The probability of identifying the cosmic web environment of galaxies around clusters motivated by the Weave Wide Field Cluster Survey

Daniel J. Cornwell [1]★ Alfonso Aragón-Salamanca [1], Ulrike Kuchner [1], Meghan E. Gray,[1] Frazer R. Pearce[1] and Alexander Knebe [2,3,4]

[1]*School of Physics and Astronomy, University of Nottingham, Nottingham NG7 2RD, UK*
[2]*Departamento de Física Teórica, Módulo 15, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain*
[3]*Centro de Investigación Avanzada en Física Fundamental (CIAFF), Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain*
[4]*International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia*

ABSTRACT

Upcoming wide-field spectroscopic surveys will observe galaxies in a range of cosmic web environments in and around galaxy clusters. In this paper, we test and quantify how successfully we will be able to identify the environment of individual galaxies in the vicinity of massive galaxy clusters, reaching out to $\sim 5R_{200}$ into the clusters' infall region. We focus on the WEAVE Wide Field Cluster Survey (WWFCS), but the methods we develop can be easily generalized to any similar spectroscopic survey. Using numerical simulations of a large sample of massive galaxy clusters from THETHREEHUNDRED project, we produce mock observations that take into account the selection effects and observational constraints imposed by the WWFCS. We then compare the 'true' environment of each galaxy derived from the simulations (cluster core, filament, and neither core nor filament, 'NCF') with the one derived from the observational data, where only galaxy sky positions and spectroscopic redshifts will be available. We find that, while cluster core galaxy samples can be built with a high level of completeness and moderate contamination, the filament and NCF galaxy samples will be significantly contaminated and incomplete due to projection effects exacerbated by the galaxies' peculiar velocities. We conclude that, in the infall regions surrounding massive galaxy clusters, associating galaxies with the correct cosmic web environment is highly uncertain. However, with large enough spectroscopic samples like the ones the WWFCS will provide (thousands of galaxies per cluster, out to $5R_{200}$), and the correct statistical treatment that takes into account the probabilities we provide here, we expect we will be able to extract robust and well-quantified conclusions on the relationship between galaxy properties and their environment.

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

1 INTRODUCTION

Galaxies are not distributed through the Universe randomly. Instead, they are constituents of the larger-scale cosmic web. This network features nodes – peaks in the matter density field, as well as sheets, filaments, walls, and voids (Aragón-Calvo, van de Weygaert & Jones 2010). First described in Zel'dovich (1970), these structures are ubiquitous and form from the anisotropic gravitational collapse of dark matter.

As galaxies assemble their mass and evolve, they may experience a range of environments which can significantly influence their properties. Most notably, galaxies in the densest cosmic environments, (i.e. galaxy clusters), frequently interact with other galaxies, the intracluster medium (ICM), and the cluster tidal field. During their lifetimes, galaxies may also travel through cosmic sheets and filaments, where their properties can be affected long before they reach the clusters themselves. This is often referred to as 'pre-processing' (Zabludoff & Mulchaey 1998). Processing and pre-processing can lead to strong environmental differences in the properties of galaxies, such as the Morphology–Density relation (Dressler 1980) that describes the finding of an increased fraction of early-type galaxies in denser environments.

Since galaxy clusters are part of the cosmic web, they co-evolve with it and with the galaxies that live in them. As clusters grow through galaxy infall and mergers with other clusters and groups, their dynamical state changes, and this can also influence the environmental effects their galaxies experience (Ribeiro, Lopes & Rembold 2013; Morell et al. 2020; De Luca et al. 2021). Galaxy groups and their central galaxies can also be significantly affected by the presence of filamentary structures that feed into them (Poudel et al. 2017). It is therefore clear that a complete understanding of how the environment affects galaxy evolution requires a thorough mapping and study of the cosmic web around galaxy clusters.

The importance of clusters and filaments is emphasized by the fact that, even though filaments and clusters only contain 6 and

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

0.1 per cent, respectively of the volume of the present-day universe, they harbour 50 and 11 per cent of the mass (Cautun et al. 2014). Since the physical processes galaxies experience depend on environment, it is vital not only to map these environments accurately, but also to be able to find well-defined subsets of galaxies in each environment.

The identification of cosmic filaments has been rigorously tested and applied in large-area surveys where a variety of detection methods have been adopted, such as the widely used discrete persistent structure extractor tool `DisPerSE` (Sousbie 2011; Sousbie, Pichon & Kawahara 2011). `DisPerSE` has been successfully applied to optical surveys such as SDSS (Malavasi et al. 2020) and GAMA (Kraljic et al. 2017). Cosmic filaments are also detectable in X-rays (Vernstrom et al. 2021) and follow-up X-ray studies of filaments found in the SDSS has resulted in a 5σ detection of their X-ray emission (Tanimura et al. 2022). Several other geometrical web extractors exist that use alternative methods (see Libeskind et al. (2017) for a review on cosmic web tracing algorithms). Whilst these cosmic web finders are effective at mapping these environments over scales of hundreds of Mpc, mostly far from galaxy clusters, observationally characterizing the environment in the vicinity of massive clusters remains a challenge. This is due primarily to the complexity of the infall regions around clusters – where multiple filaments converge – and the large peculiar velocities induced by the cluster dynamics (Tempel et al. 2016; Kuchner et al. 2021). It is therefore clear that mapping the cosmic web in the vicinity of clusters requires special attention.

This will be addressed by next-generation wide-field spectroscopic surveys covering regions of tens of Mpc around galaxy clusters. Examples of such surveys include the WEAVE Wide Field cluster Survey (WWFCS; Jin et al. 2023, Kucher et al. in prepartion) and the 4MOST CHileAN Cluster galaxy Evolution Survey (CHANCES, Haines et al. in preparation). This paper focuses on the first of these survey, which is an upcoming wide-field spectroscopic study of ∼20 galaxy clusters to be undertaken at the 4.2 m William Hershel telescope with the newly commissioned WEAVE spectrograph (Jin et al. 2023). The WWCS will obtain thousands of spectra for galaxies in and around each cluster, reaching out to several virial radii from the cluster cores. This will allow a thorough investigation of the cosmic web around galaxy clusters and the properties of the galaxies in it.

To fully exploit the extensive data generated by the WWFCS, we are developing specific analysis techniques to test and optimize the detection and characterization of the filamentary networks around the clusters (Kuchner et al. 2020, 2021; Cornwell et al. 2022; Kuchner et al. 2022) using simulated clusters from THETHREEHUNDRED project (Cui et al. 2018, 2022, see below for more details). Specifically, Kuchner et al. (2021) investigated the practicalities of extracting filament networks around the clusters in the presence of the observed 'Fingers of God' (FoG) due to the galaxies' peculiar velocities. They concluded that, because the distance uncertainties induced by these peculiar velocities are comparable with the depth of the volume explored, filament extraction near galaxy clusters need to rely on two-dimensional projections on the sky. This does not mean that the spectroscopic redshifts are not necessary – they are crucial to reliably select the galaxies that belong to the relevant volume around the cluster with an accuracy of ∼ 10 Mpc. Such accuracy cannot be achieved with photometric redshifts.

Bringing the THETHREEHUNDRED simulations one step closer to observations, Cornwell et al. (2022) demonstrated that filaments can be successfully extracted from data sets similar to those expected from the WWFCS. They did that by creating mock observations from the simulated galaxy samples and considering the effects of sample selection and completeness resulting from the spectrograph's

fibre allocation and spatial coverage. They found that the filament networks extracted in the simulations are well reconstructed in the mock observations. In this paper, we go an additional step forward and quantify our ability to allocate galaxies to the different environments around the simulated mock-observed clusters. The resulting statistics will be necessary when studying the properties of galaxies as a function of environment in a robust statistical way.

This paper is structured as follows. Section 2 describes the (expected) observational and simulated data sets. Section 3 presents the reference filament networks derived from the simulated galaxy samples and defines the 'true' environment of each galaxy. Section 4 describes how the filaments are found in the mock WWFCS-like observations, and how the 'observed' environment of each galaxy is determined. Section 5 compares the 'true' and 'observed' environments of the galaxies, and provides the necessary statistics to quantify the success (or otherwise) of the comparison. Section 6 summarizes the conclusions.

2 DATA SET

2.1 WEAVE Wide Field Cluster Survey

WEAVE (the William Herschel Telescope Enhanced Area Velocity Explorer) is a new optical multi-object spectrograph on the 4.2 m William Herschel Telescope (Balcells et al. 2010; Dalton et al. 2014; Dalton 2016; Jin et al. 2023). The WEAVE Wide-Field Cluster Survey is one of the surveys that will be carried out with WEAVE. It is designed to yield thousands of galaxy spectra in and around galaxy clusters (Kuchner et al. in preparation). The WWFCS will take advantage of the multiplex capability provided by ∼1000 individual fibres deployable over a 2-degree-diameter field-of-view to observe up to 20 low-redshift ($0.04 < z < 0.07$) galaxy clusters. The r-band magnitude limit of the WWFCS spectroscopic observations is 19.75, corresponding to an approximate galaxy stellar mass limit of $10^9 \, \mathrm{M_\odot}$. The WWFCS will use the low spectral resolution mode ($R \sim 5000$) to obtain optical spectra in the 366 nm $< \lambda <$ 959 nm range. For each cluster, we will obtain spectra for several thousands of potential cluster members, reaching beyond ∼$5R_{200}$ from the cluster centre. These spectra will not only provide accurate redshifts, but also yield key information on the stellar populations and AGN properties of the galaxies.

