WEAVE field pattern for each WWFCS cluster. The caption in each panel states the name of the cluster, $5R_{200}$ in degrees and the number of fields N used for each observation.

then we can remove one field, (we use 97 per cent to ensure that we are still covering a significant area within $5R_{200}$). We iterate through this process by removing fields in the outer region of the cluster, manually inspecting each time one is removed, until the 97 per cent threshold is exceeded. Starting from a 'naive' geometric tiling pattern (illustrated in Fig. A1, top), the total number of WEAVE fields required to cover the 16 clusters was 155, adding up to 147 250 fibre hours. Using the new optimized tiling method (illustrated in Fig. A1, bottom) the total number of fibre hours is reduced to 130 390. For the example shown in Fig. A1, even though we have removed four fields, we are still covering out to $5R_{200}$. Of the clusters, 12 out of 16 have full coverage out to $5R_{200}$, while overall we have lost a total of 0.06 per cent area within $5R_{200}$. The final field positions for all 16 WWFCS clusters is shown in Fig. A2.

APPENDIX B: GALAXY CLUSTER SCALING

This sections details our method for scaling down the mass of the clusters from TheThreeHundred to match the WWFCS selected clusters, as mentioned in Section 3.3.1.

First, we arbitrarily chose three mass-scaling factors: $M_F = 1, 2$, and 5, which the simulated cluster mass is divided by. We chose the most-massive halo in the corresponding cluster catalogue to act as a proxy for the cluster. Also, we increase our cluster catalogue sample size by a factor of 3 by including each 2D-plane (*xy*, *xz*, *yz*). The resulting mass distributions of the clusters that have been scaled down by M_F are in Fig. 2, where we have demonstrated that

by choosing these mass factors, we have covered the entire WEAVE mass range.

For each scaling factor M_F, we divide the mass range into 20 mass bins. To create a statistically significant sample, we draw from these bins with the aim of identifying 10 mass-matched simulated analogue clusters for each of the 16 WEAVE clusters. Our total sample of analogue clusters is thus 160. The presence of companion clusters within $5 < R < 15$ Mpc of the WWFCS clusters (the radius of the simulation volume) is unknown, therefore we do not exclude analogue clusters with secondary clusters within this distance. Within the sample of 160 analogue clusters there are six such configurations.

In scaling the mass of the clusters from the simulations to match WEAVE, we have to also individually scale the mass of all the associated haloes for each cluster. Fig. B1 displays the cumulative number of simulated haloes that lie above a mass interval for one cluster. We see that for higher mass scaling factors (M_F) we are shifting the masses of all the haloes associated with the cluster to lower values.

We are limited in our recovery of haloes by two thresholds: The 'scaled' simulation mass resolution and the observational mass resolution. In the case where we do not scale the simulations, ($M_F = 1$), the mass resolution limit is that of the dark matter particles in the simulations (Kuchner et al. 2021a), given by the dotted grey line in Fig. 4. For $M_F > 1$, we reduce the simulation mass resolution by dividing it by the mass scaling factor M_F. While we can change the simulated mass resolution threshold, there is a hard limit on the observational mass limit. For $F > 3$, the mass resolution stays at the
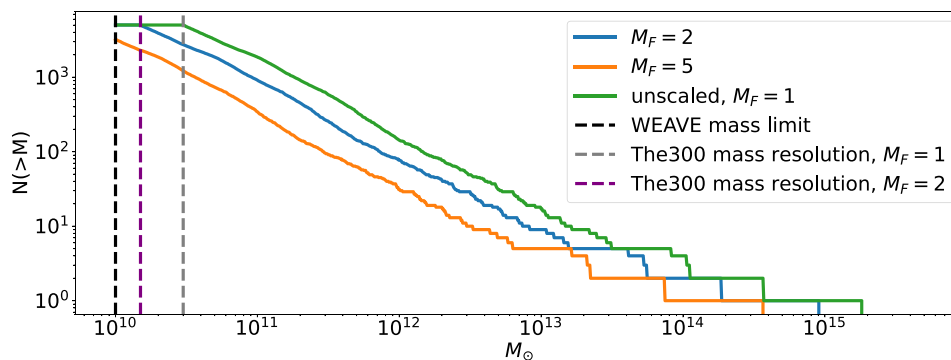
Figure B1. Cumulative sum of individual halo masses of simulated galaxies for one galaxy cluster in `TheThreeHundred` that has been modified using the three M_F. All haloes follow the criterion set for selecting 'high-quality data' from the simulations, as described in Section 3.3. Different colours represent different scale factors that have been applied to the simulated catalogues. The vertical dotted lines represent different mass resolutions.

observational limit $M_{obs} = 10 \times 10^{10} M_\odot$ (corresponding to the r band limit used for WEAVE: $r_{total} < 19.75$ which is equivalent to a stellar mass of $10^9 M_\odot$ (Kuchner et al. in preparation). At $M_F > 3$, we artificially lose haloes that are massive enough to be simulated, but are too small to meet the WWFCS observational criteria. However, as shown by the top panel in Fig. 2, we are still obtaining thousands of cluster galaxies per cluster. This reduction in the cluster member population mimics the expected relation of lower mass cluster's hosting fewer subhaloes (see e.g. Poggianti et al. 2010).