Even though each WEAVE field can target close to one thousand objects in a single pointing, and we will observe each cluster with several pointings, it is not feasible to target every single galaxy in the cluster region due to instrumental limitations that include potential fibre collisions and overlap. WEAVE uses an algorithm named `Configure` (Terrett et al. 2014) to optimize fibre allocation. It was shown by Cornwell et al. (2022) that `Configure` is able to allocate fibres to ∼ 75 per cent of the target galaxies overall, reaching 81 per cent outside R_{200}. As shown by these authors, this is enough to reliably reconstruct the filament networks around the clusters. In the current paper, we base our analysis on mock-observed (i.e. 'configured') simulated galaxy samples in clusters and their outskirts.

2.2 THETHREEHUNDRED simulations of galaxy clusters

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

---

[1] https://the300-project.org/

cube of comoving length $1\,h^{-1}$ Gpc containing 3840^3 dark matter particles, each with mass $1.5 \times 10^9\,h^{-1}\mathrm{M}_\odot$. MDPL2 uses *Planck* cosmology ($\Omega_M = 0.307$, $\Omega_B = 0.048$, $\Omega_\Lambda = 0.693$, $h = 0.678$, $\sigma_8 = 0.823$, $n_s = 0.96$; Planck Collaboration XIII 2016).

This project locates the 324 most massive haloes ($M_{200} > 8 \times 10^{14}\,h^{-1}\,\mathrm{M}_\odot$) at $z = 0$, it then follows them back to their initial conditions, and resimulates the hydrodynamics in a $15\,h^{-1}$ Mpc comoving sphere around the identified massive halo. The highest resolution dark matter particles are then split into dark matter and gas, following the cosmological baryonic mass fraction using the Planck 2015 cosmology $m_{DM} + m_{gas} = 1.5 \times 10^9\,h^{-1}\,\mathrm{M}_\odot$. We use the outputs of the zoom resimulations ran using GADGET-X, which 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 and we take the redshift $z = 0$ snapshot which is comparable to the low redshifts of the WWFCS clusters. For the purpose of this work, we take the galaxy's stellar mass to be one tenth of the dark matter halo mass, as justified in Kuchner et al. (2020). Our team has successfully exploited the THETHREEHUNDRED simulations to plan the WWFCS observations and make predictions about the properties of the observed galaxy samples (Kuchner et al. 2020, 2021; Cornwell et al. 2022), demonstrating that reliable mock observations that mimic the WWFCS can be generated from this suite of simulations. Specifically, in Cornwell et al. (2022) we mass-matched each of the 16 main clusters targeted by the WWFCS with 10 clusters in TheThreeHundred simulations, and mock WWFCS observation were created for all 160 of them. As explained above, we use Configure to mock-observe the galaxies in and around the mass-matched clusters. Going forward, we refer to these cluster galaxy samples as the 2D mock-observations. Throughout this paper, we will also use the galaxies in the full 3D simulated clusters before 'configurations' and refer to these as the 'true simulated cluster galaxy samples'.

In the next section, we outline how we identify the 'true' cosmic web structures in the simulations that we will later compare with the ones extracted from the mock observations.

3 DEFINING COSMIC WEB ENVIRONMENTS

3.1 Cosmic web extraction

Following Cornwell et al. (2022), we employ the widely used structures extractor algorithm DisPerSE (Sousbie 2011; Sousbie et al. 2011) to identify filaments in the simulation boxes at $z = 0$. To match the expected depth of the WWFCS, we use as 'mock galaxies' all the halos in the simulations with a dark matter mass larger than $10^{10}\,\mathrm{M}_\odot$, approximately corresponding to stellar masses larger than $10^9\,\mathrm{M}_\odot$[2] DisPerSE identifies persistent topological features in the underlying density field defined by the mock galaxies. These features include critical points such as peaks and segments that link the peaks along the local geometry of ridges, which are the filaments. In Cornwell et al. (2022), we explain in detail how we use DisPerSE to extract robust filamentary networks around the simulated clusters, including our choice of input parameters. Following Kuchner et al. (2020), we use mass-weighting since this option makes the filamentary networks more reliable. This choice has some observational support from the fact that galaxies in filaments

tend to be more massive (as well as redder and less star-forming) than field galaxies away from them (Malavasi et al. 2016; Chen et al. 2017; Kraljic et al. 2017; Laigle et al. 2017; Sarron et al. 2019).

3.1.1 3D reference filament networks in simulations

In order to get consistent and reliable 'true' filament networks from the simulations, we need to choose some critical input parameters for DisPerSE. Building on the work of Cornwell et al. (2022) and Cornwell et al. (in preparation), we use a *persistence* value of 4.6σ and a *smoothing* of 5 (see Sousbie 2011 and Sousbie et al. 2011 for a definition of these parameters). With the persistence ratio threshold, we vary the robustness or significance of filaments to local variations in the density field. In our case, we intend to extract the primary filaments that are responsible for the majority of cluster accretion. Note that, the use of mass-weighting requires a higher *persistence* value to obtain networks that are similar to non-mass-weighted networks obtained with a lower *persistence*. The second parameter, *smoothing*, creates smoother networks by averaging the position of each vertex with that of its direct neighbours.

3.1.2 2D mock-observational filament networks

We also need to define filament networks using a mock-observational data set similar to the one WEAVE will provide (cf. Cornwell et al. 2022). The main two differences from the original simulations are that we will not observe every single galaxy (see Section 2.1) and that we will not have radial distances but only redshifts (or radial velocities). As explained in Cornwell et al. (2022), despite these limitations, reliable 2D filamentary networks (projected on the plane of the sky) can be extracted from the simulated data sets. Again, we use the mass-weighted DisPerSE algorithm to identify the components of the cosmic web in 2D and chose a *persistence* of 2.6σ and a *smoothing* of 5.

The 3D and 2D filament networks around a simulated example cluster are shown in Fig. 1.

3.2 Filament thickness

One important issue to consider when allocating galaxies to filaments is the thickness of the filaments themselves. In other words, how close to the spine of a filament does a galaxy have to be in order to be considered a 'filament galaxy'? This is not trivial since filaments do not have sharp boundaries. Some studies show that the thickness of filaments may depend on their length: longer filaments may be thinner, on average, than shorter ones (Malavasi et al. 2020). Additionally, Rost et al. (2021) used the gas and dark matter distributions from THETHREEHUNDRED to suggest that filaments are the thickest closest to the nodes. Kuchner et al. (2020) used the transverse gas density profile of filaments in the same simulations and derived a characteristic filament radius ~ 0.7–$1\,h^{-1}$ Mpc for massive clusters. However, there is no theoretical or observational motivation for holding the filament thickness constant for the networks surrounding different clusters. The mass range of the clusters selected for the WWFCS spans more than an order of magnitude (Cornwell et al. 2022; Jin et al. 2023; Kuchner et al. in preparation). If the thickness of the filaments surrounding clusters with different masses is kept constant, the fraction of the total volume occupied by filaments inside a sphere of radius $\sim 5R_{200}$ would be significantly larger for less massive clusters. Self-similarity considerations on the dark matter distributions suggest that this should not be the case. In the absence

---

[2]We note that the halo and stellar mass limits are slightly higher for more massive clusters, as described in Cornwell et al. (2022)
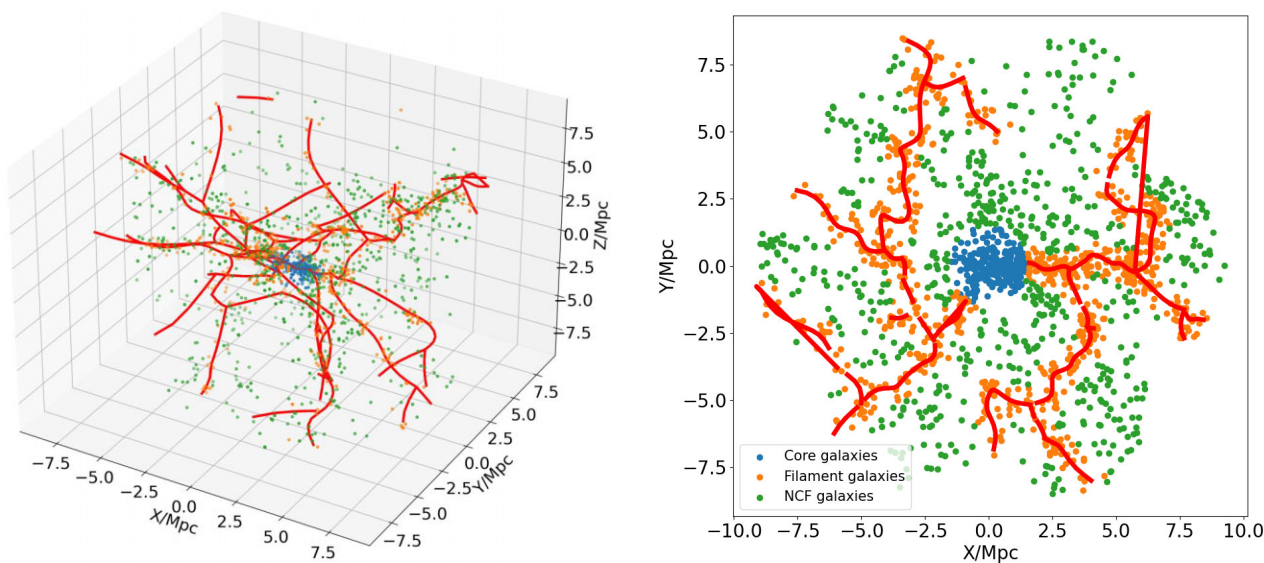
Figure 1. Illustration of an example cluster with the different cosmic web environments identified in 3D (directly from the simulations) and in 2D (from the mock observations based on WWFCS observational constraints). The environment assigned to a given galaxy may not be the same in 2D and 3D due to projection effects. Cluster core galaxies are shown in blue, filament galaxies in orange, and the rest of the galaxies (neither core nor filament, NCF) in green. We also overlay the associated filament network around this cluster, shown by the red lines. This simulated cluster has similar mass to RX0058, one of the targets in the middle of the mass range of the WWFCS sample.

of stronger evidence, it is reasonable to assume that the thickness of filaments surrounding a given cluster scales with the cluster's R_{200}, and therefore with $M_{cluster}^{1/3}$. Hence,

$$R_{fil} = R_0 * \left(\frac{M_{cluster}}{\langle M_{THETHREEHUNDRED} \rangle} \right)^{1/3}, \qquad (1)$$

where $R_0 = 0.7\, h^{-1}$ Mpc is the average filament radius calculated using the full THETHREEHUNDRED cluster simulations (Kuchner et al. 2020), $\langle M_{THETHREEHUNDRED} \rangle$ is the average cluster mass in the THETHREEHUNDRED sample ($6.5 \times 10^{14}\, M_\odot$), and $M_{cluster}$ is the mass of the cluster at the centre of a specific filament network. This thickness is kept constant for all filaments surrounding this particular cluster. As discussed above, this is an oversimplification, but alternative recipes could be easily applied. We keep this one for simplicity throughout the paper.

4 ASSIGNING GALAXIES TO COSMIC WEB ENVIRONMENTS

In this section, we provide a framework for assigning mock galaxies to the different cosmic web environments associated with galaxy clusters. This is an essential but non-trivial process if we wish to understand how each environment affects the galaxies' properties and evolution.

Simulations provide full 3D positional information, allowing a precise 'true' environmental assignments to each galaxy. Observations will provide accurate sky positions and radial velocities, but the radial distances will have uncertainties that are large in comparison with the physical dimensions of the volume considered (a sphere with a radius $\sim 5R_{200}$), meaning that the environmental assignments from the observations can only be done in 2D (see Section 3.1.2). This means that the environment associated to a galaxy from the

observational data may not coincide with the 'true' environment. We need to quantify statistically how often that occurs.

We assign each galaxy to different environments according to the following criteria

(i) Cluster core galaxy: galaxies that lie at a radial distance from the cluster centre $r < R_{200}$ in 3D or 2D. The blue points in Fig. 1 show these galaxies.

(ii) Filament galaxy: galaxies that lie outside the cluster core ($r > R_{200}$) and close to the filament spine ($r_{spine} < R_{fil}$), where r_{spine} is the perpendicular distance from the spine of the filament and R_{fil} is the filament thickness determined using equation (1). These galaxies are represented by the orange points, whilst the filaments' spines are shown as red lines in the figure.

(iii) NCF galaxy (neither core nor filament): galaxies that lie outside of the cluster core and outside of filaments ($r > R_{200}$ and $r_{spine} > R_{fil}$). They are shown as green points..[3]

The proportion of galaxies identified in each environment, together with the fraction of the total volume/surface area that each environment occupies around the simulated galaxy clusters in 3D and 2D are presented in Table 1. Not surprisingly, we find that the densest environments are the cluster cores, followed by filaments, with the NCF region being much less dense. Within $5R_{200}$, the cluster cores contain ~ 10 per cent of the galaxies, while they occupy only 1 per cent of the volume, and filaments contain ~ 36 per cent of the galaxies in ~ 6 per cent of the volume. As a comparison, over much larger regions (~ 500 Mpc scales), Cautun et al. (2014) found that

---

[3]We note that is not an exhaustive list of cluster substructure and we focus in a follow up paper (Cornwell et al. in prep.) on the detection of galaxy groups in mock-observations.

Table 1. Fraction of galaxies in each environment and fraction of the total volume/surface area occupied by each environment within $r < 5R_{200}$, in 3D and 2D.

Environment	3D population	3D volume	2D population	2D surface area
Cluster	10 %	1 %	15 %	4 %
Filament	38 %	6 %	45 %	19 %
NCF	52 %	93 %	40 %	77 %

filaments contain ∼6 per cent of the volume, whilst accounting for half of the total mass budget. Similarly, they also found that nodes, which are proxies for cluster cores (Cohn 2022), contain ∼10 per cent of the mass but only 0.1 per cent of the volume. Table 1 also shows that the density contrast between the different environments is significantly reduced when we move from 3D to 2D due to projection effects.

5 RESULTS

In what follows, we will quantify statistically how often the 'true' environment assigned to a galaxy in 3D (directly from the simulations) agrees with the one identified in 2D, once the limitations imposed by observation such as the WWFCS are taken into account (see Section 2.1). This information is essential in order to interpret the observational data correctly when trying to infer how different environments affect galaxy properties and evolution.

5.1 Overall performance of environment allocation

In Fig. 2, we display three confusion matrices to assess how well we can allocate galaxies to different environments using WWFCS-like data. A confusion matrix is a way of visualizing the success of binary classification. In our case, we allocate galaxies to different environment using the mock observations (2D) and compare them to the 'true' environment determined from the full 3D simulations, as described in previous sections. We use the following standard definitions of true-positive, true-negative, false-positive, and false-negative for environment X (where X can be core, filament, or NCF):

(i) TP = number of true-positives = number of galaxies identified as belonging to environment X in the mock observations (2D) and to the same environment X in the simulations (3D).

(ii) TN = number of true-negatives = number of galaxies identified as belonging to environment Y ($\neq$ X) in the mock observations (2D) and to environment Y in the simulations (3D).

(iii) FP = number of false-positives = number of galaxies identified as belonging to environment X in the mock observations (2D) and to environment Y in the simulations (3D).

(iv) FN = number of false-negatives = number of galaxies identified as belonging to environment Y in the observations (3D) and to environment X in the simulations (3D).

Furthermore, we define accuracy as

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN},\qquad(2)$$

and precision as

$$Precision = \frac{TP}{TP + FP}.\qquad(3)$$

In the left-hand panel of Fig. 2, we illustrate the success of classifying cluster core galaxies. For this environment, we get an accuracy of 95 per cent and a precision of 68 per cent. The true-positive identification rate (fraction of galaxies correctly identified

in 2D as belonging to the cluster core) is 67 per cent. Because of projection effects, about one third of the galaxies identified as belonging to the core in 2D are either in front or behind the core itself. For obvious reasons, the 'outside' row of the confusion matrix shows that all galaxies that are outside the core in 2D are also outside in 3D.

The second and third panels evaluate our success (or lack thereof) at identifying filament and NCF galaxies in the 2D mock observations. The true-positive rate for filament galaxies (fraction of galaxies correctly identified in 2D as belonging to filaments) is relatively small (51 per cent), while we are able to correctly identify galaxies not belonging to filaments in 73 per cent of the case true-negative rate. Conversely, we have a higher success at identifying NCF galaxies (true-positive rate of 68 per cent) than at rejecting them (true-negative rate of 59 per cent). This can be easily understood by considering projection effects and the fraction of the total surface area covered by each one of these environments.

These results indicate that identifying the correct environment of a galaxy in the vicinity of a cluster is not straightforward, and the resulting statistical uncertainties cannot be ignored when interpreting the observations. We will investigate next how these uncertainties depend on the distance of a galaxy to the cluster centre and its mass.

5.2 The dependence of the environmental identification success on galaxy mass and clustercentric distance

In this section, we will quantify our success at assigning environments to galaxies using the information provided by spectroscopic surveys like the WWFCS. The goal is to be able to answer the question: if we assign a given environment to a galaxy based on the observations, what is the probability that it is truly in that environment? And, importantly, how does this change with a galaxy's position and mass?