APPENDIX C: DISPERSE INPUT PARAMETERS

As mentioned in Section 3.4, to extract the filament networks with DisPerSE we need to set a persistence threshold. We introduce a metric Ψ with three key parameters:

(i) D_{skel} median (the positional difference in the reference network spine and the configured network spine),
(ii) D_{skel} ratio: the ratio of the two methods of calculating D_{skel},
(iii) Cluster connectivity of 'configured' network (C, number of filaments that stem from the main node and terminate outside R_{200}) comparison to reference network.

All of the above parameters are normalized against their maximum output given a persistence and are equally weighted. We compute this metric in a suitable range of persistences $2 < \sigma < 3$, in 0.1σ intervals, and the persistence that minimizes Ψ returns the most accurate reconstruction of the filamentary network. A Ψ of three corresponds to the worst possible reconstruction of the network, while Ψ values close to zero represent the most accurate filamentary mapping. We have automated a process in determining the best persistence given a reference skeleton. Our scientific rationale requires a high completeness, therefore, we explore low persistence values that map out not only the most robust structure but also filaments that connect nodes with smaller persistence ratios. After analysing Ψ for different networks for different clusters, we selected a persistence of $\sigma = 2.5$ for the reference network and $\sigma = 2.1$ for the configured network. In the process of 'configuring' a cluster, we are effectively altering the underlying density field and therefore, it is necessary to change the input persistence, as demonstrated in Fig. C1. Although varying the persistence cluster by cluster can change D_{skel}, the median and PDFs of D_{skel} do not vary significantly with changes in persistence when averaged over all 160 clusters.
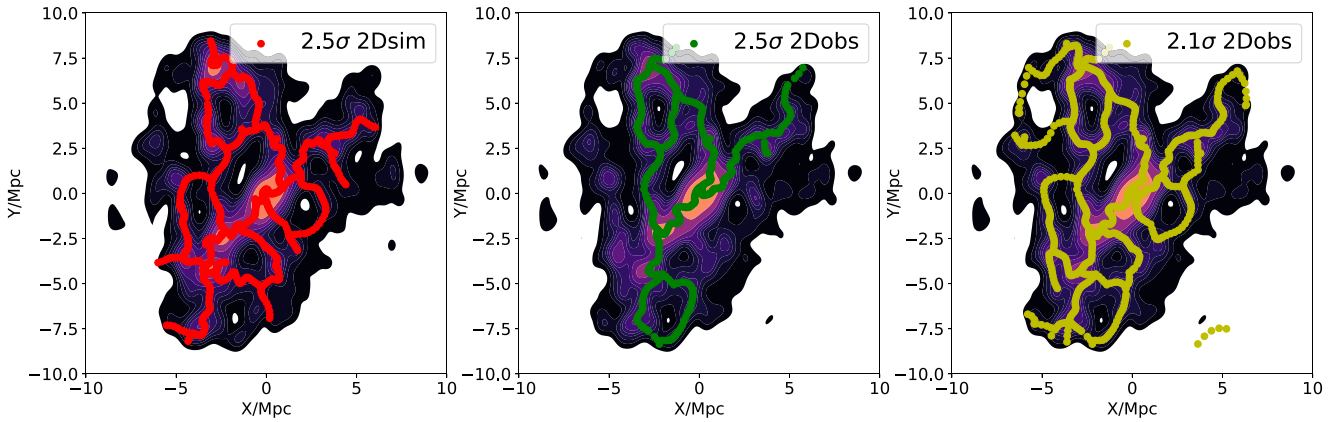
Figure C1. This figure illustrates how the change in the density field when 'configuring' a cluster makes a change in the persistence parameter necessary when extracting the filament network. Left-hand panel: KDE-smoothed density field of a simulated cluster analogue to WWFCS target RX0058, with the filaments identified by the topological structures extractor DisPerSE traced on top (using our choice of persistence for the 'pre-configured' model, 2.5σ. Middle panel: The 'mock-observational' cluster with filaments extracted using the same persistence, 2.5σ. Right-hand panel: The same but with our choice of persistence for the 'configured' model, 2.1σ. Lowering the persistence yields a more accurate reconstruction of the cosmic web around galaxy clusters.

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