5.2.1 Evaluating a single cluster

We start by using a typical cluster to describe the process. We choose a simulated cluster with a mass similar to WWFCS cluster RX0058, the cluster shown in Fig. 1. We first divide all the galaxies in mass and 2D clustercentric radial distance bins. For all the galaxies in bin (i, j), where i corresponds to a mass bin and j to a radial distance bin, we calculate the probability P_{XXij} that a galaxy has been *correctly* allocated to environment X in the mock observations (2D) as the ratio of the number of galaxies allocated to environment X in 2D whose 'true' 3D environment is also X, N_{XXij}, and the total number of galaxies allocated to environment X in 2D N_{Xij}. In other words,

$$P_{XXij} = \frac{N_{XXij}}{N_{Xij}}.\qquad(4)$$

Conversely, we calculate the probability P_{XYij} that a galaxy has been *incorrectly* allocated to environment X in 2D when its 'true' 3D environment is Y ($\neq$X) as the ratio of the number of galaxies allocated to environment X in 2D whose 'true' 3D environment is Y, N_{XYij},
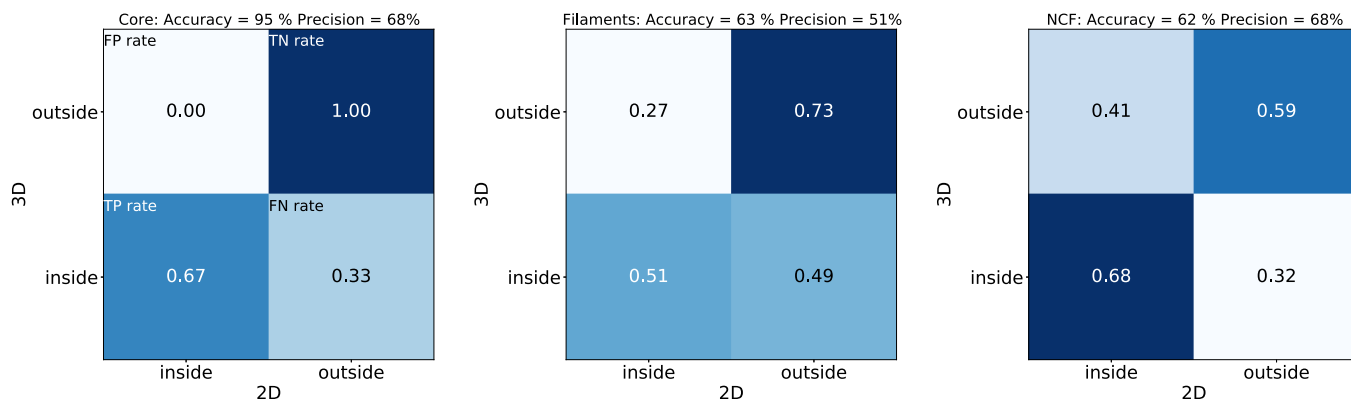
Figure 2. Confusion matrices comparing the environment identification for galaxies in 3D (models) and 2D (simulated mock observations). In each panel, the top left box represents the false positive rate, the bottom left box is the true positive rate, the top right box is the true negative rate, and the bottom right is the false negative rate. Left-hand panel: galaxies identified as core galaxies. Middle panel: galaxies identified as filament galaxies. Right-hand panel: galaxies identified as neither core nor filament galaxies (NCF). The accuracy (equation 2) and precision (equation 3) are displayed at the top of each panel.

and the total number of galaxies allocated to environment X in 2D $N_{\mathrm{X}ij}$. Hence,

$$P_{\mathrm{XY}ij} = \frac{N_{\mathrm{XY}ij}}{N_{\mathrm{X}ij}}. \tag{5}$$

We calculate this probability in four galaxy stellar mass bins and five clustercentric distance bins, covering the ranges $10^9 \, \mathrm{M}_\odot < M_* < 10^{11} \, \mathrm{M}_\odot$ and $0 < r < 6R_{200}$ (note that, all cluster regions are fully covered out to $5R_{200}$ by the WWFCS pointings, and in some cases, we reach beyond $6R_{200}$). The resulting probabilities in each mass and radial bins for the chosen cluster are shown in Fig. 3.

The first row of this figure displays the probability of identifying a galaxy as belonging to the cluster core in 2D when its true 3D environment is the core, a filament or NCF (left-hand, centre and right-hand panels). For example, if we look in the $0 < r < 1R_{200}$ and $10^9 \, \mathrm{M}_\odot < M_* < 10^{9.5} \, \mathrm{M}_\odot$ bin in the first row, we assign a probability of 0.76 for correctly identifying core galaxies as such, a probability of 0.15 for misidentifying a core galaxy as a filament galaxy, and a probability of 0.09 for misidentifying a core galaxy as an NCF galaxy. Reassuringly, the probability of correctly identifying a core galaxy is the highest by a large margin – although there is some contamination due to projection effects, identifying core galaxies is relatively easy. This is illustrated in the bottom left-hand panel of Fig. 3, where 'true' core galaxies correctly identified in 2D as belonging to the core are plotted as blue dots, core galaxies misidentified as belonging to filaments in red, and those misidentified as NCF in green. The incorrect identifications are purely a product of the contamination of the true cluster core sample due to projecting the 3D galaxy distribution in 2D.

The second row of Fig. 3 shows the probability of identifying a galaxy as belonging to a filament in 2D when in 3D it is a core galaxy, a filament galaxy, or an NCF galaxy (left-hand, centre, and right-hand panels, respectively). The left-hand panel is blank because it is not possible to identify a true core galaxy as a filament galaxy in 2D since in our framework the 2D filament networks inside a circle with projected radius of R_{200} are not taken into account (see Cornwell et al. 2022). The middle panel shows the probabilities of correctly classifying filament galaxies, while the right-hand panel present the probability of misidentifying them as NCF galaxies. Notwithstanding the statistical fluctuations, the likelihood of a correct identification for filament galaxies is generally higher than that of a misidentification, but not by much (see Section 5.2.2). The middle panel of the bottom

row in Fig. 3 shows the spatial distribution of the 'true' filament galaxies correctly identified in 2D as belonging to filaments (red dots), and the filament galaxies misidentified in 2D as NCF (red dots). The 3D filament network is shown as black lines.

In a similar way, the third row of Fig. 3 presents the probability of classifying a galaxy in 2D as an NCF galaxy when it is truly in the cluster core, in a filament, or correctly identified as an NCF galaxy. Again, the left-hand panel is blank because, by our definition, a true core galaxy cannot be classified as NCF. As before, the likelihood of correct identification of NCF galaxies is generally a bit higher than that of misidentification. The right-hand panel of the bottom row in this figure shows the spatial distribution of the 'true' NCF galaxies correctly identified in 2D as NCF (green dots), and the NCF galaxies misidentified in 2D as filament galaxies (green dots). The 3D filament network is also shown as black lines.

5.2.2 Evaluating the entire simulated cluster sample

We repeat the process described in Section 5.2.1 for the whole sample of 160 simulated clusters presented in Cornwell et al. (2022), see Section 2.2, and compute the classification probabilities for each galaxy mass and radial bin in all the simulated clusters. These probabilities are then averaged in order to improve the statistics, and the standard deviation for each bin calculated. With this information, we are now able to evaluate robustly any possible dependence on galaxy mass and radial position. Moreover, with such a large simulated cluster sample, we will explore cluster-to-cluster variations in Section 5.3. The average probabilities for each mass and radial bin are displayed in Fig. 4 using the same format as Fig. 3, with the corresponding standard deviations shown in Fig. A1.

When assessing our ability to identify the environment of a galaxy correctly, we need to compare it with that of a random allocation; after all, if all our machinery does not perform better than random, no statistical inference will be possible. We calculate the probability of randomly allocating an environment to a galaxy by shuffling the environment labels for all galaxies in the 3D simulations and recalculating the probabilities for all clusters individually, and then averaging them. In doing so, we are using the 3D population statistics from Table 1 as a prior in the random allocation since, in the absence of any other information, this is our 'best guess' distribution of galaxies in each cosmic web environment. The average random probabilities are displayed in the titles of each panel of Fig. 4 for easy
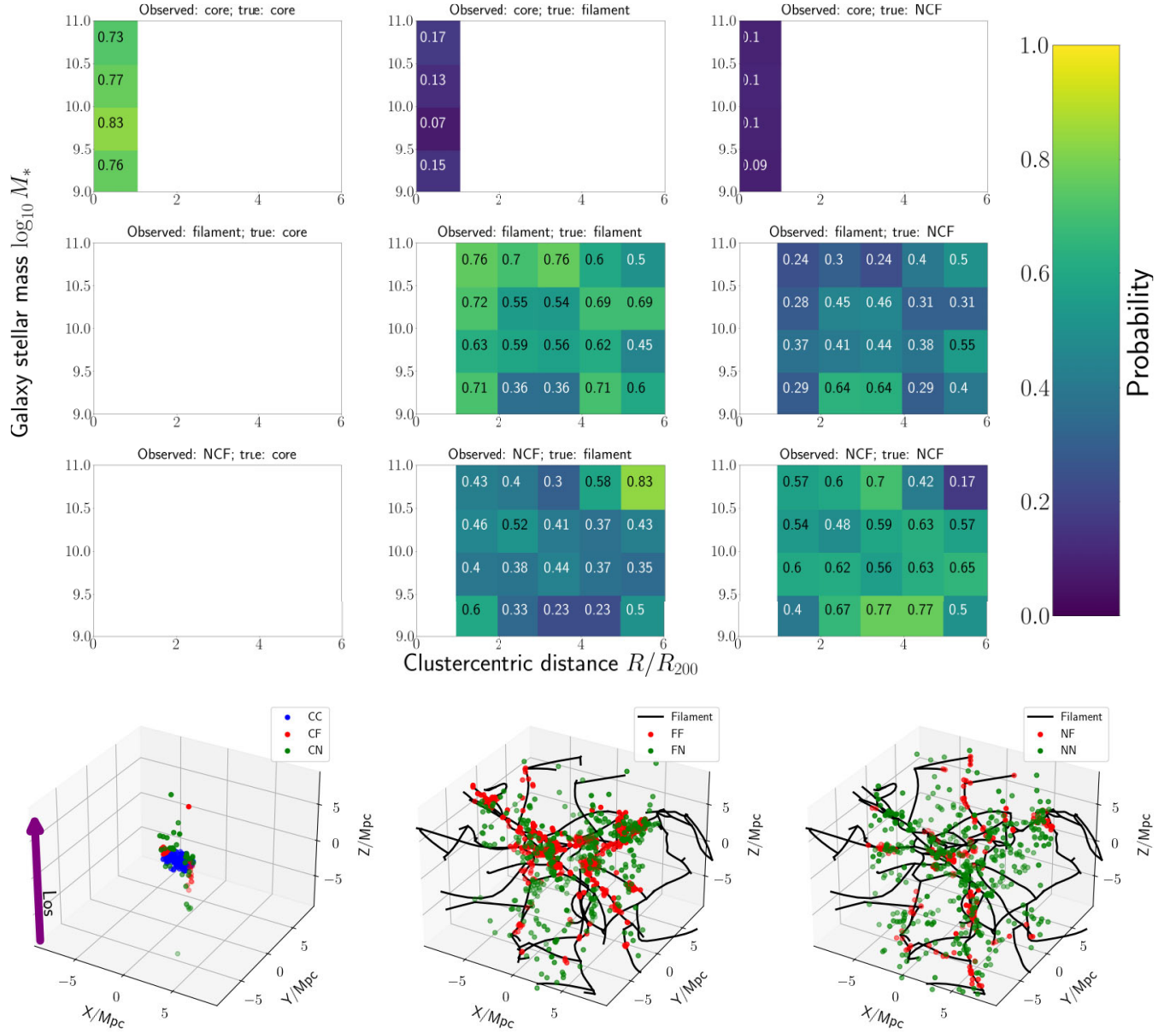
Figure 3. The first 9 panels (3 × 3 grid) display the probabilities of identifying galaxies in different environments (cluster core, cosmic filaments, or neither core nor filament, NCF) as a function of galaxy stellar mass and clustercentric distance in one example cluster. For every bin, we calculate the probability of correct identification $P_{\mathrm{XX}ij}$ (panels in the top-right to bottom-left diagonal) as well as incorrect identifications $P_{\mathrm{XX}ij}$ (other panels) using equations (4) and 5. For illustration we show this for one model cluster, the same one shown in Fig. 1. In the bottom row, we show the spatial distribution of galaxies in the different environments for the same model cluster. The left-hand panel in the bottom row shows all the galaxies that are identified as core galaxies in 2D, with the ones correctly identified as belonging to the core in blue, those misidentified as filament galaxies in red, and the ones misidentified as NCF in green. Similarly, the middle panel of the bottom row shows all the galaxies identified as filament galaxies in 2D, with the ones correctly identified as filament galaxies in red, and those misidentified as NCF galaxies in green. The filament network is shown as black lines. Finally, the right-hand panel in the bottom row shows galaxies that are identified as NCF galaxies in 2D, with the ones correctly identified NCF galaxies in green and those misidentified as filament galaxies in red. The line of sight of the cluster is indicated by the purple arrow, parallel to the *z*-axis.

reference. Reassuringly, the random probabilities thus calculated are very close to the true fractions shown in Table 1, as expected. Note that, when interpreting the results, we want the estimated probabilities to be higher than those drawn from random distributions for the panels on the top-left to bottom-right diagonal (corresponding to correct identification), and lower than the random ones for the rest of the panels (incorrect identifications).

The first row of Fig. 4 clearly indicates that the probability of identifying core galaxies does not depend significantly on galaxy

mass or distance to the cluster centre. The average probability of success is 0.67, clearly showing that we perform much better than random when identifying cluster core galaxies.

The second row, middle panel, shows that the probability of correctly identifying filament galaxies varies from ∼0.44 to ∼0.57. These probabilities are always better than the random chance (∼0.57), but in some cases not by much – thereby once again emphasizing that identifying filament galaxies in the vicinity of massive clusters is difficult, but with large enough samples the
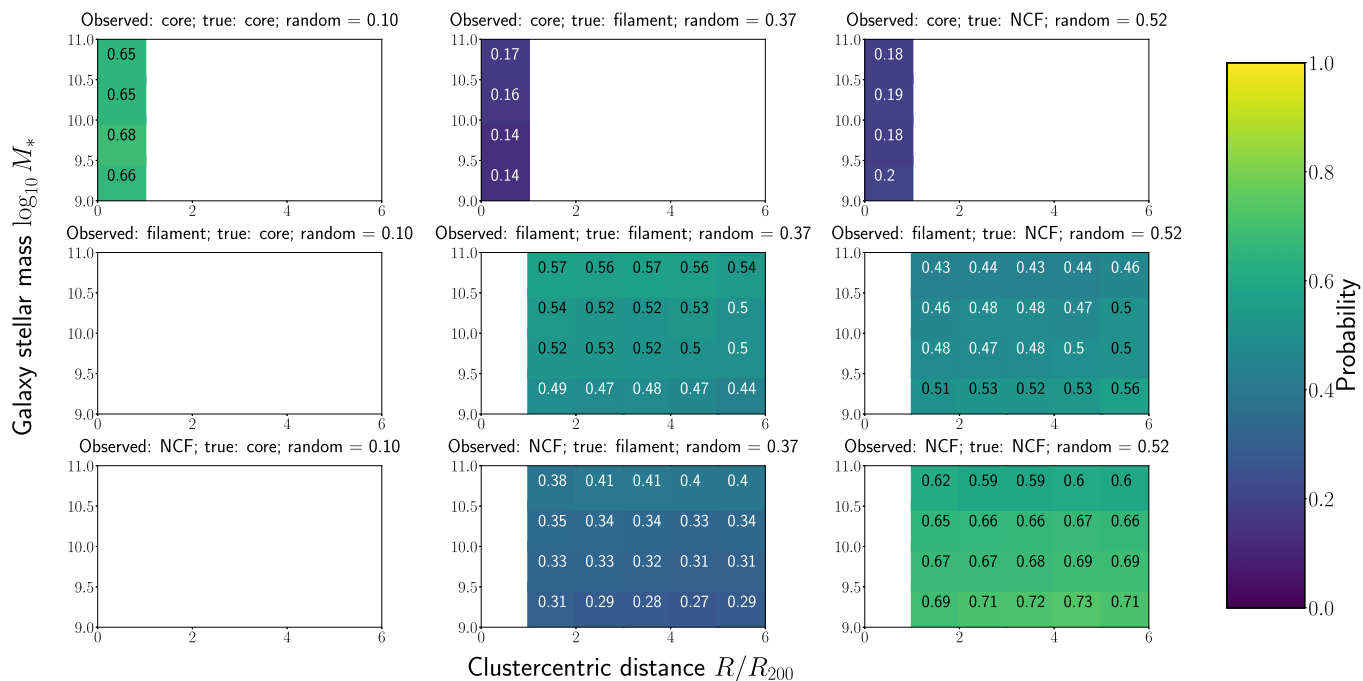
Figure 4. The same as the top 3×3 panels of Fig. 3 but averaged over all clusters. The probability of randomly classifying a galaxy correctly in the title of each plot (see the text for details). The information in this Figure is also included in Table A1.

statistics will be able to beat the noise. The likelihood of success increases significantly with galaxy mass. The most likely explanation is that filament galaxies are, on average, more massive than field galaxies outside filaments, as previously found in simulations (e.g. Ganeshaiah Veena et al. 2018) and observations (e.g. Alpaslan et al. 2016; Malavasi et al. 2016; Kraljic et al. 2017). These works indicate the galaxy stellar mass increases when getting closer to the spine of the filaments. In our own simulations, the filament galaxy samples contain more massive galaxies than the NCF samples (the median galaxy mass in filaments is ~ 17 per cent higher), enhancing the probability of correct identification at high masses. There is also a small increase in the probability of correctly identifying galaxies when the distance to the cluster centre decreases. This can be explained by the fact that closer to the cluster core the fraction of the volume (and projected area) contained in filaments is larger when compared to that occupied by NCF galaxies. As a result, the fraction of filament galaxies misidentified as NCF due to projection effects is smaller closer to the core.

The right-hand panel in the middle row shows that we are a bit better than random at preventing NCF galaxies from contaminating the filament sample at most masses and radial distances, but for the lowest mass galaxies and largest radial distances, we fare a bit worse (by up to ~ 4 per cent). At very large distances from the cluster, the volume occupied by NCF galaxies is so much larger than that occupied by the filaments that projection effects become too strong. When interpreting the properties of filament galaxy samples identified this way it is very important to be aware that in some regions of the parameter space the filament samples will suffer from low purity.

Finally, the bottom row of Fig. 4 completes the picture for the galaxies that are allocated to the NCF category. The probabilities very much mirror what was found for filament galaxies. We are quite successful (in some cases, 20 percentage points above random) at distinguishing NCF galaxies from filament ones, although the

NCF sample will contain significant contamination from filament galaxies. There is very little radial dependence, but we find some galaxy mass dependence in the opposite sense to the one found for filament galaxies: we are slightly more successful at identifying NCF galaxies at lower galaxy masses, as expected from the discussion above.

5.3 The dependence of the environmental identification success on cluster mass

In Appendix A (see Fig. A1), we show that there are significant cluster-to-cluster variations in the probabilities discussed above. This suggests that there may be a systematic dependence of our ability to identify galaxy environments on the properties of the central cluster and the filamentary network that surrounds it. The most obvious cluster property to consider is the cluster mass since the galaxy clusters in the WWFCS sample span over 1 order-of-magnitude in mass (from $\sim 7 \times 10^{13}\,\mathrm{M_\odot}$ to over $\sim 10^{15}\,\mathrm{M_\odot}$). We consider two properties of the filamentary network that may also have an effect: the total length of all filaments in the network, and number of nodes identified by `DisPerSE`. These properties encode the complexity and extent of the network, and may therefore influence the probabilities we calculate. Fortunately, both network length and number of nodes correlate reasonably well with cluster mass (Fig. 5), and therefore, for the purpose of parametrizing the relatively small cluster-to-cluster systematic variations, it suffices to use the cluster mass as main parameter.

We divide the clusters into three different mass bins, spanning the full WWFCS mass range: $7.3 \times 10^{13}\,\mathrm{M_\odot} < M_{\mathrm{cluster}} < 2.5 \times 10^{14}\,\mathrm{M_\odot}$, $2.5 \times 10^{14}\,\mathrm{M_\odot} < M_{\mathrm{cluster}} < 5.0 \times 10^{14}\,\mathrm{M_\odot}$, and $5.0 \times 10^{14}\,\mathrm{M_\odot} < M_{\mathrm{cluster}} < 1.2 \times 10^{15}\,\mathrm{M_\odot}$. Each bin contains ~ 50 simulated clusters. To determine if our success in environmental classification correlates systematically with cluster mass, we calculate the difference in average probabilities between the most massive cluster bin and the
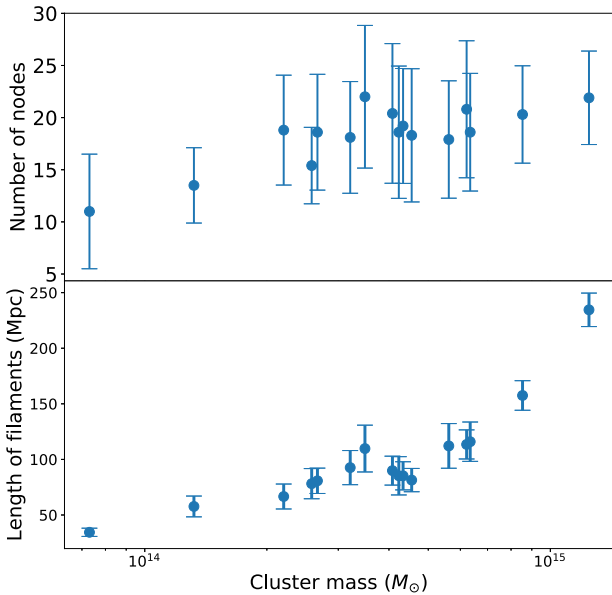
Figure 5. Correlation between filament network properties and the mass of the central cluster for the 160 simulated clusters mass-matched to the WWFCS sample based on `TheThreeHundred` simulations. The top panel shows the number of nodes in the network as a function of cluster mass. The lower panel shows the total length of all filaments in each network also as a function of cluster mass. Points correspond to the average for the 10 simulated clusters in each mass bin, with the error bars showing the corresponding scatter (standard deviation). Clear positive correlation are found between cluster mass and the number of nodes or the length of the filament network.

least massive one. This is computed for every mass and radial distance bin, and shown in Fig. 6.

For the most massive clusters, there is a small but systematic excess in the probability of correctly identifying cluster core galaxies. The difference comes from the fact that the probability of contamination from NCF galaxies is systematically larger for the least massive clusters. This can be understood because the length of the filament network is smaller for clusters with low masses (Fig. 5), and therefore the fraction of the volume occupied by NCF galaxies is larger, making projection effects worse. The contamination of the core sample from filament galaxies changes very little across cluster masses given the relatively small volume occupied by filaments.

The second row of Fig. 6 shows that for high mass clusters, we are systematically more successful at identifying filament galaxies correctly, while still performing generally better than random for lower mass clusters. This result is statistically significant since the differences in probabilities are generally and systematically larger that the 0.01–0.02 uncertainties. It is likely that the reason for this is simply that in more massive clusters the length of the filament network and thus the fraction of volume occupied by them is larger[4] (Fig. 5), implying that the the projection effects leading to the misidentification of filament galaxies are smaller for high mass clusters.

Reciprocally, similar arguments explain how the probability of correctly identifying NCF clusters is higher in low-mass cluster regions (third row of Fig. 6). The fractional volume occupied for

NCF galaxies increases as the mass of the cluster decreases due to the opposite trend shown by filament galaxies.

We conclude that relatively small but systematic variations with cluster mass (and correlated filamentary network properties) exist in the probabilities of correctly identifying the environment of galaxies, and it is therefore useful to calculate separate tables for different masses. Given the size of the variations, it suffices to divide the clusters in three mass bins – further granularity would reduce the statistical accuracy of the calculated probabilities without significantly altering the results. Table A1 presents in numerical form the probabilities calculates for the complete cluster sample, as shown in Fig. 4. Similarly, Tables A2, A3, and A4 contain the probabilities for low-, intermediate-, and high-mass clusters.

6 CONCLUSIONS

Galaxies experience different physical processes in different environments. Next generation wide-field spectroscopic surveys will be able to accurately map out in detail the distribution of galaxies in the cosmic web around galaxy clusters. In Cornwell et al. (2022), we laid down the framework for developing mock observations to accurately forecast the success in reconstructing cosmic filaments around galaxy clusters for one such survey, the WEAVE wide-field cluster survey (WWFCS). In this paper, we assess the feasibility and accuracy of assigning individual galaxies to different cosmic web environments using a large sample of simulated galaxy clusters from THETHREEHUNDRED project (Cui et al. 2018). In order to do so, we compare the 'true' environments we assign to galaxies using the 3D information provided by the simulations with the 'observed' environment we assign to the same galaxies using mock observations that take into account the observational constraints and selection effects of the planned WWFCS. We summarize our main findings below.

(i) Filaments occupy only ~ 6 per cent of the volume enclosed in a sphere with radius $5R_{200}$ around massive galaxy clusters, but contain ~ 38 per cent of the galaxies with masses above $10^9\,\mathrm{M_\odot}$. This is calculated using filament thicknesses that decrease with the mass of the main halo (see Section 3.2). In comparison, galaxies that are neither in the cluster core nor in filaments (NCF) make up ~ 52 per cent of the galaxy population, whilst occupying 93 per cent of the volume. The cluster core itself (defined as the sphere with radius R_{200}) contains ~ 1 per cent of the volume and ~ 10 per cent of the galaxies. To understand how these different environments affect the properties and evolution of the galaxies that inhabit them we need to be able to associate galaxies to the correct environment and to quantify statistically the uncertainties involved.

(ii) When we allocate galaxies to different environments in the mock observations and compare them to the allocations from the 'true' simulations, we find that, combining all cluster and galaxy masses, and at all clustercentric distances, we are able to identify core, filament, and NCF galaxies with statistical accuracies (precisions) of 95 per cent (68 per cent), 63 per cent (51 per cent) and 62 per cent (68 per cent), respectively (see definitions in equations 2 and 3). This indicates that, while cluster core galaxy samples can be built with a high level of completeness and moderate contamination, the filament and NCF galaxy samples will be significantly contaminated and incomplete due to projection effects, even with good-quality spectroscopic redshifts.

(iii) In our framework, we calculate the probabilities of galaxies being correctly assigned to a given environment, together with the probabilities of misidentifying them as a belonging to a different
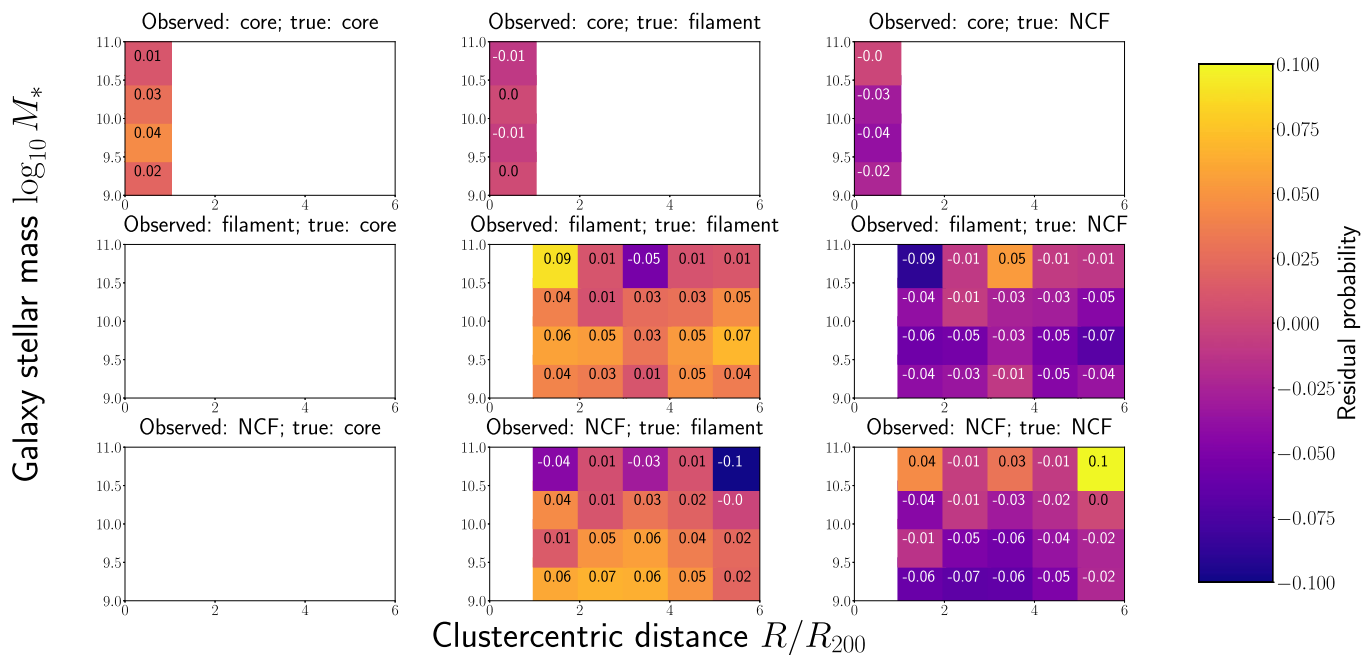
---

[4]As a consequence of the correlation between cluster mass and total length of the filament network shown in Fig. 5 we also find a clear positive correlation between the cluster mass and the fraction of volume occupied by filaments inside $r < 5R_{200}$.

Figure 6. Residual probabilities of the environmental classification showing the dependency on cluster mass. These are calculated as the difference between the average probabilities for the most massive third of the clusters ($M_{\text{cluster}} > 5 \times 10^{14}\,M_\odot$) and the least massive third ($M_{\text{cluster}} < 2.5 \times 10^{14}\,M_\odot$). A positive residual indicates that a given probability is higher for the more massive clusters.

one (Fig. 4 and Table A1). We do that as a function of galaxy mass and clustercentric distance. We find that, outside the cluster core (beyond $\sim R_{200}$), identifying filament galaxies is marginally more successful at high galaxy masses and low clustercentric distances, while the reciprocal is true for NCF galaxies. Generally, the success of the environment allocation is significantly better than random, but sometimes only marginally so. We conclude that identifying the cosmic web environments of galaxies in the vicinity of massive clusters (within a sphere of radius $\sim 5R_{200}$ from the cluster centre) is remarkably difficult due to projection effects exacerbated by the magnitude of the galaxies' peculiar velocities (Fingers-of-God).

(iv) We also find that the calculated probabilities vary with the mass of the central cluster and, by association, with properties of the filamentary network such as the number of nodes or the total length of the filaments. We therefore calculate the probabilities for different cluster mass ranges (Tables A2, A3, and A4), and find that identifying filament galaxies is marginally more successful around the most massive clusters because their filament networks occupy a relatively larger fraction of the total volume considered.

We conclude that, in the infall regions surrounding masssive galaxy clusters, associating galaxies with the correct cosmic web environment is highly uncertain. However, applying our statistical framework and probabilities to large spectroscopic samples like the WWFCS will allow us to observationally extract robust and well-defined conclusions on relationships between galaxy properties and their environments.

ACKNOWLEDGEMENTS

---

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

DATA AVAILABILITY

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

REFERENCES

Alpaslan M. et al., 2016, MNRAS, 457, 2287

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

Balcells M. et al., 2010, in Ground-based and Airborne Instrumentation for Astronomy III. p. 2659

Cautun M., van de Weygaert R., Jones B. J. T., Frenk C. S., 2014, MNRAS, 441, 2923

Chen Y.-C., et al., 2017, MNRAS, 466, 1880

Cohn J. D., 2022, MNRAS, 513, 624

Cornwell D. J. et al., 2022, MNRAS, 517, 1678

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, in Ground-based and airborne instrumentation for astronomy V. p. 177

De Luca F., De Petris M., Yepes G., Cui W., Knebe A., Rasia E., 2021, MNRAS, 504, 5383

Dressler A., 1980, ApJ, 236, 351

Ganeshaiah Veena P., Cautun M., van de Weygaert R., Tempel E., Jones B. J. T., Rieder S., Frenk C. S., 2018, MNRAS, 481, 414

Gill S. P., Knebe A., Gibson B.K., 2004, MNRAS, 351, 399

Jin S. et al., 2023, MNRAS, in press

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., 2017, MNRAS, 474, 547

Kuchner U. et al., 2020, MNRAS, 494, 5473

Kuchner U. et al., 2021, MNRAS, 503, 2065

Kuchner U. et al., 2022, MNRAS, 510, 581

Laigle C. et al., 2017, MNRAS, 474, 5437

Libeskind N. I. et al., 2017, MNRAS, 473, 1195

Malavasi N., Aghanim N., Tanimura H., Bonjean V., Douspis M., 2020, A&A, 634, A30

Malavasi N. et al., 2016, MNRAS, 465, 3817

Malavasi N., Aghanim N., Tanimura H., Bonjean V., Douspis M., 2020, A&A, 634, A30

Morell D. F., Ribeiro A. L. B., de Carvalho R. R., Rembold S. B., Lopes P. A. A., Costa A. P., 2020, MNRAS, 494, 3317

Planck Collaboration XIII, 2016, A&A, 594, A13

Poudel A., Heinämäki P., Tempel E., Einasto M., Lietzen H., Nurmi P., 2017, A&A, 597, A86

Ribeiro A. L. B., Lopes P. A. A., Rembold S. B., 2013, A&A, 556, A74

Rost A. et al., 2021, MNRAS, 502, 714

Sarron F., Adami C., Durret F., Laigle C., 2019, A&A, 632, A49

Sousbie T., 2011, MNRAS, 414, 350

Sousbie T., Pichon C., Kawahara H., 2011, MNRAS, 414, 384

Tanimura H., Aghanim N., Douspis M., Malavasi N., 2022, A&A

Tempel E., Kipper R., Tamm A., Gramann M., Einasto M., Sepp T., Tuvikene T., 2016, A&A, 588, A14

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

Vernstrom T., Heald G., Vazza F., Galvin T., West J., Locatelli N., Fornengo N., Pinetti E., 2021, MNRAS, 505, 4178

Zabludoff A. I., Mulchaey J. S., 1998, ApJ, 496, 39

Zel'dovich Y. B., 1970, A&A, 5, 84

APPENDIX A: CLUSTER-TO-CLUSTER PROBABILITY VARIATION

To investigate whether there are significant cluster-to-cluster variations in the probabilities of correctly associating galaxies with their environments, we calculate the standard deviation of these probabilities over all clusters and present them in Fig. A1. The measured scatter is generally well above the statistical uncertainty (the standard error for the probability values is between 0.01 and 0.02), demonstrating that there are significant cluster-to-cluster variations. We explore these in Section 5.3.

Table A1. Probability of galaxies being identified in different cosmic web environments for model clusters of all masses. These probabilities are shown for each environment as a function of galaxy stellar mass and clustercentric distance in a similar arrangement as in Fig. 4. For each mass and distance bin, three numbers are given, each corresponding to each one of the columns in Fig. 4. A number '−1' indicates that a given situation is not possible, and corresponds to a white cell in Fig. 4. For instance, we cannot have galaxies classified as 'core' at projected radial distances larger than R_{200}, and therefore the corresponding probabilities are not defined. Note that, the numbers and information presented in this table are exactly the same as those presented in Fig. 4. We provide them here in tabular form for easy access.

Core	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	0.66/0.14/0.20	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{9.5} < M < 10^{10.0}$	0.68/0.14/0.18	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{10.0} < M < 10^{10.5}$	0.65/0.16/0.19	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{10.5} < M < 10^{11.0}$	0.65/0.17/0.18	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
Filaments	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	−1/−1/−1	−1/0.49/0.51	−1/0.47/0.53	−1/0.48/0.52	−1/0.47/0.53	−1/0.44/0.56
$10^{9.5} < M < 10^{10.0}$	−1/−1/−1	−1/0.52/0.48	−1/0.53/0.47	−1/0.52/0.48	−1/0.50/0.50	−1/0.50/0.50
$10^{10.0} < M < 10^{10.5}$	−1/−1/−1	−1/0.54/0.46	−1/0.52/0.48	−1/0.52/0.48	−1/0.53/0.47	−1/0.50/0.50
$10^{10.5} < M < 10^{11.0}$	−1/−1/−1	−1/0.57/0.43	−1/0.56/0.44	−1/0.57/0.43	−1/0.56/0.44	−1/0.54/0.46
NCF	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	−1/−1/−1	−1/0.31/0.69	−1/0.29/0.71	−1/0.28/0.72	−1/0.27/0.73	−1/0.29/0.71
$10^{9.5} < M < 10^{10.0}$	−1/−1/−1	−1/0.33/0.67	−1/0.33/0.67	−1/0.32/0.68	−1/0.31/0.69	−1/0.31/0.69
$10^{10.0} < M < 10^{10.5}$	−1/−1/−1	−1/0.35/0.65	−1/0.34/0.66	−1/0.34/0.66	−1/0.33/0.67	−1/0.34/0.66
$10^{10.5} < M < 10^{11.0}$	−1/−1/−1	−1/0.38/0.62	−1/0.41/0.59	−1/0.41/0.59	−1/0.40/0.60	−1/0.40/0.60

Table A2. Probability of galaxies being identified in different cosmic web environments for low mass model clusters ($7.3 \times 10^{13}\,\mathrm{M_\odot} < M_{\mathrm{cluster}} < 2.5 \times 10^{14}\,\mathrm{M_\odot}$). The format is the same as in Table A1.

Core	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	0.67/0.13/0.20	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{9.5} < M < 10^{10.0}$	0.65/0.14/0.20	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{10.0} < M < 10^{10.5}$	0.63/0.16/0.21	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{10.5} < M < 10^{11.0}$	0.63/0.18/0.19	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
Filaments	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	−1/−1/−1	−1/0.48/0.52	−1/0.47/0.53	−1/0.47/0.53	−1/0.45/0.55	−1/0.43/0.57
$10^{9.5} < M < 10^{10.0}$	−1/−1/−1	−1/0.49/0.51	−1/0.50/0.50	−1/0.50/0.50	−1/0.48/0.52	−1/0.46/0.54
$10^{10.0} < M < 10^{10.5}$	−1/−1/−1	−1/0.52/0.48	−1/0.53/0.47	−1/0.51/0.49	−1/0.53/0.47	−1/0.48/0.52
$10^{10.5} < M < 10^{11.0}$	−1/−1/−1	−1/0.51/0.49	−1/0.57/0.43	−1/0.63/0.37	−1/0.57/0.43	−1/0.55/0.45
NCF	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	−1/−1/−1	−1/0.29/0.71	−1/0.26/0.74	−1/0.26/0.74	−1/0.26/0.74	−1/0.28/0.72
$10^{9.5} < M < 10^{10.0}$	−1/−1/−1	−1/0.32/0.68	−1/0.29/0.71	−1/0.29/0.71	−1/0.29/0.71	−1/0.30/0.70
$10^{10.0} < M < 10^{10.5}$	−1/−1/−1	−1/0.33/0.67	−1/0.35/0.65	−1/0.33/0.67	−1/0.33/0.67	−1/0.33/0.67
$10^{10.5} < M < 10^{11.0}$	−1/−1/−1	−1/0.42/0.58	−1/0.42/0.58	−1/0.45/0.55	−1/0.39/0.61	−1/0.50/0.50

Table A3. Probability of galaxies being identified in different cosmic web environments for intermediate mass model clusters ($2.5 \times 10^{14}\,\mathrm{M_\odot} < M_{\mathrm{cluster}} < 5.0 \times 10^{14}\,\mathrm{M_\odot}$). The format is the same as in Table A1.

Core	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	0.64/0.15/0.21	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{9.5} < M < 10^{10.0}$	0.68/0.13/0.19	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{10.0} < M < 10^{10.5}$	0.64/0.15/0.21	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{10.5} < M < 10^{11.0}$	0.66/0.16/0.17	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
Filaments	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	−1/−1/−1	−1/0.49/0.51	−1/0.47/0.53	−1/0.49/0.51	−1/0.49/0.51	−1/0.44/0.56
$10^{9.5} < M < 10^{10.0}$	−1/−1/−1	−1/0.50/0.50	−1/0.51/0.49	−1/0.51/0.49	−1/0.48/0.52	−1/0.49/0.51
$10^{10.0} < M < 10^{10.5}$	−1/−1/−1	−1/0.51/0.49	−1/0.50/0.50	−1/0.49/0.51	−1/0.49/0.51	−1/0.47/0.53
$10^{10.5} < M < 10^{11.0}$	−1/−1/−1	−1/0.53/0.47	−1/0.52/0.48	−1/0.55/0.45	−1/0.54/0.46	−1/0.51/0.49
NCF	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	−1/−1/−1	−1/0.33/0.67	−1/0.32/0.68	−1/0.33/0.67	−1/0.27/0.73	−1/0.30/0.70
$10^{9.5} < M < 10^{10.0}$	−1/−1/−1	−1/0.32/0.68	−1/0.33/0.67	−1/0.31/0.69	−1/0.31/0.69	−1/0.31/0.69
$10^{10.0} < M < 10^{10.5}$	−1/−1/−1	−1/0.32/0.68	−1/0.31/0.69	−1/0.31/0.69	−1/0.31/0.69	−1/0.36/0.64
$10^{10.5} < M < 10^{11.0}$	−1/−1/−1	−1/0.39/0.61	−1/0.37/0.63	−1/0.38/0.62	−1/0.39/0.61	−1/0.36/0.64

Table A4. Probability of galaxies being identified in different cosmic web environments for high mass model clusters ($5.0 \times 10^{14}\,\mathrm{M_\odot} < M_{\mathrm{cluster}} < 1.2 \times 10^{15}\,\mathrm{M_\odot}$). The format is the same as in Table A1.

| Core	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	0.69/0.14/0.18	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{9.5} < M < 10^{10.0}$	0.70/0.14/0.17	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{10.0} < M < 10^{10.5}$	0.66/0.17/0.18	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
$10^{10.5} < M < 10^{11.0}$	0.64/0.17/0.19	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1	−1/−1/−1
Filaments	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	−1/−1/−1	−1/0.52/0.48	−1/0.50/0.50	−1/0.48/0.52	−1/0.50/0.50	−1/0.47/0.53
$10^{9.5} < M < 10^{10.0}$	−1/−1/−1	−1/0.55/0.45	−1/0.55/0.45	−1/0.53/0.47	−1/0.54/0.46	−1/0.53/0.47
$10^{10.0} < M < 10^{10.5}$	−1/−1/−1	−1/0.56/0.44	−1/0.53/0.47	−1/0.54/0.46	−1/0.56/0.44	−1/0.53/0.47
$10^{10.5} < M < 10^{11.0}$	−1/−1/−1	−1/0.60/0.40	−1/0.58/0.42	−1/0.58/0.42	−1/0.58/0.42	−1/0.56/0.44
NCF	$0 < r < 1R_{200}$	$1R_{200} < r < 2R_{200}$	$2R_{200} < r < 3R_{200}$	$3R_{200} < r < 4R_{200}$	$4R_{200} < r < 5R_{200}$	$5R_{200} < r < 6R_{200}$
$10^{9.0} < M < 10^{9.5}$	−1/−1/−1	−1/0.35/0.65	−1/0.33/0.67	−1/0.32/0.68	−1/0.31/0.69	−1/0.31/0.69
$10^{9.5} < M < 10^{10.0}$	−1/−1/−1	−1/0.34/0.66	−1/0.34/0.66	−1/0.35/0.65	−1/0.33/0.67	−1/0.33/0.67
$10^{10.0} < M < 10^{10.5}$	−1/−1/−1	−1/0.37/0.63	−1/0.36/0.64	−1/0.36/0.64	−1/0.34/0.66	−1/0.33/0.67
$10^{10.5} < M < 10^{11.0}$	−1/−1/−1	−1/0.37/0.63	−1/0.43/0.57	−1/0.42/0.58	−1/0.40/0.60	−1/0.40/0.60
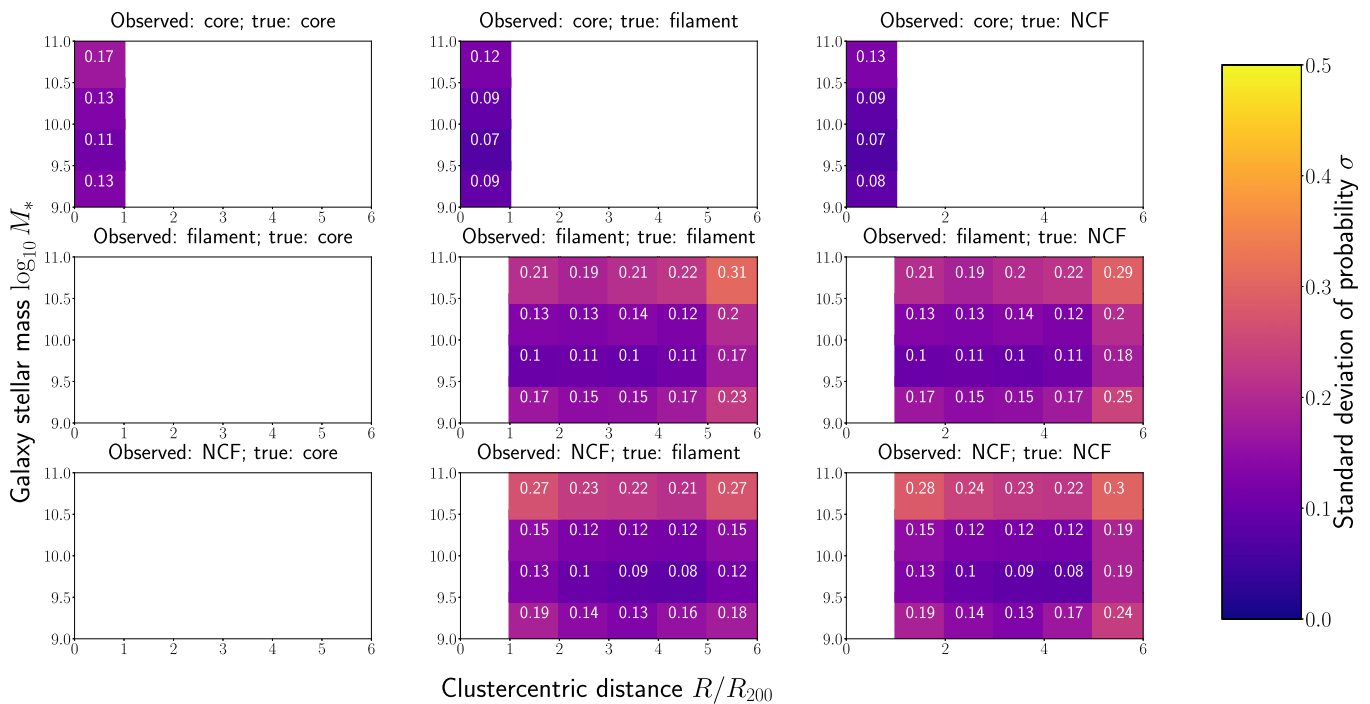
Figure A1. The cluster-to-cluster scatter (standard deviation) of the probabilities shown in Fig. 4.

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